A Risk Explanation for the R&D Anomaly

Woon Sau Leung[*†], Khelifa Mazouz[*], and Kevin P. Evans[*]

This version: March 2015

† Corresponding author. Email: LeungWS@cardiff.ac.uk. Tel: +44 (0)29 208 76736.

* Cardiff Business School, Cardiff University, Cardiff CF10 3EU, United Kingdom.

We thank seminar participants at Cardiff Business School and Nottingham Trent Business School for helpful comments.

A Risk Explanation for the R&D Anomaly

Abstract

This study provides new evidence on the Research and Development (R&D) anomaly that supports a risk explanation. An R&D-related factor explains both the abnormal returns on R&D intensive stocks and the long-term abnormal returns after R&D increases. Using investor sentiment to detect mispricing, we find no evidence of mispricing correction in R&D stocks. Consistent with systematic risk, exposure to an R&D factor predicts returns and subsumes R&D intensity. Furthermore, R&D factor loadings are predicted by risk indicators relating to liquidation costs, shareholder recovery, financial distress risk, and information asymmetry.

Keywords: R&D anomaly; Rational factor pricing; Covariance risk; Mispricing; Distress risk.

JEL Classification: G12; G14.

1. Introduction

It is widely documented in the literature that R&D investment is followed by significant stock returns that asset pricing models are unable to explain. Higher future abnormal returns are associated with a larger stock of R&D capital (Lev and Sougiannis, 1996), higher levels of R&D expenditure (Chan et al., 2001; Li, 2011) and increases in R&D expenditure (Eberhart et al., 2004).[1] These patterns represent the R&D anomaly, which is one of the most important challenges to rational asset pricing theories (Fama and French, 2008). Despite a large literature confirming the anomaly, there is far less consensus on why it exists. This is the research question that we address in this study.

Many studies argue that the R&D anomaly is due to investor mispricing and these purport two justifications. The first relates to investor attention constraints (Barber and Odean, 2008). R&D projects are long term whilst investors' horizons are short term. Due to their attention constraint, investors fail to incorporate the longer term relative benefits of R&D into current financial information, which causes R&D intensive stocks to be undervalued (Chan et al., 2001).[2] The second mispricing argument suggests that investors are misled by the conservative accounting treatment of R&D. Accounting standards stipulate that R&D must be fully expensed. Therefore, earnings are understated (overstated) when R&D is increased (decreased), causing investors to underprice (overprice) these stocks (Lev et al., 2005; Penman and Zhang, 2002; and Eberhart et al., 2004). Under both mispricing arguments, higher future returns on R&D intensive

[1] For further evidence see Chan et al. (1990), Lev and Sougiannis (1999), Chambers et al. (2002), Al-Horani et al. (2003), Guo et al. (2006), Szwejczewski et al. (2006), Ciftci et al. (2011), and Donelson and Resutek (2012).

[2] This effect is also known as the functional fixation hypothesis (Chan et al., 2001) where investors fixate on the face value of financial statements information rather than incorporating future benefits.

stocks represent correction of initial mispricing. A mispricing explanation has important implications for demanding greater disclosure surrounding intangible assets and changes to accounting regulations to help investors evaluate the relative benefits of R&D and managers to allocate resources more efficiently (Lev, 2004).

In direct contrast, Lev et al. (1999), Chambers et al. (2002), and Donelson and Resutek (2012) argue that the R&D anomaly represents rational compensation for risk. If capital markets are efficient and competition among rational investors generates equilibrium, then expected returns are determined only by systematic risk. The presence of irrational investors is offset quickly by the exploitation of arbitrage opportunities so that mispricing cannot persist (Baker and Wurgler, 2006; Ciftci et al., 2011). According to this risk explanation, the returns to R&D stocks are driven by covariance with a systematic risk factor. The persistence of the anomaly implies that this factor is unidentified as it has yet to be incorporated into asset pricing models. Support for a risk explanation helps investors to understand the risks inherent in R&D firms, to improve portfolio allocation and to adjust performance more accurately for risk.

We contribute to this debate by presenting new arguments and evidence in support of a risk explanation. We show that R&D stock returns covary and that firms' exposure to a systematic R&D risk factor predicts future returns, after including other asset pricing controls. More importantly, we argue that this covariation arises from R&D firms sharing a number of risk characteristics that relate to the nature of R&D investment.[3] First, R&D projects are long term, irreversible, inflexible and have uncertain outcomes which drive higher business risk (Ciftci et al., 2011). Second, Aboody and Lev (2000) show that R&D investments are a major source of

[3] This is consistent with Donelson and Resutek (2012) who show that higher future excess returns to R&D stocks are driven by a component of returns that is not related to R&D.

information asymmetry. A large literature shows that higher information asymmetry increases cost of capital and that information risk operates as a systematic risk factor.[4] Therefore, information asymmetry likely contributes to R&D-related return comovement. Third, R&D intensive firms are smaller and more financially constrained. Li (2011) shows that the premium required by investors for holding financially constrained firms increases with R&D intensity, suggesting that financial constraint is an R&D-related characteristic shared by these firms. Fourth, R&D firms exhibit common properties associated with financial distress risk. The intangibility and specificity of assets, particularly R&D, correspond to higher liquidation costs and lower shareholder recovery in the event of distress. Garlappi et al. (2008) find that investors require higher expected returns from R&D firms when default risk is higher. Opler and Titman (1994) find that R&D investments suffer relatively more in distress. Related studies document that financial distress risk is a priced risk factor (Vassalou and Xing, 2004; George and Hwang, 2010) and it is proposed as a potential explanation of the ubiquitous book-to-market effect.[5] In summary, business risk, information asymmetry, financial constraints and financial distress risk represent particular risk characteristics shared by R&D firms. We argue that these common risk characteristics combine to drive the systematic covariation in R&D stock returns, but which have not been captured adequately by existing pricing models. The R&D anomaly therefore reflects a rational risk premium and we contribute to the literature by investigating the extent to which these risk characteristics relate to this premium.

[4] See for example Easley and O'Hara (2004); Lambert et al. (2007); and Armstrong et al. (2010).

[5] See Fama and French (1992); Lakonishok et al. (1994) and Daniel and Titman (1997) for the early evidence on this debate.

We first confirm the presence of the R&D anomaly in decile portfolios. To extract a factor that mimics the latent R&D risk, we construct a zero cost portfolio based on anomalous returns that goes long R&D intensive stocks and short stocks with low R&D intensity. This R&D factor fully explains the abnormal returns related to R&D expenditure, R&D capital and the long term abnormal returns after R&D increases. Portfolio returns load more heavily onto this factor as R&D increases across firms. We extend the existing literature by treating R&D risk as a latent systematic factor and constructing a return based factor to mimic it. Our evidence of return comovement related to this R&D factor satisfies the first necessary condition for a rational risk explanation (Hirshleifer et al., 2012).

To distinguish between risk and mispricing, we perform two tests. The first investigates the effect of investor sentiment on R&D decile portfolios. Baker and Wurgler (2006) argue that shifts in investor sentiment have cross-sectional effects on returns that reveal correction of mispricing. Certain stocks that are hard to value and more costly to arbitrage are subject to larger mispricing, and this should manifest through larger correction of mispricing in the cross section following shifts in investor sentiment. Under a mispricing argument, R&D intensive stocks should exhibit precisely such properties. We contribute to the mispricing and investor sentiment literatures by extending this approach to detect correction of mispricing in R&D stocks by adjusting explicitly for size and other factors. Our results reveal that sentiment has no effect on R&D portfolios, inconsistent with a mispricing explanation. Specifically, cross-sectional patterns in portfolio returns and spreads, both within and in comparison across sentiment states show evidence inconsistent with mispricing correction. More revealing are the patterns in decile portfolio return differences across states, conditional on R&D, which refute a mispricing explanation in favour of a systematic risk interpretation. Furthermore, the scant evidence of mispricing correction is annihilated when controlling for size, suggesting that any mispricing detected in R&D stocks is due to small firms. Controlling for size explicitly, a strong R&D effect remains and including a systematic R&D risk factor captures this residual R&D effect, consistent with a risk explanation.

The second test, which represents our most important contribution, evaluates the importance of stocks' covariance with the R&D factor (risk) compared to R&D intensity (mispricing characteristic) for explaining the cross section of expected stock returns. A rational risk explanation requires that R&D factor loadings explain future stock returns after controlling for the R&D characteristic (Daniel and Titman, 1997; Hirshleifer et al., 2012). In Fama and MacBeth (1973) regressions, we find that both R&D factor loadings and R&D intensity predict stock returns independently in the pricing tests. However, when included jointly, the significance of the R&D characteristics is subsumed by the R&D factor loadings. Controlling for the R&D characteristic and other variables including book-to-market equity, coefficients on the loadings remain positive and highly significant for predicting stock returns. This evidence clearly opposes the mispricing explanation suggesting instead that investors require a premium for exposure to this R&D-related systematic risk.

Following our argument that R&D firms share common risk characteristics, we investigate whether these features predict a firm's exposure to R&D systematic risk, thereby providing our most innovative contribution to the literature. In support of our contention, we find that higher R&D expenditure, liquidation costs, shareholder recovery, financial distress risk, and information asymmetry contribute significantly to firms' future R&D factor loadings. We are careful to control for many other variables, including book-to-market equity. The insignificance of this ratio shows that R&D risk is not part of a wider book-to-market equity effect, however explicit measures of financial distress risk are significant determinants of R&D risk. This evidence confirms the financial distress risk and information risk mechanisms by which R&D activities determine future systematic risk exposure. Furthermore, the commonality of these risks generates exposure to a systematic R&D risk factor that is priced in the cross section of expected returns, thus providing economic foundation to the systematic risk explanation for the R&D anomaly.

This paper is motivated by the persistence of the R&D anomaly and the debate as to whether this reflects risk or mispricing. We show evidence that refutes mispricing and is consistent with a risk explanation. There are persuasive arguments that R&D intensive firms exhibit common risk properties that drive R&D related return covariation. Our main contributions show that the R&D factor loading subsumes the R&D characteristic in explaining the cross section of expected returns and that this loading is determined by firm characteristics relating to asset tangibility, financial constraints, distress risk and information asymmetry. The remainder of the paper is organized as follows. Section 2 describes the sample and presents some descriptive statistics. Section 3 documents the R&D anomaly and presents the R&D factor mimicking portfolio. Section 4 presents our evidence in favour of a risk explanation and Section 5 investigates the determinants of R&D factor risk exposure. Section 6 explains additional results from robustness checks and Section 7 concludes.

2. Data and Descriptive Statistics

Our sample contains all NYSE, AMEX and NASDAQ stocks with accounting data available from COMPUSTAT and securities data from the Centre for Research in Security Prices (CRSP) for the period 1976-2013. All accounting variables as at fiscal year-end in calendar year t-1 are matched with monthly returns from July *t* to June *t*+1, allowing sufficient time for the public information to be made available to investors. Any unmerged data and firm-month observations with negative book values of equity are discarded and financial stocks are excluded. We adjust returns for stock delisting to avoid survivorship bias using the approach of Shumway (1997).[6] Our final sample consists of 1,627,201 firm-month observations (15,114 firms), with 770,929 firm-month observations (7,612 firms) having non-zero R&D expenditures.

Following Fama and French (1993), we use firm book equity value as at fiscal year-end in calendar year *t*-1 and its market equity at the end of December *t*-1 to compute the book-to-market equity ratio (*BM*) and market value (*SIZE*) at the end of June of year *t*. To control for price momentum (Jegadeesh and Titman, 1993), we compute a past return variable, *RET*(-12,-2), which is the compounded gross return from month *j*-12 to *j*-2. The excess return on the CRSP value-weighted index (*MKT_RF*), monthly Treasury Bill yield (*RF*), Fama and French (1993) Small-Minus-Big (*SMB*) and High-Minus-Low (*HML*) factors, and Carhart's (1997) Up-Minus-Down (*UMD*) factor are obtained from the Kenneth French data library.[7]

The most important variable of this study is R&D intensity. This is defined in the literature as R&D expenditure relative to market value, total assets or sales, which we denote as *RD-MV*, *RD-A* and *RD-S*, respectively. Table 1 shows descriptive statistics of these variables over time and across sectors. The average values of *RD-MV*, *RD-A*, and *RD-S* are 31.7, 9.1, and

[6] When a stock delists, the last return is the delisting return if it is available. Following Shumway (1997), if this is not available, we assign a return of -30% for stocks that delist for performance reasons and -100% for those that delist for other reasons.

[7] We thank Prof. French for making the factors available online. We verify our results throughout to the inclusion of the liquidity risk factor of Pastor and Stambough (2003). Since this has no impact on our results, we do not report the analysis.

254.2 percent, respectively. R&D intensity has increased sharply recently and in most sectors, with the highest levels observed during 1996-2005, which included the "Internet Bubble". Although firms invest in R&D in all sectors, the most R&D intensive firms are in computer software, electronic equipment and healthcare. *RD-MV* varies considerably across sectors, ranging from slightly over 100 percent in durable goods and telecommunication to 0.4 percent in utilities in the full sample period. R&D expenditures relative to total assets (*RD-A*) range from 18.9 percent in the healthcare to 1.7 in utilities. The highest (lowest) *RD-S* is 1,270.3 (5.0) percent in healthcare (non-durable goods). Following Chan et al. (2001), we assign R&D firms into high or low technology categories. Both *RD-MV* and *RD-A* confirm that high-tech firms invest more heavily in R&D than low-tech firms. Separating firms according to their *BM* ratios around the threshold of one, *RD-MV* is consistently higher for value stocks, whereas *RD-A* and *RD-S* are larger for growth stocks. We perform all our analysis using the three measures and find qualitatively similar results. We only report results for *RD-MV* because it is analogous to price multiples and can be readily applied to practical investment analysis. Also, it is not as volatile across time and sectors as *RD-S*, is not as persistent as *RD-A*, and is less likely to be influenced by creative accounting. [8]

Insert Table 1 about here

3. R&D and Future Returns

Firms with higher R&D intensity enjoy higher future stock returns. Our univariate portfolio analysis on an updated sample shows that this relationship cannot be explained by conventional

[8] Full details are of course available upon request.

asset pricing risk factors. However, augmentation of asset pricing models with an R&D factor explains almost all of the anomalous return and shows significant covariation among R&D stocks. This reflects either a premium for systematic risk or mispricing by investors; these conflicting explanations motivate our subsequent analysis.

3.1. The R&D Anomaly

To identify the cross-sectional relation between R&D and returns, we form decile portfolios based on R&D intensity (*RD-MV*) at the end of June each year. Monthly equally weighted returns are calculated for each portfolio, which is rebalanced annually. Table 2 presents the average returns and other characteristics of these portfolios. As expected, we find a monotonic increase in average returns with R&D intensity. The increase in return from 0.79 percent for Portfolio 1 to 2.48 percent for Portfolio 10 represents a statistically and economically significant return to the zero-cost spread portfolio (10-1). Our results are consistent with the literature (Chan et al., 2001; Li, 2011) and show that this relationship persists in our more recent sample.

Insert Table 2 about here

Adjusting for risk using the Carhart (1997) four-factor model, we find that increasing R&D intensity leads to a similar positive relation in abnormal returns and also a declining pattern in adjusted R^2. Together, these suggest that more R&D intensive stocks are more difficult to price, which is entirely consistent with the uncertain, opaque and intangible nature of R&D investments that make them difficult to value. The zero-cost spread portfolio (10-1) yields an alpha of 1.48 percent per month, significant at the 1% level, and has a particularly low R^2. This could indicate that R&D intensive stocks are exposed to risks that are not captured sufficiently by existing models.

The analysis shows that firms with higher R&D intensity are smaller, have higher book-to-market equity ratios and performed better in the recent past, indicative of size, value and momentum effects. As an alternative to the four-factor model, we apply the characteristics-based benchmark portfolio matching approach of Daniel et al. (1997) (henceforth DGTW) to control for these potential effects. [9] Table 2 shows that DGTW adjusted returns also increase monotonically with R&D intensity and that the (10-1) portfolio earns a significant return of 0.98 percent per month. As a final and very conservative check, we re-estimate the four-factor model using the DGTW adjusted returns and find an identical pattern.[10]

3.2. The R&D Factor

The findings above signal the presence of common variation in returns to R&D stocks that is not captured by existing risk factors. This motivates us to construct a factor mimicking portfolio related to R&D intensity to improve the pricing performance for these stocks. The addition of

[9] Specifically, at the beginning of each month between July of year t and June of year t+1, we sort R&D stocks into 125 characteristics-based benchmark portfolios along *SIZE*, *BM* and *RET(-12,-2)* dimensions. For each R&D stock, the characteristic-adjusted return (denoted as the DGTW return) is the difference between its raw return and its DGTW return provided by the characteristic portfolio that best matches that particular stock. We also apply independent benchmark portfolio sorts as a robustness check and find that both independent and dependent sorts generate remarkably similar results.

[10] As a final robustness check we use bivariate portfolio sorts to control for size, value and momentum effects. The R&D anomaly remains clearly evident so is not a manifestation of these effects. In addition, it is reassuring that patterns in returns are not driven by the role of the denominator in our *RD-MV* variable.

new candidate risk factors to conventional asset pricing models is not without problems. It is difficult to ensure that a new factor structure captures all relevant risks and it is possible for incremental factors to seem to perform well by explaining the anomaly even if the factor captures behavioural mispricing rather than a rational risk premium. The vast literature on detecting anomalies therefore proxies for risk factors by using the anomaly itself to create a portfolio that mimics the underlying risk factor (see the important works of Fama and French, 1993; Carhart, 1997; Pastor and Stambaugh, 2003). This serves an important purpose; it replicates a risk factor that is highly correlated with the anomaly such that stocks most exposed to the risk factor load most heavily onto this portfolio, and this occurs even when the latent risk factor is unobserved. We argue that under a rational factor risk explanation, R&D intensive firms share common properties that manifest as covariance with this factor, which represents systematic risk that should be compensated.

A factor-mimicking portfolio is constructed based on *RD-MV*. At the end of June of year *t*, we sort R&D stocks into three portfolios with breakpoints at the 30^{th} and 70^{th} percentiles. The R&D factor, which we denote *RD-HML*, is the equally weighted average return to the portfolio that buys the high and sells the low *RD-MV* portfolios.[11] Firms with higher R&D intensity will load more heavily onto this factor in time series regressions.

The summary statistics and pairwise correlations for the *RD-HML* factor and the Carhart (1997) four factors are reported in Table 3. Panel A shows that the R&D factor yields higher

---

[11] We also consider alternative approaches for the construction of this factor including bivariate sorts that adjust for size, book-to-market, momentum and DGTW returns. Our conclusions are not sensitive to these alternatives and details are available on request. We use the univariate version throughout because it generates the lowest adjusted R^2 when regressed on the Carhart (1997) four factors, implying that it contains the most information unexplained by these factors (Lamont et al., 2001).

average returns, lower standard deviation and a higher Sharpe ratio than other factors, suggesting that an investment strategy based on R&D intensity is superior to the passive strategy of tracking the market index or other strategies based on size, value or price momentum. More specifically, the *RD-HML* factor generates an average return of 1.15 percent per month, which is considerably higher than the 0.63, 0.28, 0.34 and 0.69 percent earned on *MKT_RF*, *SMB*, *HML* and *UMD* factors, respectively. The standard deviation of returns on *RD-HML* is 3.44 percent per month, which is lower than for *MKT_RF* and *UMD*. These properties combine to deliver an ex-post Sharpe ratio of 0.33 for *RD-HML*, which is more than double that for other factors, suggesting that an R&D strategy delivers more impressive performance than implied by higher returns.

Insert Table 3 about here

The pairwise correlations in Panel B suggest a relatively high correlation between *RD-HML* and *SMB* of 0.54 and the correlations with other factors are moderate. We follow Lamont et al. (2001) and investigate the extent of the variation in the *RD-HML* factor that is not captured by the other factors using the following time series regression:

$$RD-HML_t = \alpha + \beta_{MKT} MKT\_RF_t + \beta_{SMB} SMB_t + \beta_{HML} HML_t + \beta_{UMD} UMD_t + \varepsilon_t \ . \qquad (1)$$

The estimation results are shown in Panel C. In both single and four-factor versions, the intercept is positive and statistically significant, implying that *RD-HML* generates abnormal returns beyond those predicted by established factors. The significant positive loading on *SMB* reflects the descriptive patterns noted above that R&D-intensive firms tend to be small. The

R&D anomaly appears to be separate from value and momentum effects as evidenced by insignificant coefficients on the other factors. The low adjusted R^2 measures (28.5 percent for the four-factor model) confirm that a considerable amount of variation in the *RD-HML* factor remains unexplained by conventional risk factors.

Given the inability of the Carhart (1997) model to price R&D portfolios appropriately, we evaluate whether including an R&D factor improves this pricing performance using the following R&D-augmented model.[12]

$$R_t - RF_t = \alpha + \beta_{MKT} MKT\_RF_t + \beta_{SMB} SMB_t + \beta_{HML} HML_t + \beta_{UMD} UMD_t + \beta_{RD-HML} RD - HML_t + \varepsilon_t \,. \qquad (2)$$

Significant loadings on the *RD-HML* factor would show evidence of the return covariation attributable to *RD-MV*. We also expect the inclusion of *RD-HML* to eliminate alpha and increase adjusted R^2.

The estimation results are presented in Table 4. Panel A shows statistically significant loadings on *RD-HML* that increase with *RD-MV* showing that return covariance increases with R&D intensity. The alphas are indistinguishable from zero confirming that the *RD-HML* factor captures the R&D anomaly. Adjusted R^2 values are larger compared to Table 2, and for the higher *RD-MV* portfolios in particular, emphasising the improved pricing performance of the augmented model. The final column shows that the return of the (10-1) spread portfolio also loads heavily on *RD-HML* as we would expect. Although its alpha is significant, its value has reduced from 1.48 to only 0.3 percent per month, and the adjusted R^2 has increased substantially.

[12] We also estimate equation (2) adding the Pastor and Stambough (2003) liquidity risk factor and our conclusions remain unchanged. Further details are available upon request.

In terms of economic significance, the R&D factor commands a 1.37 percent per month higher expected return on average.[13]

Insert Table 4 about here

The Gibbons et al. (1989) (GRS) test of whether alphas are jointly zero is reported in Panel B. Whilst the GRS F-test rejects the null hypothesis, it is reassuring that both the test statistic and average alpha decline sharply after including *RD-HML*. In addition, average adjusted R^2 increases by 5 percentage points and the Sharpe ratio of the alphas declines from 0.42 to 0.24. This evidence suggests that the *RD-HML* factor captures the covariation in returns on R&D intensive stocks and thus eliminates the R&D anomaly.

3.3. The R&D Increase Anomaly

The success of the R&D factor in pricing R&D stocks raises the interesting question of whether the factor also explains longer term abnormal returns following R&D increases. Eberhart et al. (2004) document significant long-term abnormal stock returns in the 60 months following R&D expenditure increases, which they interpret as evidence of investors' undervaluation of the benefits of R&D investments. However, as shown above, their use of a four-factor model may not be the most appropriate specification for R&D stocks. This has important implications for the explanation underlying the anomaly, the interpretation of results and risk adjustment methods.

[13] This is calculated by multiplying the loading on *RD-HML* for the (10-1) portfolio by the average monthly return of the *RD-HML* factor (1.19×1.15=1.37). This represents 1.37/1.69=81.1 percent of the *RD-MV* premium on average.

Whilst the prevailing interpretation is one of mispricing, abnormal returns found to reflect an omitted R&D factor give rise to the possibility that the anomaly could be explained by systematic risk. Evidence in support of a risk explanation would suggest that the R&D increase anomaly results from an inadequate adjustment for risk. As a first step towards investigating the important risk versus mispricing debate, we test the extent to which the abnormal returns documented by Eberhart et al. (2004) are captured by our R&D factor.

We follow Eberhart et al. (2004) very closely in selecting only firms experiencing economically significant R&D increases.[14] Stocks are pooled into a portfolio whenever an observation is within 60-months after the R&D increase and the monthly portfolio return is calculated in calendar time.[15] Our sample contains 10,326 firm-year observations of significant R&D increases from 3,448 firms. The equally and value weighted average portfolio returns are the dependent variables when estimating the four-factor and R&D augmented models.

Table 5 replicates quantitatively similar results to those of Eberhart et al. (2004). In Row (1) using equally weighted returns, an abnormal return of 0.80 percent per month is significant at the 1 percent level and confirms the anomaly. This becomes insignificant in Row (2) when incorporating the *RD-HML* factor into the model. The estimated loading on the *RD-HML* factor is 0.7 and is highly significant at the 1 percent level. This is synonymous with higher expected returns of 0.80 percent on average, which is comparable to alpha in Row (1). The adjusted R^2 increases by 6 percentage points reflecting the incremental explanatory power of the additional

[14] These firms must have at least 5 percent *RD-A* and must have R&D intensity increasing by at least 5 percent in any year. This increase must be across R&D expenditure, *RD-A* and *RD-S* simultaneously.

[15] We skip the three months immediately after the reporting of the R&D increase becomes publicly available to allow sufficient time for the accounting information to be incorporated into prices, and require firms to have a complete 60 months of returns after the increase. Relaxing this 60 months requirement does not affect our results.

factor. These findings combine to suggest that the seemingly abnormal returns from R&D increases that are not explained by the four-factor model are fully subsumed by the R&D factor. Rows (3) and (4) show the results from value-weighted returns, which are qualitatively similar. Specifically, alpha is positive and significant in Row (3) becoming insignificant in Row (4) when the R&D factor is included and the loading on *RD-HML* is again significant.[16]

Insert Table 5 about here

The evidence reported in this section demonstrates that R&D anomalies are eliminated when including an R&D factor mimicking portfolio in asset pricing models. While unable to eliminate a mispricing explanation, this evidence proposes the possibility of a rational risk factor pricing explanation for these anomalies. The remaining sections address this debate.

4. Evidence for a Risk Explanation

The prior section clearly demonstrates higher returns to R&D intensive firms that are not explained by size, book-to-market equity and momentum. Furthermore, loadings on an R&D factor increase with R&D intensity and explain these returns, demonstrating covariance with this additional factor. In a frictionless, rational, multifactor asset pricing framework, these higher returns are compensation for systematic risk, as measured by the R&D factor loading, but which are not captured by prevailing risk factors. This covariance risk could arise from firms sharing common risk characteristics that are related to the underlying nature of R&D, such as being

[16] As an even more stringent robustness test, we also estimate using DGTW adjusted returns that are averaged using both equal and value weighting schemes and find entirely consistent results, although these are not reported.

option-like (Berk et al., 2004), long term, inflexible (Li, 2011), and a source of business risk (Ciftci et al., 2011). Furthermore, R&D investment contributes to other well known risks in firms including lower liquidation value of intangibility and specificity of assets (Shleifer and Vishny, 1992; Garlappi et al., 2008), financial constraints (Li, 2011), financial distress (Opler and Titman, 1994; George and Hwang, 2010) and information asymmetry (Aboody and Lev, 2000). In empirical asset pricing, if existing risk factors fail to capture these traits explicitly, it is unlikely that risk in R&D intensive firms will be measured accurately. Another interpretation of our results is mispricing caused by investors who are unable to value accurately the benefits and risks of R&D investment, due to its intangible and opaque nature, the uncertainty of its success, its conservative treatment in financial reporting regulations and investor attention constraints. This leads to undervaluation and higher subsequent stock returns when the mispricing is corrected.

To distinguish between risk and mispricing interpretations, we adopt two testing strategies. First, we examine the correction of mispricing after waves of investment sentiment, when R&D stocks that are difficult to value may be more exposed to mispricing. Second, we test the relative importance of the R&D factor loading versus the mispricing characteristic for explaining the cross section of expected stock returns.

4.1. Investor Sentiment

In classical finance, the cross section of expected returns depends only on exposure to systematic risk and the presence of arbitrageurs eliminates any deviation from equilibrium prices caused by irrational investors. However, irrational speculation and constraints to arbitrage give conditions for mispricing to persist. Sentiment measures the propensity for investors to speculate and Baker and Wurgler (2006) show that shifts in sentiment can create cross-sectional differences in returns for certain stocks. Theoretically, stocks that are more sensitive to speculative demand, which are more difficult to value, are also more costly to arbitrage. The demand and arbitrage mechanisms are reinforcing so that these stocks are more affected by shifts in sentiment. The empirical evidence provided by Baker and Wurgler (2006) supports this hypothesis with largest sentiment effects observed for younger, smaller, highly volatile, low dividend and unprofitable firms. R&D stocks share these characteristics and are hard to value, however Baker and Wurgler (2006) do not find the same convincing evidence of sentiment effects for R&D stocks.

As Baker and Wurgler (2006) note, mispricing is hard to identify. They prefer to detect evidence of the correction of mispricing after sentiment states that cannot be explained by compensation for systematic risk. At the end of high (low) sentiment states, hard-to-value and costly-to-arbitrage stocks, sharing similar characteristics with R&D intensive firms, are subject to more (less) speculative demand and are mispriced. The subsequent correction of the mispricing generates lower (higher) returns, the extent of the returns being proportional to the degree of subjectivity of the firm's value. Therefore, a mispricing explanation of the R&D anomaly implies monotonically declining (increasing) returns to *RD-MV* decile portfolios following high (low) sentiment states. Furthermore, the magnitude of the spread across decile portfolios should be symmetric across sentiment states. However, Baker and Wurgler (2006) show increasing returns with R&D intensity following both sentiment states. This absence of a correction in returns following the high state is inconsistent with mispricing. Meanwhile, the higher returns following low sentiment states is consistent with both undervaluation and risk explanations, but may also be driven by a failure to control for other characteristics associated with sentiment effects. To address these concerns, we extend the analysis by controlling explicitly and more appropriately for characteristics, conventional risk factors and size.

We use the Baker and Wurgler (2006) sentiment index proxy, defining high and low beginning of period sentiment states according to its sample median. Subsequent equally weighted returns to R&D decile portfolios are then averaged within sentiment states.[17] Panel A of Table 6 replicates the analysis of Baker and Wurgler (2006) for unadjusted portfolio returns. We confirm a monotonic increase in returns across R&D deciles following both sentiment states. Difficult to value stocks subject to sentiment effects should be overvalued in high sentiment states with returns correcting in subsequent periods. Our pattern in returns is entirely opposite to this prediction and so is inconsistent with mispricing. The monotonic pattern in returns following low states gives rise to competing interpretations. The mispricing explanation is that the most R&D intensive stocks will be most undervalued in low sentiment states, with subsequent returns correcting relatively higher. Alternatively, increasing returns along decile portfolios are entirely consistent with compensation for systematic risk shared by R&D stocks.

Insert Table 6 about here

The information held in the average returns across the deciles allows us to address this debate. For each decile, the spread between portfolio returns following high and low states is conditional on R&D intensity. Under a risk explanation, these conditional differences are

[17] Specifically, sentiment state is measured at the end of year *t-1* and subsequent R&D portfolio returns are calculated from July of year *t* to June of year *t+1*. This maintains consistency with our earlier methods. Calculating returns for the twelve months from January of year *t* instead makes no difference to our results.

expected to display a flat pattern across R&D deciles since sentiment should not affect the R&D-return relation. We find negative spreads that decrease across R&D deciles. This suggests that high R&D stocks are affected disproportionately by sentiment, which intimates mispricing. However, an important concern is that univariate portfolio sorts fail to control for other characteristics that cause mispricing, but which are also shared by R&D stocks, the most notable being size. Indeed, Baker and Wurgler (2006) show the same prominent size effect only after low sentiment states, implying correction for mispricing. Our results in Panel A may therefore be driven by a similar size effect, especially given its role in constructing our sorting characteristic. The remaining panels of Table 6 extend the analysis to distinguish between these competing interpretations.

Panel B reports average DGTW adjusted returns on the decile portfolios, which control for size, book-to-market and momentum via characteristic-matched benchmark portfolios. Consistent with Panel A, DGTW returns increase with R&D following both sentiment states. For high sentiment states, this is again inconsistent with the correction of mispricing. Across deciles, the conditional pattern of returns (high – low) is far less pronounced, but appears to be sloping downwards albeit much less steeply. In confirmation, the conditional return difference for the (10-1) portfolio reduces from 0.76 percent in Panel A to 0.44 percent in Panel B. This suggests that any correction for mispricing following the low state is dampened when controlling for characteristics, which is likely the result of controlling for mispricing caused by size. As an alternative approach, Panel C reports average abnormal returns for decile portfolios calculated using the Carhart (1997) four-factor model and documents similar patterns of increasing returns. More importantly, the pattern in conditional differences across deciles is barely noticeable and for the (10-1) portfolio it reduces to only 0.14 percent.

Our findings in Panel B and C suggest that the monotonic increase in returns following low sentiment states as well as the pattern of increasing conditional differences with R&D reflects size related mispricing. This is partly a consequence of our measure of R&D intensity (*RD-MV*) being a size related multiple. This means that parts of the variations in *RD-MV* and the corresponding decile portfolio returns are related to size. It is therefore imperative that we control for size explicitly using bivariate sorted quintile portfolios. Panel D confirms strong R&D and size effects in the 25 portfolios. However, after adjusting for size, returns increase monotonically with R&D following both sentiment states and both 5-1 spread portfolio returns (0.71 and 0.95 percent for high and low states) are significant, demonstrating the persistence of the R&D effect. Although size adjusted returns are higher following the low sentiment state, the conditional differences between high and low sentiment states show no apparent pattern. The conditional difference in the 5-1 spread portfolio is only 0.24 percent per month. These findings show that any effects of mispricing in R&D portfolios are negligible when controlling for size appropriately.

Panel D also reports the four-factor alpha on size adjusted R&D quintiles separated by sentiment states. Controlling for size and conventional risk factors results in a similar increasing pattern in abnormal returns and significant 5-1 portfolio returns, confirming the familiar R&D effect. However, more important is the change in the patterns in conditional differences. These revert from negative to positive and display no discernible increasing pattern across quintiles, and the difference across states in the 5-1 portfolios almost vanish at 0.02 percent. Together, these findings show that there is no remaining sentiment effect in R&D, that the weak results in Panels A to C are annihilated when we make more appropriate adjustments for size and that the residual R&D-return relation is not consistent with mispricing.

Finally, we investigate the impact on our results when augmenting the four-factor model with an R&D factor. Controlling for size-related mispricing creates closer alignment of the cross-sectional R&D effect across sentiment states, which does not support a mispricing explanation. Removing mispricing effects leaves returns that are likely determined by systematic R&D risk. A required condition for a risk explanation is the ability of a factor mimicking portfolio to fully explain these residual returns. Our evidence in Panel D supports this requirement. Alphas from the augmented model are mostly insignificant and loadings on the R&D factor are large and significant showing that any abnormal returns remaining after removing size-related effects are fully explained by the R&D factor. Furthermore, the size adjusted 5-1 spread portfolios returns become indistinguishable from zero showing that this factor captures the cross-sectional R&D effect, whilst conditional differences in alphas across states offer identical conclusions to those above that are inconsistent with mispricing. In summary, Table 6 provides new evidence that eliminates support for a mispricing explanation of the R&D anomaly in favour of a more plausible systematic risk explanation.

4.2. Fama and MacBeth (1973) Cross-sectional Regressions

An alternative explanation for the R&D anomaly is compensation for systematic risk. A sufficient condition for this requires R&D portfolio returns to load onto an R&D factor and for this factor to explain the cross section of abnormal returns. More importantly, a convincing risk explanation requires evidence for the importance of this factor loading in predicting future cross-sectional returns. Critically, the explanatory power must come from this covariance risk at the expense of the R&D characteristic. To test this we follow Hirshleifer et al. (2012) by including both the R&D factor loading and the R&D characteristic, along with other control variables, in firm-level Fama and MacBeth (1973) regressions. Evidence for a risk explanation should find R&D intensity to be subsumed by covariance risk.

We estimate cross-sectional regressions of monthly stock returns on pre-estimated *RD-HML* factor loadings, the *RD-MV* characteristic, pre-estimated factor loadings on *SMB*, *HML* and *UMD*, and a number of other firm characteristics. All factor loadings are pre-estimated for each of 25 size-*RD-MV* portfolios over a 60-month window using the R&D-augmented model. Estimated factor loadings are then assigned to each stock constituent of that portfolio to alleviate measurement error. Under the Fama and MacBeth (1973) procedure, the pre-estimated loadings are used to explain stock returns in the subsequent month with the window rolled forward at monthly iterations.[18]

We are careful to control for a number of firm characteristics that are known to determine expected returns, including log size (*Ln(SIZE)*) (Banz, 1981), log book-to-market equity ratio (*Ln(BM)*) (Fama and French, 1992), intermediate past return (*RET(-12,-2)*) (Jegadeesh and Titman, 1993), log turnover ratio (*Ln(TURN)*) (Chordia et al., 2001), stock illiquidity (*ILLIQ*) (Amihud, 2002) and log idiosyncratic volatility (*Ln(IVol*) (Ang et al., 2006).[19] As further controls, we also include the pre-estimated factor loadings on *MKT_RF*, *SMB*, *HML* and *UMD* and industry dummies according to the Fama and French 49-industry classification. The time series averages of the coefficient estimates, time series Newey and West (1987) robust *t*-statistics, and average R^2 values from the cross-sectional regressions are reported in Table 7.

Insert Table 7 about here

[18] We analyse the sensitivity of results to 36- and 48-month windows and find that all our results are robust.

[19] Detailed definitions of the independent variables are provided in Table A1 of the Appendix.

In column (1), the R&D factor loading is positively and significantly priced, consistent with a rational risk factor pricing explanation of the R&D anomaly. In column (2), *RD-MV* is included on its own and also has a significant and positive coefficient, which is entirely expected since the characteristic identifies the anomaly. Column (3) provides the direct test of the source of the R&D anomaly. When including both the *RD-HML* factor loading and *RD-MV* characteristic, the factor loading is significantly priced whilst the explanatory power of the characteristic is subsumed, lending strong support to the risk explanation. All remaining columns of Table 7 add control variables and this conclusion remains strong and robust. This significant price of R&D risk is also consistent with an economic interpretation. Using an estimated price of R&D risk of 0.7% (from column (7)), we measure the extent to which the 0.98% spread in ex-post DGTW returns across R&D deciles (Table 2) can be attributed to exposure to this risk. The spread in estimated R&D factor loadings between Portfolio 10 and Portfolio 1 is 1.09-(-0.10) = 1.19, which corresponds to a difference in average returns of 0.83% (=0.7%×1.19). This represents a large proportion of the 0.98% spread in DGTW returns showing that the majority of R&D anomalous returns reflect compensation for systematic R&D risk captured by the R&D factor.

For completeness, the estimated coefficients on other explanatory variables in Table 7 are consistent with prior literature, extending support to our specification and results. There is a significant negative size effect and positive but insignificant value and momentum effects. Consistent with Chordia et al. (2001) higher turnover predicts lower returns; more illiquid stocks show higher expected returns in accordance with Amihud (2002); and higher idiosyncratic volatility generates lower returns as shown by Ang et al. (2006). When incorporating pre-estimated factor loadings, *MKT_RF*, *SMB*, and *HML* show coefficients with the correct sign, whilst that on *UMD* is ambiguous, and all are indistinguishable from zero.

In conclusion, analysis of investor sentiment effects on the cross section of expected returns shows evidence that refutes mispricing as an explanation for R&D returns. In a direct test in firm level Fama and MacBeth (1973) regressions, R&D covariance risk subsumes the R&D characteristic and is statistically significant and economically important. Together, this evidence lends more convincing support for a risk explanation for the R&D anomaly, which motivates us to examine the determinants of firms' exposure to this risk.

5. Firm Level Determinants of R&D Risk

There are very few studies investigating R&D risk; some existing research examines whether R&D investment contributes to firm specific business risk and its components, as measured by earnings and cash flow volatility, information risk, competition risk and operations risk.[20] R&D expenditure is associated with higher firm specific risk according to these measures, which is inferred as a risk explanation for the anomaly. In this paper, we adopt a very different approach. We argue, supported by our empirical evidence above, that R&D risk is partly systematic, motivating us to measure it explicitly and examine its determinants. Most closely related to our approach, Lev and Sougiannis (1999) and Chambers et al. (2002) find support for R&D risk, but without applying an R&D factor they are unable to measure firms' exposure to this R&D-related systematic risk. We argue that this systematic risk is driven by two factors: (i) the nature of R&D investment and (ii) risk related characteristics that are more prominent in R&D intensive firms. Investigating these determinants provides an economic rationale for the R&D anomaly, enriches

[20] See Shi (2003), Ciftci et al. (2011) and Lev et al. (2012).

interpretations for portfolio management and identifies alternative mechanisms through which R&D investment impacts firm value. We extend the literature to examine the firm level characteristics that predict R&D factor loadings.

We estimate the following regression to investigate the extent to which the R&D factor loading can be predicted by firm characteristics

$$\hat{\beta}_{RD-HML,i,t} = \alpha_0 + \sum_{k=1}^{K} \alpha_k x_{k,i,t-1} + \mu_h + \varepsilon_{i,t} \, , \tag{3}$$

where $\hat{\beta}_{RD-HML,i,t}$ is the estimated loading on *RD-HML* for firm *i* obtained from Equation (2), $x_{k,i}$ is a set of *K* firm characteristics and μ_h are industry fixed effects. Specifically, we estimate a firm's R&D factor loading using a 36-month window from July of year *t* to June *t*+3 with a minimum of ten monthly observations. This is predicted by year *t*-1 accounting variables.[21]

Equation (3) is estimated on 68,025 firm-year observations, which comprises 6,871 firms with positive R&D investments. To help reduce the problem of outliers, variables are winsorised at the 1st and 99th percentiles. Following Campbell et al. (2010), all variables are market adjusted, by subtracting year specific means. Independent variables are then normalised to have unit variance. Each coefficient is then easily interpreted as the effect of a standard deviation change in the independent variable on the future R&D factor loading. The reciprocals of the number of cross-sectional observations in each year are used as weights in the estimation of the pooled weighted least squares (WLS) regressions. This ensures that each cross section is applied an

[21] We test the sensitivity of our results to 12-month, 24-month and 60-month windows and find qualitatively similar conclusion.

equal weight, which is similar in spirit to a Fama and MacBeth (1973) approach. Robust standard errors are clustered at the firm level.

Explanatory variables are constructed from accounting data at fiscal year-end *t*-1 and detailed definitions of these variables are provided in Table A1 of the Appendix. We include R&D intensity (*RD-MV*) throughout to confirm the significant role of R&D investment on return covariation. In addition, controlling for R&D intensity helps to evaluate the contribution of other characteristics to R&D-related systematic risk. The important variables measuring these characteristics relate to shareholder recovery, financial constraints, distress and information asymmetry.

Property, plant and equipment to total assets (*PPE-A*) measures the tangibility of assets. Garlappi et al. (2008) argue that an increase in asset tangibility results in a reduction in liquidation cost such that the transparency of value boosts the recovery rate of shareholders in the event of distress. Since R&D investment contributes to firms' intangible assets, the opacity of its value raises the uncertainty of shareholder recovery. An R&D firm with relatively less tangible assets should therefore load more heavily on to the *RD-HML* factor.

Shleifer and Vishny (1992) argue that firms suffer "fire sale" discounts during distress. Furthermore, the liquidation value of a firm's asset in distress is related to their specificity (Acharya et al., 2011). We follow Garlappi et al. (2008) in using the Herfindahl-Hirschman index (*HHI*), which measures the degree of concentration of an industry, as a proxy for asset specificity. R&D investment is likely to create highly specific assets, which exposes them to higher liquidation discounts. Such liquidation fire sales occur in distress, which is more likely for R&D intensive firms as R&D projects are highly uncertain. Moreover, Opler and Titman (1994) argue that firms with high levels of investment in R&D suffer the most in financial distress. We adopt the popular Altman (1968) *Z-score* to measure the likelihood of distress. We test the extent to which asset specificity and distress characteristics predict R&D-related systematic risk.

The third variable measures firms' financial constraint. As documented by Li (2011), R&D investment is inflexible. If firms rely on external financing, this inflexibility imposes risks on the firm when this financing is constrained. More specifically, R&D projects are more likely to be suspended or scrapped the more financially constrained the firm is. Financial constraint and R&D intensity therefore display a seemingly symbiotic relationship in both risk and return. Following Li (2011), we use the Kaplan and Zingales (1997) *KZ index* and firm size (*Ln(MCAP)*) to measure financial constraint and examine their abilities to predict a firm's R&D factor loading.

Finally, analyst forecast dispersion (*Ln(Disp)*) is an established proxy for information asymmetry (Zhang, 2006). The intangible and opaque nature of R&D increases both the difficulty in evaluating the benefits of R&D and information asymmetry. R&D intensive firms are therefore more likely to experience wider analyst forecast dispersions, which we expect to predict their loading on the R&D factor. Further to these characteristics, we include profitability (*ROA*), capital investment to total assets (*INV-A*) and industry fixed effects as control variables.

Table 8 presents the summary statistics for the estimated R&D factor loadings and firm characteristics in Panel A and the estimation results of the WLS regression in Panel B. We note that the sample size fluctuates according to the availability of data on variables, but emphasise that our results are not sensitive to this variation. The coefficients on *RD-MV* are positive and significant at the 1 percent level across all specifications. This confirms the role of R&D expenditure in predicting the factor loading.

Insert Table 8 about here

Column (2) investigates asset tangibility. The negative and significant coefficient shows that lower asset tangibility predicts a higher R&D factor loading. Column (3) includes the Herfindahl-Hirschman index (*HHI*) that measures industry concentration and proxies for asset specificity. The more specific a firm's assets are, the higher the loading on the R&D factor as shown by the positive and significant coefficient. This suggests that higher liquidation cost and a corresponding lower recovery rate for shareholders as measured by asset intangibility and specificity influence return covariation among R&D stocks.

The Altman (1968) *Z-score* in column (4) has a negative and significant coefficient, which means that R&D firms with higher financial distress risk have more exposure to the R&D risk factor. This suggests that, even after controlling for R&D intensity, the R&D factor loading is predicted by financial distress risk. This is consistent with George and Hwang (2010) who argue that financial distress is a priced systematic risk. Our evidence shows that the R&D factor captures at least part of this systematic distress risk among R&D stocks.

Financial constraint is measured by the *KZ index* (column (5)) and market value (included throughout), following Li (2011). We find that the *KZ index* has no effect, but that size has a significantly negative coefficient. To the extent that smaller firms are more constrained, the risk relating to financial constraint appears to contribute to the return comovement driven by R&D, consistent with Li (2011). In column (6) we investigate the role of information asymmetry and find a strong and significant effect. Wider analyst forecast dispersion predicts a higher R&D factor loading, suggesting that asymmetric information may be a source of systematic risk captured by exposure to the R&D factor.

Return on assets and capital investment are included as control variables in all specifications. *ROA* shows a strong negative effect on future loadings in all columns. This suggests that the returns to less profitable firms are more correlated with the R&D factor. The extent of the premium on R&D stock returns therefore partly reflects poorer operating performance and its associated risks. *INV-A* shows a positive coefficient, but this not significant in all models. In column (7), when we include all variables, our results remain similar and robust and the adjusted R^2 increases.

The results of previous sub-sections show that rational investors require a premium for holding stocks with higher R&D risk. This sub-section shows that the R&D risk exposure of firms is a function of firm characteristics including R&D intensity, asset tangibility and specificity, financial distress, financial constraint, and information asymmetry. Since R&D intensity is controlled for throughout, these additional common characteristics shared by R&D firms represent further contributors to the R&D loading. We argue that these variables have risk-based motivations. Financial distress risk is already recognised as a priced risk factor (George and Hwang, 2010) and our findings confirm that R&D exposure is partly driven by this risk. It is interesting to note that the remaining characteristics also capture some degree of financial distress, particularly those relating to liquidation costs and shareholder recovery. Our evidence shows that these indicators of distress risk determine firms' R&D factor loading, which are subsequently priced in the cross section of expected returns. This confirms that at least part of the R&D-return relation is driven by compensation for distress risk.

6. Robustness Checks

We perform a number of robustness checks to verify our results. These are detailed below with only the important results presented in the online Appendix.

We first confirm the R&D anomaly for alternative measures of R&D intensity: R&D expenditure relative to Total Assets (*RD-A*) or Sales (*RD-S*). Consistent with the extant literature, R&D intensive stocks earn higher abnormal returns irrespective of the measure. Since the benefits of R&D may accrue over long horizons, the stock of R&D capital may capture this accrued value better than expenditure. We follow Chan et al. (2001) in using the perpetual inventory method to measure the stock of R&D capital as follows:

$$RDC_{i,t} = RD_{i,t} + 0.8 \times RD_{i,t-1} + 0.6 \times RD_{i,t-2} + 0.4 \times RD_{i,t-3} + 0.2 \times RD_{i,t-4}. \tag{4}$$

Using *RDC-MV*, we confirm the anomaly (Table IA.1), replicate the success of an augmented model that uses an *RDC-MV-HML* factor, and show that its loading explains future cross-sectional returns (Tables IA.3 to IA.5). Thus, our conclusions are not sensitive to the choice of measure.

Next, we test whether our results are affected by a number of empirical issues relating to portfolio construction. These include alternative techniques for sorting portfolios, weighting schemes, factor construction and return adjustment. First, to confirm the robustness of the R&D anomaly, we adopt bivariate sorts to control for size, book-to-market equity and momentum (Table IA.2). The results from bivariate sorts are robust to both equal and value weighting schemes and independent and dependent sorts (not reported). Second, the R&D anomaly is confirmed for value weighted portfolios, although the magnitude is less extreme, but this is entirely expected given that R&D intensive stocks tend to be smaller (Table IA.1). Third, the success of the *RD-HML* factor is robust to variations in its method of construction. These include adjustments for size, size and book-to-market, DGTW characteristics matching, independent and dependent portfolio sorting, value weighting and the use of alternative breakpoints (not reported). Our results remain consistent throughout.

Our conclusions rely on appropriate risk adjustment of returns. For all estimations using portfolio returns we check whether our findings are robust to the inclusion of the Pastor and Stambaugh (2003) liquidity factor and find confirmatory evidence (not reported). For firm level Fama and MacBeth (1973) regressions, we check alternative specifications by including additional risk controls. These comprise liquidity shocks (Bali et al., 2014), maximum daily returns in a month (Bali et al., 2011), short-term return reversal (Jegadeesh, 1990), and analyst earnings forecast dispersion (Zhang, 2006) (Table IA.5). The significance of the R&D factor loading, and hence the support for a rational risk pricing explanation for the R&D anomaly, remain unchanged following these inclusions. Indeed, we find further support for a risk explanation when separating these regressions between good and bad states measured by GDP growth (not reported), which is entirely consistent with Lev and Sougiannis (1999).

Finally, we perform a number of checks on our analysis of determinants of firms' R&D factor loadings. We experiment with 12 and 24 month windows for estimating stock exposures to factors and the results are insensitive. Our results are also robust to alternative estimation methods, such as OLS and the Fama and MacBeth (1973) procedure. More importantly, although we already include characteristics that directly or indirectly measure risk of financial distress, we investigate the role of the book-to-market ratio in this context (Table IA.6). We find its coefficient is not significant, whilst existing characteristics maintain their importance. This shows the book-to-market equity ratio, which is often interpreted as associated with distress, does not help to explain stocks' exposures to R&D risk. Rather, we confirm that an element of R&D risk exposure is related to financial distress, which we capture more explicitly by shareholder recovery, liquidation costs, likelihood of failure and information asymmetry.

7. Conclusion

There is considerable evidence of a significant positive relation between R&D investment and subsequent stock returns, which is not explained by empirical asset pricing models. There is much less literature devoted to understanding the cause and implications of this anomaly. A common explanation based on simple patterns in abnormal returns is mispricing, which argues that investor attention constraints and conservative accounting treatment of R&D expenditure mean that investors are unable to recognise and price accurately the benefits of R&D. The positive R&D-stock relation over intermediate horizons represents correction of mispricing as tangible benefits of R&D are realised. Such behavioural mispricing and information asymmetry explanations demand greater disclosure and changes to accounting practices to help investors evaluate the relative benefits of R&D and managers to allocate resources more efficiently.

The alternative explanation is that R&D related expected returns represent entirely rational compensation for systematic risk. The persistent evidence detecting the anomaly implies that conventional empirical asset pricing models have yet to identify and proxy for such a latent risk factor. Support for a risk explanation is important for understanding risk pricing and the cost of capital in R&D stock and improving portfolio allocation and performance measurement. We contribute to this debate by providing arguments and new empirical support for this risk explanation.

More specifically, we contribute to the literature by pursuing three mutually reinforcing tests. First, we extend the analysis of Baker and Wurgler (2006) to test for correction of mispricing in R&D stocks via the cross-sectional effects of investor sentiment. We find that any evidence of mispricing is due to size rather than R&D. Controlling appropriately for size annihilates any evidence for mispricing and refutes this explanation for the anomaly. Second, whilst we meet the necessary condition for a systematic risk explanation of covariance in returns in association with R&D intensity, the risk versus mispricing debate revolves around detecting whether this covariance successfully prices the cross section of expected returns at the expense of the mispricing characteristic. Our evidence confirms that R&D covariance risk is priced and subsumes the R&D mispricing characteristic.

The economic rationale for covariance risk argues that R&D intensive firms share a number of common risk characteristics that relate to the nature of R&D investment and cause comovement in stock returns. For example, R&D projects are long term, irreversible, inflexible and have uncertain outcomes which drive higher business risk, and are a major source of information asymmetry that increases the cost of capital. Also, R&D firms are likely exposed to financial constraints, financial distress risk and lower shareholder recovery. Our final and more innovative contribution investigates whether these characteristics determine firm's future exposure to R&D covariance risk. We find that liquidation costs, shareholder recovery, financial distress risk, and information asymmetry contribute significantly to firms' future R&D factor loadings, which confirms the common risk characteristics shared to some extent by all R&D firms. In conclusion, we show strong evidence in support of a risk explanation for the R&D anomaly, which has at least some foundation in financial distress risk and information uncertainty.

References

Aboody, D. and Lev, B. (2000) Information Asymmetry, R&D, and Insider Gain. *Journal of Finance* 55, 2747-2766.

Acharya, V., Sundaram, R. K., and John, K. (2011) Cross-Country Variations in Capital-Structures: The Role of Bankruptcy Codes. *Journal of Financial Intermediation* 20, 25-54.

Al-Hourani, A., Pope, P. F., and Stark, A. W. (2003) Research and Development Activity and Expected Returns in teh Uinted Kingdom. *European Economic Review* 7, 27-46.

Altman, E. I. (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* 23, 589-609.

Amihud, Y. (2002) Illiquidity and Stock Returns: Cross-Section and Time-Series Effects. *Journal of Financial Markets* 5, 31-56.

Ang, A., Hodrick, R., Xing, Y. and Zhang, X. (2006) The Cross-Section of Volatility and Expected Returns. *Journal of Finance* 61, 259-299.

Armstrong, C. S., Core, J. E., Taylor, D. J. and Verrecchia, R. E. (2010) When Does Information Asymmetry Affect the Cost of Capital? *Journal of Accounting Research* 49, 1-40.

Bali, T. G., Cakici, N. and Whitelaw, R. (2011) Maxing Out: Stocks as Lotteries and the Cross-Section of Expected Returns. *Journal of Financial Economics* 99, 427-446.

Bali, T. G., Peng, L., Shen, Y. And Tang, Y. (2014) Liquidity Shocks and Stock Market Reactions. *The Review of Financial Studies* 27, 1434-1485.

Baker, M. and Wurgler, J. (2006) Investor Sentiment and the Cross-Section of Stock Returns. *Journal of Finance* 61, 1645-1680.

Banz, R. W. (1981) The Relationship between Return and Market Value of Common Stocks. *Journal of Financial Economics* 9, 3-18.

Barber, B. and Odean, T. (2008) All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies* 21, 785-818.

Berk, J. B., Green, R. C. and Naik, V. (2004) Valuation and Return Dynamics of New Ventures. *The Review of Financial Studies* 17, 1-35.

Carhart, M. (1997) On Persistence in Mutual Fund Performance. *Journal of Finance* 52, 57-82.

Campbell, J. Y., Polk, C., and Vuolteenaho, T. (2010) Growth or Glamour? Fundamentals and Systematic Risk in Stock Returns. *The Review of Financial Studies* 23, 305-344.

Chambers, D., Jennings, R., and Thompson II, R. B. (2002) Excess Returns to R&D Intensive Firms. *The Review of Accounting Studies* 7, 133-158.

Chan, S. H., Martin, J. D., and Kensinger J. W. (1990) Corporate Research and Development Expenditures and Share Value. *Journal of Financial Economics* 26, 255-276.

Chan, L. K. C., Lakonishok, J., and Sougiannis, T. (2001) The Stock Market Valuation of Research and Development Expenditures. *Journal of Finance* 56, 2431-2456.

Chordia, T., Subrahmanyam, A., and Anshuman, V. R. (2001) Trading Activity and Expected Stock Returns. *Journal of Financial Economics* 59, 3-32.

Ciftci, M., Lev, B., and Radhakrishnan, S. (2011) Is Research and Development Mispriced or Properly Risk Adjusted? *Journal of Accounting, Auditing & Finance* 26, 81-116.

Daniel, K. and Titman, S. (1997) Evidence on the Characteristics of Cross-Sectional Variation in Common Stock Returns. *Journal of Finance* 52, 1-33.

Daniel, K. Grinblatt, M., Titman, S., and Wermers, R. (1997) Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. *Journal of Finance* 52, 1035-1058.

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

Easley, D. and O'Hara, M. (2004) Information and the Cost of Capital. *Journal of Finance* 59, 1553-1583.

Eberhart, A. C., Maxwell, W. F., and Siddique, A. R. (2004) An Examination of Long-Term Abnormal Stock Returns and Operating Performance Following R&D Increases. *Journal of Finance* 59, 623-650.

Fama, E. F. and French, K. R. (1992) The Cross Section of Expected Stock Returns. *Journal of Finance* 47, 427-465.

Fama, E. F. and French, K. R. (1993) Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33, 3-56.

Fama, E. F. and French, K. R. (2008) Dissecting Anomalies. *Journal of Finance* 63, 1653-1678.

Fama, E. F. and MacBeth, J. (1973) Risk, Return and Equilibrium: Empirical Tests. *Journal of Political Economy* 81, 607-636.

Gao, X. and Ritter, J. R. (2010) The Marketing of Seasoned Equity Offerings. *Journal of Financial Economics* 97, 33-52.

Garlappi, L., Shu, T., and Yan, H. (2008) Default Risk, Shareholder Advantage, and Stock Returns. *The Review of Financial Studies* 21, 2743-2778.

George, T. J. and Hwang, C. Y. (2010) A Resolution of the Distress Risk and Leverage Puzzles in the Cross-Section of Stock Returns. *Journal of Financial Economics* 96, 56-79.

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989) A Test of the Efficiency of a Given Portfolio. *Econometrica*, 57, 1121-1152.

Guo, R. J., Lev, B., and Shi, C. (2006) Explaining the Short- and Long-Term IPO Anomalies in the US by R&D. *Journal of Business, Finance & Accounting* 33, 550-579.

Hirshleifer, D., Hou, K., and Teoh, S. H. (2012) The Accrual Anomaly: Risk or Mispricing? *Management Science 58,* 320-335.

Jegadeesh, N. (1990) Evidence of Predictable Behavior of Security Returns. *Journal of Finance* 45, 881-898.

Jegadeesh, N. and Titman, S. (1993) Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance* 68, 65-91.

Kaplan, S. N. and Zigales, L. (1997) Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financial Constraints? *Quarterly Journal of Economics* 112, 169-215.

Lakonishok, J., Shleifer, A., and Vishny, R. W. (1994) Contrarian Investment, Extrapolation, and Risk. *Journal of Finance* 49, 1541-1578.

Lambert, R., Leuz, C., and Verrecchia. R. (2007) Accounting Information, Disclosure, and the Cost of Capital. *Journal of Accounting Research* 45, 385-420.

Lamont, O., Polk, C., and Saa-Requejo, J. (2001) Financial Constraints and Stock Returns. *The Review of Financial Studies* 14, 529-554.

Lev, B. (2004) Sharpening the intangibles edge. *Harvard Business Review,* June, 109-116.

Lev, B. and Sougiannis, T. (1996) The Capitalization, Amortization, and Value-Relevance of R&D. *Journal of Accounting and Economics* 21, 107-138.

Lev, B. and Sougiannis, T. (1999) Penetrating the book-to-market black box: The R&D effect. *Journal of Business Finance & Accounting* 26, 419–449.

Lev, B., Sarath, B., and Sougiannis, T. (2005) R&D Reporting Biases and Their Consequences. *Contemporary Accounting Research* 22, 977–1026.

Lev, B., Radhakrishnan, S., and Tong, J. Y. (2012) Risk Management for Tangible and Intangible Investments: The Relationship between R&D and Capital Expenditures and Risk Components. Working Paper (Dec).

Li, D. (2011) Financial Constraints, R&D Investment, and Stock Returns. *The Review of Financial Studies* 24, 2974-3007.

Newey, W. K. and West, K. D. (1987) A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55, 703-708.

Opler, T. C. and Titman, S. (1994) Financial Distress and Corporate Performance. *Journal of Finance* 49, 1015-1040.

Pastor, L. and Stambaugh, R. F. (2003) Liquidity Risk and Expected Stock Returns. *Journal of Political Economy* 111, 642-685.

Penman, S. H. and Zhang, X. J. (2002) Accounting Conservatism, the Quality of Earnings and Stock Returns. *The Accounting Review* 77, 237-264.

Shi, C. (2003) On the Trade-Off between the Future Benefit and Riskiness of R&D: A Bondholders' Perspective. *Journal of Accounting & Economics* 35, 227-254.

Shleifer, A. and Vishny, R. W. (1992) Liquidation Values and Debt Capacity: A Market Equilibrium Approach. *Journal of Finance* 47, 1343-1366.

Shumway, T. (1997) The Delisting Bias in CRSP Data. *Journal of Finance* 72, 327-340.

Szwejczewski, M., Mitchell, R., and Lemke, F. (2006) A Study of R&D Portfolio Management Among UK Organizations. *International Journal of Management and Decision Making* 7, 604-624.

Vassalou, M. and Xing, Y. (2004) Default Risk in Equity Returns. *Journal of Finance* 59, 831-868.

Zhang, X. (2006) Information Uncertainty and Stock Returns. *Journal of Finance* 61, 105-137.

TABLE 1
Descriptive summary

This table reports the average level of R&D intensity for our sample of R&D stocks over the full period and four sub-periods. We include three measures of R&D intensity: R&D expenditure to market value (*RD-MV*); total assets (*RD-A*) and total sales (*RD-S*). Panel A reports the average R&D intensity for the full sample of R&D stocks. Panel B groups R&D stocks into low-tech and high-tech categories according to the classification of Chan et al. (2001) and reports the average intensity. Panel C groups R&D stocks into low-growth (*BM*≥1) and high-growth (*BM*<1) categories and reports the average intensity. Panel D reports average R&D intensity by Fama-French 12-industry classification (excluding the financials industry).

Period		1976-2012			1976-1985			1986-1995			1996-2005			2006-2012		
R&D Intensity	Obs.	*RD-MV*	*RD-A*	*RD-S*	*RD-MV*	*RD-A*	*RD-S*	*RD-MV*	*RD-A*	*RD-S*	*RD-MV*	*RD-A*	*RD-S*	*RD-MV*	*RD-A*	*RD-S*
Panel A: Full sample																
All R&D firms																
(7,612 firms)	770,929	31.7%	9.1%	254.2%	10.2%	4.2%	29.4%	39.1%	8.2%	132.3%	45.0%	11.3%	209.6%	22.1%	11.0%	653.9%
Panel B: High-tech and low-tech classification																
Low-Tech	401,847	27.3%	7.9%	412.3%	9.4%	2.9%	6.5%	32.0%	6.5%	222.4%	43.1%	10.2%	325.5%	19.7%	12.0%	1240.4%
High-Tech	369,082	36.5%	10.5%	82.1%	11.7%	6.7%	72.4%	46.9%	10.1%	33.3%	46.7%	12.4%	108.1%	24.4%	10.1%	97.6%
Panel C: High-growth and low-growth classification																
Low-growth (*BM*≥1)	194,028	107.1%	5.3%	48.4%	18.2%	2.6%	2.2%	143.3%	5.6%	16.8%	199.5%	7.2%	57.8%	80.9%	6.7%	158.0%
High-growth (*BM*<1)	576,901	6.4%	10.4%	323.5%	4.6%	5.3%	48.6%	6.1%	9.0%	168.8%	6.7%	12.4%	247.2%	7.1%	12.1%	780.4%
Panel D: Fama-French 12-industry classification																
1 Non Durables	35,045	25.0%	2.4%	5.0%	5.8%	1.5%	2.4%	68.0%	3.1%	11.4%	21.5%	3.2%	4.3%	6.7%	1.9%	2.2%
2 Durables	35,415	107.5%	4.1%	6.0%	10.0%	2.9%	4.6%	129.6%	4.0%	5.0%	208.8%	4.8%	7.8%	47.3%	4.7%	6.6%
3 Manufacturing	157,340	20.7%	3.6%	13.3%	9.7%	2.7%	2.4%	24.5%	3.9%	9.7%	31.0%	4.3%	25.9%	17.0%	3.5%	16.7%
4 Energy	15,968	10.8%	1.7%	28.5%	2.9%	1.4%	8.7%	7.3%	2.7%	32.4%	25.3%	1.6%	5.8%	9.0%	1.1%	92.9%
5 Chemicals	37,628	24.0%	3.7%	15.8%	23.8%	3.4%	5.9%	21.4%	4.4%	31.0%	37.5%	3.5%	15.1%	9.1%	3.4%	8.6%
6 Business Equipment	294,131	37.4%	10.9%	54.9%	11.8%	7.2%	10.5%	50.9%	10.5%	28.1%	44.4%	12.9%	100.0%	27.3%	10.5%	37.2%
7 Telecommunications	10,489	102.6%	3.9%	25.4%	5.0%	3.0%	5.4%	104.2%	3.3%	8.7%	187.2%	4.6%	54.0%	32.5%	3.9%	8.1%
8 Utilities	397	0.4%	1.7%	18.7%	0.4%	2.4%	26.5%	0.1%	0.0%	0.0%	0.5%	0.7%	2.8%	0.0%	0.0%	0.0%
9 Shops	13,802	8.2%	4.0%	10.3%	4.8%	2.3%	1.8%	15.9%	4.8%	12.3%	8.1%	5.6%	22.2%	3.2%	3.3%	4.0%
10 Healthcare	135,744	20.9%	18.9%	1270.3%	9.0%	6.7%	301.1%	13.5%	14.7%	694.5%	28.4%	20.7%	746.9%	20.3%	23.4%	2688.3%
12 Other	34,970	12.0%	5.4%	98.9%	5.0%	3.3%	10.1%	7.1%	5.7%	55.2%	23.7%	6.1%	206.5%	9.1%	6.5%	99.8%

TABLE 2

The R&D anomaly

This table reports the results of univariate portfolio analysis using monthly returns from 1976 to 2013. We sort stocks with non-zero R&D expenditure into ten equally weighted portfolios according to their R&D expenditure to market value ratio (*RD-MV*). Portfolio 1 (10) contains stocks with the lowest (highest) *RD-MV*. The first set of columns report the average portfolio returns (*RET*) and the intercepts (α) and adjusted R^2 from the Carhart (1997) four-factor model based on raw returns. The second group of columns reports the same values when using characteristics-adjusted returns (*DGTW RET*). The final set of columns reports the average log size (*Ln(SIZE)*), log book-to-market equity ratios (*Ln(BM)*), past intermediate term cumulative returns (*RET(-12,-2)*), R&D expenditure to market value ratio (*RD-MV*) and number of stocks in each portfolio. The (10-1) portfolio is a zero-cost portfolio that goes long Portfolio 10 funded by going short Portfolio 1. The Newey and West (1987) robust *t*-statistics are reported in brackets.

	Simple			*DGTW*							
Portfolio	*RET*	α	Adj. R^2	*RET*	α	Adj. R^2	*Ln(Size)*	*Ln(BM)*	*RET(-12,-2)*	*RD-MV*	N
1 (Low)	0.79%	-0.26%	87.7%	-0.51%	-0.93%	18.4%	5.55	0.39	14.16%	0.37%	169
		[-2.18]			[-10.38]						
2	1.14%	0.07%	90.9%	-0.24%	-0.66%	25.8%	5.62	0.42	15.69%	1.05%	169
		[0.81]			[-10.27]						
3	1.28%	0.20%	91.6%	-0.13%	-0.54%	16.1%	5.45	0.43	16.03%	1.77%	169
		[1.87]			[-8.57]						
4	1.28%	0.20%	91.6%	-0.20%	-0.58%	2.2%	5.33	0.44	16.17%	2.62%	169
		[1.78]			[-9.43]						
5	1.46%	0.30%	91.8%	-0.10%	-0.51%	1.9%	5.18	0.47	16.75%	3.71%	169
		[3.04]			[-7.95]						
6	1.61%	0.47%	89.7%	0.01%	-0.38%	0.4%	5.03	0.50	17.38%	5.11%	169
		[3.86]			[-5.65]						
7	1.81%	0.60%	87.1%	0.15%	-0.29%	2.9%	4.76	0.53	17.59%	7.15%	169
		[4.10]			[-4.57]						
8	1.99%	0.81%	85.1%	0.20%	-0.19%	17.1%	4.44	0.59	19.12%	10.32%	169
		[4.35]			[-2.29]						
9	2.20%	0.94%	80.8%	0.33%	-0.11%	21.8%	4.11	0.67	19.98%	16.51%	169
		[4.50]			[-1.19]						
10 (High)	2.48%	1.22%	74.9%	0.46%	0.06%	12.3%	3.66	1.44	24.39%	245.21%	169
		[5.41]			[0.53]						
(10-1)	1.69%	1.48%	19.0%	0.98%	0.99%	20.9%	-1.88	1.04	10.23%	244.84%	
	[6.87]	[6.94]		[5.10]	[6.09]						

TABLE 3

Summary statistics, correlations and pricing tests of factors

This table reports the summary statistics, the pairwise correlations and the results of time series regressions of the R&D factor on other pricing factors. Panel A reports the mean, standard deviation, ex post Sharpe ratio, and 25%, 50% and 75% percentile statistics for the market (*MKT_RF)*, *SMB*, *HML*, *UMD* and the R&D factor (*RD-HML*). To construct the R&D factor, we first sort R&D stocks into three portfolios according to *RD-MV* using 30% and 70% breakpoints. The R&D risk factor is then calculated as the return on an equally weighted, zero-cost portfolio that goes long the portfolio of the top 30% of *RD-MV* stocks and goes short the portfolio of the lowest 30%. Panel B reports the pairwise correlations between the factors. Panel C reports the results of the time series regressions of *RD-HML* on the *MKT_RF*, *SMB*, *HML* and *UMD* factors. The Newey and West (1987) standard errors are reported in parenthesis. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Panel A: Summary statistics

Factors	Mean	Stdev	SR	25%	50%	75%	N
RD-HML	1.15%	3.44%	0.33	-0.57%	0.80%	2.27%	456
MKT_RF	0.63%	4.51%	0.14	-1.93%	1.07%	3.61%	456
SMB	0.28%	3.04%	0.09	-1.31%	0.19%	2.05%	456
HML	0.34%	2.98%	0.11	-1.23%	0.31%	1.72%	456
UMD	0.69%	4.46%	0.15	-0.78%	0.77%	2.87%	456

Panel B: Pairwise correlations

	RD-HML	*MKT_RF*	*SMB*	*HML*	*UMD*
RD-HML	1.00	—	—	—	—
MKT_RF	0.13***	1.00	—	—	—
SMB	0.54***	0.26***	1.00	—	—
HML	-0.18***	-0.32***	-0.27***	1.00	—
UMD	0.08*	-0.09*	0.08*	-0.17***	1.00

Panel C: Pricing RD-HML

Variables	α	*MKT_RF*	*SMB*	*HML*	*UMD*	Adj. R^2	N
(1)	0.011***	0.100**	—	—	—	1.5%	456
S.E.	[0.002]	[0.040]					
(2)	0.010***	-0.010	0.601***	-0.035	0.020	28.5%	456
S.E.	[0.002]	[0.042]	[0.122]	[0.105]	[0.063]		

TABLE 4

The R&D augmented model

Panel A reports the estimated coefficients, Newey and West (1987) robust t-statistics and adjusted R^2 of the time series regressions of decile *RD-MV* portfolio returns and the (10-1) return spread on the R&D augmented model (equation (2)). Panel B reports the results of the Gibbons et al. (1989) (*GRS*) tests across decile portfolios, including mean intercepts, *GRS* F-statistics, *p*-values, average adjusted R^2 and the Sharpe ratio of the intercepts. These are reported for both the four-factor and R&D augmented models.

Panel A: The augmented model for RD-MV deciles

Portfolios	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	(10-1)
α	-0.16%	0.15%	0.14%	0.02%	0.06%	0.08%	0.09%	0.06%	-0.07%	0.15%	0.30%
t(α)	[-1.14]	[1.51]	[1.24]	[0.18]	[0.58]	[0.72]	[0.67]	[0.50]	[-0.57]	[1.15]	[3.43]
MKT_RF	1.06	1.03	1.02	1.03	1.03	1.05	1.07	1.04	1.07	1.00	-0.06
t(*MKT_RF*)	[32.99]	[39.17]	[42.15]	[37.42]	[41.83]	[38.09]	[39.60]	[37.08]	[38.53]	[33.64]	[-2.75]
SMB	0.73	0.72	0.80	0.79	0.83	0.75	0.83	0.82	0.77	0.67	-0.06
t(*SMB*)	[10.54]	[9.97]	[13.67]	[11.64]	[14.38]	[11.92]	[12.53]	[11.05]	[12.87]	[9.23]	[-1.81]
HML	-0.15	-0.08	-0.11	-0.18	-0.09	-0.14	-0.10	-0.19	-0.12	-0.03	0.12
t(β_{HML})	[-1.80]	[-1.13]	[-2.04]	[-2.80]	[-1.86]	[-2.53]	[-1.49]	[-3.04]	[-1.72]	[-0.40]	[3.78]
UMD	-0.25	-0.23	-0.25	-0.25	-0.21	-0.23	-0.24	-0.27	-0.26	-0.21	0.04
t(*UMD*)	[-4.34]	[-4.09]	[-5.28]	[-4.65]	[-6.4]	[-5.48]	[-4.77]	[-5.28]	[-4.88]	[-3.38]	[1.79]
RD-HML	-0.10	-0.08	0.06	0.18	0.24	0.39	0.53	0.76	1.03	1.09	1.19
t(*RD-HML*)	[-1.93]	[-2.44]	[1.56]	[5.59]	[7.05]	[9.37]	[11.22]	[19.14]	[27.92]	[18.96]	[35.74]
N	456	456	456	456	456	456	456	456	456	456	456
Adj. R^2	87.9%	91.0%	91.7%	92.1%	93.0%	92.5%	91.5%	93.3%	94.0%	91.0%	84.2%

Panel B: GRS statistics

	Mean α	GRS Stat	*p*-value	Mean Adj. R^2	α SR
Four-factor model	0.46%	6.96	0.00	87%	0.42
R&D augmented model	0.05%	2.07	0.03	92%	0.24

TABLE 5

The R&D increase anomaly and the *RD-HML* factor

This table reports the results of time series regressions of portfolio returns of stocks exhibiting significant R&D increases on the Carhart (1997) four-factor and R&D augmented models. Using a calendar time approach, we pool stocks into a portfolio whenever they are within a 60-month window following economically significant R&D increases and calculate the portfolio return. The portfolio returns used for rows 1 and 2 are calculated using an equal weighting scheme and those in rows 3 and 4 use a value weighting scheme. Estimated coefficients, Newey and West (1987) robust standard errors (in parentheses), number of observations and adjusted R^2 of the regressions are reported. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Models	α	*MKT_RF*	*SMB*	*HML*	*UMD*	*RD-HML*	N	Adj. R^2
Equal weighting								
Row 1	0.008***	1.104***	1.303***	-0.492***	-0.180***	—	456	0.877
S.E.	(0.002)	(0.040)	(0.089)	(0.073)	(0.055)			
Row 2	0.001	1.111***	0.883***	-0.468***	-0.194***	0.700***	456	0.937
S.E.	(0.001)	(0.026)	(0.062)	(0.064)	(0.051)	(0.045)		
Value weighting								
Row 3	0.004***	1.008***	-0.043	-0.548***	-0.135***	—	456	0.862
S.E.	(0.001)	(0.029)	(0.035)	(0.058)	(0.035)			
Row 4	0.002	1.010***	-0.177***	-0.540***	-0.139***	0.224***	456	0.875
S.E.	(0.001)	(0.028)	(0.040)	(0.061)	(0.037)	(0.040)		

TABLE 6

Investor sentiment and the R&D anomaly

This table reports *RD-MV* decile portfolio returns conditional on investor sentiment states. Sentiment state is defined as high (low) if the beginning-of-period Baker and Wurgler (2006) sentiment index is higher (lower) than its sample median. We divide R&D stocks into decile portfolios according to *RD-MV* and calculate the equally weighted average monthly portfolio returns within each state. Panels A and B report the simple and DGTW characteristics-adjusted returns and the Newey and West (1987) robust *t*-statistics for each decile portfolio and the zero-cost spread portfolio (10-1). Panel C reports the estimated alpha from the Carhart (1997) four-factor model within sentiment state. The bottom rows of Panels A to C (H-L) show the differences in average returns across sentiment states conditional on *RD-MV* decile. Panel D presents bivariate portfolio analysis after adjusting for size by sorting by size then *RD-MV*. For each sentiment state, we report the equally weighted average portfolio returns on each 25 size-*RD-MV* portfolio, size averaged *RD-MV* quintiles (Avg) and size-adjusted R&D spread portfolio (H-L). For these size-adjusted portfolios, we also report alpha from the Carhart (1997) four-factor model, and alpha and *RD-HML* factor loadings from the R&D augmented model separated by sentiment state.

<u>*Panel A: Simple Returns*</u>

Portfolio	Sentiment	1	2	3	4	5	6	7	8	9	10	(10-1)
RET	High	0.20%	0.61%	0.77%	0.68%	0.79%	0.96%	1.08%	1.31%	1.34%	1.53%	1.32%
t(*RET*)		[0.36]	[1.20]	[1.48]	[1.29]	[1.50]	[1.76]	[1.86]	[2.09]	[1.98]	[2.46]	[5.56]
RET	Low	1.26%	1.56%	1.66%	1.76%	2.03%	2.17%	2.46%	2.60%	2.97%	3.34%	2.08%
t(*RET*)		[3.72]	[4.67]	[4.58]	[4.45]	[4.97]	[4.80]	[4.97]	[4.76]	[4.89]	[5.27]	[4.99]
RET	(H-L)	-1.05%	-0.95%	-0.89%	-1.08%	-1.25%	-1.21%	-1.38%	-1.29%	-1.63%	-1.81%	-0.75%

<u>*Panel B: DGTW Returns*</u>

Portfolio	Sentiment	1	2	3	4	5	6	7	8	9	10	(10-1)
DGTW RET	High	-0.48%	-0.15%	0.03%	-0.15%	-0.13%	0.01%	0.09%	0.23%	0.22%	0.29%	0.76%
t(*DGTW RET*)		[-4.28]	[-1.50]	[0.40]	[-2.22]	[-1.62]	[0.12]	[1.24]	[2.14]	[1.76]	[2.09]	[3.54]
DGTW RET	Low	-0.56%	-0.33%	-0.31%	-0.28%	-0.07%	0.02%	0.21%	0.20%	0.45%	0.64%	1.20%
t(*DGTW RET*)		[-3.87]	[-3.03]	[-3.38]	[-3.23]	[-0.98]	[0.29]	[2.57]	[1.79]	[2.81]	[3.36]	[3.84]
DGTW RET	(H-L)	0.09%	0.18%	0.34%	0.12%	-0.06%	-0.02%	-0.12%	0.03%	-0.22%	-0.35%	-0.44%

Panel C: Four-factor α

Portfolio	Sentiment	1	2	3	4	5	6	7	8	9	10	(10-1)
α	High	-0.29%	0.07%	0.25%	0.18%	0.21%	0.43%	0.56%	0.86%	0.85%	1.09%	1.38%
t(α)		[-1.62]	[0.54]	[1.70]	[1.11]	[1.63]	[2.75]	[3.08]	[3.93]	[3.30]	[4.05]	[5.55]
α	Low	-0.28%	0.02%	0.07%	0.16%	0.39%	0.49%	0.62%	0.71%	0.99%	1.24%	1.52%
t(α)		[-2.01]	[0.19]	[0.53]	[1.28]	[2.64]	[2.86]	[2.84]	[2.74]	[3.33]	[4.02]	[5.11]
α	(H-L)	-0.01%	0.05%	0.18%	0.02%	-0.18%	-0.05%	-0.05%	0.15%	-0.14%	-0.16%	-0.14%

Panel D: Controlling for size

High Sentiment

								Four-factor		*Augmented model*			
RD-MV	Small	2	3	4	Big	Avg	*t*(*RET*)	*α*	*t*(*α*)	*α*	*t*(*α*)	*RD-HML*	*t(RD-HML)*
Low	0.45%	0.60%	0.52%	0.41%	0.54%	0.50%	[0.98]	0.02%	[0.14]	0.25%	[1.46]	-0.25	[-3.70]
2	0.76%	0.87%	0.76%	0.85%	0.71%	0.79%	[1.63]	0.29%	[2.23]	0.37%	[2.53]	-0.09	[-1.99]
3	1.00%	0.84%	0.80%	0.95%	0.85%	0.89%	[1.82]	0.35%	[3.19]	0.21%	[1.77]	0.15	[3.98]
4	1.26%	1.01%	1.18%	1.14%	0.96%	1.11%	[2.17]	0.57%	[4.46]	0.29%	[2.25]	0.31	[5.67]
High	1.63%	1.25%	1.03%	1.11%	1.07%	1.22%	[2.22]	0.66%	[4.20]	0.15%	[1.18]	0.56	[11.03]
(H-L)	1.18%	0.66%	0.51%	0.69%	0.53%	0.71%	[3.56]	0.64%	[2.79]	-0.10%	[-0.63]	0.80	[10.79]

Low Sentiment

								Four-factor		*Augmented model*			
RD-MV	Small	2	3	4	Big	Avg	*t*(*RET*)	*α*	*t*(*α*)	*α*	*t*(*α*)	*RD-HML*	*t(RD-HML)*
Low	1.64%	1.14%	1.51%	1.30%	1.14%	1.34%	[4.56]	-0.11%	[-1.14]	0.02%	[0.24]	-0.12	[-2.95]
2	2.04%	1.55%	1.72%	1.44%	1.16%	1.58%	[5.04]	0.11%	[0.90]	0.06%	[0.61]	0.04	[0.93]
3	2.65%	1.87%	1.90%	1.39%	1.33%	1.83%	[5.43]	0.34%	[3.01]	0.22%	[2.14]	0.11	[3.16]
4	2.90%	1.90%	1.97%	1.93%	1.44%	2.03%	[5.00]	0.39%	[2.33]	0.03%	[0.22]	0.34	[8.04]
High	3.48%	2.26%	2.28%	1.94%	1.51%	2.30%	[5.33]	0.55%	[3.12]	0.02%	[0.16]	0.51	[13.27]
(H-L)	1.84%	1.13%	0.77%	0.64%	0.38%	0.95%	[4.11]	0.66%	[3.49]	0.00%	[-0.03]	0.64	[17.54]
					Sent. H-L	-0.24%		-0.02%		-0.10%		0.16	

TABLE 7
Covariance risk versus characteristic mispricing

This table reports the results of the firm-level monthly Fama and MacBeth (1973) regression procedure to test whether the *RD-HML* factor loading is priced in the cross section of expected stock returns, after controlling for the *RD-MV* characteristic. To reduce the errors-in-variables problem, we assign pre-estimated portfolio factor loadings to individual stocks. The portfolios used in the pre-estimation are 25 equally weighted size-*RD-MV* portfolios. Firm-specific variables include *RD-MV*, log market capitalization (*Ln(MCAP)*), log book-to-market equity ratio (*Ln(BM)*), past intermediate term cumulative returns (*RET(-12,-2)*), log turnover ratio (*Ln(TURN)*), stock market illiquidity (*ILLIQ*) and log idiosyncratic volatility (*Ln(IVol)*). For details of variable construction, please see Table A1 in the Appendix. Columns (6) and (7) include pre-estimated factor loadings on *MKT_RF*, *SMB*, *HML* and *UMD* as further controls. Column (7) also includes industry dummy variables according to the Fama-French 49-industry classification. We report the average coefficients, Newey and West (1987) robust standard errors and the average R^2 for each model. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Independent variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\beta_{RD\text{-}HML}$	0.005***	—	0.005***	0.003***	0.003***	0.008***	0.007***
	(0.001)		(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
RD-MV	—	0.002**	0.000	-0.000	-0.001	-0.001	-0.001*
		(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Ln(MCAP)	—	—	—	-0.001**	-0.001**	-0.001***	-0.001***
				(0.001)	(0.000)	(0.000)	(0.000)
Ln(BM)	—	—	—	0.003*	0.002*	0.001	0.002*
				(0.001)	(0.001)	(0.001)	(0.001)
RET(-12,-2)	—	—	—	0.002	0.003	0.003	0.002
				(0.002)	(0.002)	(0.002)	(0.002)
Ln(TURN)	—	—	—	—	-0.002**	-0.002**	-0.002***
					(0.001)	(0.001)	(0.001)
ILLIQ	—	—	—	—	0.002***	0.002***	0.002**
					(0.001)	(0.001)	(0.001)
Ln(IVol)	—	—	—	—	-0.154***	-0.168***	-0.171***
					(0.039)	(0.037)	(0.035)
β_{MKT_RF}	—	—	—	—	—	0.001	0.001
						(0.002)	(0.002)
β_{SMB}	—	—	—	—	—	-0.002	-0.003*
						(0.002)	(0.002)
β_{HML}	—	—	—	—	—	0.000	0.001
						(0.001)	(0.001)
β_{UMD}	—	—	—	—	—	-0.001	-0.000
						(0.003)	(0.003)
Intercept	0.011***	0.016***	0.011***	0.014**	0.018***	0.018***	0.020***
	(0.003)	(0.004)	(0.003)	(0.006)	(0.005)	(0.005)	(0.007)
Industry FE	N	N	N	N	N	N	Y
N	695,853	769,939	695,853	689,644	634,774	634,774	634,774
Average R^2	0.004	0.001	0.005	0.024	0.045	0.049	0.088

TABLE 8

The determinants of R&D factor loadings

This table reports the summary statistics of firm-specific variables and the results of pooled weighted least squares (WLS) regressions that investigate the determinants of individual stocks' exposure to the R&D risk factor. Panel A reports the number of firm-year observations, mean, standard deviation and 25%, 50% and 75% percentile statistics of the firm-specific variables. Firm-specific variables with fiscal year end in year *t*-1 are matched with individual stocks' subsequent R&D factor loadings estimated using the monthly returns over the period from July year *t* to June year *t*+3 (36-month window and requires a minimum of 10 months). To reduce the problem of outliers, we winsorise firm-year observations at the 1st and 99th percentiles. Following Campbell et al. (2010), we de-mean all variables at each cross-section and normalize each independent variable to have unit variance. We use the inverse of the number of cross-sectional observations in each year as weights. The independent variables include R&D expenditure to market value (*RD-MV)*, property, plant and equipment to total assets (*PPE-A*), Herfindahl-Hirschman index (*HHI*), Altman's (1968) Z-score (*Z-score*), Kaplan and Zingales' (1997) KZ index (*KZ index*), analyst forecast dispersion (*Ln(Disp)*), log market capitalization (*Ln(MCAP)*), return on assets (*ROA*) and capital expenditure to total assets (*INV-A*). We also include industry dummy variables constructed according to the Fama-French 49-industry classification. Estimated coefficients, robust standard errors (in parentheses) clustered at the firm-level, the number of observations and adjusted R^2 are reported in Panel B. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Panel A: Summary statistics

				Percentiles		
Variable	Obs.	Mean	Stdev.	25%	50%	75%
$\beta_{RD\text{-}HML}$	68,025	0.47	2.03	-0.49	0.26	1.27
RD-MV	59,964	13.02%	39.39%	1.70%	4.13%	9.67%
PPE-A	66,967	22.58%	16.45%	9.61%	19.11%	31.67%
HHI	68,025	8.62%	7.67%	4.31%	5.74%	10.36%
Z-score	29,403	0.11	0.51	0.00	0.01	0.02
KZ index	60,157	-8.72	25.89	-6.50	-1.89	0.30
Ln(Disp)	20,726	0.69%	1.46%	0.05%	0.19%	0.62%
Ln(MCAP)	61,059	5.06	2.22	3.39	4.83	6.51
ROA	66,742	3.68%	27.80%	0.42%	11.28%	18.04%
INV-A	66,080	5.99%	5.43%	2.34%	4.43%	7.79%

Panel B: WLS regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
RD-MV	0.070***	0.070***	0.069***	0.078***	0.071***	0.078***	0.066***
	(0.010)	(0.010)	(0.010)	(0.015)	(0.010)	(0.016)	(0.017)
PPE-A	—	-0.041**	—	—	—	—	-0.069***
		(0.018)					(0.026)
HHI	—	—	0.062**	—	—	—	0.026
			(0.025)				(0.045)
Z-score	—	—	—	-0.024**	—	—	-0.040***
				(0.012)			(0.014)
KZ index	—	—	—	—	-0.013	—	-0.022
					(0.011)		(0.023)
Ln(Disp)	—	—	—	—		0.090***	0.097***
						(0.022)	(0.025)
Ln(MCAP)	-0.163***	-0.159***	-0.162***	-0.140***	-0.163***	-0.148***	-0.120***
	(0.014)	(0.014)	(0.014)	(0.019)	(0.014)	(0.020)	(0.022)
ROA	-0.270***	-0.267***	-0.271***	-0.268***	-0.278***	-0.234***	-0.277***
	(0.021)	(0.021)	(0.021)	(0.031)	(0.020)	(0.032)	(0.041)
INV-A	0.019	0.038**	0.016	0.021	0.024*	0.073***	0.106***
	(0.014)	(0.016)	(0.014)	(0.019)	(0.014)	(0.020)	(0.024)
Intercept	0.160	0.171	0.037	0.013	0.135	-0.105	-0.041
	(0.290)	(0.289)	(0.293)	(0.570)	(0.294)	(0.318)	(0.390)
Industry FE	Y	Y	Y	Y	Y	Y	Y
N	57,358	57,350	57,358	27,647	56,737	19,975	16,302
Adj. R^2	4.5%	4.5%	4.5%	4.9%	4.7%	6.5%	6.8%

APPENDIX
Table A1
Detailed variable definitions

Note: Following Fama and French (1993, 2008), accounting variables from Compustat with fiscal year end in calendar year *t*-1 are matched with monthly CRSP returns from July of year t to June of year *t*+1.

Variable	Definition	Source
R&D intensity		
RD-MV	Total R&D expenditure in fiscal year end year *t*-1 divided by total market capitalization as at the end of December year *t-1*.	Compustat and CRSP
RD-A	Total R&D expenditure divided by the total book value of assets at end of fiscal year *t-1*.	Compustat
RD-S	Total R&D expenditure divided by total sales revenue at end of fiscal year *t-1*.	Compustat
RDC-MV	R&D capital in fiscal year end *t-1* divided by total market capitalization at end of December year *t-1*. Following Chan et al. (2001), we assume that the productivity of each dollar of R&D spending declines linearly by 20 percent per year and compute the stock of R&D capital (RDC_{it}) as follows: $$RDC_{i,t} = RD_{i,t} + 0.8\times RD_{i,t-1} + 0.6\times RD_{i,t-2} + 0.4\times RD_{i,t-3} + 0.2\times RD_{i,t-4}$$ where $RD_{i,t}$ refers to the R&D expenditure of firm *i* in year *t*.	Compustat
Asset pricing variables		
Ln(BM)	Log of one plus the book-to-market equity ratio. Following Fama and French (1993), book equity is total assets for year *t*−1, minus liabilities, plus balance sheet deferred taxes and investment tax credit if available, minus preferred stock liquidating value if available, or redemption value if available, or carrying value. Market equity is shares outstanding at end of December year *t*−1 times share price.	Compustat and CRSP
Ln(MCAP)	Log of one plus the market capitalization (shares outstanding times share price) at end of June year *t*.	CRSP
RET(-12,-2)	Momentum is calculated as the cumulated compounded stock return from month *j−12* to month *j−2*, which is updated monthly.	CRSP
Ln(TURN)	Following Chordia et al. (2001), a stock's turnover ratio is share trading volume divided by the number of shares outstanding. We use the log turnover ratio of month *j*-2 to explain returns in month *j*.	CRSP

ILLIQ	Following Amihud (2002), a stock's market illiquidity in month *j* is measured as the average of the daily ratio of absolute stock return to trading volume: $$ILLIQ_{i,j}=Avg[(\lvert R_{i,d} \rvert)/VOLD_{i,d}]$$ where $R_{i,d}$ and $VOLD_{i,d}$ are the daily return and trading volume of stock *i* on day *d* in month *j*. Following Gao and Ritter (2010), institutional features are taken into account by dividing the NASDAQ volume by 2.0, 1.8, 1.6 and 1 for the periods prior to Feb 2001, between Feb 2001 and Dec 2001, between January 2002 and Dec 2003 and Jan 2004 and later years, respectively. Similar to Bali et al. (2014), we require a stock to have at least 15 daily return observations within month *j* to compute this illiquidity measure and scale this variable by 106.	CRSP
Ln(IVol)	Following Ang et al. (2006), a stock's monthly idiosyncratic volatility (*IVol*) is defined as the log of one plus the standard deviation of residuals from the Fama and French (1993) three-factor model, estimated using daily returns within that month. We require a stock to have at least 15 daily stock return observations for the estimation.	CRSP
LIQU	Following Bali et al. (2014), a stock's illiquidity shock is computed as: $$LIQU_{i,j} = ILLIQ_{i,j} - AVGILLIQ_{i,j-12,j-1}$$ where $ILLIQ_{i,t}$ is the monthly stock illiquidity measure of Amihud (2002) of stock *i* in month *j* and $AVGILLIQ_{i,j-12,j-1}$ is the mean $ILLIQ_{i,t}$ over the past 12 months (from month *j-12* to *j-1*).	CRSP
MAX	The maximum daily return in the previous month following Bali et al. (2011).	CRSP
REV	Short-term reversal is the stock return in the previous month following Jegadeesh (1990).	CRSP
Firm characteristics		
$\beta_{RD\text{-}HML}$	Loadings on the *RD-HML* factor. For a given stock in year *t*, the loading is estimated from a time series regression of monthly stock returns on the R&D augmented model from July of year *t* to June of year *t+3* (36 months). We require a stock to have at least 10 consecutive monthly return observations in the 36 month window. $\beta_{RD\text{-}HML}$ is the estimated coefficient on the *RD-HML* factor.	Estimated
PPE-A	Net property, plant & equipment to total book value of assets, as a measure of asset tangibility.	Compustat

HHI	Herfindahl-Hirschman Index of the industry that firm *i* belongs to, based on the Fama and French 49-industry classification.	Compustat
Z-score	The Altman (1968) Z-score measure for bankruptcy prediction. Following Altman (1968), we estimate the *Z-score* as: $$Z\text{-}score = 1.2\times(WC/TA) + 1.4\times(RE/TA) + 3.3\times(EBIT/TA) + 0.6\times(MVEQ/DEBT) + 0.999\times(SALE/TA),$$ where *WC*, *TA*, *RE*, *EBIT*, *MVEQ*, *DEBT* and *SALE* are working capital, total assets, earnings before interest and taxes, market equity value, book value of debts and total sales. For easier interpretation of our regression coefficient we divide the *Z-score* by 1000.	Compustat
KZ index	Financial constraint index, computed according to Kaplan and Zingales (1997) as follows: $$KZ\ Index = -1.002\times(Cash\ flow) + 0.283\times(Tobin's\ Q) + 3.139\times(Leverage) - 39.368\times(Dividends) - 1.315\times(Cash\ holdings),$$ *Cash flow* is the sum of income before extraordinary items and depreciation and amortization, divided by total net property, plant & equipment. *Tobin's Q* is computed as ((book value of assets + market equity (at fiscal year end))-book equity-Deferred Taxes)/book value of assets). *Leverage* is the sum of long-term debt and debt in current liabilities, divided by the sum of long-term debt, debt in current liabilities and total stockholders' equity. *Dividend* is the sum of common and preferred dividends, divided by total net property, plant & equipment. *Cash holdings* is the cash and short-term investments to total net property, plant & equipment. The total net property, plant & equipment used in computing this *KZ index* is lagged by one year.	Compustat and CRSP
Ln(Disp)	Analyst forecast dispersion is the log of the cross-sectional standard deviation of analysts' one-year ahead earning per share forecasts. Following Zhang (2006), we calculate the time series of monthly standard deviations in each fiscal year as analysts update their forecasts and then average these to obtain a measure for each firm's fiscal year. To reduce the problem of heteroskedasticity, this average is scaled by the fiscal year end market price before taking the log.	IBES
ROA	Return on assets, computed as the operating income before depreciation to total book value of assets.	Compustat
INV-A	Capital investment, computed as the capital expenditure to total book value of assets.	Compustat

A Risk Explanation for the R&D Anomaly

Woon Sau Leung   Khelifa Mazouz   Kevin P. Evans

Version: March 2015

****Internet Appendix****

****Supporting Documents****

****Not included in the paper****

TABLE IA.1

The R&D anomaly

This table compares univariate portfolio analysis using *RD-MV* and the alternative *RDC-MV* measure. The equally (*EW*) and value weighted (*VW*) average portfolio returns are reported for each decile and the (10-1) zero cost portfolios. We report average portfolio returns using raw returns (*RET*), DGTW characteristic-adjusted returns, and Carhart (1997) four-factor alphas. Newey and West (1987) robust *t*-statistics are reported in square brackets. See Table 2 for more details on the portfolio construction and the Appendix for definitions of the R&D intensity measures.

	RD-MV						*RDC-MV*					
	RET		*DGTW*		*α*		*RET*		*DGTW*		*α*	
Portfolio	*EW*	*VW*	*EW*	*VW*	*EW*	*VW*	*EW*	*VW*	*EW*	*VW*	*EW*	*VW*
Low	0.79%	0.82%	-0.51%	-0.13%	-0.26%	-0.03%	0.97%	0.98%	-0.45%	-0.07%	-0.10%	0.10%
2	1.14%	1.10%	-0.24%	0.03%	0.07%	0.32%	1.16%	0.97%	-0.18%	0.02%	0.07%	0.14%
3	1.28%	1.05%	-0.13%	0.00%	0.20%	0.18%	1.31%	1.12%	-0.18%	0.03%	0.22%	0.30%
4	1.28%	1.06%	-0.20%	-0.05%	0.20%	0.24%	1.27%	1.05%	-0.19%	-0.06%	0.16%	0.14%
5	1.46%	1.08%	-0.10%	-0.05%	0.30%	0.19%	1.45%	1.07%	-0.01%	0.05%	0.27%	0.24%
6	1.61%	1.17%	0.01%	0.10%	0.47%	0.27%	1.62%	1.17%	-0.04%	0.05%	0.45%	0.23%
7	1.81%	1.52%	0.15%	0.14%	0.60%	0.41%	1.76%	1.49%	0.15%	0.20%	0.57%	0.47%
8	1.99%	1.45%	0.20%	0.25%	0.81%	0.39%	2.00%	1.22%	0.15%	0.11%	0.75%	0.16%
9	2.20%	1.68%	0.33%	0.17%	0.94%	0.52%	2.15%	1.69%	0.31%	0.28%	0.96%	0.52%
High	2.48%	1.53%	0.46%	0.08%	1.22%	0.35%	2.46%	1.51%	0.44%	0.03%	1.16%	0.27%
H-L	1.69%	0.71%	0.98%	0.22%	1.48%	0.38%	1.49%	0.53%	0.90%	0.10%	1.27%	0.17%
t(H-L)	[6.87]	[3.15]	[5.10]	[1.50]	[6.94]	[1.95]	[5.72]	[2.34]	[4.68]	[0.69]	[5.72]	[0.85]

TABLE IA.2
Bivariate portfolio sorting

This table reports returns to bivariate sorted portfolios that control explicitly for size (Panel A), book-to-market equity ratios (Panel B) and momentum (Panel C) effects. At the end of June each year, we first sort R&D stocks into five portfolios according to either log size (*Ln(Size)*), log book-to-market equity ratios (*Ln(BM)*) or past returns (*RET(-12,-2)*). Within each sorted portfolio, we further divide stocks into five portfolios according to its *RD-MV*. The equally weighted average portfolio returns on each of the 25 portfolios are reported. Average returns to the zero cost spread portfolios (5-1) are reported for each sorting dimension. The rightmost column reports the R&D portfolio returns averaged across the size, book-to-market equity ratio or momentum dimensions. We also report the Carhart (1997) four-factor alphas for the (5-1) portfolios with Newey and West (1987) robust *t*-statistics shown in the square brackets.

Panel A: Bivariate portfolio analysis sorting by Ln(Size) and RD-MV

RET	Size quintiles						
RD-MV	1 (Small)	2	3	4	5 (Big)	(5-1)	Avg.
1 (Low)	1.12%	0.94%	1.09%	0.92%	0.87%	-0.25%	0.99%
2	1.45%	1.27%	1.32%	1.19%	0.99%	-0.46%	1.24%
3	1.87%	1.39%	1.42%	1.21%	1.14%	-0.73%	1.40%
4	2.12%	1.48%	1.62%	1.60%	1.25%	-0.87%	1.61%
5 (High)	2.61%	1.80%	1.70%	1.59%	1.34%	-1.27%	1.81%
(5-1)	1.49%	0.85%	0.61%	0.67%	0.48%	-1.02%	0.82%
t(5-1)	[6.63]	[4.20]	[3.41]	[3.64]	[2.43]	[-4.03]	[5.38]
(5-1) α	1.36%	0.75%	0.34%	0.47%	0.26%	1.10%	0.64%
$t(\alpha)$	[6.81]	[3.41]	[1.67]	[2.53]	[1.32]	[4.40]	[4.15]

Panel B: Bivariate portfolio analysis sorting by Ln(BM) and RD-MV

RET	*BM* quintiles						
RD-MV	1 (Low)	2	3	4	5 (High)	(5-1)	Avg.
1 (Low)	0.35%	1.08%	1.38%	1.58%	1.72%	1.37%	1.22%
2	0.92%	1.26%	1.51%	1.57%	2.05%	1.12%	1.46%
3	1.03%	1.54%	1.75%	1.73%	2.28%	1.25%	1.67%
4	1.22%	1.73%	1.89%	2.05%	2.35%	1.13%	1.85%
5 (High)	1.76%	2.06%	2.46%	2.50%	2.07%	0.31%	2.17%
(5-1)	1.41%	0.98%	1.08%	0.92%	0.35%	-1.06%	0.95%
t(5-1)	[4.96]	[2.93]	[3.08]	[2.82]	[1.76]	[3.51]	[3.75]
(5-1) α	1.12%	0.97%	1.12%	0.85%	0.47%	-0.66%	0.90%
$t(\alpha)$	[4.59]	[3.83]	[4.21]	[3.25]	[2.37]	[-2.39]	[4.85]

Panel C: Bivariate portfolio analysis sorting by RET(-12,-2) and RD-MV

RET	Momentum quintiles						
RD-MV	1 (Loser)	2	3	4	5 (Winner)	(5-1)	Avg.
1 (Low)	0.50%	0.82%	0.95%	1.23%	1.56%	1.06%	1.01%
2	0.89%	1.18%	1.23%	1.38%	1.88%	0.98%	1.31%
3	1.35%	1.43%	1.44%	1.66%	1.94%	0.59%	1.56%
4	1.88%	1.58%	1.83%	1.83%	2.31%	0.43%	1.89%
5 (High)	2.68%	2.06%	2.06%	2.48%	2.45%	-0.23%	2.35%
(5-1)	2.18%	1.24%	1.11%	1.25%	0.89%	-1.29%	1.33%
t(5-1)	[6.91]	[4.85]	[4.90]	[4.91]	[4.06]	[-5.12]	[6.07]
(5-1) α	2.01%	0.99%	0.99%	1.13%	0.90%	-1.11%	1.21%
$t(\alpha)$	[6.98]	[4.56]	[5.17]	[5.02]	[4.77]	[-4.33]	[6.68]

TABLE IA.3

Summary statistics, pairwise correlations and pricing tests of the *RDC-HML* factor

This table reports summary statistics, pairwise correlations and the results of time series regressions of *RDC-HML* on other pricing factors. Panel A reports the mean, standard deviation, ex post Sharpe ratio and 25%, 50% and 75% percentiles of the market (*MKT_RF)*, *SMB*, *HML* and *UMD* factors and the alternative R&D capital factor (*RDC-HML*). To construct this R&D capital factor, we first sort R&D stocks into three portfolios according to *RDC-MV* using 30% and 70% breakpoints. The R&D capital factor is calculated as the equally weighted zero-cost portfolio that goes long the portfolio of the top 30% and goes short the portfolio of the lowest 30%. Panel B reports the pairwise correlations between factors. Panel C reports the results of time series regressions of *RDC-HML* on the *MKT_RF*, *SMB*, *HML* and *UMD* factors. Newey and West (1987) standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Panel A: Summary statistics

Factors	Mean	Stdev	SR	25%	50%	75%	N
RDC-HML	1.06%	3.41%	0.31	-0.64%	0.71%	2.19%	456
MKT_RF	0.63%	4.51%	0.14	-1.93%	1.07%	3.61%	456
SMB	0.28%	3.04%	0.09	-1.31%	0.19%	2.05%	456
HML	0.34%	2.98%	0.11	-1.23%	0.31%	1.72%	456
UMD	0.69%	4.46%	0.15	-0.78%	0.77%	2.87%	456

Panel B: Pairwise correlations

Correlation	*RDC-HML*	*MKT_RF*	*SMB*	*HML*	*UMD*
RDC-HML	1.00	—	—	—	—
MKT_RF	0.16***	1.00	—	—	—
SMB	0.55***	0.26***	1.00	—	—
HML	-0.20***	-0.31***	-0.27***	1.00	—
UMD	0.07	-0.09	0.08*	-0.17***	1.00

Panel C: Pricing the RDC-HML factor

Variables	α	*MKT_RF*	*SMB*	*HML*	*UMD*	Adj. R^2	N
(1)	0.010***	0.120***	—	—	—	2.3%	456
S.E.	(0.002)	(0.036)					
(2)	0.009***	0.004	0.602***	-0.055	0.015	30.2%	456
S.E.	(0.002)	(0.043)	(0.124)	(0.096)	(0.060)		

TABLE IA.4
The R&D augmented model using *RDC-HML*

Panel A reports the estimated coefficients, Newey and West (1987) robust t-statistics and adjusted R^2 of the time series regressions of decile *RCD-MV* portfolio returns and the (10-1) return spread on the *RDC-HML* augmented model. Panel B reports the results of the Gibbons et al. (1989) (*GRS*) tests across decile portfolios, including mean intercepts, *GRS* F-statistics, *p*-values, average adjusted R^2 and the Sharpe ratio of the intercepts. These are reported for both the four-factor and *RDC-HML* augmented models.

Panel A: The augmented model for RDC-MV deciles

Portfolios	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	(10-1)
α	0.00%	0.13%	0.22%	0.04%	0.06%	0.12%	0.12%	0.12%	0.11%	0.11%	0.11%
t(α)	[0.03]	[1.37]	[2.3]	[0.45]	[0.59]	[1.12]	[1.23]	[1.19]	[0.98]	[0.95]	[1.35]
MKT_RF	1.02	1.04	1.00	1.01	1.03	1.04	1.03	1.05	1.02	0.99	-0.04
t(MKT_RF)	[33.59]	[38.93]	[40.98]	[48.87]	[45.03]	[40.29]	[38.19]	[36.7]	[33.36]	[34.1]	[-1.73]
SMB	0.73	0.69	0.72	0.75	0.69	0.76	0.77	0.76	0.77	0.62	-0.12
t(SMB)	[10.25]	[10.81]	[12.18]	[13.55]	[10.86]	[11.87]	[13.33]	[12.85]	[9.99]	[9.52]	[-3.32]
HML	-0.02	-0.04	-0.02	-0.03	0.00	-0.08	-0.02	-0.03	-0.09	0.04	0.06
t(HML)	[-0.30]	[-0.73]	[-0.29]	[-0.63]	[-0.01]	[-1.41]	[-0.41]	[-0.45]	[-1.14]	[0.74]	[1.68]
UMD	-0.25	-0.22	-0.22	-0.23	-0.17	-0.20	-0.23	-0.22	-0.29	-0.18	0.07
t(UMD)	[-4.78]	[-5.29]	[-4.43]	[-6.46]	[-6.04]	[-4.82]	[-6.37]	[-5.12]	[-4.41]	[-3.96]	[2.66]
RDC-HML	-0.12	-0.07	0.00	0.13	0.24	0.36	0.50	0.70	0.94	1.18	1.29
t(RDC-HML)	[-2.51]	[-2.02]	[0.11]	[3.36]	[7.19]	[6.55]	[12.25]	[17.92]	[20.56]	[24.78]	[37.82]
N	456	456	456	456	456	456	456	456	456	456	456
Adj. R^2	0.88	0.92	0.91	0.93	0.93	0.92	0.93	0.93	0.92	0.92	0.85

Panel B: GRS Statistics

	Mean α	GRS Stat	*p*-value	Mean Adj. R^2	α SR
Four-factor model	0.45%	5.07	0.00	87%	0.35
RDC-HML augmented model	0.10%	0.97	0.47	92%	0.16

TABLE IA.5
Covariance risk versus characteristic mispricing

This table reports results of robustness tests of the firm-level monthly Fama and MacBeth (1973) regression procedure to test whether the *RD-MV* and *RDC-HML* factor loadings are priced in the cross section of expected stock returns. Columns (1) and (2) ((3) and (4)) report the results for the *RD-HML* (*RDC-HML*) factor. To reduce the errors-in-variables problems, we assign pre-estimated portfolio level factor loadings to individual stocks. The portfolios used in the pre-estimation are 25 equally weighted size-*RD-MV* (or *RDC-MV*). The firm-specific variables include *RD-MV (RDC-MV)*, log size (*Ln(SIZE)*), log book-to-market equity ratio (*Ln(BM)*), past intermediate term cumulative returns (*RET(-12,-2)*), log turnover ratio (*Ln(TURN)*), stock market illiquidity (*ILLIQ*), log idiosyncratic volatility (*Ln(IVol)*), liquidity shocks (*LIQU*), maximum daily returns in the previous month (*MAX*), log analyst dispersions (*Ln(Disp)*) and short-term reversal (*REV*). Pre-estimated factor loadings on *MKT_RF*, *SMB*, *HML*, *UMD* and industry fixed effects constructed according to the Fama-French 49-industry classification are included throughout as further controls. We report the time series average coefficients, Newey and West (1987) robust standard errors and average R^2 for each model. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Independent variables	(1)	(2)	(3)	(4)
$\beta_{RD\text{-}HML}$	0.007***	0.009***	—	—
	(0.001)	(0.001)		
RD-MV	-0.001*	-0.000	—	—
	(0.001)	(0.001)		
$\beta_{RDC\text{-}HML}$	—	—	0.007***	0.009***
			(0.002)	(0.001)
RDC-MV	—	—	-0.000	0.000
			(0.000)	(0.000)
Ln(Size)	-0.001***	-0.000	-0.002***	-0.001
	(0.000)	(0.000)	(0.001)	(0.001)
Ln(BM)	0.002*	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
RET(-12,-2)	0.002	0.002	0.001	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Ln(TURN)	-0.002***	-0.001	-0.002**	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
ILLIQ	0.002**	0.011**	0.002**	0.014***
	(0.001)	(0.004)	(0.001)	(0.004)
Ln(IVol)	-0.171***	-0.368***	-0.187***	-0.393***
	(0.035)	(0.060)	(0.038)	(0.061)
β_{MKT_RF}	0.001	-0.003*	0.014***	0.011***
	(0.002)	(0.002)	(0.003)	(0.003)
β_{SMB}	-0.003*	0.002	-0.006***	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
β_{HML}	0.001	0.001	-0.002	-0.003
	(0.001)	(0.001)	(0.002)	(0.002)
β_{UMD}	-0.000	0.001	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
LIQU	—	0.016***	—	0.019***
		(0.005)		(0.005)

MAX	—	0.067***	—	0.073***
		(0.014)		(0.015)
Ln(Disp)	—	-0.003***	—	-0.002**
		(0.001)		(0.001)
REV	—	-0.058***	—	-0.055***
		(0.005)		(0.005)
Intercept	0.020***	0.019***	0.019**	0.010
	(0.007)	(0.006)	(0.008)	(0.008)
Industry FE	Y	Y	Y	Y
N	634,774	381,164	554,558	341,398
Avg. R^2	0.088	0.143	0.095	0.152

TABLE IA.6
The determinants of R&D factor loadings

This table reports the results of robustness tests of the pooled weighted least squares (WLS) regressions that investigate the determinants of stocks' exposure to the *RD-HML* factor. Columns (1) and (2) report the results using pre-estimated R&D factor loadings from a 12-month window. Columns (3) and (4) use a 24-month window and columns (5) and (6) use a 36-month window. The treatment of other variables is identical to that in Table 8. The independent variables include R&D expenditure to market value (*RD-MV)*, property, plant and equipment to total assets (*PPE-A*), Herfindahl-Hirschman index (*HHI*), Kaplan and Zingales' (1997) KZ index (*KZ index*), Altman's (1968) Z-score (*Z-score*), analyst forecast dispersion (*Ln(Disp)*), log market capitalization (*Ln(MCAP)*), return on assets (*ROA*), capital expenditure to total assets (*INV-A*) and log book-to-market equity ratio (*Ln(BM)*). Industry effects are constructed based on the Fama-French 49-industry classification and are included in each regression. Estimated coefficients, robust standard errors clustered at the firm level (in parentheses), the number of observations and adjusted R^2 are reported. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively.

Window	12-month		24-month		36-month	
	(1)	(2)	(3)	(4)	(5)	(6)
RD-MV	0.088***	0.084***	0.080***	0.097***	0.063***	0.082***
	(0.022)	(0.027)	(0.018)	(0.023)	(0.017)	(0.021)
PPE-A	-0.074**	-0.076**	-0.099***	-0.090***	-0.067**	-0.060**
	(0.031)	(0.032)	(0.025)	(0.026)	(0.026)	(0.027)
HHI	0.056	0.057	0.013	0.010	0.023	0.023
	(0.059)	(0.059)	(0.044)	(0.044)	(0.045)	(0.045)
Z-score	-0.040*	-0.040*	-0.041***	-0.042***	-0.041***	-0.042***
	(0.023)	(0.023)	(0.015)	(0.015)	(0.014)	(0.014)
KZ index	-0.001	-0.001	-0.004	-0.004	-0.022	-0.022
	(0.038)	(0.038)	(0.026)	(0.026)	(0.024)	(0.023)
Ln(Disp)	0.097**	0.097**	0.095***	0.097***	0.093***	0.095***
	(0.038)	(0.038)	(0.028)	(0.028)	(0.026)	(0.026)
Ln(MCAP)	-0.200***	-0.199***	-0.151***	-0.154***	-0.120***	-0.124***
	(0.028)	(0.028)	(0.022)	(0.022)	(0.022)	(0.022)
ROA	-0.260***	-0.260***	-0.269***	-0.265***	-0.300***	-0.296***
	(0.045)	(0.045)	(0.037)	(0.037)	(0.041)	(0.041)
INV-A	0.126***	0.127***	0.115***	0.110***	0.108***	0.102***
	(0.034)	(0.034)	(0.026)	(0.026)	(0.024)	(0.024)
Ln(BM)	—	0.006	—	-0.028	—	-0.031
		(0.028)		(0.021)		(0.021)
Intercept	0.157***	0.157***	0.088**	0.090**	-0.033	-0.037
	(0.059)	(0.059)	(0.044)	(0.044)	(0.388)	(0.383)
Industry FE	Y	Y	Y	Y	Y	Y
N	17,119	17,119	16,740	16,740	16,098	16,098
Adj. R^2	0.023	0.023	0.052	0.052	0.069	0.069

References (Internet Appendix)

Altman, E. I. (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* 23, 589-609.

Carhart, M. (1997) On Persistence in Mutual Fund Performance. *Journal of Finance* 52, 57-82.

Fama, E. F. and MacBeth, J. (1973) Risk, Return and Equilibrium: Empirical Tests. *Journal of Political Economy* 81, 607-636.

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989) A Test of the Efficiency of a Given Portfolio. *Econometrica*, 57, 1121-1152.

Kaplan, S. N. and Zingales, L. (1997) Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financial Constraints? *Quarterly Journal of Economics* 112, 169-215.

Newey, W. K. and West, K. D. (1987) A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55, 703-708.