The Value of Software

Collin Dursteler Roberto Gómez-Cram Alastair Lawrence*

October 9, 2022

Abstract

Software companies have steadily become key pillars of the digital economy, representing upwards of 12 percent of U.S. market capitalization. A simple buy-and-hold strategy of pure-play software companies over the past three decades produced annual alphas of over 7.1 percent. We document that these firms are growing at 13.9 percent annually and that both management and analysts systematically underestimate over a third of this growth. We show that these expectation errors appear to largely explain the foregoing outperformance of software companies and that management, analysts, short sellers, and other market participants ignore key performance indicators that describe these pure-play software firms and signal future growth. Together, the study underscores the value of software to the economy and how its economic impact has been significantly under-appreciated for the past two decades.

*We thank Itamar Drechsler and Julian Franks for helpful discussions and comments, and Ricardo Martinez for research-assistance support. We also thank the AQR Asset Management Institute for generous financial support. The authors are at the London Business School and can be reached at cdursteler@london.edu, rgomezcram@london.edu, and alastairlawrence@london.edu.

1 Introduction

Software is the set of instructions and programs that tells computers what to do (IBM, 2022). Over the past few decades, these instructions and programs have not only become integrally woven into every traditional industry, but they have also supported profound advances in the modern economy through important digital innovations. Marc Andreessen penned the now infamous phrase "software is eating the world" and stated, "My own theory is that we are in the middle of a dramatic and broad technological and economic shift in which software companies are poised to take over large swathes of the economy" (Andreessen, 2011). In 2020, software companies provided 15.8 million U.S. jobs and contributed a direct value-added GDP of $933 billion to the U.S. economy, representing a 15.2 percent growth in GDP from 2019 to 2020, which substantially outpaced the 2 percent growth of total U.S. GDP over the same period (Software.org, 2021). Further supporting the notion that software is continuing to eat the world, Satya Nadella, Microsoft CEO, stated in Microsoft's 2022Q3 earnings call that "Going forward, digital technology will be the key input that powers the world's economic output."

In this paper, we examine the importance of software to the economy and attempt to tease out whether market participants are adequately appreciating the dramatic technological shift that software is supporting and disrupting in the global economy. We speak to this broader question by empirically examining how stock market participants and analysts evaluate and price the performance of pure-play software companies. We categorize pure-play software companies as those whose main offering is selling some form of software. These firms can be easily identified using the FactSet's Revere Business Industry Classifications System. The majority of firms in our sample are B2B (business-to-business) software companies (e.g., Salesforce, Adobe, DataDog, Microsoft, etc.). Customer technology companies that have a software or software platform component such as Netflix, Uber, Apple, or Amazon are not included in our software categorization as their main offerings relate mostly to consumer goods, even if they may be supported by software platforms. We show that pure-play software companies have grown from about 0.7 percent of total U.S. market capitalization in 1990 to upwards of 12.7 percent in 2021, while only representing 5.7 percent of total listed companies. Additionally, the market capitalization of pure-play software companies is greater than that of any other Fama-French 49 industry capitalization and would make pure-play software the third largest industry if it were included in the Fama-French 20 industry classification.

Software companies benefit from an immense level of technological innovation. In terms of fundamentals, we highlight their previously unseen blend of growth combined with high gross margins and strong free cash flows. To evaluate how the market prices these companies, we begin by documenting the raw returns of software firms based on a simple buy-and-hold strategy starting in 1990 and ending in February 2022. We document that $1 invested in an equally-weighted portfolio of software companies in January 1990 translates to $405 return by 2022, whereas $1 invested in January 1990 in an equally-weighted portfolio of non-software companies translates to just $30.60 over the same time period. Hence, software company returns have outpaced those of non-software companies by more than 13 times. In terms of excess returns, we find annualized mean excess returns of over 19 percent for software and 10 percent for non-software companies. We find that classical asset pricing models can only explain a portion of these substantial returns. Specifically, we document Fama-French risk-adjusted annual returns of 7.16 and 0.96 percent for software and non-software companies, respectively, indicating that the excess returns of software companies are not explained by standard risk factors.

In an attempt to explain these alphas, we explore the notion that the market incorrectly assesses the growth rates of these companies. We find striking differences in analysts' abilities to forecast the growth rates of software and non-software companies. Specifically, we document systematic and persistent one-year ahead revenue growth forecast errors over the past two decades of 563 basis points for software companies, indicating that analysts systematically underestimate these firms' annual growth rates by over 40 percent on average. Conversely, we find one-year-ahead analyst forecast errors for non-software companies of -137 basis points, suggesting that analysts overestimate the annual growth rates of non-software companies, consistent with prior literature showing that such forecasts are on average overly optimistic for one-year-ahead forecast horizons (Bouchaud, Krüger, Landier, and Thesmar, 2019). We explore cross-sectional and time-series variation in these forecast errors and document that they are more pronounced for higher growth pure-play software companies and are also evident in the years following software company

IPO.

The persistent and systematic errors in software company growth rates appear to relate to the fact that analysts do not learn from their own previous forecast errors and from previously disclosed key firm performance indicators: remaining performance obligation (RPO) (i.e., the dollar amount of multi-year signed contracts) and net revenue retention. To investigate the possibility that analysts' forecasts herd around management's forecasts (Matsumoto, 2002; Cotter, Tuna, and Wysocki, 2006), we examine management one-quarter-ahead revenue growth forecast errors. We find that management forecast errors are greater than analyst errors and that management also systematically underestimates own-firm revenue growth rates, even when considering the most optimistic upper bound of their forecasts.

We next analyze the implications of our findings for stock return predictability. Using the NYSE Trade and Quote (TAQ) database, we document a large asymmetric price reaction to negative and positive revenue surprises within minutes after revenue announcements. In particular, we document a strong overreaction to negative revenue news: that is, initially large negative post-announcement returns drift which then reverse to significant positive abnormal returns in the subsequent three quarters. In contrast, for positive revenue surprises, we document a significant underreaction to information, as shown by persistently positive returns over several quarters, closely following the analyst forecast errors during the same period. Overall, the observed stock price drifts following revenue announcements supports the inference that stock market participants underestimate software company growth rates. In turn, this underestimation of growth largely explains these firms' outperformance of the market.

Finally, we examine whether informed traders appreciate the systematic underestimation of software company growth rates. Specifically, we show that short-seller positions in software companies do not vary with the magnitude of analyst growth forecast errors. In turn, we find that while short sellers are able to profit from their positions in non-software companies, they are unable to profit from their positions in software companies. Due to short-interest-related borrowing costs, it is likely that short sellers in fact lose money on short positions in software firms. These findings suggest that even investors who are perceived to be more informed under-appreciate the growth rate and scalability of these businesses.

This study directly illustrates the importance of pure-play software companies to the U.S. public markets and, in turn, the need to understand the unique operating and valuation nuances that distinguish these companies from traditional enterprises. More importantly, it highlights that the economy as a whole continues to be naive to both the importance of technological shifts resulting from advances in software and to the fact that software continues to take more and more of a leading role than is expected by market participants.

Moreover, as software creates and houses data, it is the key mechanism that is driving the production and value of digital data. In this regard, the study contributes to the recent and growing literature that speaks to the value of data (Farboodi and Veldkamp, 2021; Farboodi, Singal, Veldkamp, and Venkateswaran, 2022)—these studies argue that the most valuable firms in the United States are valued mostly for their data and their ability to use this data to create market power (Eeckhout and Veldkamp, 2022). While data and software are distinct concepts, in practice, digital data does not exist without software: it is nearly impossible to have one without the other. Because attempting to value data is inherently complex, an advantage of focusing on software is that its value and impact can be directly measured through companies that sell only software. In turn, we can identify these companies using granular industry classifications. That said, a portion of software firm market values will likely also reflect the data the firms hold and are privy to, which often supports the firm's future strategic directions.

Additionally, the paper also contributes to the literature examining analyst expectations and the predictability of their forecast errors. Specifically, our work adds to the growing literature on the role of subjective beliefs in asset pricing (Barberis, 2018; Bouchaud et al., 2019; Bordalo, Gennaioli, Porta, and Shleifer, 2019; Bordalo, Gennaioli, Ma, and Shleifer, 2020; Coibion and Gorodnichenko, 2015) and to recent papers that find predictable bias in macroeconomic variables (Bianchi, Ludvigson, and Ma, 2020), interest rates (Cieslak, 2018) and analyst expectations about future firm cash-flows (Gómez-Cram, 2022; van Binsbergen, Han, and Lopez-Lira, 2020).

2 The anatomy of pure-play software companies

2.1 Why is pure-play software different from other industries?

To illustrate why software companies are different, it is helpful to demonstrate the unique aspects of software business models and contrast them with those of traditional models. In Appendix A1, we compare the business models of Salesforce and Nike using somewhat hypothetical numbers to ease tractability while still capturing the essence of the two models. Appendix A1 begins with a customer buying a pair of trainers for $120 from Nike, one of the world's most successful apparel brands. It costs Nike $60 to make the trainers, which includes costs for materials, freight, insurance, duty, and any merchandiser fees. Next year, the customer might buy another pair of trainers from Nike, they might choose to buy from Adidas, or they might choose to not buy any trainers at all. Nike grows revenues at 7 percent year-over-year (YoY), implying that Nike's revenue will be 40 percent higher in five years' time.

Moving to Salesforce, a customer purchases a monthly software subscription for $10 per month, leading to the same annual revenue as Nike of $120. However, it only costs Salesforce $2 per month to sell each additional software subscription. Salesforce has a net revenue retention (NRR) rate of approximately 115 percent, which means each customer will likely spend $11.50 per month next year. Salesforce grows revenues at 30 percent YoY, implying that in five years' time, its revenues will have almost quadrupled. With gross margins of over 80 percent, many software companies choose to be unprofitable in the short-run (as Salesforce largely did in its first 20 years of operations) in order to heavily invest in long-term growth opportunities. This growth is in part financed through higher levels of sales and marketing and research and development expenses.

As these two examples illustrate, software firms have the potential for previously unseen levels of scalability and growth. Once a software product is created and operational, the software can be immediately deployed through cloud services to all users simultaneously (Govindarajan, Rajgopal, and Srivastava, 2018).[1] Additionally, any

[1] By way of example, in Snowflake's fiscal-year 2022 10-K report, they note that their platform can be "deployed across multiple public clouds and regions" and that they have the ability "to elastically scale up and scale down," emphasizing the "ease of deployment, implementation, and use" of its product offering. Additionally, in an Okta conference call on March 28, 2022, the CEO highlighted the speed of

product updates are deployed in the same manner. Hence, software companies typically have low levels of inventory and are able to avoid the logistical and input-pricing concerns typical of other manufacturing-based industries. Software firms' primary expenses are typically associated with research and development costs related to software offering development. Because the marginal cost of providing an additional unit software is minimal relative to the sales price, software companies' immense potential for scalability affords them previously unseen levels of gross margin. The increasing returns to scale on intangible investment of software companies not only affords high gross margins, it also supports high revenue growth. Software firms are able to avoid revenue bottlenecks associated with product or service supply issues. Therefore, revenue growth is mainly a function of demand, unlike traditional business models for which revenue growth is often constrained by supply. Software company revenues are persistent and are based on long-term subscription contracts which in part makes software firm revenue growth easier to predict than for traditional enterprises. While many are aware of the relative importance and market impact of software companies in the economy, the full extent of their uniqueness is generally discounted in both industry and academic contexts.

2.2 The growing importance of software companies

We next illustrate the increasing importance of pure-play software companies to the U.S. economy and capital markets over the past three decades. We identify pure-play software companies using FactSet's Revere Business Industry Classifications System (RBICS) as it is an extremely granular industry classification system. Appendix A2 includes the list of FactSet industries that we have used to identify the pure-play software companies included in our sample. We identify 457 pure-play software companies that were publicly listed over the period January 1990 through February 2022.

In order to speak to the growing importance of software companies, Figure 1 charts the total market capitalization of pure-play software firms scaled by total market capitalization over the sample period. The figure also plots the number of pure-play software firms as a percentage of all public firms. The total number of firms has been increasing at a relatively

software distribution via cloud, saying: "It just works. You turn it on. There's no deployment with cloud SaaS (software as a service) companies. So, simplicity is a big, big differentiator."

constant rate over the time period.[2] Over this period, the relative market capitalization of software companies increased from 0.7 percent in 1990 to 12.7 percent in 2022. Apart from the dot-com bubble, until 2008, the relative market capitalization of software firms grew largely in tandem with the proportion of software companies in the market. However, since 2008, while the trend in number of software companies has remained largely unchanged, software companies have consistently captured a disproportionately larger share of total market capitalization. As of 2022, software companies represent only 5.7 percent of the total number of market firms but account for 12.7 percent of total market capitalization. [3]

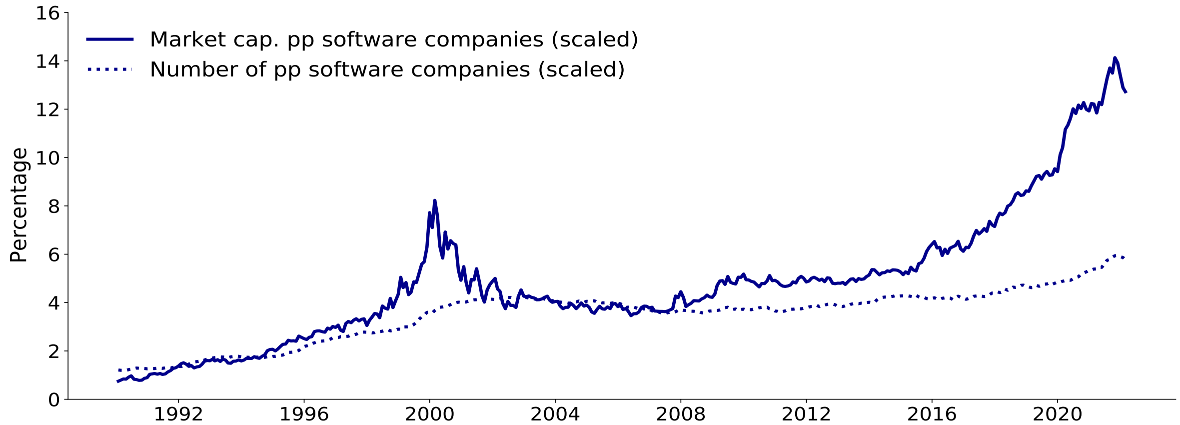
Figure 2 provides evidence on the frequency of software mentions in quarterly earnings calls of U.S. public companies between 2008 and 2022 to quantify the influence of the software industry on other public firms. We show the frequency of software mentions in quarterly earnings calls of U.S. public companies between 2008 and 2022. The figure reports the percentage of earnings calls including discussion of "software," "digital," or "cloud" related topics and illustrates that the frequency of such discussions increased from being mentioned in approximately 16 percent of earnings calls in 2008 to over 35 percent by 2022. While descriptive, these findings shed light on the increasing influence of the software industry on other industries.

Building on the inferences from the Nike and Salesforce business model comparison above, we examine the financial performance of software companies relative to the broader market. Specifically, we investigate how pure-play software enterprises compare to enterprises from other industries (defined following Fama and French's 49 Industry Portfolio specifications) along the dimensions of gross margin, revenue growth, and free-cash-flow margin (in the top, middle, and bottom panels of Figure 3). If the pure-play software industry portfolio were constructed as a new and distinct industry, it would

[2]Figure A1 in the Appendix charts the actual number of firms in each year over the same period. There were 118 software firms in 1990 and 390 in 2022.

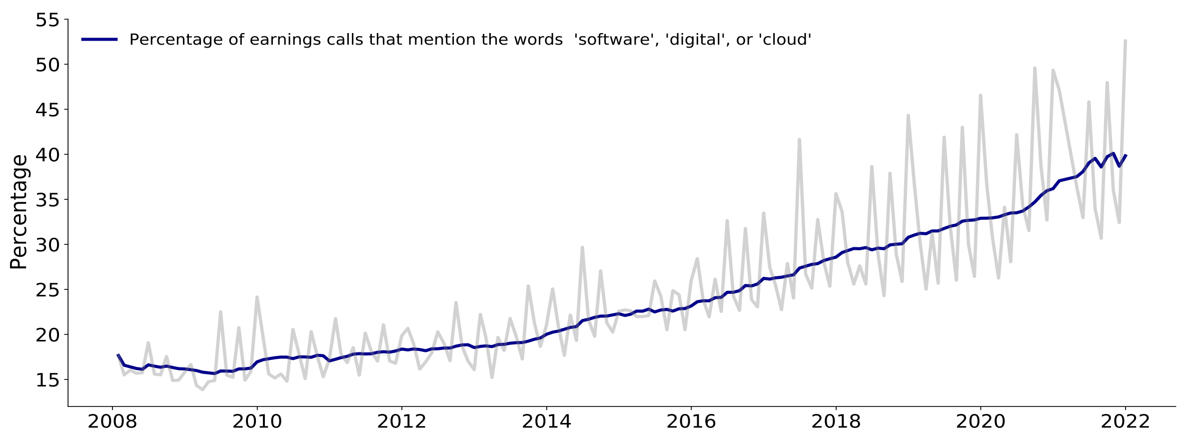
[3]Part of their rise since 2008 is due to the fact that software companies have benefited from an industry transition to cloud computing. Since the sub-prime crisis, more and more software companies have been able to further leverage the power of their software offerings through switching from perpetual licenses to cloud-based annual subscriptions. In line with the quotes from industry titans included above, not only has the software industry firmly established its position in the market, the increasing proportion of total market capitalization captured by software firms suggests that software is spreading its influence, impacting other industries as well.

Fig. 1. **Software firm market capitalization**



Notes: This figure shows the market capitalization of software firms scaled by the total NYSE, AMEX, and Nasdaq market capitalization. The dotted line charts the number of software firms scaled by the total number of market firms over the same period. The figure spans the period from January 1990 through February 2022.

Fig. 2. **Software mentions in earnings calls**



Notes: This figure shows the percentage of earnings calls that mention at least one of the following words: *software*, *digital*, or *cloud*. We present average values at both quarterly (gray line) and yearly (blue line) frequencies. The figure spans the period from January 2008 through February 2022.

rank first in terms of median gross margin (68.3 percent), third in year-over-year revenue growth (13.9 percent), and fifth in free-cash-flow margin (7.16 percent). No other industry has such a strong combination of gross margins, revenue growth, and free cash flows over the period from 1990 through 2022, suggesting that their unique combination of technological innovation, scalability, and growth clearly differentiates software companies

Table 1. **Proportion of software companies in portfolio sorts**

Panel A: Book-to-market portfolios									
Value	2	3	4	5	6	7	8	9	**Growth**
2.84	3.87	4.73	4.43	5.77	7.22	9.18	11.69	18.88	31.39
Panel B: Size portfolios									
Small	2	3	4	5	6	7	8	9	**Large**
35.01	15.11	10.08	8.89	7.43	6.06	4.99	5.44	4.03	2.95
Panel C: Momentum portfolios									
Losers	2	3	4	5	6	7	8	9	**Winners**
11.48	12.17	9.95	8.20	7.20	6.98	7.47	9.17	11.67	15.72
Panel D: Revenue growth portfolios									
Low	2	3	4	5	6	7	8	9	**High**
7.10	8.10	8.61	8.45	9.99	11.94	13.51	14.07	11.12	7.10

Notes: This table reports the proportion of software firms that would be allocated to Fama-French value, size, and momentum-sorted decile bins (Panels A–C) and decile bins sorted based on revenue growth (Panel D). Portfolio sortings are based on annual data and decile proportions are presented in percentages. Panel A portfolios are sorted based on quarter-end book-to-market ratios. Panel B portfolios are sorted based on quarter-end market capitalization. Panel C portfolios are sorted based on quarter-end return momentum. Panel D portfolios are sorted based on quarter-end year-over-year revenue growth. The figures spans the period from January 1990 through February 2022.

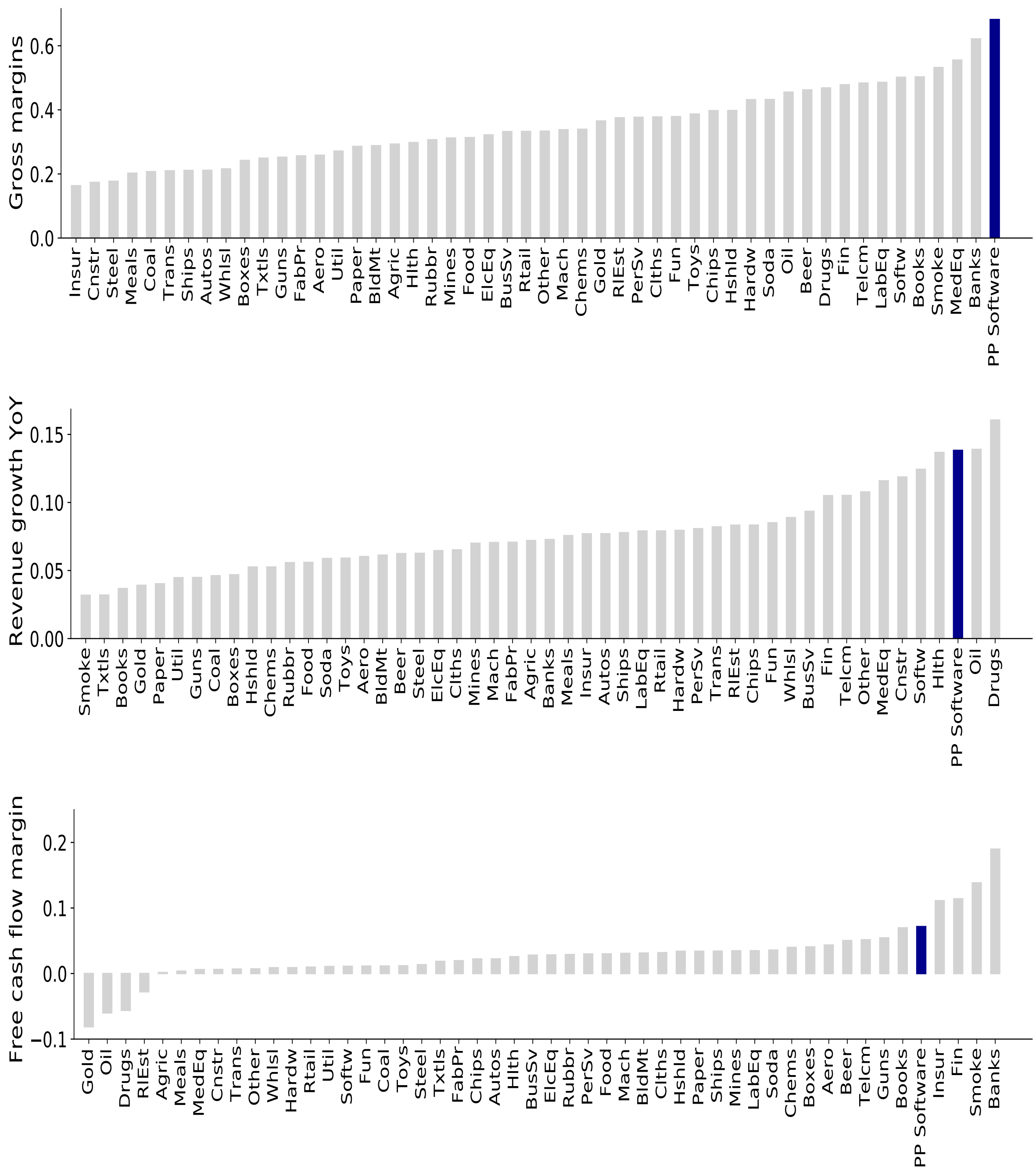
from more traditional enterprises.[4]

We next sort software companies based on commonly used risk factors, presented in Table 1, to obtain a better understanding of their risk profiles. Panels A through C show the portfolio sorts of software firms relative to the Fama-French book-to-market, size, and momentum portfolios. Panel D shows the portfolio sorts of software firms relative to revenue growth portfolios for all publicly-listed firms. Firms are sorted every quarter based on their quarter-over-quarter revenue growth. Portfolios in all panels are rebalanced monthly.

We find that software firms are generally high-growth firms: 61.9 percent of software firms fall into the three highest book-to-market portfolio bins. In terms of size, software firms are generally smaller firms: 69.1 percent of software firms fall into the four smallest size-sorted decile bins. Regarding return momentum, the distribution of software

[4]The Fama-French 49 Software industry classification (Softw), which is based on SIC codes, misses around one quarter of the pure-play software companies we have identified using FactSet's detailed industry classification and includes many other equipment and non-software companies whose main operations do not constitute developing and distributing software.

Fig. 3. **Median gross margin, revenue growth, and free-cash-flow margin**



Notes: This figure ranks the median quarterly gross margin, year-over-year revenue growth, and free-cash-flow margin for each Fama-French 49 Industry classification and pure-play software companies (highlighted in blue) in the top, middle, and bottom panels, respectively. Gross margin is calculated as the fiscal-quarter gross margin (revenue minus cost of goods sold) over total fiscal-quarter revenue. Revenue growth is calculated using annual revenue growth as of each fiscal quarter. Free-cash-flow margin is calculated as fiscal-quarter free cash flows (operating cash flows minus capital expenditures) over total fiscal-quarter revenues. The sample spans the period from January 1990 through December 2020. Variables are winsorized at the 1^{st} and 99^{th} percentiles.

companies is a bit more varied. Software firms are generally skewed both left and right, indicating that software firms exhibit both high and low return momentum, but are less likely to fall in the middle momentum portfolios. Finally, consistent with earlier industry rankings, we find that software firms are skewed towards higher revenue growth portfolios with 57.7 percent of firms falling within the five highest revenue growth portfolios. Taken together, software firms are generally smaller and high-growth firms, with both extreme winners and losers. In the following section, we consider how this unique combination of factors may contribute to the relative financial performance of software firms over our sample period.

2.3 Portfolio performance

In this section, we consider the relative market performance of software firms compared to other firms. The foregoing analyses suggest that software companies have unseen levels of innovation and performance which have driven their financial performance. If these firms offer unparalleled growth opportunities, we would expect this potential to be reflected in market returns if this growth was not fully anticipated by the market.

We first look at the mean excess returns, presented in Panel A of Table 2, and compare pure-play software portfolio returns against those of a non-software portfolio. While we present results for both equal- and value-weighted portfolios, we will focus our discussion here on the equal-weighted portfolio results. Focusing on the equal-weighted portfolio results addresses the potential concern that returns may be driven by a few large firms. The results are estimated over the period from 1990 through 2022. An equally weighted portfolio of software companies earned an annualized mean excess return of 20.5 percent with an annualized standard deviation of 28.9. In comparison, the non-software company portfolio earned an annualized return of 10.0 percent with an annualized standard deviation of 18.8. The difference in returns, 10.5 percent, is economically significant, as pure-play software firms substantially outperformed non-software firms over the 1990–2022 period. These results imply a software portfolio Sharpe ratio (relative to the market return) of 23.3 percent, compared to the non-software portfolio Sharpe ratio of -7.2 percent. So, when adjusting for risk, non-software companies actually earned

Table 2. **Performance evaluation and alphas**

Panel A: Performance evaluation	Software companies		Non-software companies	
Portfolio weights	Equal (1)	Value (2)	Equal (3)	Value (4)
Mean excess returns	20.49	15.53	10.01	8.27
Std. dev. returns	28.91	25.28	18.76	14.84
Sharpe ratio	23.27	6.88	-7.19	-3.03
Mean max. drawdowns	10.61	21.41	6.86	7.44
Skewness	0.27	-0.00	-0.27	-0.70

Panel B: Alphas	Software companies		Non-software companies	
Portfolio weights	Equal (1)	Value (2)	Equal (3)	Value (4)
Market	7.16	3.42	0.96	-0.44
	[2.59]	[2.82]	[0.65]	[-1.46]
FF3	8.73	5.44	0.31	-0.64
	[3.78]	[2.98]	[0.28]	[-2.12]
FF3 + Mom	10.47	5.24	2.17	-0.48
	[4.05]	[2.83]	[1.91]	[-1.49]
FF3 + Mom + CF(4)	13.57	7.71	3.04	-0.68
	[4.71]	[3.84]	[2.44]	[-2.20]

Notes: This table presents performance evaluation measures and alphas for software and non-software portfolios. Returns are presented for both equal- and value-weighted portfolios. Panel A reports measures of portfolio performance. Mean excess returns are calculated using returns in excess of risk-free rates obtained from French's website. Sharpe ratios are presented as percentages and are calculated relative to the total market return. Panel B presents portfolio alphas based on the market factor, Fama-French three-factors (FF3), Moskowitz, Ooi, and Pedersen (2012) momentum factor (Mom), and four cash-flow factors (CF(4)). These cash-flow factors capture earnings persistence (Francis, LaFond, Olsson, and Schipper, 2004), sales growth (Lakonishok, Shleifer, and Vishny, 1994), profit margins (Soliman, 2008), and change in gross margin minus change in sales (Abarbanell and Bushee, 1998). t-statistics are reported in brackets. The table spans the period from January 1990 through February 2022.

negative returns over the same period.

Panel B of Table 2 presents the alphas for both software and non-software portfolios. The alphas are presented for the regression of portfolio returns on the market; on Fama and French's three factors (FF3); on FF3 and the Moskowitz et al. (2012) momentum factor (Mom); and on FF3, Mom, and four pricing factors related to firm cash flows. Specifically, we consider the earnings persistence factor of Francis et al. (2004), the sales growth factor of Lakonishok et al. (1994), the profit margin factor of Soliman (2008), and the change in gross margin minus change in sales factor of Abarbanell and

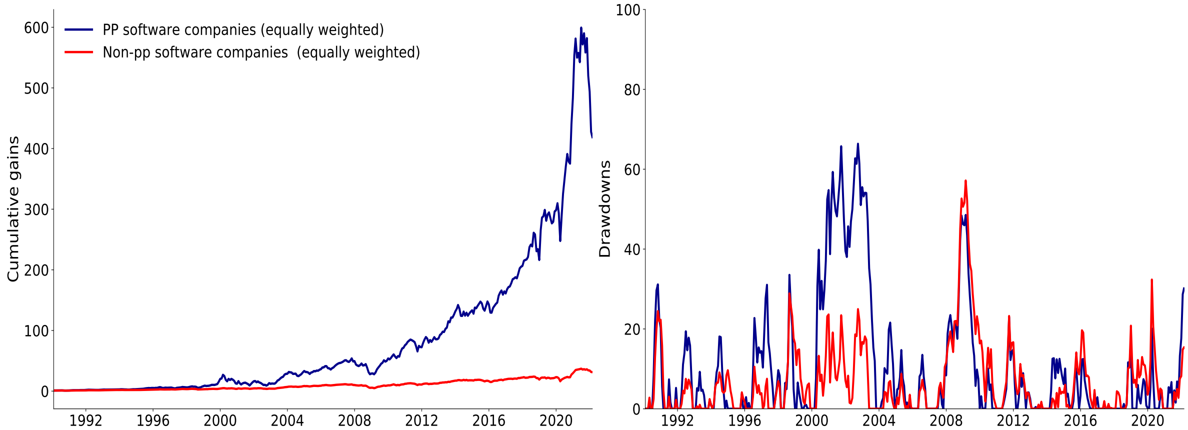
Bushee (1998).[5] In each specification, an equally-weighted portfolio of software companies consistently earns large annualized alphas ranging from 7.16 percent (t-statistic = 2.59) to 13.57 percent (t-statistic = 4.71), depending on the factor specification. In comparison, non-software portfolio alphas are mostly close to zero and statistically insignificant, except in the $FF3 + Mom + CF(4)$ specification. Taken together, even when controlling for traditional risk factors, the portfolio of software companies significantly outperforms its non-software company counterpart over the period from 1990 to 2022. Moreover, traditional asset pricing models fail to explain software company returns over the past three decades.

To confirm the findings in Table 2 and further highlight the relative outperformance of software companies, we present the cumulative gains and total drawdowns for both software and non-software portfolios. The left panel of Figure 4 charts the buy-and-hold equally weighted cumulative gains of a \$1 investment in both portfolios from 1990 to 2022. Both strategies invest \$1 at the beginning of 1990 and close their position at the end of February 2022. Comparing the relative performance of both portfolios, gains on software companies were \$418.84, while non-software companies gained \$31.37 over the same period. Again, we see a large disparity in the cumulative dollar gains from a simple buy-and-hold investment in the software company portfolio. The right panel of Figure 4 charts portfolio drawdowns. Except during the 2000–2004 period (the aftermath of the dot-com bubble) and the beginning of 2022, software company portfolio drawdowns are largely similar to those of the non-software company portfolio. As presented in Panel A of Table 2, average software company drawdowns are 10.61 percent and average non-software company drawdowns are 6.84 percent, suggesting that software companies have a greater crash risk than non-software companies, and that drawdowns are a risk associated with the software industry. Figure A3 of the Appendix presents the same investment strategies using value-weighted portfolios, and the results are generally consistent with those presented in Figure 4.

In Tables A4 and A5 in the Appendix, we examine whether the outperformance of software companies is robust to different time periods. We split the sample period in

[5] We downloaded the data from `https://jkpfactors.com/` and thank Jensen, Kelly, and Pedersen (2021) for making these data publicly available.

Fig. 4. **Cumulative gains and drawdowns**



Notes: The left panel plots equally weighted cumulative monthly buy-and-hold returns for software firms (blue line) and non-software firms (red line), respectively. The right panel depicts the drawdowns associated with each strategy. The sample spans from January 1990 through February 2022.

December 2002 to see whether the outperformance continues following the dot-com bubble. Mean excess return results are largely unchanged in both sub-periods. However, Sharpe ratios are significantly different for portfolios in both sub-periods. The average software company portfolio Sharpe ratio is 49.6 percent from 1990–2002 and 25.4 percent from 2003–2022. Additionally, the average non-software company portfolio Sharpe ratios are 31.8 percent during the first sub-period and -17.3 percent during the second. Comparing portfolio alpha results, software company portfolio alphas are larger during the first sub-period and, while still positive and significant, are lower during the second sub-period. On the other hand, only the first sub-period $FF3 + Mom$ and $FF3 + Mom + CF(4)$ average portfolio alphas are significant for the non-software portfolio. These results suggest that while software companies outperform non-software companies in both sub-periods, the outperformance is more pronounced pre-2002. The reduction in outperformance post-2002 suggests that the market may be raising its expectations of software company performance.

3 Analyst and management expectations

3.1 Analysts systematically underestimate revenue growth rates

As shown in the previous section, the pure-play software industry has grown to be one of the most important industries in the market. Moreover, these companies have experienced pronounced gains and alphas that are not fully explained by standard risk factors or traditional asset pricing models. In this section, we examine potential explanations for these firms' dramatic market outperformance. We begin by examining analyst growth forecasts to understand the market's perception and assessment of these companies and the impact these expectations may have on software firm pricing.

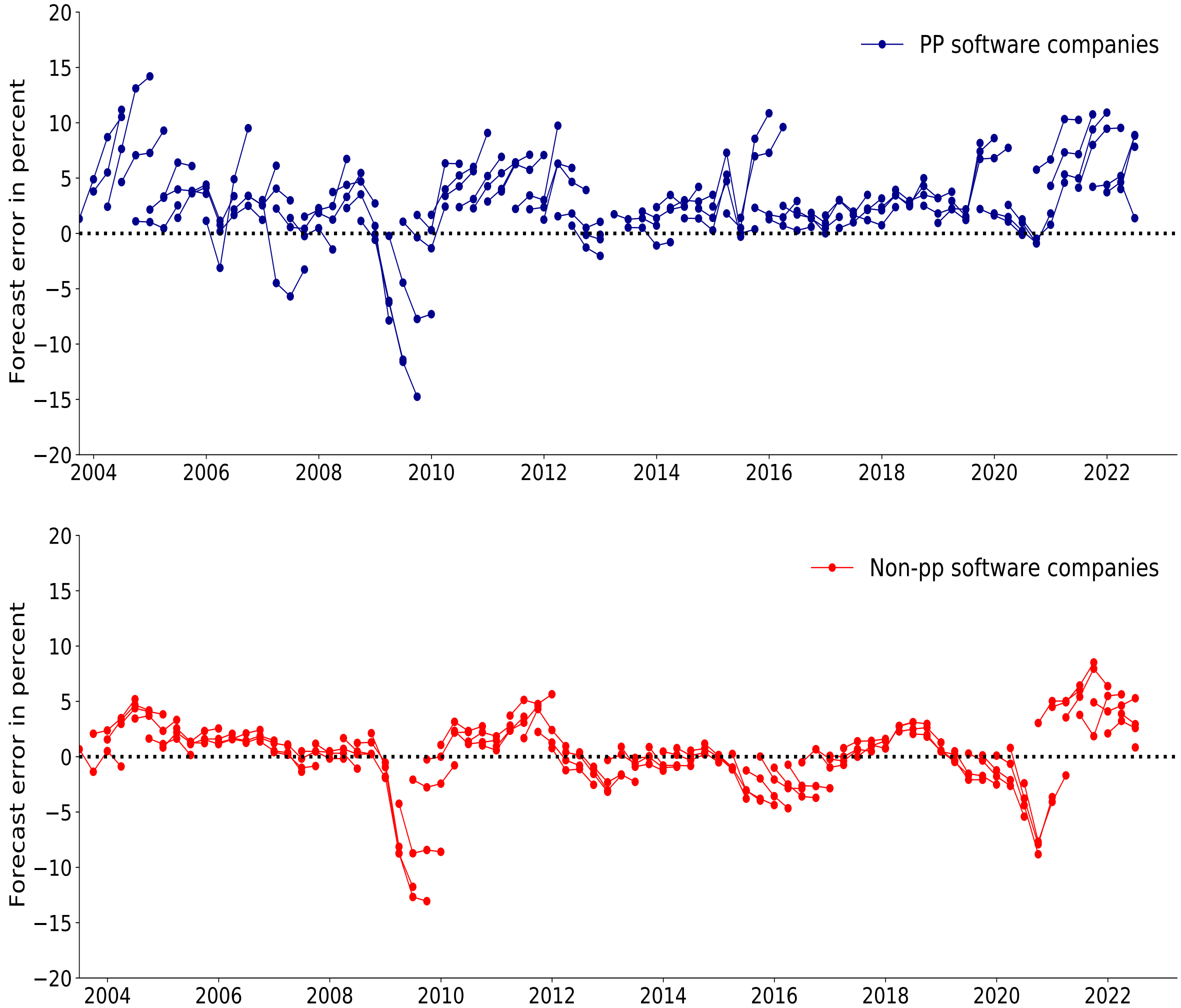
We first compare the relative term structure of cumulative analyst revenue forecast errors for software and non-software companies, presented in Figure 5. We define forecast errors as the difference between realized revenue growth and analyst forecast growth, where a positive value indicates that the consensus analyst forecast was too low relative to observed values. Measured for every quarter t between 2003Q1 and 2022Q1, each string in the figure represents the average conditional cumulative analyst revenue forecast error for the subsequent $t+h$-quarter-ahead forecast, where $h = 1, 2, 3, 4$.[6] At each quarterly earnings announcement date, analysts make forecasts of one-, two-, three-, and four-quarter-ahead revenue. To construct our measure of analyst expectations, we obtain analyst revenue forecasts from the I/B/E/S Detail History File. See Figure A4 in the Appendix for a more detailed explanation of the forecast timeline and our process of calculating the cumulative forecast errors presented in Figure 5.

The results in Figure 5 show marked differences between software and non-software cumulative forecast errors. Perhaps most strikingly, the forecast errors of software companies are consistently positive whereas the forecast errors for non-software companies generally oscillate between -5 and 5 percent over the course of our sample. The forecasts errors for both groups were abnormally negative following the 2008 financial crisis and somewhat more positive during the COVID-19 pandemic.[7] Additionally, the cumulative

[6] We begin our analysis in 2003 as the availability of analyst forecast data before this time is more limited.

[7] The result indicating negative forecast errors around the time of the 2008 recession is consistent with results presented in Gómez-Cram (2022), who shows that analysts are too optimistic at the onset

Fig. 5. **Conditional cumulative forecast errors**



Notes: This figure presents the term structure of cumulative forecast errors for every fiscal quarter-end in the sample for software and non-software companies. At each quarter t, h-quarter-ahead mean consensus forecasts are calculated based on all analyst forecasts made within 15 days of earnings announcement for forecast horizons $h = 1, 2, 3, 4$. Each point in the figure represents the cumulative percentage forecast error at each forecast horizon. The sample spans from 2003Q1 to 2022Q1.

four-quarter-ahead forecast errors of non-software companies are on average negative. However, software company cumulative forecast errors are consistently positive for each quarterly forecast horizon and are much greater than those for their non-software counterparts. These findings show that software companies have consistently outperformed analyst expectations and that their outperformance on average increases with the quarterly forecast horizon. In turn, they provide initial evidence that analysts systematically

of recessions.

Table 3. **Properties of forecasts errors**

	Software companies				Non-software companies			
	Forecast horizon in quarters							
	One	Two	Three	Four	One	Two	Three	Four
Mean	2.34	3.20	4.00	5.63	0.74	0.34	-0.15	-1.37
t-stat	[31.14]	[5.49]	[5.10]	[4.89]	[40.59]	[14.55]	[-5.22]	[-39.82]
Std. dev.	6.62	50.99	66.58	95.48	8.07	10.14	11.71	13.72
5%	-5.20	-12.53	-18.06	-22.04	-11.27	-16.60	-20.61	-29.13
25%	-0.01	-2.01	-3.76	-4.92	-1.96	-3.89	-5.53	-7.78
50%	1.83	1.85	1.73	1.85	0.67	0.22	-0.05	-0.82
75%	4.33	5.89	7.14	8.74	3.69	4.66	5.24	5.66
95%	11.22	17.03	22.51	28.61	13.06	17.62	20.33	21.93
nObs	7,787	7,645	7,223	6,869	193,378	189,290	169,998	158,766

Notes: This table presents summary statistics of software and non-software company quarterly forecast errors for each $t + h$ forecast horizon, where $h = 1, 2, 3, 4$. At each quarter t, h-quarter-ahead mean consensus forecasts are calculated based on all analyst forecasts made within 15 days of earnings announcement dates. The sample spans from 2003Q1 to 2022Q1.

underestimate the growth rates of software companies.

In Table 3, we present summary statistics of the same $t + h$-quarter-ahead forecasts presented in Figure 5. The mean software company forecast error monotonically increases in the forecast horizon, from a mean value of 2.34 (t-statistic = 31.14) at a one-quarter horizon to 5.63 (t-statistic = 4.89) at an annual horizon. This pattern is in contrast to non-software mean forecast errors, which monotonically decrease from 0.74 (t-statistic = 40.59) to -1.37 (t-statistic = -39.82) for the same one-quarter and annual horizons, respectively. Additionally, these results confirm the initial inferences from Figure 5 that the mean forecast errors of software companies are cumulatively positive and, in turn, that analyst forecasts fail to accurately predict growth, suggesting one potential reason for the outperformance of software stocks. The left panel in Figure 6 illustrates these trends in graphical form.

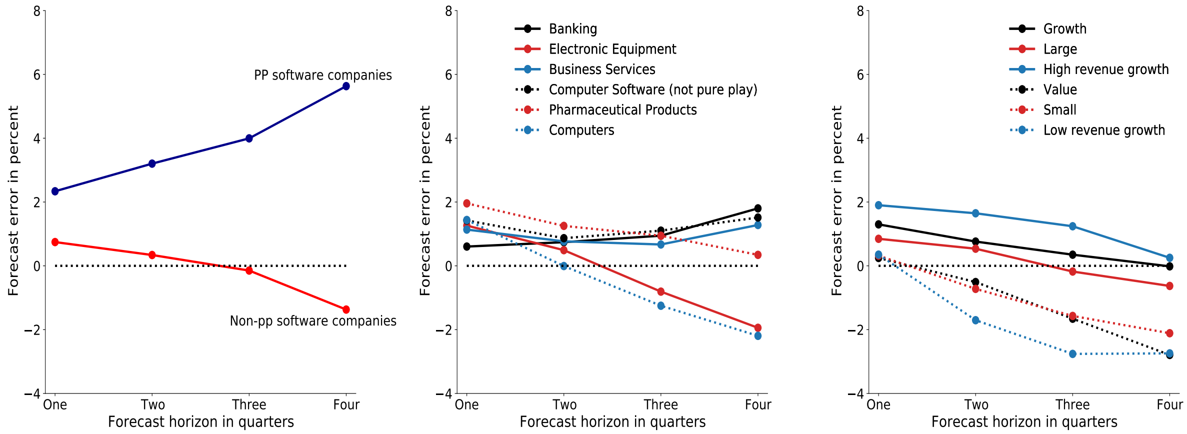
In the following analysis, presented in the center and right panels of Figure 6, we explore whether other industries possessing characteristics comparable to those of pure-play software companies also consistently outperform analyst expectations. While not perfect peer comparisons, we examine forecast errors in the banking, electronic equipment, business services, non-pure-play software, pharmaceutical, and computer industries (based on Fama-French 49 Industry Portfolio specifications) as these industries share some similarities with the pure-play software industry.[8] For example, some of these industries also have traditionally high gross margins (e.g., banking) or have highly scalable business models (e.g., pharmaceutical).

In the center panel of Figure 6, we chart the cumulative quarterly forecast errors for the industries noted above. While the banking, business services, and non-pure-play software industries exhibit cumulatively positive forecast errors, the magnitude of error is significantly less than that of pure-play software forecast error (reported in Table 3). Additionally, the pharmaceutical, electronic equipment, and computer industry cumulative forecast errors are negative at the annual horizon, which is consistent with the patterns reported for non-software companies. Hence, the plots in Figure 6 clearly illustrate the unparalleled magnitude of the pure-play software industry's outperformance of analyst expectations.

We next consider whether portfolios constructed along other dimensions might be able to mimic the consistent forecast errors observed for software companies. In the right panel of Figure 6, we compare the forecast errors for portfolios constructed based on size, book-to-market, and revenue growth properties. Both the size and book-to-market portfolios are based on Fama-French specifications. The revenue growth portfolio is constructed based on quarter-over-quarter revenue growth rebalanced monthly. As shown in Table 1, software companies are generally smaller companies with higher growth, in terms of book-to-market ratio and revenue. Hence, it is reasonable to expect that portfolios constructed along these same dimensions might exhibit similar forecast error patterns to those of software companies. However, as seen in the right panel, this is not the case: five of the six specified portfolios have negative cumulative forecast errors, while only the high revenue growth portfolio exhibits marginally positive forecast errors, yet to nowhere near the same extent as those for software companies. If anything, the portfolio errors exhibit similar trends to non-software company forecast errors. Analysts appear to systematically underestimate only pure-play software company performance and not the

[8] We note that the charted forecast errors for the six Fama-French portfolios do not include the effect of the pure-play software firms in our sample. We have removed these firms from their respective industry portfolios and include them in a separate pure-play software industry portfolio.

Fig. 6. **Unconditional cumulative forecast errors**



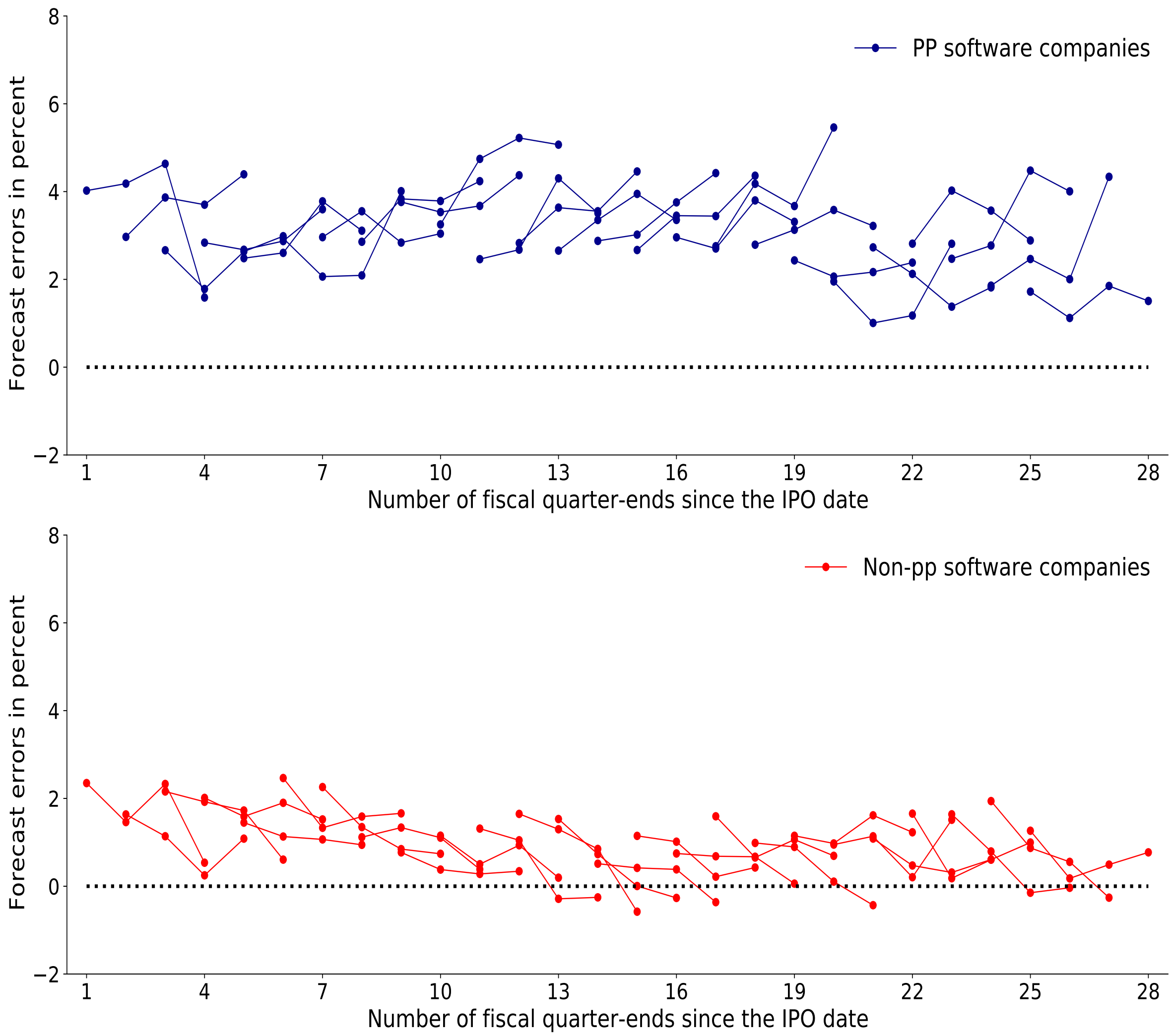
Notes: This figure shows the unconditional term structure of cumulative forecast errors. As above, at each quarter t, h-quarter-ahead forecasts are made for $h = 1, 2, 3, 4$. The left panel presents the unconditional average cumulative forecast error. The central panel shows the unconditional forecast error structure for a selection of Fama-French 49 Industries. The right panel presents unconditional average forecast errors for value-growth, small-large, and high-low revenue growth sorted portfolios. The figure spans the period 2003Q1 to 2022Q1.

performance of peer industries or of portfolios constructed based on attributes similar to those of pure-play software companies.

Finally, we examine whether an age-related factor might explain the disparity in observed forecast errors. Figure 7 plots the the difference in analyst revenue forecast errors between software and non-software companies for the 28 fiscal quarters following a firm's initial public offering (IPO). As software companies have a greater proportion of younger firms than public markets as a whole, it could be that analysts lack a sufficiently long time series of data on which to base their forecasts.

Figure 7 rules out the notion that an age-related factor may be responsible for the systematic underestimation of software company growth rates. The figure illustrates that, while there is a slight downward trend in software company forecast errors over time, the forecast errors for each quarterly horizon remain consistently positive and are of similar magnitude to those in Figure 5, even after seven years of public listing. However, the same trend is not observed for non-software companies: while the initial forecast errors for all four quarterly forecast horizons are positive for the first few years after IPO, the magnitude of error is far lower than that of software company forecast error. Furthermore, over time, the forecast errors of non-software companies decrease and are

Fig. 7. **Post-IPO average cumulative forecast error**



Notes: This figure presents the term structure of cumulative forecast errors since firm IPO date. At each quarter t, h-quarter-ahead mean consensus forecasts are calculated based on all analyst forecasts made within 15 days of earnings announcement, for forecast horizons $h = 1, 2, 3, 4$. Each point represents the cumulative forecast error for h-quarter-ahead forecasts made at each period t following firm IPO. The sample spans the period from 2003Q1 to 2022Q1.

generally centered around zero, also consistent with the inferences from Figure 5. The fact that initially positive forecast errors ameliorate over time suggests analyst learning for non-software companies; while analysts may become more familiar with software companies, they continue to systematically under-estimate their growth rates even as the firm matures.

3.2 Predictability of forecast errors

In Section 3.1 above, we show that analysts have systematically underestimated the annual growth rates of software companies over the past two decades. We next investigate whether these forecast errors are predictable and, in turn, the potential implications of this predictability for future stock returns. Using the following set of regressions, we examine whether lagged analyst forecast errors and revenue growth predict future realized cumulative revenue growth (Equation (1a)), analyst estimates of cumulative revenue growth (Equation (1b)), and resulting analyst forecast errors (Equation (1c)).

$$\Delta Revenue_{t,t+h} = \rho_0^e + \rho_1^e \Delta Revenue_t + \rho_2^e FE_t + \epsilon_{t+h}^e \tag{1a}$$

$$E_t^s \left[\Delta Revenue_{t,t+h}\right] = \rho_0^s + \rho_1^s \Delta Revenue_t + \rho_2^s FE_t + \epsilon_{t+h}^s \tag{1b}$$

$$\Delta Revenue_{t,t+h} - E_t^d \left[\Delta Revenue_{t,t+h}\right] = \rho_0^d + \rho_1^d \Delta Revenue_t + \rho_2^d FE_t + \epsilon_{t+h}^d \tag{1c}$$

Equation (1a) regresses the realized cumulative revenue growth for quarter $t+h$ on the revenue growth and revenue growth forecast error for quarter t. Equation (1b) replaces realized revenue growth in Equation (1a) with quarter-t consensus analyst cumulative revenue growth forecast for quarter $t+h$, retaining the same control variables. Equation (1c) follows the foregoing structure but replaces the dependent variable with the realized quarter-$t+h$ analyst forecast error based on quarter-t expectations. Using this set of equations, we are able to examine the extent to which trailing revenue growth and forecast error relate to future actual revenue growth, the degree to which analysts appreciate these relations, and the predictability of analyst forecast errors.

For brevity, Table 4 presents regression estimates only for software companies. The regression estimates for non-software companies are included in Table A6 of the Appendix. For each regression equation, we present the cumulative regression estimates for each $t+h$-quarter-ahead horizon for $h = 1, 2, 3, 4$. Panel A presents regression estimates for realized cumulative revenue growth values (Equation (1a)), Panel B presents estimates for analyst forecasts of cumulative revenue growth (Equation (1b)), and Panel C provides estimates for the resulting cumulative forecast errors (Equation (1c)).

First, Panel A of Table 4 shows that lagged revenue growth in quarter t is negatively

Table 4. **Predictability of realized revenue, analyst expectations, and forecast errors**

	Panel A: Realized values				**Panel B**: Analyst expectation				**Panel C**: Cumulative forecast error			
	Forecast horizon in quarters											
	One	Two	Three	Four	One	Two	Three	Four	One	Two	Three	Four
Constant	4.64	8.83	13.92	14.74	3.21	7.89	13.17	13.66	1.43	0.94	0.76	1.08
	[8.36]	[15.61]	[17.68]	[18.81]	[6.37]	[17.70]	[21.16]	[24.02]	[10.14]	[3.26]	[1.77]	[2.07]
$\Delta Revenue_t$	-0.13	-0.24	-0.58	0.09	-0.12	-0.24	-0.61	0.08	-0.01	0.00	0.03	0.00
	[-2.97]	[-4.57]	[-8.67]	[2.14]	[-2.71]	[-4.90]	[-9.69]	[2.49]	[-1.79]	[0.33]	[2.35]	[0.22]
FE_t	0.21	0.47	1.21	1.00	-0.23	-0.07	0.52	0.27	0.44	0.53	0.69	0.73
	[1.13]	[1.45]	[9.23]	[7.48]	[-1.22]	[-0.37]	[4.66]	[2.69]	[14.07]	[3.68]	[9.90]	[7.67]
R^2	0.02	0.04	0.16	0.06	0.03	0.06	0.26	0.03	0.17	0.07	0.06	0.04
nObs	7,711	7,371	6,943	6,538	7,711	7,371	6,943	6,538	7,711	7,371	6,943	6,538

Notes: This table reports software firm linear regression coefficient estimates and t-statistics for the following equations for each forecast horizon $h = 1, 2, 3, 4$:

$$\Delta Revenue_{t,t+h} = \rho_0^e + \rho_1^e \Delta Revenue_t + \rho_2^e FE_t + \epsilon_{t+h}^e$$
$$E_t^s \left[\Delta Revenue_{t,t+h}\right] = \rho_0^s + \rho_1^s \Delta Revenue_t + \rho_2^s FE_t + \epsilon_{t+h}^s$$
$$\Delta Revenue_{t,t+h} - E_t^d \left[\Delta Revenue_{t,t+h}\right] = \rho_0^d + \rho_1^d \Delta Revenue_t + \rho_2^d FE_t + \epsilon_{t+h}^d$$

in Panels A–C, respectively. Dependent variables $\Delta Revenue_{t,t+h}$, $E_t^s[\Delta Revenue_{t,t+h}]$, and $\Delta Revenue_{t,t+h} - E_t^d\left[\Delta Revenue_{t,t+h}\right]$ denote the h-quarter-ahead expected revenue, realized revenue, and the difference between the two, respectively. Independent variables, $\Delta Revenue_t$ and FE_t, represent the change in revenue from the prior period and the prior-period forecast error, respectively. Panel C is the difference between Panels A and B and represents the forecast error for each horizon. The table spans the period from January 1990 through February 2022.

associated with the one-, two-, and three-quarter-ahead cumulative revenue growth but is positively associated with the four-quarter-ahead cumulative revenue growth. Specifically, the coefficient estimates on $\Delta Revenue_t$ are -0.13, -0.24, and -0.58 for the one-, two-, and three-quarter-ahead cumulative revenue growth, respectively: all three coefficients are significant at the one-percent level. However, the coefficient on $\Delta Revenue_t$ for the four-quarter-ahead revenue growth is 0.09 and is significant at the five-percent level. These findings suggest that period-$t+h$ revenue growth for software companies mean reverts over the immediately subsequent quarters but then increases after one year. Looking at the lagged forecast errors, the coefficients on FE_t are positive for all one- to four-quarter-ahead cumulative revenue growth estimates. However, the estimates are statistically significant for only the three- and four-quarter-ahead realized revenue growth estimates, with coefficient values of 1.21 and 1.00, respectively. In terms of economic significance, the foregoing coefficients for the three- and four-quarter-ahead realized revenue growth horizons imply that 121 and 100 percent of the quarter-t forecast error persists into realized revenue growth three and four quarters ahead. These findings suggest that software firm growth surprises in quarter t are durable and appear to persist well into future quarters.

In Panel B of Table 4, we examine whether analyst forecasts reflect an appreciation of the relation between prior-period revenue growth and revenue growth forecast errors and quarter-$t+h$ revenue growth. The results indicate that analyst forecasts almost perfectly anticipate the relation between past and future revenue growth observed in Panel A. Specifically, the coefficient estimates for $\Delta Revenue_t$ of -0.12, -0.24, -0.61, and 0.08 for the one-, two-, three-, and four-quarter-ahead forecast horizons, respectively, almost perfectly reflect the relation noted in Panel A between lagged revenue growth in quarter t and quarter-$t+h$ cumulative revenue growth. On the other hand, analyst forecasts only partially reflect the durability of software firm revenue growth: prior-period analyst forecast errors do not predict current-period forecasts to the extent of the realized relation. The coefficient estimates for FE_t are lower than the observed relation in Panel A between lagged forecast errors in quarter t and quarter-$t+h$ cumulative revenue growth for all forecast horizons. These findings provide initial evidence that analyst forecasts do not adequately reflect the future durability or persistence of software company growth surprises observed at time t (especially for the three- and four-quarter-head horizons).

Panel C of Table 4 presents regression results with the cumulative forecast error as the dependent variable. Panel C confirms the combined inferences from Panels A and B. Panel C shows that analyst forecasts correctly anticipate the implications quarter-t revenue growth has for cumulative revenue growth in subsequent quarters. Specifically, $\Delta Revenue_t$ coefficient estimates are close to zero and are insignificant at conventional levels, except for the three-quarter-ahead estimate. Moreover, as also evidenced in Panel B, analysts' forecasts systemically underestimate the persistence of quarterly revenue growth surprises for future cumulative revenue growth in each of the following four quarters, evidenced by the positive and significant (at one percent) FE_t estimates across all four horizons. Specifically, the FE_t coefficient values are 0.44, 0.53, 0.69, and 0.73 for the one-, two-, three-, and four-quarter-ahead cumulative revenue growth forecast errors estimates, respectively. These results indicate that analyst forecasts do not incorporate the fact that revenue growth forecast errors in quarter t are persistent and continue to predict revenue growth rates in subsequent quarters, especially at the three- and four-quarter-ahead horizons. As highlighted in Panel A above, the coefficient estimates for FE_t for the three- and four-quarter-ahead realized revenue growth specifications imply that 121 and 100 percent of the quarter-t forecast error persists in realized revenue growth three and four quarters ahead. However, the FE_t coefficient in Panel C of 0.69 and 0.73 for the three- and four-quarter-ahead revenue growth forecast error specifications suggest that analyst revenue forecasts do not incorporate 57.0 (0.69/1.21) and 73.0 (0.73/1.00) percent of the persistence of quarter-t revenue growth forecast errors in predicting three- and four-quarter-ahead cumulative revenue growth rates, respectively.

The results in Table 4 stand in contrast to those for non-software companies presented in Table A6 of the Appendix. Firstly, non-software company revenue growth forecast errors in quarter t do not statistically relate to future cumulative revenue growth in any of the four subsequent $t + h$ quarters, as the estimated coefficients are close to zero in Panel A. Quarter-t growth surprises for non-software companies are not durable nor do they persist into subsequent quarters. Moreover, while non-software company analyst forecast errors in quarter t are positive and significantly relate to cumulative forecast errors for all horizons, these forecasts errors are relatively small given that analysts largely anticipate the actual relation noted in Panel A. Moreover, the foregoing findings do not result from the durability of revenue growth forecast errors, but rather result from analysts overestimating the negative reversals of quarter-t forecast errors.

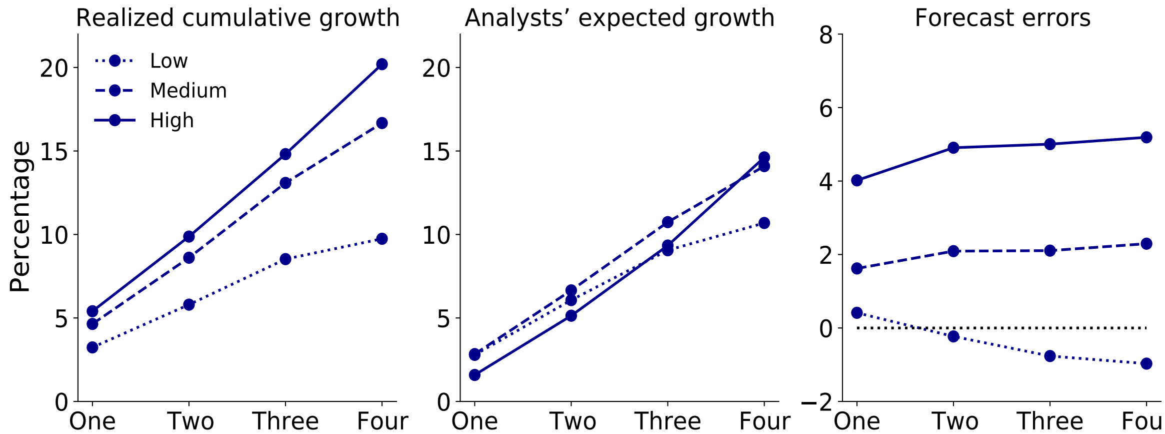
Figure 8 graphically presents the analyses of Table 4. Companies are sorted into terciles based on quarter-t revenue growth forecast errors. The left panel plots quarter-$t + h$ realized cumulative revenue growth rates for each tercile and provides evidence that realized cumulative growth rates in quarters $t + h$ for software companies are monotonically increasing in quarter-t revenue growth forecast errors. This confirms the inference that quarter-t software company growth surprises are durable and appear to persist well into future quarters. The middle panel plots quarter-$t + h$ analyst consensus cumulative revenue growth forecasts for each tercile. In strong contrast to the actual relation between quarter-t revenue growth forecast errors and four-quarter-ahead realized cumulative revenue growth rates presented in the left panel, analyst cumulative revenue growth forecasts across all terciles are almost identical and are independent of quarter-t revenue growth forecast errors, as presented in the central panel.

Moving to the plot on the right, which presents analyst cumulative revenue growth forecast errors for each $t + h$-quarter-ahead horizon, it is clear that software company four-quarter cumulative forecast errors are a function of quarter-t forecast errors. These findings confirm the inference that analyst forecasts do not incorporate the fact that quarter-t revenue growth surprises for software companies are durable and persist well into future quarters, leading to *ex post* forecast error predictability.

We next examine how key software company performance indicators relate to quarter-$t + h$ realized cumulative revenue growth rates and whether quarter-t analyst revenue growth forecasts reflect this relation. We consider two of the arguably most important leading measures of software company growth: net revenue retention (NRR) and remaining performance obligation (RPO).[9] The net revenue retention rate is the increase in spending

[9]Firm executives and analysts both have indicated the importance of RPO as a key indicator of software firm performance. For example, during Anaplan's fiscal-year 2020Q4 earnings conference call, company CFO David Horton remarked: "As a reminder, our remaining performance obligations, or RPO, represents the total booked or signed business within a quarter, and we believe this provides a more accurate commercial view into the underlying momentum of our business." Additionally, in Survey Monkey's fiscal-year 2020Q4 earnings conference call, CFO Deborah Clifford said: "I would just highlight that we look at RPO growth as the leading indicator of our business performance and the exit RPO growth rate was 17 percent. ... So I encourage you to look at RPO as the indicator of our

Fig. 8. **Predictability of forecast errors**



Notes: This figure presents h-quarter-ahead return predictions based on lagged forecast errors. Each quarter, we sort firms into terciles based on lagged forecast errors and plot the cumulative realized revenue growth, analyst expected growth, and forecast error, in each panel from left to right, respectively.

each year by existing customers (SoftwareStackInvesting.com, 2022). Or, said differently, NRR is the growth in software company revenue before considering revenue from new clients. It provides an indication of the durability of revenue growth and cash flows and illustrates a firm's ability to generate new contracts by up-selling, cross-selling, and renewing existing customer contracts. Moreover, the degree to which existing clients increase spending highlights the importance of the software and the likelihood of uptake by new customers and overall growth prospects. The remaining performance obligation (RPO) represents the total dollar value of signed contracts to be recognized in revenue over future periods. This measure is especially relevant for software firms, as the majority of software subscriptions relate to multi-year contracts that are required to be recorded as part of the firm's RPO balance. RPO represents future cash flows to the firm and is a very important indicator of the one- to two-year growth rates for pure-play software firms.

For this analysis, we consider a similar set of equations to Equations (1a), (1b), and (1c), but we add lagged net revenue retention ($NetRet_t$) and RPO growth (ΔRPO_t),

performance. That's where we're most focused on."

and drop lagged revenue growth ($\Delta Revenue_t$) as explanatory variables:

$$\Delta Revenue_{t,t+h} = \rho_0^e + \rho_1^e NetRet_t + \rho_2^e \Delta RPO_t + \rho_3^e FE_t + \epsilon_{t+h}^e \quad \text{(2a)}$$

$$E_t^s\left[\Delta Revenue_{t,t+h}\right] = \rho_0^s + \rho_1^s NetRet_t + \rho_2^s \Delta RPO_t + \rho_3^s FE_t + \epsilon_{t+h}^s \quad \text{(2b)}$$

$$\Delta Revenue_{t,t+h} - E_t^d\left[\Delta Revenue_{t,t+h}\right] = \rho_0^d + \rho_1^d NetRet_t + \rho_2^d \Delta RPO_t + \rho_3^d FE_t + \epsilon_{t+h}^d \quad \text{(2c)}$$

Table 5 Panel A reports the coefficient estimates for Equation (2a), Panel B the coefficient estimates for Equation (2b) and Panel C the coefficient estimates for Equation (2c). We run these regressions using a sub-sample of firm observations from 2018 through 2021, as RPO reporting only began in 2018 with the adoption of ASC 606 (Accounting Standards Codification) as U.S. GAAP. Table A7 in the Appendix presents descriptive statistics of NRR ($NetRet_t$) and RPO (ΔRPO_t) for this sub-sample and shows that between 2018 and 2021, the mean software company had net revenue retention rates of 118.5 and quarter-over-quarter RPO growth rates of 8.6 percent.

Looking first at Panel A, the $NetRet_t$ coefficient estimates for one-, two-, three-, and four-quarter-ahead cumulative revenue growth are monotonically increasing, with values of 0.23, 0.43, 0.63, and 0.86, respectively (each coefficient is statistically significant at conventional levels). However, while positive for all horizons, the ΔRPO_t estimates are only statistically significant for the four-quarter-ahead realized cumulative revenue growth, with a value of 0.86 (t-statistic = 2.41). The FE_t coefficient estimates are all close to zero and are not statistically significant. It appears that the relation between lagged forecast errors and quarter-ahead cumulative revenue growth rates is partly subsumed by the effect of net revenue retention and RPO growth rates, as well as by the low power resulting from the size of the sub-sample.

Panel B illustrates that analyst forecasts incorporate approximately half of the positive relation between net revenue retention rates and realized cumulative revenue growth rates for all horizons. Specifically, the $NetRet_t$ coefficient estimates for one-, two-, three-, and four-quarter-ahead cumulative revenue growth are again monotonically increasing and are 0.13, 0.22, 0.30, and 0.37, respectively (each coefficient is statistically significant at conventional levels). However, it appears that analyst forecasts do not incorporate RPO growth rates given that the coefficient estimates for ΔRPO_t are all close to zero and are not statistically significant at conventional levels. Additionally, to some degree, analyst forecasts negatively incorporate lagged forecast errors but the economic significance of this association is small in the sub-sample. Panel C, which presents the regression estimates for the cumulative forecast errors, shows that net revenue retention rates in quarter t positively relate to one- and two-quarter-ahead analyst cumulative revenue growth forecast errors. Specifically, the $NetRet_t$ coefficient estimates for the one- and two-quarter-ahead horizons are 0.10 and 0.21, respectively, and are statistically significant at conventional levels. Moreover, the coefficient estimate for ΔRPO_t is 0.25 (t-statistic = 2.99) for the four-quarter-ahead horizon. Hence, net revenue retention rates in quarter t relate to analyst forecast errors in the immediately following quarters, while RPO growth rates in quarter t relate to the one-year-ahead analyst growth forecast errors.

Figure 9 presents results similar to those in Table 5 but in graphical form. We sort firms into terciles based on lagged RPO in Panel A and net revenue retention rates in Panel B. As in Table 5, we present the relation between RPO and NRR with cumulative realized growth, cumulative analyst expected revenue growth, and cumulative revenue growth analyst forecast errors (from the left to right sub-panels, respectively). Panel A depicts the predictive ability of lagged RPO growth (left sub-panel) and analysts' cumulative revenue growth forecast errors (right sub-panel) on future realizations of cumulative revenue growth. However, the middle sub-panel shows that analyst forecasts of quarter-$t+h$ cumulative revenue growth do not incorporate RPO growth rates, as all three terciles have almost identical growth expectations. Moreover, analyst forecasts for each quarter-ahead horizon are much lower than the realized amounts, confirming the underestimation of software company growth rates in the sub-sample. Overall, Panel A illustrates that quarter-t RPO growth relates to quarter-ahead realized cumulative revenue growth and that this relation is more pronounced for the four-quarter-ahead horizon. However, analyst forecasts do not appear to consider quarter-t RPO growth rates, and, in turn, analyst forecast errors for quarters $t+h$ monotonically sort based on quarter-t RPO growth rates.

Panel B, of Figure 9 shows similar inferences for terciles sorted based on quarter-t net revenue retention rates. Specifically, analysts' quarter-$t+h$ forecasts only partially incorporate net revenue retention rates into their quarter-$t+h$ forecasts, despite the fact

Table 5. **Predictability of forecast errors: Net retention rate and RPO**

	Panel A: Realized values				**Panel B**: Analyst expectation				**Panel C**: Cumulative forecast error			
	Forecast horizon in quarters											
	One	Two	Three	Four	One	Two	Three	Four	One	Two	Three	Four
Constant	-18.91	-34.69	-49.10	-67.74	-11.29	-16.47	-18.45	-19.37	-7.62	-18.22	-30.65	-48.37
	[-5.58]	[-3.19]	[-2.08]	[-1.72]	[-3.01]	[-2.31]	[-1.57]	[-1.84]	[-2.21]	[-1.63]	[-1.36]	[-1.29]
$\Delta NetRet_t$	0.23	0.43	0.63	0.86	0.13	0.22	0.30	0.37	0.10	0.21	0.33	0.49
	[9.23]	[4.65]	[3.04]	[2.41]	[4.38]	[3.83]	[3.09]	[4.34]	[3.48]	[2.13]	[1.62]	[1.44]
ΔRPO_t	0.02	0.07	0.12	0.32	-0.01	-0.02	-0.03	0.07	0.03	0.09	0.15	0.25
	[0.33]	[0.78]	[1.28]	[2.63]	[-0.17]	[-0.50]	[-0.64]	[1.21]	[1.26]	[1.70]	[1.86]	[2.99]
FE_t	-0.01	0.04	0.06	0.02	-0.05	-0.05	-0.03	-0.05	0.04	0.09	0.09	0.07
	[-0.11]	[0.50]	[0.52]	[0.13]	[-3.26]	[-6.24]	[-2.23]	[-1.86]	[0.77]	[1.03]	[0.73]	[0.47]
R^2	0.08	0.08	0.08	0.10	0.05	0.08	0.11	0.30	0.09	0.06	0.05	0.05
nObs	365	356	313	276	365	356	313	276	365	356	313	276

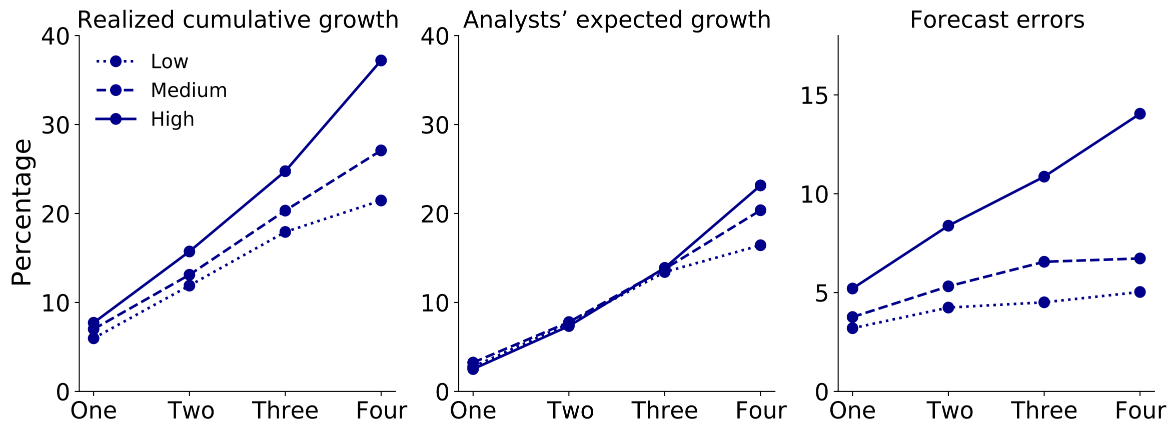
Notes: This table reports linear regression coefficient estimates and t-statistics for the following equations for each forecast horizon $h = 1, 2, 3, 4$:

$$\Delta Revenue_{t,t+h} = \rho_0^e + \rho_1^e \Delta NetRet_t + \rho_2^e \Delta RPO_t + \rho_3^e FE_t + \epsilon_{t+h}^e$$
$$E_t^s\left[\Delta Revenue_{t,t+h}\right] = \rho_0^s + \rho_1^s \Delta NetRet_t + \rho_2^s \Delta RPO_t + \rho_3^s FE_t + \epsilon_{t+h}^s$$
$$\Delta Revenue_{t,t+h} - E_t^s\left[\Delta Revenue_{t,t+h}\right] = \rho_0^d + \rho_1^d \Delta NetRet_t + \rho_2^d \Delta RPO_t + \rho_3^d FE_t + \epsilon_{t+h}^d$$

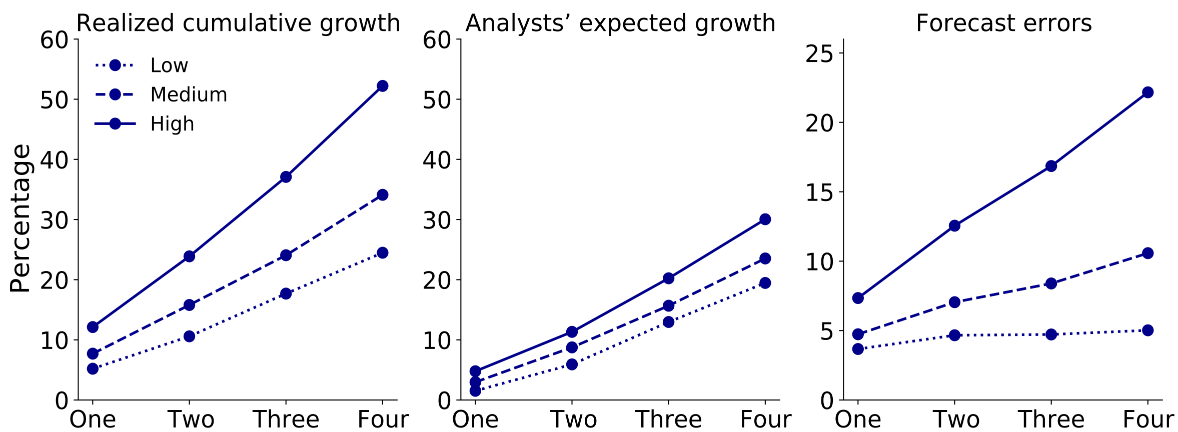
in Panels A–B, respectively. Dependent variables $\Delta Revenue_{t,t+h}$, $E_t^s[\Delta Revenue_{t,t+h}]$, and $\Delta Revenue_{t,t+h} - E_t^s[\Delta Revenue_{t,t+h}]$ denote the h-quarter-ahead expected revenue, realized revenue, and the difference between the two, respectively. $\Delta NetRet_t$ is the prior-period net retention rate, ΔRPO_t is the prior-period remaining performance obligation, and FE_t is the prior-period forecast error. Panel C is the difference between Panels A and B and represents the forecast error for each horizon.

Fig. 9. **Predictability of forecast errors based on lagged RPO and net retention rate**

Panel A: RPO



Panel B: NRR



Notes: This figure presents h-quarter-ahead return predictions based on lagged remaining performance obligation amounts (RPO) and net retention rates (NRR). Each quarter, we sort firms into terciles based on lagged RPO and NRR and plot the cumulative realized revenue growth, analyst expected growth, and forecast error for each tercile, in each panel from left to right, respectively.

that net revenue retention rates appear to effectively and monotonically sort software companies based on quarter-$t + h$ realized cumulative revenue growth rates. Taken together, these analyses illustrate that analyst forecasts do not fully incorporate the information contained in previously disclosed key performance indicators, indicators that describe software company businesses and growth trajectories.

3.3 Management systematically underestimate revenue growth rates

In this section, we turn our attention towards management-issued revenue guidance. Previous studies have highlighted the nature of management-analyst relationships, and that both parties face incentives to maintain positive working relationships with each other, incentives which are reflected in both management guidance and analyst forecasts (Williams, 1996; Yu, 2008; Green, Jame, Markov, and Subasi, 2014). Accordingly, we examine whether management also systematically underestimates the growth rate of software companies. It could be that analysts simply herd around management forecasts to support positive relationships, and this may explain the pattern of forecast errors described above.

In Table 6, we present results for the regression of quarterly management forecast errors on prior-period revenue growth and analyst forecast errors, based on the following equation:

$$FE^{m}_{t+h} = \rho_0 + \rho_1 \Delta Revenue_t + \rho_2 \Delta FE^{a}_t + \epsilon_{t+h}$$

Examining the mean forecast specifications, we see that management forecast errors are on average positive and significant for both software and non-software firms. Specifically, in Columns (1), (2), (5), and (6), the *Constant* coefficient values are 3.08, 1.23, 1.91, and 0.96 (t-statistics = 15.84, 8.02, 14.45, 9.85), respectively. These findings are consistent with previous empirical studies which have shown that management guidance is generally conservative and that management looks to provide guidance in order to "beat and raise" market growth expectations (Bartov, Givoly, and Hayn, 2002; Skinner and Sloan, 2002; Matsumoto, 2002). It is important to note two key takeaways here. First, the unconditional revenue growth forecast errors of 3.08 in Column (1) for software companies are over 60 percent larger in magnitude than those of 1.91 in Column (5) for non-software companies. Second, the unconditional one-quarter-ahead management revenue growth forecast errors of 308 basis points are larger than the corresponding analyst forecast errors reported in Table 3 of 234 basis points. Management forecasts often include a range of expected values. We next consider the upper bound of such guidance (i.e., the most optimistic forecast) in Columns (3), (4), (7), and (8) of Table 6. In contrast to the results discussed above, while the most optimistic guidance continues to underestimate revenue growth for software firms by 158 basis points, there is zero effect for non-software firms. Specifically, the coefficient estimates for $Constant$ are 1.58 (t-statistic = 7.63) in Column (3) and -0.00 (t-statistic = -0.01) in Column (7).

While forecast errors are predictable for both sets of firms, as evidenced by the positive and statistically significant FE_t coefficients in Columns (2), (4), (6), and (8) of 0.43, 0.43, 0.30, and 0.28 for software and non-software companies, respectively, the forecast errors are approximately 40 percent larger in magnitude for software companies than for non-software companies. In Figure 10, we plot the time series of one-quarter-ahead management guidance forecast errors for both sets of firms. Other than around the time of the sub-prime mortgage crisis in 2008, guidance errors are consistently positive for software companies and are generally positive for non-software companies. Additionally, the magnitude of error for software companies is consistently greater than that of non-software companies. These findings, taken together with those from above, indicate that management forecast errors are greater than those of analysts and that management also systematically underestimates firm revenue growth rates, even when considering the most optimistic upper bound of management forecasts.

4 Market participant expectations

4.1 Stock returns

In this section, we consider the implications of the foregoing results for stock return predictability. We first focus on the immediate stock price reactions within minutes of quarterly revenue announcements. We then look at the long-term price dynamics over the course of the year following the same revenue announcements.

Immediate stock price reaction. We use high-frequency data to show the differential price reaction to revenue announcements for software and non-software companies in a narrow window around the time of revenue announcement. The high-frequency data come from the NYSE Trade and Quote (TAQ) dataset, and we consider all stocks traded

Table 6. **Forecast errors in management guidance**

Panel A: Quarterly frequency	Software companies				Non-software companies			
	Mean forecast		Upper bound		Mean forecast		Upper bound	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	3.08	1.23	1.58	-0.22	1.91	0.96	-0.00	-0.86
	[15.84]	[8.02]	[7.63]	[-1.35]	[14.45]	[9.85]	[-0.01]	[-9.51]
$\Delta Revenue_t$		10.13		9.26		9.24		7.81
		[6.72]		[6.15]		[16.83]		[13.77]
FE_t		0.43		0.43		0.30		0.28
		[14.38]		[12.73]		[22.48]		[20.58]
R^2		0.28		0.27		0.19		0.15
nObs	3,945	3,945	3,945	3,945	17,225	17,225	17,225	17,225

Notes: This table presents linear regression results and t-statistics for the following equation:

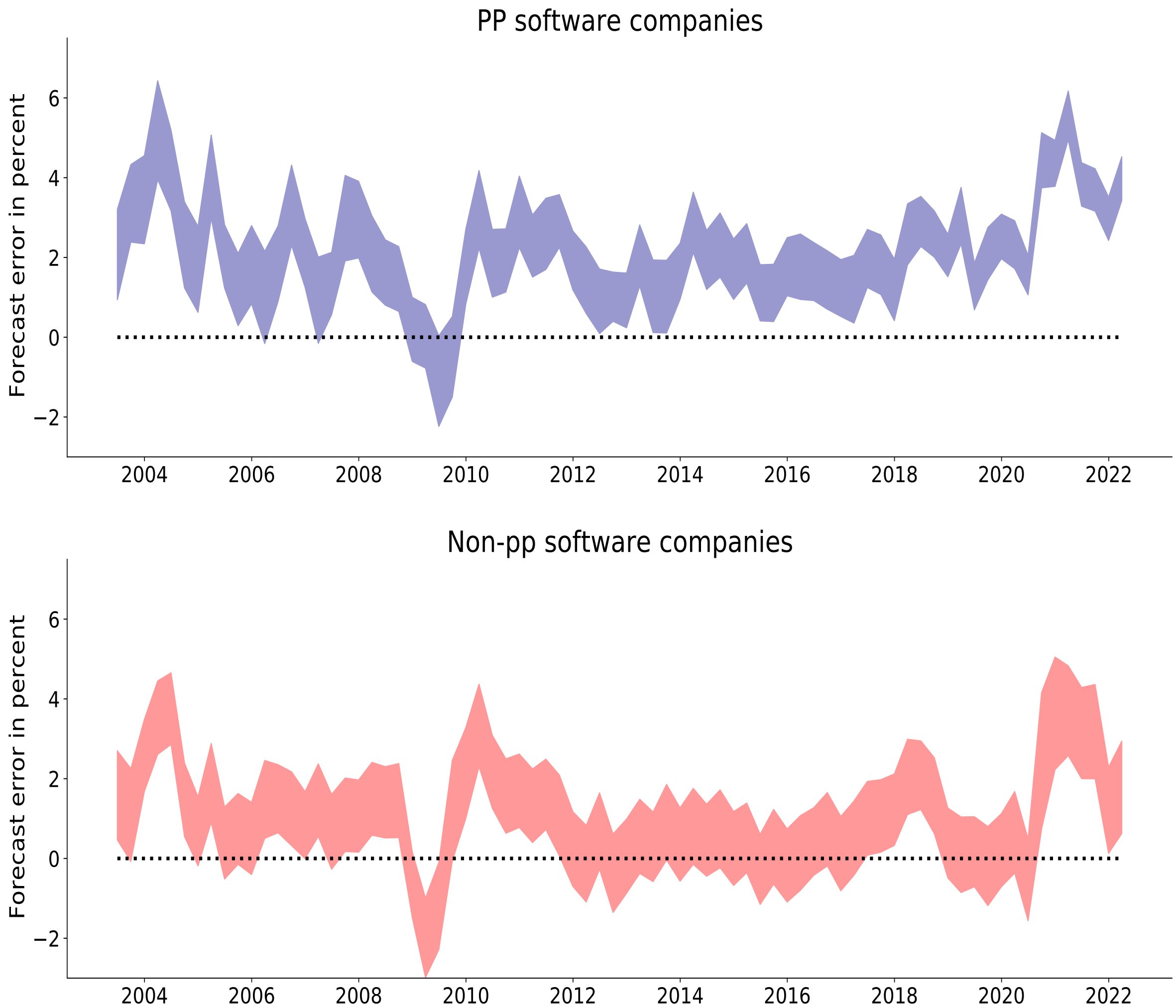
$$FE^m_{t+h} = \rho_0 + \rho_1 \Delta Revenue_t + \rho_2 \Delta FE^a_t + \epsilon_{t+h}$$

for software and non-software firms. Dependent variable FE^m_{t+h} denotes the quarter-ahead management forecast error. $\Delta Revenue_t$ is the prior-period revenue and FE^a_t is the prior-period analyst forecast error. Management guidance is often provided over a range of values, so we consider both the mean and upper bound of forecast values, as labeled. t-statistics are presented in brackets. The table spans the period January 2003 through February 2022.

in the New York Stock Exchange, American Stock Exchange, and Nasdaq National Market System stock markets.

Each quarter, we sort firms into low, medium, and high tercile portfolios based on quarter-t revenue announcement news. Figure 11 shows the cumulative stock price response in the 20 minutes immediately preceding and the 100 minutes following revenue announcement for the low and high terciles. While both sets of firms experience mostly similar price reactions to positive news, pure-play software firms experience an asymmetrically larger negative price response to negative revenue news. Software firms (blue line) experience an approximately -4 percent price reaction in the ten minutes following earnings announcement, about 2 percent more than the -2 percent price reaction experienced by non-software firms (red line). Notably, the price reaction to bad news continues to drift downward in the subsequent minutes, reaching values of around -6 and -3 percent for software and non-software companies, respectively.

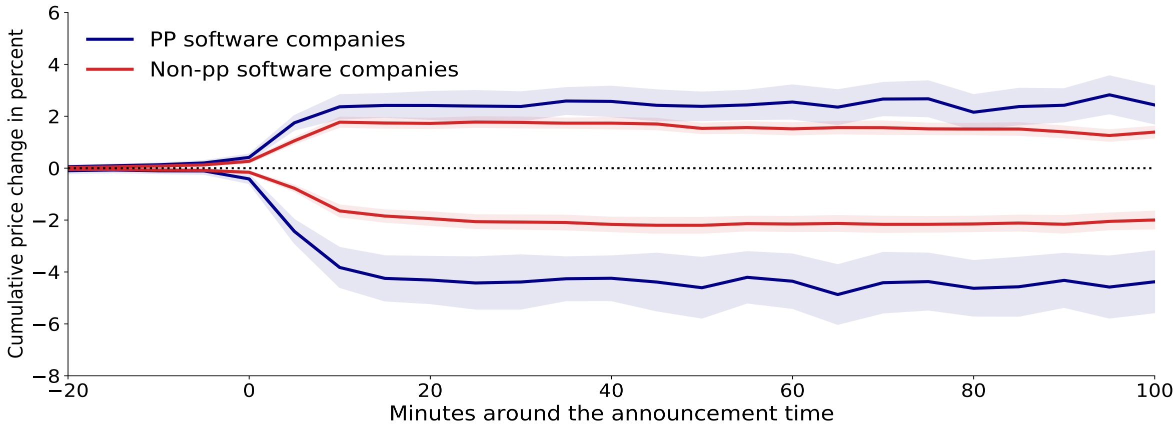
Fig. 10. **Mean management guidance forecast errors**



Notes: This figure reports the mean forecast errors associated with one-quarter-ahead management guidance values for software and non-software firms, in percentages. The bands represent the range of forecasts provided by management. Management guidance data are obtained from the I/B/E/S Detail History File and span the period from January 2003 through February 2022.

Long-term stock price dynamics. We now look at long-term price dynamics by focusing on returns over the year subsequent to revenue announcement. Because we focus on longer time windows, we control for aggregate market movements by computing CAPM-adjusted returns. As before, we stratify firm observations into terciles based on time-t revenue surprises and compute mean CAPM-adjusted realized returns for a total of four quarters following the time-t revenue announcement. Each quarter starts the day of the $t + h$ revenue announcement and ends one day before the following revenue

Fig. 11. **High-frequency stock price reaction to revenue announcements**



Notes: This figure shows the differential price reaction to revenue announcements for both software and non-software firms, in percentages. We sort firms into terciles based on revenue announcement news in quarter t. We show the average percent change in price in the 20 minutes before and the 100 minutes following firm revenue announcement with 99% confidence intervals. We obtain NYSE Trade and Quote (TAQ) intraday transaction data for all firms listed on the New York Stock Exchange, American Stock Exchange, and Nasdaq National Market System stock markets and measure the percent change in price over the noted window of time for the period spanning January 2003 through February 2022.

announcement date.

Table 7 presents our findings. Similar to the high-frequency results in Figure 11, we see that the average quarterly abnormal return for software firms reporting low revenue surprises is -4.34 percent, while the corresponding return for non-software firms is -3.14 percent. However, over the course of four quarters, the negative returns for software firms turn positive and are 1.43 percent (t-statistic = 2.12) by the third quarter following the time-t announcement and remain positive and high thereafter. In contrast, for non-software firms, mean abnormal returns remain negative. Column (3) of Table 7 shows the quarterly mean abnormal returns for software companies with the highest revenue surprises. The mean abnormal return is positive and high for the first and second quarters, with values of 4.94 (t-statistic = 11.54) and 2.14 percent (t-statistic = 5.13), but are zero thereafter. Column (6) shows that mean abnormal returns for non-software companies are positive for only the first quarter, with a value of 1.90 percent (t-statistic = 23.14), then decline monotonically for the next quarter, turning negative in the third and fourth quarter.

Figure 12 shows these results graphically by plotting the cumulative returns over the

Table 7. **Predictability of quarterly CAPM-adjusted abnormal returns**

Panel A: CAPM-adjusted returns

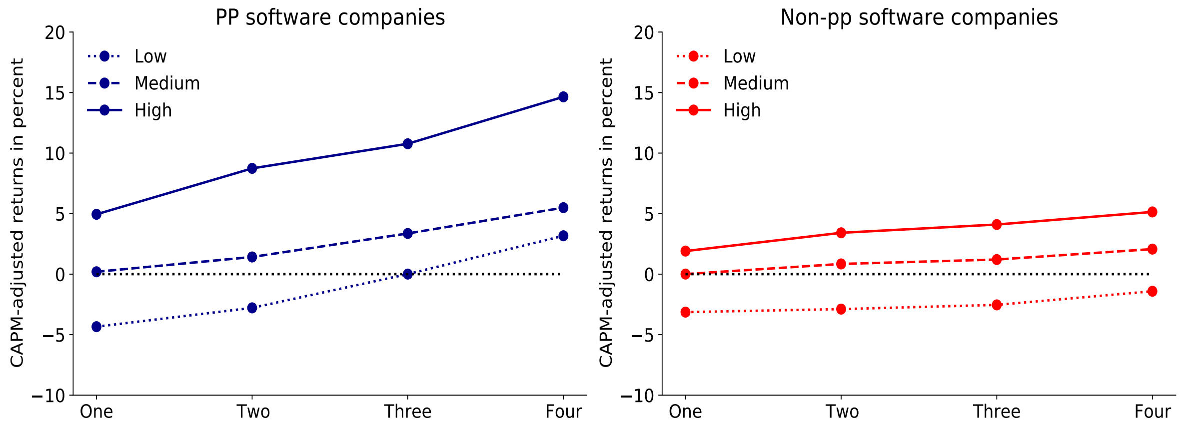
	Software companies			Non-software companies		
	Low	Medium	High	Low	Medium	High
Horizon	(1)	(2)	(3)	(4)	(5)	(6)
1Q	-4.34	0.19	4.94	-3.14	0.01	1.90
	[-7.14]	[0.60]	[11.54]	[-37.35]	[0.07]	[23.14]
2Q	-0.38	0.51	2.14	-1.39	-0.18	0.15
	[-0.60]	[1.55]	[5.13]	[-16.36]	[-2.68]	[1.80]
3Q	1.43	1.17	0.19	-0.94	-0.18	-0.27
	[2.12]	[3.55]	[0.46]	[-10.75]	[-2.64]	[-3.35]
4Q	1.87	0.94	0.50	-0.78	-0.09	-0.44
	[2.68]	[2.87]	[1.23]	[-8.78]	[-1.24]	[-5.36]

Notes: This table presents individual quarter abnormal returns for portfolios sorted based on revenue news for quarters $h = 1, 2, 3, 4$. We sort firms into terciles based on revenue announcement news in quarter t. To compute abnormal returns, we use the CAPM-factor model and a rolling 252-daily estimation window (with a minimum data availability requirement of 126 days) to estimate market betas. The return for each quarter is measured starting at time t through the day before the subsequent earnings announcement day at $t + h$. Returns are presented as individual CAPM-adjusted abnormal returns and the table spans the period from January 2003 through February 2022.

first four quarters following a revenue surprise. Results are presented for software and non-software firms in the left and right panels, respectively. The pattern of cumulative abnormal returns coincides with results in Table 7. The lowest tercile of software firms experiences initially negative returns; by the third quarter, returns reverse and are cumulatively positive. On the other hand, non-software firm returns exhibit a much flatter price response in the quarters following earnings announcement and show no correction. Table A8 and Figure A9 in the Appendix present robustness results using four-factor-model adjusted returns.

The initial price dynamics documented—which may in part explain persistently conservative management revenue growth guidance—relate to findings in Lakonishok et al. (1994) and Skinner and Sloan (2002), who show that growth firms experience an asymmetrically large negative price reaction to negative surprises. These papers argue that the large price decline for growth firms indicates overly optimistic expectations, resulting in subsequently negative returns when those expectations are not met. However, one important difference in our findings is that the initial large negative price reaction for

Fig. 12. **CAPM-adjusted return predictions based on revenue announcement news**



Notes: This figure shows unconditional abnormal $t+h$-period-ahead returns for portfolios sorted based on revenue news for quarters $h = 1, 2, 3, 4$. We sort firms into terciles based on revenue announcement news in quarter t. Returns are presented as cumulative CAPM-adjusted abnormal returns and span the period from January 2003 through February 2022.

software firms reverses within two quarters, suggesting that investors were too pessimistic and ended up overreacting to revenue surprises. This result also differentiates our findings from the large literature on post earnings announcement drift, as the price drift in our findings persists for several quarters, likely due to the documented persistence in analyst forecast errors, leading to strong return predictability that lasts for a year rather than a single quarter.

4.2 Evidence from informed traders

We now turn our attention towards short sellers in order to further understand the extent to which important market participants systematically underestimate pure-play software firm growth rates. While analysts surely are important channels of information diffusion and market efficiency, they may lack the same financial incentives as short sellers. Short sellers face various financial risks associated with their short positions (Gargano, 2020; Engelberg, Reed, and Ringgenberg, 2018), and are therefore, within the literature, widely considered to be some of the most informed traders. Various previous studies have empirically shown that higher short interest (the quantity of shares shorted as a fraction of total shares outstanding) signals overpricing and is a strong predictor of future negative stock returns at various horizons, levels of aggregation, and across different countries (Boehmer, Jones, and Zhang, 2008; Rapach, Ringgenberg, and Zhou, 2016; Wang, Yan, and Zheng, 2020; Boehmer, Huszár, Wang, Zhang, and Zhang, 2022). In the following paragraphs, we explore the ways in which short seller expectations for software companies might differ from those of analysts, and, in particular, whether short sellers are able to better predict pure-play software firm growth.

We first consider the level to which firm revenue growth and forecast errors predict short interest, using the following regression equation:

$$si_{i|t,t+90} = a + d \cdot I_{i,sw} + (b + e \cdot I_{i,sw})\,\Delta Revenue_t + (c + f \cdot I_{i,sw}) \cdot FE_t + \epsilon_t \qquad (3)$$

where $si_{i|t,t+90}$ measures the average level of short interest in the 90 days following revenue announcement. Column (1) of Table 8 presents regression results for the sample of all firms and indicates that revenue growth, $\Delta Revenue_t$, and forecast errors, FE_t, are incorporated into short-sale decisions. Specifically, significant coefficient estimates for b and c of 0.003 and -0.005 show that high levels of revenue growth (forecast error) positively (negatively) predict short interest. The results for b and c are consistent with investor expectations of revenue growth reversals and persistent forecast errors, respectively.

In Column (2), we look to see whether there is a differential effect for our sample of software firms by interacting the measures of revenue growth and forecast error with an indicator variable, $I_{i,sw}$. First, the coefficient estimate for d of 0.22 is positive but insignificant. Turning to coefficients e and f, we see that both revenue growth and forecast errors positively predict short interest. The estimate f is in contrast to the findings for all firms above, given that prior-period forecast errors negatively predict short interest. However, software-specific coefficient estimates—d, e, and f—are across the board insignificant, based on t-statistic estimates of 1.19, 1.70, and 1.53, respectively. This statistical insignificance suggests that short sellers do not treat software firms any differently than all other firms.

We next consider the relation between short interest and future stock returns for our portfolio of pure-play software firms. In Figure 13, we chart the abnormal returns for twenty portfolios sorted based on firm-specific short interest. CRSP provides short

Table 8. **Short interest, revenue growth, and forecast errors**

Coefficient	Variable	(1)	(2)
a	*Constant*	4.88	4.88
		[79.31]	[79.32]
b	$\Delta Revenue_t$	0.003	0.003
		[8.06]	[7.82]
c	FE_t	-0.005	-0.005
		[-7.46]	[-7.56]
d	$I_{i,sw}$		0.22
			[1.19]
e	$I_{i,sw} \cdot \Delta Revenue_t$		0.003
			[1.70]
f	$I_{i,sw} \cdot FE_t$		0.007
			[1.53]
R^2		0.16	0.17
nObs		233,057	233,057

Notes: This table presents regression results for the equation:

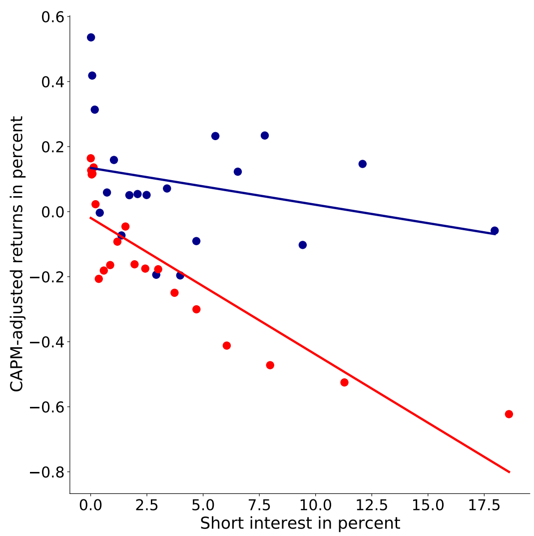
$$si_{i|t,t+90} = a + d \cdot I_{i,sw} + (b + e \cdot I_{i,sw}) \cdot \Delta Revenue_t + (c + f \cdot I_{i,sw}) \cdot FE_t + \epsilon_t$$

where t is the day of actual revenue announcement. Dependent variable $si_{i|t,t+90}$ is the average short interest over the 90 days following the day of actual announcement. Independent variables $\Delta Revenue_t$ and FE_t denote the quarterly revenue growth and quarterly forecast error, respectively. Column 1 presents results for all firms, while Column 2 includes software firm-specific results by interacting the dependent variables, $\Delta Revenue_t$ and FE_t, with an indicator for software firms, $I_{i,sw}$. t-statistics are presented in brackets.

interest data for each firm every two weeks. Based on these data, we assign firms to bins and calculate the average short interest for each portfolio. Additionally, we calculate cumulative returns for each firm over the subsequent two weeks. The left panel of Figure 13 charts market risk-adjusted abnormal returns for each software portfolio (blue) and each non-software portfolio (red). The right panel presents the same information using Fama-French three-factor plus momentum-adjusted returns.

Looking first at the non-software relation, we find similar results to those in empirical studies cited above for both return measures: that is, we find that higher levels of short interest predict increasingly negative returns, evidenced by the downward sloping red line in both panels. Alternatively, for software firms, we do not find the same negative association between short interest and returns. Rather, we find a mostly flat relation in both panels, with positive returns across all levels of short interest.

Fig. 13. **Short interest and future abnormal returns**



Notes: This figure charts abnormal returns for portfolios ranked by short interest. We first sort firms into 20 bins based on their relative degree of short interest, which is the quantity of shares shorted expressed as a fraction of shares outstanding. We then compute both average short interest and two-week cumulative returns for each of the 20 portfolios. The left panel uses CAPM-adjusted returns, while the right panel uses four-factor adjusted returns (based on Fama-French three factors plus a momentum factor). The blue circles denote results for pure-play software companies, while the red denote for non-software companies. Both panels plot regression lines of best fit.

In conjunction with Figure 13, we provide regression-based results in Table 9, based on the following pooled OLS regression equation:

$$r^e_{i|t,t+h} = a \cdot I_{i,sw} + b \cdot I_{i,not\ sw} + c \cdot I_{i,sw} \cdot si_{it} + d \cdot I_{i,not\ sw} \cdot si_{it} + \epsilon_{t+h} \qquad (4)$$

where $r^e_{i|t,t+h}$ is the two-week cumulative abnormal return for stock i. Abnormal returns are measured using the CAPM, Fama-French three-factor model (FF3), and FF3 plus a momentum factor (FF3 + Mom). si_{it} is a measure of short interest (the number of shares shorted scaled by total shares outstanding). $I_{i,sw}$ and $I_{i,not\ sw}$ are dummy variables indicating whether firm i is a pure-play software company or not, respectively. Standard errors are clustered at the firm-day level.

We again find results for non-software firms similar to those of previous empirical studies: coefficient d is significant and negative for all return measures. Specifically, for each return measure, coefficient d OLS estimates are -5.34, -4.24, -3.42, and -3.26 (t-statistic = -3.87, -4.43, -4.81, -4.51), respectively. On the other hand, we see that

Table 9. **Short interest and future abnormal returns**

			Adjusted return measures		
Coeff.	Variable	Raw returns	CAPM	FF3	FF3 + Mom
a	$I_{i,sw}$	90.10	13.73	10.41	4.40
		[3.35]	[0.96]	[0.90]	[0.37]
b	$I_{i,not\ sw}$	140.76	-2.20	-2.41	-5.90
		[5.66]	[-0.21]	[-0.31]	[-0.77]
c	$I_{i,sw} \cdot si_{it}$	-0.41	-1.25	-0.79	0.34
		[-0.24]	[-0.89]	[-0.63]	[0.27]
d	$I_{i,not\ sw} \cdot si_{it}$	-5.34	-4.24	-3.42	-3.26
		[-3.87]	[-4.43]	[-4.81]	[-4.51]
R^2		0.03	0.03	0.02	0.02
nObs		2,431,141	2,338,356	2,338,291	2,338,321

Notes: This table presents regression results for the equation:

$$r^e_{i|t,t+h} = a \cdot I_{i,sw} + b \cdot I_{i,not\ sw} + c \cdot I_{i,sw} \cdot si_{it} + d \cdot I_{i,not\ sw} \cdot si_{it} + \epsilon_{t+h}$$

where dependent variable $r^e_{i|t,t+h}$ measures the two-week cumulative abnormal return for stock i. We present raw returns and risk-adjusted returns based on the CAPM, Fama-French three-factor model (FF3), and FF3 plus a momentum factor (FF3 + Mom), in Columns 2 through 4, respectively. si_{it} measures the level of short interest and is the number of shares shorted scaled by total shares outstanding. $I_{i,sw}$ and $I_{i,not\ sw}$ are dummy variables indicating whether firm i is a pure-play software company or not, respectively. The short interest variable is expressed as a percentage and the returns as basis points. Standard errors are clustered at the firm-day level and t-statistics are presented in brackets.

estimates for coefficient c are insignificant across all specifications, indicating that short interest is not a significant predictor of pure-play software firm returns. Taken together with Figure 13, these results suggest that, while short sellers may be some of the most sophisticated traders in the market, their expectations of software firm overpricing are on average inaccurate. In fact, due to short interest borrowing costs, it is likely that short sellers are in fact losing money based on software firm short positions.

5 Conclusion

In this study, we examine and document how market participants evaluate and price pure-play software companies. We highlight the growing importance of software companies in the financial markets and illustrate the effects of their unprecedented scalability and growth. We document that over the past three decades, software companies have outperformed their non-software counterparts by more than 13 times and that classical asset pricing models cannot fully explain this outperformance. The results indicate that both management and analysts systematically underestimate the annual growth rates of software companies by over a third and that management forecast errors are greater than those of analysts. Moreover, we document a large asymmetric software company price effect for negative and positive revenue surprises. Specifically, there is a strong overreaction to negative news followed by a reversal in the subsequent quarters, and an underreaction to positive news followed by a significant downward drift over the following year. Accordingly, the identified growth forecast errors appear to be largely responsible for the documented outperformance of software companies. The findings provide evidence that market participants, even those perceived to be the most sophisticated, struggle to appreciate the nuances of these digital firms and the metrics that describe them. Taken together and most importantly, the study illustrates over the past several decades that both the overall value and disruption of software to the economy has been under-appreciated,

References

Abarbanell, J. S., Bushee, B. J., 1998. Abnormal returns to a fundamental analysis strategy. Accounting Review pp. 19–45.

Andreessen, M., 2011. Why software is eating the world. Wall Street Journal 20, C2.

Barberis, N., 2018. Psychology-based models of asset prices and trading volume. In: *Handbook of behavioral economics: applications and foundations 1*, Elsevier, vol. 1, pp. 79–175.

Bartov, E., Givoly, D., Hayn, C., 2002. The rewards to meeting or beating earnings expectations. Journal of accounting and economics 33, 173–204.

Bianchi, F., Ludvigson, S. C., Ma, S., 2020. Belief distortions and macroeconomic fluctuations. Tech. rep., National Bureau of Economic Research.

Boehmer, E., Huszár, Z. R., Wang, Y., Zhang, X., Zhang, X., 2022. Can shorts predict returns? a global perspective. The Review of Financial Studies 35, 2428–2463.

Boehmer, E., Jones, C. M., Zhang, X., 2008. Which shorts are informed? The Journal of Finance 63, 491–527.

Bordalo, P., Gennaioli, N., Ma, Y., Shleifer, A., 2020. Overreaction in macroeconomic expectations. American Economic Review 110, 2748–82.

Bordalo, P., Gennaioli, N., Porta, R. L., Shleifer, A., 2019. Diagnostic expectations and stock returns. The Journal of Finance 74, 2839–2874.

Bouchaud, J.-P., Krüger, P., Landier, A., Thesmar, D., 2019. Sticky expectations and the profitability anomaly. The Journal of Finance 74, 639–674.

Cieslak, A., 2018. Short-Rate Expectations and Unexpected Returns in Treasury Bonds. The Review of Financial Studies 31, 3265–3306.

Coibion, O., Gorodnichenko, Y., 2015. Information rigidity and the expectations formation process: A simple framework and new facts. The American Economic Review 105, 2644–2678.

Cotter, J., Tuna, I., Wysocki, P. D., 2006. Expectations management and beatable targets: How do analysts react to explicit earnings guidance?*. Contemporary Accounting Research 23, 593–624.

Eeckhout, J., Veldkamp, L., 2022. Data and market power. Tech. rep., National Bureau of Economic Research.

Engelberg, J. E., Reed, A. V., Ringgenberg, M. C., 2018. Short-selling risk. The Journal of Finance 73, 755–786.

Farboodi, M., Singal, D., Veldkamp, L., Venkateswaran, V., 2022. Valuing financial data. Tech. rep., National Bureau of Economic Research.

Farboodi, M., Veldkamp, L., 2021. A growth model of the data economy. NBER working paper .

Francis, J., LaFond, R., Olsson, P. M., Schipper, K., 2004. Costs of equity and earnings attributes. The accounting review 79, 967–1010.

Gargano, A., 2020. Short selling activity and future returns: Evidence from fintech data. Available at SSRN 3775338 .

Gómez-Cram, R., 2022. Late to recessions: Stocks and the business cycle. The Journal of Finance 77, 923–966.

Govindarajan, V., Rajgopal, S., Srivastava, A., 2018. Why financial statements don't work for digital companies. Harvard Business Review pp. 2–6.

Green, T. C., Jame, R., Markov, S., Subasi, M., 2014. Access to management and the informativeness of analyst research. Journal of Financial Economics 114, 239–255.

IBM, 2022. What is software development?

Jensen, T. I., Kelly, B. T., Pedersen, L. H., 2021. Is there a replication crisis in finance? Tech. rep., National Bureau of Economic Research.

Lakonishok, J., Shleifer, A., Vishny, R. W., 1994. Contrarian investment, extrapolation, and risk. The journal of finance 49, 1541–1578.

Matsumoto, D. A., 2002. Management's incentives to avoid negative earnings surprises. The Accounting Review 77, 483–514.

Moskowitz, T. J., Ooi, Y. H., Pedersen, L. H., 2012. Time series momentum. Journal of financial economics 104, 228–250.

Rapach, D. E., Ringgenberg, M. C., Zhou, G., 2016. Short interest and aggregate stock returns. Journal of Financial Economics 121, 46–65.

Skinner, D. J., Sloan, R. G., 2002. Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. Review of accounting studies 7, 289–312.

Software.org, 2021. Software supports us economy during covid-19 pandemic, new study finds.

SoftwareStackInvesting.com, 2022. 2021 year end review.

Soliman, M. T., 2008. The use of dupont analysis by market participants. The accounting review 83, 823–853.

van Binsbergen, J. H., Han, X., Lopez-Lira, A., 2020. Man vs. machine learning: The term structure of earnings expectations and conditional biases. Tech. rep., National Bureau of Economic Research.

Wang, X., Yan, X. S., Zheng, L., 2020. Shorting flows, public disclosure, and market efficiency. Journal of Financial Economics 135, 191–212.

Williams, P. A., 1996. The relation between a prior earnings forecast by management and analyst response to a current management forecast. Accounting Review pp. 103–115.

Yu, F. F., 2008. Analyst coverage and earnings management. Journal of financial economics 88, 245–271.

Appendix

Table A1. **Traditional v. software business model case study**

Traditional model (e.g., Nike)	Software model (e.g., Salesforce)
Nike sells customer a pair of trainers for $120	Salesforce sells customer a monthly software subscription for $10/month
It costs Nike $60 to make the shoe (materials, freight, insurance, duty, merchandiser fees, etc.)	It costs Salesforce $2 to produce and distribute an additional software subscription unit
The following year, the customer may or may not buy trainers again, or may purchase a different brand	Salesforce has a net dollar retention rate (NDR) of approximately 115 percent, meaning that each customer on average will contribute $11.50 per month over the next year towards revenue
Nike grows revenues at 7 percent year-over-year. Within 5 years, revenues will be 40 percent higher	Salesforce grows revenues at 30 percent year-over-year. Within 5 years, revenues will have almost quadrupled
	With 80 percent gross margins, software companies are able to sacrifice short-term profitability for long-term growth (through higher sales and marketing and research and development costs)

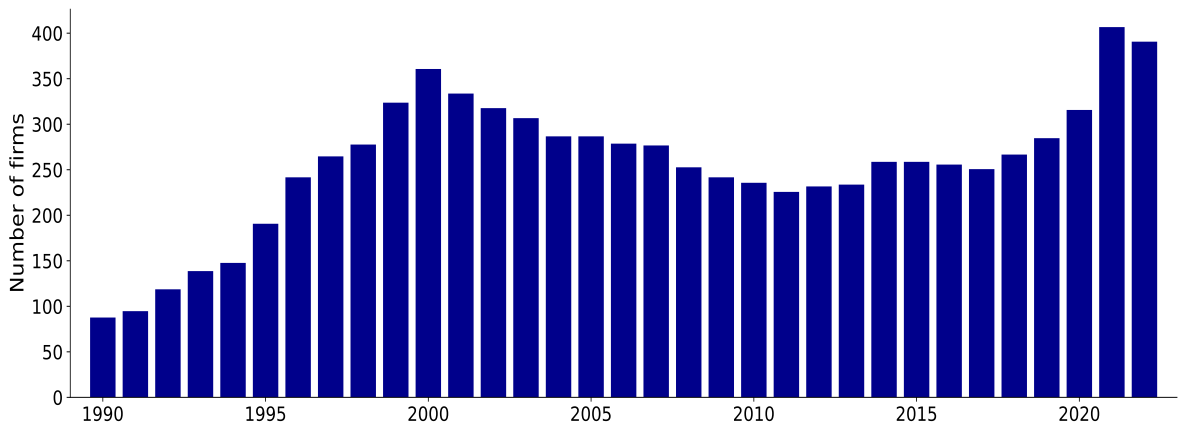
Notes: This table presents a comparison of traditional (Nike) and software-based business (Salesforce) models. While not exact, the numbers used are representative of the respective firms' sales and cost structures.

Table A2. **List of sub-industries used to identify pure-play software companies**

RBICS industry name		
Asset Management Software	Handheld and Smart Phone Games Software	Other Games Software
Automotive Industry Software	Healthcare Management Software	Other Handheld and Smart Phone Software
Business Intelligence Software	Healthcare Operations Support Software	Other Healthcare and Pharma Industry Software
Business Planning and Control ERP Software	Hospitality Industry Software	Other Network Software
Commercial Bank and Credit Union Software	Human Resources ERP Software	Other Telecommunications Industry Software
Communications Infrastructure Software	IC-Level Electronic Design Software	Patient Data Management Software
Computer Aided Design (CAD) Software	IC-Level Intellectual Property Software Libraries	Print and Prepress Industry Software
Computer and Software Stores	Insurance Software	Productivity Software
Console Games Software	Investment Management/Brokerage Software	Real Estate and Construction Industry Software
Customer Service Software	Legal, Tax and Accounting Industry Software	Retail Industry Software
Data Storage Infrastructure Software	Manufacturing Industry Software	Sales Force Automation (SFA) Software
Diversified Content Management Software	Mapping/Geographic Information Systems Software	Software Design and Engineering Consulting
Diversified Customer Relationship Software	Marketing CRM Software	Software Development Software
Diversified Enterprise Resource Planning Software	Media and Entertainment Industry Software	Software Distributors
Diversified IT Infrastructure Software	Mobile Platform Applications Software	Supply Chain ERP Software
Document Management Software	Multi-Type Home and Office Software	Telecommunications Customer Relationship Software
Drug Development Software	Multimedia Design and Engineering Software	Telecommunications Operations Support Software
Educational Software	Multiple Industry-Specific Software	Trading Software
Enhanced Telecommunications Services Software	Network Administration Software	Transportation Industry Software
Enterprise Middleware Software	Network Security Access Policy Software	Utilities Industry Software
Enterprise Security Management Software	Network Security Software	Vehicle Autonomous Control Software
Financial and Compliance ERP Software	Not-For-Profit Industry Software	Virtual Reality Design and Engineering Software
General Enterprise Management Software	Online Game Websites and Software	Web Development Software Makers
General and Mixed-Type Software	Other Design and Engineering Software	Web Search Sites and Software
Government and Public Service Industry Software	Other Finance Industry Software	

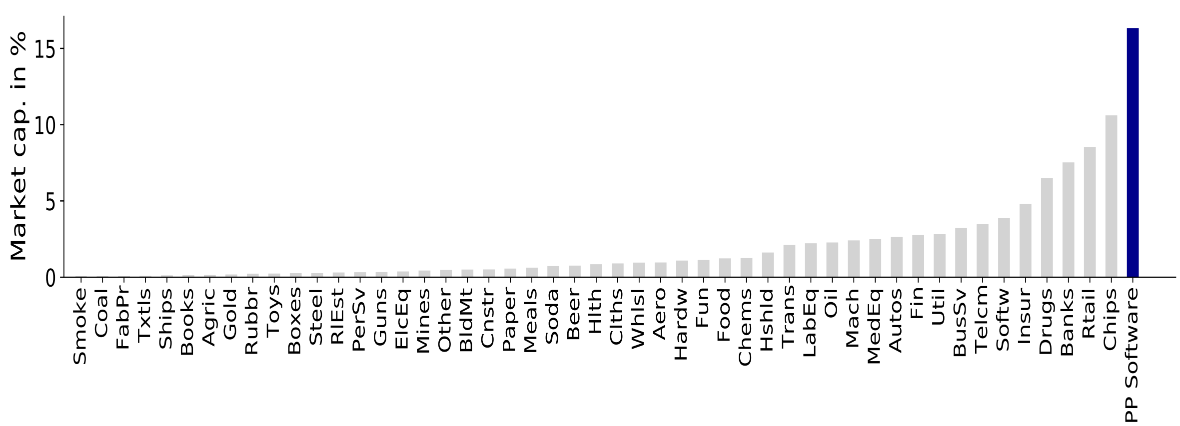
Notes: This table lists each of the Factset Revere Business Industry Classification System (RBICS) sub-industries used in our sample construction. We select all sub-industries containing the word *software* and for which the main business relates to creating and selling software or software platforms. There are 457 total firms spanning 74 distinct RBICS sub-industries. We note that for those observations which delisted during the sample period, they have not been assigned an RBICS sub-industry. For such observations, we select based on Factset industry *Prepackaged Software*.

Fig. A1. **Number of pure-play software firms**



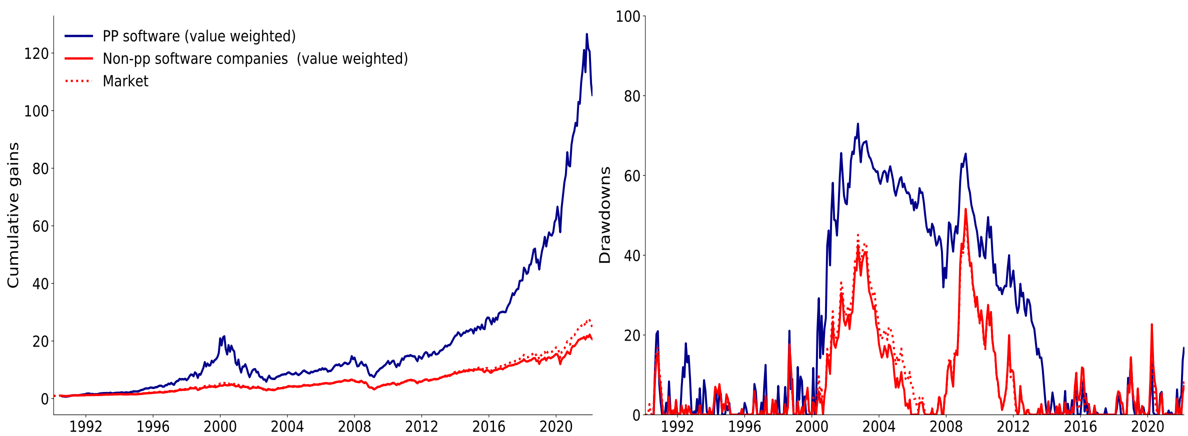
Notes: This figure shows the total number of firms per year in our sample. The figure spans the period 1990–2022.

Fig. A2. **Proportion of total market capitalization**



Notes: This figure shows the relative proportion of total market capitalization for each Fama-French industry classification and our portfolio of software firms from 1990 through 2022.

Fig. A3. **Cumulative gains and drawdowns**



Notes: The left panel plots value-weighted cumulative monthly gains for software firms (blue line), non-software firms (red line), and the total market portfolio (dotted red line). The right panel depicts the drawdowns associated with each strategy and for the market. The sample spans from January 1990 through February 2022.

Table A3. **Where are they located in FF49 industries?**

Industry Name	**Number of firms**	**Percent**
Computer Software	652	77.71
Business Services	37	4.41
Electronic Equipment	25	2.98
Computers	25	2.98
Almost Nothing	18	2.15
Trading	13	1.55
Wholesale	13	1.55
Communication	8	0.95
Measuring and Control Equipment	8	0.95
Personal Services	6	0.72
Machinery	4	0.48
Banking	4	0.48
Printing and Publishing	4	0.48
Retail	3	0.36
Healthcare	3	0.36
Insurance	2	0.24
Chemicals	2	0.24
Pharmaceutical Products	2	0.24
Real Estate	2	0.24
Entertainment	2	0.24
Electrical Equipment	1	0.12
Automobiles and Trucks	1	0.12
Recreation	1	0.12
Steel Works	1	0.12
Business Supplies	1	0.12
Construction Materials	1	0.12

Notes: This table presents the number and proportion of software firms in each of the Fama-French industry portfolios. The period spans from January 1990 through February 2022.

Table A4. **Performance evaluation: sub-sample analysis**

Panel A: Jan. 1990 to Dec. 2002	Software companies		Non-software companies	
Portfolio weights	Equal (1)	Value (2)	Equal (3)	Value (4)
Mean excess returns	19.85	14.86	8.77	5.22
Std. dev. returns	37.07	33.67	18.60	14.82
Sharpe ratio	49.62	23.28	31.76	-1.64
Mean max. drawdowns	15.09	13.45	6.07	6.05
Skewness	0.27	-0.01	-0.09	-0.69

Panel B: Jan. 2003 to Feb. 2022	Software companies		Non-software companies	
Portfolio weights	Equal (1)	Value (2)	Equal (3)	Value (4)
Mean excess returns	19.90	14.63	11.40	9.97
Std. dev. returns	21.77	17.85	18.91	14.88
Sharpe ratio	25.44	12.52	-17.26	-8.02
Mean max. drawdowns	6.18	5.71	7.15	5.89
Skewness	-0.00	-0.23	-0.38	-0.72

Notes: This table presents performance evaluation measures for software and non-software portfolios for two sub-periods: January 1990 through December 2002 (Panel A) and January 2003 through February 2022 (Panel B). Returns are presented for both equal- and value-weighted portfolios. Mean excess returns are calculated using returns in excess of risk-free rates obtained from French's website. Sharpe ratios are presented as percentages and are calculated relative to the total market return. The table spans the period from January 1990 through February 2022.

Table A5. **Alphas: sub-sample analysis**

Panel A: Jan. 1990 to Dec. 2002	Software companies		Non-software companies	
Portfolio weights	Equal (1)	Value (2)	Equal (3)	Value (4)
Market	12.03	3.26	5.65	0.25
	[2.10]	[3.31]	[1.27]	[-0.34]
FF3	16.21	14.24	1.60	-0.68
	[3.51]	[4.60]	[0.93]	[-1.69]
FF3 + Mom	21.03	13.91	5.27	-0.30
	[4.36]	[3.99]	[2.88]	[-0.60]
FF3 + Mom + CF(4)	23.99	16.59	6.55	-0.62
	[4.19]	[5.43]	[2.59]	[-1.53]
Panel B: Jan. 2003 to Feb. 2022				
Market	4.83	4.36	-2.49	-0.99
	[3.97]	[2.55]	[-1.47]	[-1.71]
FF3	4.62	2.91	-1.12	-0.84
	[5.97]	[2.65]	[-1.35]	[-1.84]
FF3 + Mom	4.78	2.75	-0.56	-0.79
	[5.96]	[2.58]	[-0.68]	[-1.79]
FF3 + Mom + CF(4)	7.30	3.01	0.92	-0.71
	[9.25]	[1.99]	[1.65]	[-1.56]

Notes: This table presents presents portfolio alphas based on the market factor, Fama-French three-factors (FF3), Moskowitz et al. (2012) momentum factor (Mom), and four cash-flow factors (CF(4)) for the sub-sample periods January 1990 through December 2002 (Panel A) and January 2003 through February 2022 (Panel B). These cash-flow factors capture earnings persistence (Francis et al., 2004), sales growth (Lakonishok et al., 1994), profit margins (Soliman, 2008), and change in gross margin minus change in sales (Abarbanell and Bushee, 1998). t-statistics are reported in brackets.

Table A6. **Predictability of forecast errors: non-software companies**

	Panel A: Realized values				**Panel B**: Analyst expectation				**Panel C**: Cumulative forecast error			
	Forecast horizon in quarters											
	One	Two	Three	Four	One	Two	Three	Four	One	Two	Three	Four
Constant	1.89	3.19	5.03	6.48	1.10	3.13	5.46	7.33	0.79	0.06	-0.44	-0.85
	[3.33]	[5.67]	[7.86]	[10.73]	[2.13]	[6.76]	[10.53]	[18.71]	[6.34]	[0.22]	[-1.10]	[-1.68]
$\Delta Revenue_t$	-0.06	-0.01	-0.08	-0.01	-0.06	-0.01	-0.09	-0.01	-0.00	-0.00	0.00	0.00
	[-9.05]	[-1.01]	[-7.76]	[-47.88]	[-8.39]	[-0.89]	[-7.92]	[-90.93]	[-8.50]	[-2.49]	[3.57]	[0.68]
FE_t	0.06	0.08	0.06	0.12	-0.29	-0.33	-0.38	-0.26	0.34	0.40	0.44	0.38
	[1.42]	[1.09]	[1.14]	[3.78]	[-7.47]	[-5.23]	[-8.79]	[-10.48]	[32.83]	[18.25]	[18.79]	[14.46]
R^2	0.07	0.01	0.06	0.03	0.10	0.04	0.14	0.10	0.12	0.07	0.05	0.03
nObs	187,448	179,267	162,517	153,090	187,448	179,267	162,517	153,090	187,448	179,267	162,517	153,090

Notes: This table reports software firm linear regression coefficient estimates and t-statistics for the following equations for each forecast horizon $h = 1, 2, 3, 4$:

$$\Delta Revenue_{t,t+h} = \rho_0^e + \rho_1^e \Delta Revenue_t + \rho_2^e FE_t + \epsilon_{t+h}^e$$
$$E_t^s \left[\Delta Revenue_{t,t+h}\right] = \rho_0^s + \rho_1^s \Delta Revenue_t + \rho_2^s FE_t + \epsilon_{t+h}^s$$
$$\Delta Revenue_{t,t+h} - E_t^d \left[\Delta Revenue_{t,t+h}\right] = \rho_0^d + \rho_1^d \Delta Revenue_t + \rho_2^d FE_t + \epsilon_{t+h}^d$$

in Panels A–C, respectively. Dependent variables $\Delta Revenue_{t,t+h}$, $E_t^s[\Delta Revenue_{t,t+h}]$, and $\Delta Revenue_{t,t+h} - E_t^d \left[\Delta Revenue_{t,t+h}\right]$ denote the h-quarter-ahead expected revenue, realized revenue, and the difference between the two, respectively. Independent variables, $\Delta Revenue_t$ and FE_t, represent the change in revenue from the prior period and the prior-period forecast error, respectively. Panel C is the difference between Panels A and B and represents the forecast error for each horizon. The table spans the period from January 1990 through February 2022.

Table A7. **Summary statistics for software industry performance variables**

	Mean	Std. Dev.	1%	5%	25%	50%	75%	95%	99%
$NetRet$	118.5	13.1	90.0	98.0	110.0	119.0	126.0	137.0	168.0
$RevenueGrowth$	6.6	21.5	-34.3	-23.5	-5.5	5.0	14.2	44.5	74.0
ΔRPO	8.6	30.8	-39.3	-24.5	-0.6	0.0	13.5	61.7	116.8
$GrossMargin$	71.4	10.6	43.0	50.0	66.0	73.0	78.0	86.0	91.0
$FCFMargin$	7.0	23.0	-52.6	-31.0	-5.8	7.0	20.8	41.0	59.6

Notes: This table presents summary statistics for main software industry performance measure variables for the period from January 2018 through February 2022. $\Delta NetRet$ is the percentage increase in revenues derived from existing customers. $RevenueGrowth$ is the quarter-over-quarter percentage growth in revenue. ΔRPO is quarter-over-quarter remaining performance obligation growth. $GrossMargin$ and $FCFMargin$ are as traditionally defined.

Table A8. **Predictability of cumulative abnormal returns: revenue news**

Panel A: CAPM-adjusted cumulative returns						
	Software companies			Non-software companies		
	Low	Medium	High	Low	Medium	High
Horizon	(1)	(2)	(3)	(4)	(5)	(6)
0Q to 1Q	-4.34	0.19	4.94	-3.14	0.01	1.90
	[-7.14]	[0.60]	[11.54]	[-37.35]	[0.07]	[23.14]
0Q to 2Q	-2.79	1.42	8.74	-2.89	0.84	3.41
	[-2.71]	[2.70]	[11.74]	[-20.83]	[7.34]	[24.38]
0Q to 3Q	0.00	3.36	10.77	-2.54	1.21	4.10
	[0.00]	[5.04]	[10.88]	[-13.92]	[8.21]	[22.15]
0Q to 4Q	3.17	5.49	14.65	-1.41	2.07	5.14
	[1.94]	[6.54]	[11.68]	[-6.06]	[11.17]	[22.24]

Panel B: Four-factor adjusted cumulative returns						
	Software companies			Non-software companies		
	Low	Medium	High	Low	Medium	High
Horizon	(1)	(2)	(3)	(4)	(5)	(6)
0Q to 1Q	-4.56	-0.16	4.54	-3.17	0.18	1.63
	[-7.91]	[-0.50]	[11.20]	[-39.93]	[2.70]	[20.68]
0Q to 2Q	-2.78	0.70	7.72	-3.08	1.17	2.95
	[-2.71]	[1.42]	[11.29]	[-23.73]	[10.71]	[22.23]
0Q to 3Q	-0.92	1.97	9.27	-3.08	1.64	3.55
	[-0.75]	[3.08]	[10.28]	[-18.46]	[11.82]	[20.53]
0Q to 4Q	2.34	3.59	12.45	-2.24	2.55	4.64
	[1.84]	[4.45]	[10.66]	[-10.70]	[14.73]	[21.25]

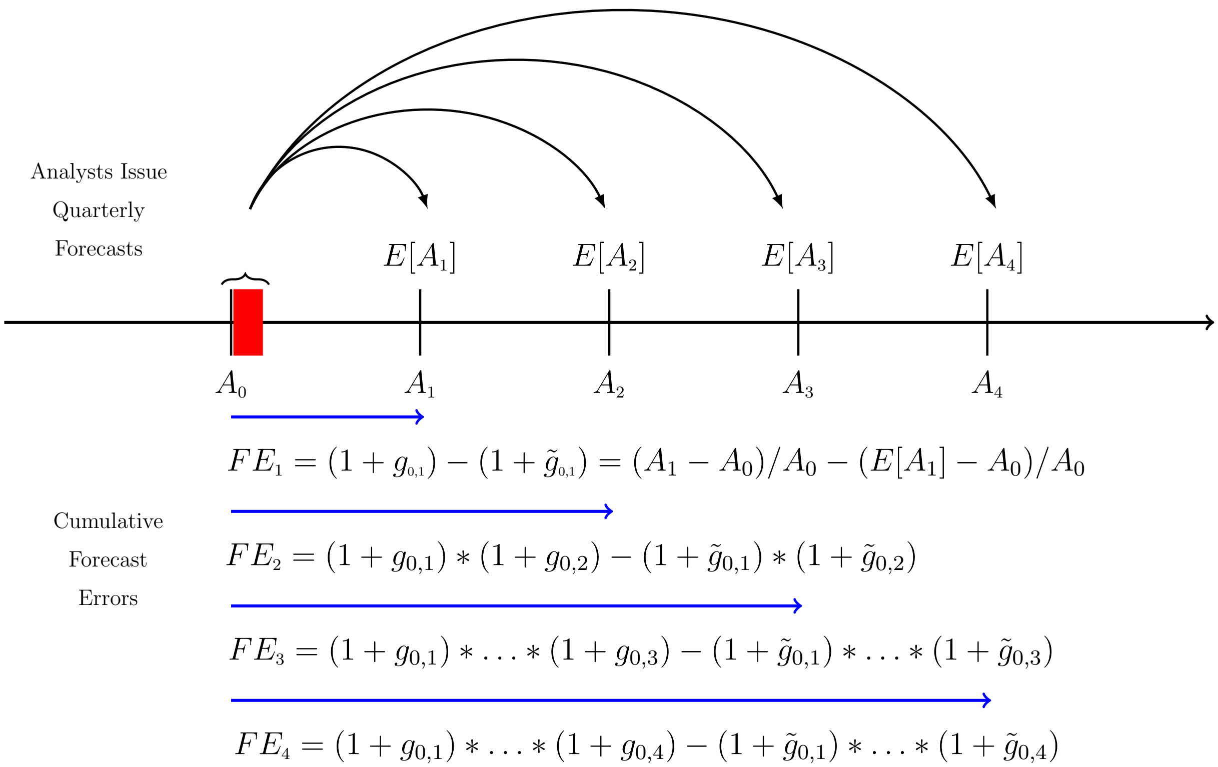
Notes: This table presents cumulative quarter abnormal returns for portfolios sorted based on revenue news for quarters $h = 1, 2, 3, 4$. We sort firms into terciles based on revenue announcement news in quarter t. To compute abnormal returns, we use a CAPM-factor model (Panel A) and a Fama-French three-factor and momentum factor-based model (Panel B) and a rolling 252-daily estimation window (with a minimum data availability requirement of 126 days) to estimate factor betas. The return for each quarter is measured starting at time t through the day before the subsequent earnings announcement date at $t + h$. Returns are presented as cumulative factor-adjusted abnormal returns and the table spans the period from January 2003 through February 2022.

Table A9. **Predictability of quarterly four-factor abnormal returns**

Panel A: Four-factor adjusted returns	Software companies			Non-software companies		
	Low	Medium	High	Low	Medium	High
Horizon	(1)	(2)	(3)	(4)	(5)	(6)
1Q	-4.56	-0.16	4.54	-3.17	0.18	1.63
	[-7.91]	[-0.50]	[11.20]	[-39.93]	[2.70]	[20.68]
2Q	-0.56	0.31	1.85	-1.45	0.03	0.01
	[-0.93]	[0.98]	[4.63]	[-18.03]	[0.39]	[0.14]
3Q	1.39	0.67	0.19	-1.12	-0.04	-0.21
	[2.15]	[2.15]	[0.48]	[-13.54]	[-0.58]	[-2.65]
4Q	1.99	0.53	0.38	-0.86	0.03	-0.45
	[2.95]	[1.70]	[0.97]	[-10.33]	[0.49]	[-5.69]

Notes: This table presents individual quarter abnormal returns for portfolios sorted based on revenue news for quarters $h = 1, 2, 3, 4$. We sort firms into terciles based on revenue announcement news in quarter t. To compute abnormal returns, we use a Fama-French three-factor and momentum factor based model and a rolling 252-daily estimation window (with a minimum data availability requirement of 126 days) to estimate factor betas. The return for each quarter is measured starting at time t through the day before the subsequent earnings announcement day at $t + h$. Returns are presented as individual factor-adjusted abnormal returns and span the period from January 2003 through February 2022.

Fig. A4. **Timeline of Analyst Quarterly Revenue Forecast Errors**



Notes: This figure shows the timing of quarterly analyst revenue forecasts and details the process for the calculation of conditional cumulative forecast errors presented in Figure 5. The forecast errors in Figure 5 are calculated by the following process:

1. Analysts observe firm quarterly revenue realizations at time A_0 and forecast future revenues for each of the following four quarters $(E[A_1], E[A_2], \ldots)$. We use the consensus forecast of all analysts' forecasts made during the 15-day period following quarter-end, represented by the red shaded region above.

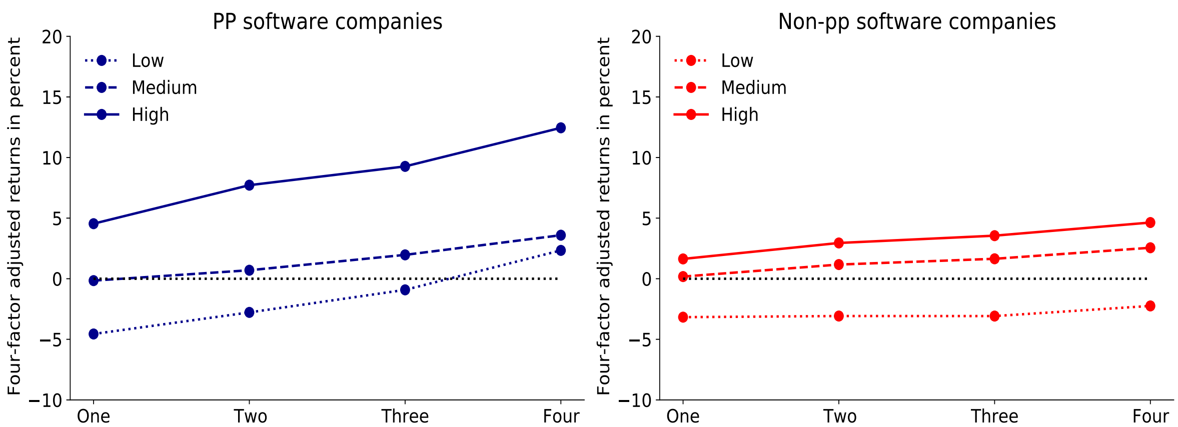
2. At the following quarter-end, realized firm revenues are observed (A_1), and the quarter-over-quarter growth rate for both realized revenue, $g_{0,1}$, and analyst forecasts, $\tilde{g}_{0,1}$, is calculated. The analyst forecast error, $FE_{0,1}$, is the realized growth in excess of the forecast amount.

3. In the following quarter, $t = 2$, the same revenue growth calculations are made for the realized and forecast revenue growth from $t = 0$ to $t = 2$, $g_{0,2}$ and $\tilde{g}_{0,2}$ respectively. The cumulative forecast error is calculated as the difference between the product of the realized gross growth rates and the product of the analyst forecast gross growth rates.

4. This process is repeated at each subsequent quarter, $t = 3$ and $t = 4$, and the cumulative forecast error in each quarter is the difference between the cumulative realized gross growth rates and gross forecast growth rates.

5. The calculated conditional forecast errors, $FE_1, FE_2, \ldots$, correspond to the dots in Figure 5 for each quarter spanning the period 2003Q1 through 2022Q1.

Fig. A5. **Four-factor-adjusted return predictions based on revenue announcement news**



Notes: This figure shows unconditional abnormal $t+h$-period-ahead returns for portfolios sorted based on revenue news for quarters $h = 1, 2, 3, 4$. We sort firms into terciles based on revenue announcement news in quarter t. Returns are presented as cumulative Fama-French three-factor plus momentum factor adjusted abnormal returns and span the period from January 2003 through February 2022.