Equity Market Views and Digital Technology Investment in Non-IT Firms

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the Committee for the PhD in Business Administration have examined a dissertation entitled

Equity Market Views and Digital Technology Investment in Non-IT Firms

Presented by Wilbur Xinyuan Chen

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature Suraj Srinivasan

Suraj Srinivasan, Chair

Signature Paul Healy

Paul Healy

Signature

Charles Wang

Signature

Shane Greenstein

Date: April 18, 2022

Equity Market Views and Digital Technology Investment in Non-IT Firms

A DISSERTATION PRESENTED
BY
WILBUR XINYUAN CHEN
TO
THE DEPARTMENT OF ACCOUNTING AND MANAGEMENT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
BUSINESS ADMINISTRATION

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
APRIL 2022

Thesis advisor: Professor Suraj Srinivasan Wilbur Xinyuan Chen

Equity Market Views and Digital Technology Investment in Non-IT Firms

Abstract

Motivated by the increasing investment in digital technologies, such as analytics, big data and artificial intelligence technologies, in non-IT firms, this dissertation studies the role that equity markets play in the digital investment in these firms. In the first chapter, I examine the market valuations of non-IT companies that disclose digital-related activities in the 10-K annual report filings. My findings show that firms that disclose digital-related activities exhibit significantly higher valuations relative to peers and exhibit significantly positive long-run returns. In the second chapter, I use digital-related questions on earnings conference calls to study the views and assessments of key market intermediaries, financial analysts, on digital technology investment in non-IT firms. I find that financial analysts' views and sentiment on digital investment are aligned with factors that predict greater success in digital technology investment. Moreover, I find evidence that their views are linked with more future investment in advanced digital technologies – AI technologies, and cross-sectional analyses suggest that analysts play an active role in encouraging firms to invest more in these technologies. In the third chapter, I examine the relationship between various capital market forces, such as short-term market incentives, and non-IT firms' investments in digital technologies. Using vesting equity as a proxy for short-term market incentives, I find that digital technology investment as measured by the proportion of IT workers and AI worker vacancies is not significantly associated with these incentives. On the other hand, digital technology investment as measured by technology acquisitions is negatively associated with vesting equity.

Contents

The following authors contributed to Chapter 1: Suraj Srinivasan

Listing of Figures

Listing of Tables

This dissertation is dedicated to my parents, Kan Chen and Gloria Yap, my siblings, Gillian, Kristi and Kowin, as well as several dear friends who have constantly provided much support throughout this PhD journey.

Acknowledgments

I thank generous feedback and guidance from my dissertation committee, Suraj Srinivasan, Charles Wang, Paul Healy and Shane Greenstein, for their time and many suggestions that have helped develop this dissertation. I also thank workshop participants at Harvard Business School, 2019 MIT-Asia Conference, The Second Conference on Intelligent Information Retrieval in Accounting and Finance, 2020 American Accounting Association Conference, Digital Initiative Doctoral Workshop, 2021 Hawaii Accounting Research Conference, Korean Accounting Information Association Webinar, for their valuable comments and suggestions on the first chapter of the dissertation. I also thank workshop participants at the Harvard Business School Accounting & Management Seminar, Accounting & Management Doctoral Workshop, Digital Initiative Doctoral Workshop, University of Hong Kong, University of Chicago, University of California (Berkeley), Chinese University of Hong Kong, Singapore Management University, Hong Kong University of Science and Technology, Boston College, University of Waterloo, NBER Productivity Lunch, Krishna Palepu, Jonas Heese, Aiyesha Dey, Yuan Zou, Eugene Soltes, Joe Pacelli, Andy Wu, Karim Lakhani, Tim DeStefano, Hubert Joly, Ariel Stern, Feng Zhu, Trang Nguyen, Elliott Tobin, Hashim Zaman, Yina Yang, Aditya Mohan, Ehsan Valavi, Trung Nguyen, Maya Balakrishnan, Bob Kaplan, Tatiana Sandino, Siko Sikochi and Jung Koo Kang for their valuable comments and suggestions on the second chapter of the dissertation. I also thank workshop participants at the Harvard Business School Accounting & Management Seminar, Krishna Palepu, Jonas Heese, Aiyesha Dey, Yuan Zou, Andy Wu and Feng Zhu for their valuable comments and suggestions on the third chapter of the dissertation.

1
Introduction

Over the past decade, the new wave of digital technologies such as analytics, big data and artificial intelligence have been increasingly used in a variety of business applications in multiple sectors of the economy. While there is some consensus that these technologies have significant benefits in terms of driving productivity gains and creating new products/services for firms in traditional, non-IT sectors, the adoption and investment process of these technologies is uncertain due to the challenges in integrating these technologies into non-IT firms. From an equity market perspective, the rise in digital technology investment leads to questions on how (1) markets are viewing and assessing these investments and (2) whether market views on these technologies are playing some role in shaping investment in these technologies. In this dissertation, I examine these questions through the empirical analyses presented in three chapters.

In the first chapter, which is joint work with Suraj Srinivasan, we examine firm value and performance implications of the growing trend of non-IT companies engaging in activities relating to digital technologies. We measure digital activities in firms based on the disclosure of digital words in the business description section of 10-Ks. Digital activities are associated with a market-to-book ratio 8-26% higher than industry peers, and only 25% of the differences in market-to-book is explained by accounting capitalization restrictions. To control for selection bias, we implement lagged dependent variable and IV regressions, and we find that our market-to-book findings are robust to these specifications. Portfolios formed on digital activity disclosure earn a DGTW-adjusted return of 30% over a 3-year horizon and a monthly alpha of 51-basis-points. On the other hand, we find weak evidence of near-term, positive improvements in fundamental performance, as we find some evidence of interim productivity increases, but declines in sales growth conditional on digital activities.

In the second chapter, I use analysts' digital questions in quarterly earnings conference calls to study how they evaluate digital investments made by non-IT firms and to examine the relationship between their views and future investment in advanced digital technologies – artificial intelligence (AI). Consistent with analysts functioning as effective information analyzers, determinant analyses show that analysts' interest and sentiment on digitization is based on various factors that drive greater digitization. Moreover, I examine the consequences of analysts' digital questions and I find that digital questions are positively associated with AI investment (as measured by job postings with AI skills) in the following year. Further cross-sectional analyses suggest that analysts are actively encouraging firms to invest more in AI as the positive relationship is stronger when there are no firm disclosures on digital topics, or when the firm is not currently investing in AI. Moreover, the positive relationship is also stronger when firms are currently investing in AI at abnormally low-levels. Thus, the overall findings suggest that analysts play a positive governance role in shaping non-IT firms' investment in AI.

In the third chapter, I study three capital market forces that are predicted from the corporate investment literature, to drive advanced digital investments in non-IT firms — (1) the expected productivity of digital investments, (2) capital constraints and (3) short-term market incentives. Consistent with prior literature, I find that firms with higher expected productivity of advanced digital investments (AI technologies) and lower capital constraints are associated with higher digital investments, as measured by technology acquisitions and the hiring of workers with AI-related skills. Moreover, I find that greater sensitivity to near-term stock prices, as measured by vesting equity, is associated with lower investment, as measured by technology acquisitions.

Overall, my research yields the following insights. First, I find that non-IT firms are increasing investment in digital technologies and that markets are positively valuing these investments. Second, I find that key market intermediaries, such as financial analysts are increasingly interested in these technologies. The views of these analysts are also aligned with factors that predict digital technology investments and also portend future investment in advanced digital technologies such as artificial technologies. Lastly, I find that CEO's equity incentives could play some role in the digital investment decision, as I find that in quarters with high CEO vesting equity, firms tend to invest less in technology acquisitions.

2

Going Digital: Implications for Firm Value and Performance

The new wave of data-driven digital technologies, such as analytics, artificial intelligence, big data, cloud computing, and machine learning, has brought substantial changes in recent years to how companies are organized, invest, and operate. In 2016 alone, a McKinsey survey estimates, large technology companies have invested a total of 20 to 30 billion USD in artificial intelligence (AI). While initial investments in new digital technologies were concentrated in IT firms, recent developments, especially in cloud computing, have also enabled non-IT firms to invest in these technologies at scale. While, in the past, firms seeking to adopt digital technology had to invest in data infrastructure and hardware, cloud-computing technologies provide firms with an alternative option of renting data infrastructure from service providers such as Amazon Web Services (AWS). As a result, digital technologies have become easier to scale-up at a lower cost (Brynjolfsson et al. 2017). Recent anecdotal evidence suggests that some non-IT firms have responded by actively engaging in digital technologies at a large-scale (Bass 2018). For example, many car manufacturers have invested in self-driving and autonomous technologies, and retail firms are investing in digital marketing and data analytics.

Our objective in this paper is to identify, characterize and examine the economic performance of firms from non-IT industries that are among the first movers in engaging in digital activities relating to analytics, automation, AI, big data, cloud computing and machine learning. To measure the deployment of these technologies, we construct a dictionary of digital terms[1], and obtain word counts of digital terms in firms' 10-K reports to proxy for the extent of digital activity.

We provide novel large-sample empirical findings, consistent with anecdotal evidence, of an increasing trend in the engagement of digital activities by non-IT firms in recent years. Our sample consists of all US-listed non-IT firms, which are identified by their industry classification[2], for the fiscal years 2010-2020. Based on digital word counts from the business description of 10-Ks, we find that firms are indeed disclosing more about digital activities as the proportion of disclosers of digital terms in our sample increased from 8% in 2010 to 30% in 2020.

Using other measures that track digital activities in firms, we perform validation tests of our textual-based proxy. We find that our proxy is associated with a higher probability of filing digital-related patents[3] and a higher proportion of IT workers by 1-3%, relative to industry peers. Moreover,

1. We define digital terms in Appendix A.1.
2. Appendix A.2 presents the list of industry codes that are used to identify IT firms.
3. We use the digital-related patent search terms provided in Bloom et al. (2018) and Webb (2020).

we also find that our proxy is associated with greater (lower) economic similarity with IT (non-IT) firms. Specifically, firms with digital activities exhibit higher return co-movement with the IT portfolio by 45-135% higher co-movement with this portfolio (i.e., a firm in the top tercile of digital disclosure exhibits 0.054 higher co-movement relative to the sample average of 0.04, or 135%)[4] and less co-movement with the non-IT portfolio by 5-15%. Thus, these analyses suggest that our proxy identifies non-IT firms that are actively engaging in digital activities.

Next, we examine the profile of firms that engage in digital activities. Our results suggest that firms that engage in digital activities are larger, younger, exhibit higher return volatility and expend more on SG&A. Past digital activities also significantly predict current digital activity, which suggests going digital is a persistent process. Moreover, we also find that firms in industries with higher digital activity also tend to be more likely to engage in digital activities, consistent with industry-wide complementarities in digital technologies.

Building on the technology adoption literature, we examine whether digital activities increase firm value. Prior studies such as Brynjolfsson et al. (2017) and Cockburn et al. (2017) have argued that digital technologies increase the growth opportunities and productivity of firms. Consequently, markets should place a higher valuation on non-IT firms that engage in digital activities due to potential future gains in performance. On the other hand, prior work has also suggested that there are various frictions associated with new technologies that may delay or limit the benefits of new technologies (Bresnahan and Greenstein 1996; Brynjolfsson et al. 2017). Consistent with digital technologies providing net benefits to firms, we find that the market-to-book ratio of non-IT firms that engage in digital activities is higher than their industry peers in an economically significant way. Notably, we estimate that a firm that adopts digital activities has a 8-26% higher market-to-book than its peers (that is, a firm in the top tercile of digital disclosure exhibits a market-to-book that is by 0.798 higher relative to the sample average of 3.05, or 26%).

4. In the rest of the text, we use the same methodology to compute the economic ranges.

We recognize that accounting rules for capitalization could play a role in explaining some of the market-to-book results as expenditures relating to digital activities are likely not capitalized thus affecting the book value of equity. We study how capitalization restrictions influence the market-to-book increases that we find in firms with digital activities in two ways. First, we address the capitalization restrictions by controlling directly for the expenditures relating to the intangibles (SG&A and R&D). With these controls, we find that differences in market-to-book remain statistically significant but falls in magnitude to about 5-14%. Second, we control for capitalization restrictions more formally, by correcting market-to-book for accounting conservatism following McNichols et al. (2014). We implement the market-to-book adjustment for a sub-sample of firms with sufficient investment histories and we show that correcting for accounting conservatism reduces the effect of digital activities on market-to-book by roughly 25%. Specifically, without the conservatism correction, firms that engage in digital activities exhibit a 13-39% higher market-to-book with the conservatism correction, the same firms exhibit a 10-29% higher market-to-book.

To further address concerns of selection bias, we implement an instrumental variable analysis to examine the robustness of the market-to-book result. We instrument for our digital textual score with a measure of AI technology exposure at the industry-level, which is estimated by the cosine similarity between the abstracts/titles of AI-related patents (which are identified using the methodology in Webb 2020) and NAICS industry descriptions (following the GloVe vector approach in Kogan et al. 2020)[5]. With this instrument, we find that the coefficient on the digital score remains positively significant for the unadjusted and adjusted market-to-book specification.

We corroborate our market-to-book results with an analysis of the Earnings and Sales Response

5. As a measure of technology exposure, this instrument proxies for industries that are more likely to benefit from AI technologies and thus firms within these industries should exhibit a greater likelihood of adoption digital technologies. Consistent with this conjecture, our first stage regressions show that the measure of AI technology overlap correlates strongly with higher values of the digital score. Moreover, we argue that the assignment of industries that are more exposed to AI technology is fairly exogenous, as much of the patents are filed in universities and other non-profit organizations.

Coefficient (ERC and SRC)[6], conditional on digital activity. If firms that engage in digital activities are more highly valued by investors, we expect that their ERCs and SRCs would increase as investors would increase their pricing multiples on earnings and sales. Consistent with this prediction, we find that ERCs for firms with digital activities are substantially higher than those of their peers. Specifically, such a firm exhibits a 16-50% higher annual ERC and a 22-66% higher decile-based ERC than its industry peers. Similarly, we find that SRCs for firms with digital activities are also substantially higher compared to peers. In particular, a firm with digital activities exhibits 62-186% higher annual SRC and a 25-75% higher decile-based SRC compared to its industry peers. The SRC results are especially noteworthy, as inferences based on the sales-based valuation metric are not confounded with earnings impact of expenses on digital activities that cannot be capitalized per accounting rules.

Next, we study the effects of digital activities on future returns. Prior work suggests that innovation activities are uncertain and difficult to value (Fitzgerald et al. 2020; Lee et al. 2019), thus digital activities for non-IT firms could predict future returns. We conduct several return predictability tests to investigate this conjecture, and we find that digital activity predicts future returns. For long-short value-weighted portfolios formed on digital activity disclosure, we find that these portfolios earn, on average, a 30% DGTW-adjusted return[7] over three years[8]. Moreover, in panel regressions, we find that these portfolios earn 0.2%, 2.9%, 5.8% and 8.0% returns at the monthly and 1-3-year horizons, after controlling for size, book-to-market, operating profit, investment, momentum, SG&A, R&D and year fixed effects[9]. Finally, in calendar portfolio tests, we find that after controlling for market, size, value, investment, profitability and momentum risk factors, the digital-based portfolios earn a

6. We study the SRC, as the valuation of sales, unlike book equity and earnings, is not confounded by capitalization restrictions. Thus, examining the effects of digital activities on the SRC is a relatively clean way of studying whether digital activities do indeed increase firm valuations.

7. Abnormal returns are estimated by deducting the firm's raw returns from the corresponding firm's size, book-to-market and momentum quintile portfolio returns, following Daniel et al. (1997).

8. These portfolios hold firms that are in the top tercile of digital disclosers in the long position and firms that do not disclose digital terms in the short position.

9. Results are also similar without risk factor controls.

monthly alpha of 51 basis points[10]. These results suggest that digital activities are associated with predictable long-run returns, and from a managerial standpoint, also suggest that firms could receive even higher valuations by providing greater disclosure about digital activities.

Moreover, we also examine whether the long-run drift in returns is explained by the reduction of uncertainty related to digital activities through continuous digital disclosure over time. Our findings suggest that the long-run returns can be explained by the continuous disclosure of digital activities. Firms that do not continuously disclose high-levels of digital terms during the 2 and 3-year return window, exhibit no statistically significant abnormal returns. While firms that continuously disclose high levels of digital activities, exhibit significantly positive returns over the 2 and 3-year horizons. Taken together, our results suggest that persistent disclosure of digital activities is associated with better pricing of digital activities.

Finally, we examine whether the increase in valuations is validated by increases in future financial performance measured by return-on-assets, asset turnover, return-on-net operating assets, net operating asset turnover (Soliman 2008), profit margins and sales growth. Consistent with productivity gains from data-driven technologies (Tambe 2014), we find that firms with digital activities exhibit ROA and asset turnover that is 13-40% and 2.7-7.5%, respectively, higher relative to industry peers. Moreover, consistent with digital technologies improving long-term productivity in firms, we also show that for firms initiating digital activities, ROA and net operating asset turnover increase by 36-110% and 7-20%, respectively, compared to industry peers over the 3-years after initial digital disclosure.

On the other hand, we find no significant differences in profit margins and 13-40% lower sales growth for firms that engages in digital activity compared to peers. Thus, given these mixed accounting performance results, we reconcile these findings with our earlier valuation analysis with three interpretations – (1) The long payoffs of digital investments. (2) The benefits of going digital are quickly

10. The long-short portfolios also yield a significant unadjusted return of 50 basis points.

eroded through market competition, as firms tend to go digital when faced with greater market pressures[11]. (3) Digital technologies could have limited performance effects, but investors are eager to be associated with companies that are disclosing that they are involved with these technologies (Cooper et al. 2001).

One limitation of the paper is that our findings are associative, and thus we cannot attribute causality to our results. We acknowledge four potential issues relating to selection bias, specifically, (1) better performing firms engaging in digital activities, (2) firms selectively disclosing only successful digital activities, (3) firms withholding proprietary information on digital activities, (4) mis-classified non-IT digital firms and (5) extrapolating our results for future adopters of digital technologies. We address the first concern by showing that our results are robust to a lagged dependent variable specification and an IV analysis. We argue that the second effect is unlikely as investment in new technologies leads to uncertain outcomes and success is difficult to predict ex ante. The third concern also seems unlikely, as we find that firms that do not disclose digital activities are associated with negative long-run returns. We evaluate the fourth concern by examining our results after dropping potentially mis-classified non-IT firms and we find no changes to our main inferences. We address the fifth concern, by cautioning readers from over-generalizing our results to future adopters of digital technologies, as the inferences that we make in this study is based on the sample of *early* adopters.

Our findings relate to two strands of research. First, we are among the first studies, to our knowledge, that provide large-sample empirical evidence at the firm level of the impact of AI and other digital technologies. Notably, a feature of our study is that our proxy for digital activity is created using publicly available data for a wide range of publicly listed firms and is easily replicable. Second, we contribute to the literature on valuation by introducing a new source of non-financial information that significantly drives prices. In particular, we find that markets slowly incorporate the value im-

11. In our determinants analysis (Table 4), we find that past sales growth and stock returns are negatively related to digital activities.

plications of digital technologies over time, as portfolios formed on the disclosure of digital activities earn statistically significant positive returns.

2.1 Literature Review

2.1.1 Digital Technology Adoption and Firm Value

The adoption of digital technology potentially enhances firm value in two ways. First, digital technologies can increase firm value by increasing productivity. For example, during the information technology (IT) revolution in the 1990s, several large and diversified organizations benefited from IT adoption by improving inventory management (Brynjolfsson and Hitt 2000). IT adoption also allows firms to produce more (Brynjolfsson and Hitt 1996) and expand more effectively (Hitt 1999). Moreover, recent studies that explore the potential consequences of adopting digital technologies, such as data analytics and AI suggest that these technologies will also improve firm productivity (Tambe 2014) and facilitates growth (Babina et al. 2020). For example, studies on FinTech have also found that adopting these technologies leads to significant improvements in productivity for financial services firms (Philippon 2016) and that these technologies are highly valued by markets (M. A. Chen et al. 2018).

Second, another value-enhancing aspect of digital technologies is that they potentially increase the value of existing investments within the firm. Recent studies that explores the potential productivity benefits of AI have argued that these technologies are general purpose technologies (GPT), which act as complements to other existing investments (Cockburn et al. 2017) and thus, increases the value of existing resources.

Fricitons in Adopting New Technology

Although technology adoption potentially introduces many benefits to the firm, these take long to be realized, lowering their value, especially in the short term. In the late 1980s, the benefits of IT adoption took so long to realize that they were not evident in the data, leading Robert Solow to coin the famous "Solow's paradox" – the observation that you can see the computer age everywhere but in the productivity statistics.

There are several reasons that explains why the benefits of IT adoption take long to realize. First, adopting technologies requires developing complementary organizational capabilities (Bresnahan and Greenstein 1996), which may be difficult to implement without sufficient managerial expertise. Bloom et al. (2012) illustrate this point as they show that managerial capabilities explain the US-Europe productivity gap in IT adoption. Specifically, the authors find that US firms have better "people-management" practices[12] that allow US firms to more effectively implement the necessary organizational changes that complement IT adoption. Notably, these findings also generalize to the adoption of digital technologies. Organizational changes are also required to generate value from these technologies as these technologies require employees with data expertise and the creation of new organizational structures that emphasize knowledge sharing (Cockburn et al. 2017). These changes are difficult to implement quickly and could explain why we do not observe immediate changes in firm performance from digital technology adoption (Brynjolfsson et al. 2017).

Second, new technology adoption incurs high fixed costs of implementation and also of creating new markets. Consistent with this view, several empirical studies show that the benefits of technology adoption are higher for firms in geographical regions or industries that have already adopted the technology (Dranove et al. 2014) as shared fixed costs – in the form of developing human capital and physical infrastructure – are lower for later entrants. Moreover, the cost of creating new markets is

12. For example, better reward-punishment practices, performance evaluations.

also a shared fixed cost borne by early entrants. For example, Brynjolfsson and Smith (2000) found that early internet retailers had to provide lower prices and spend more on advertising to convince consumers to trust internet retailing. Similarly, new business products and services that are based on digital technologies may be unfamiliar to consumers, and additional investments must be made by early adopters to create markets for these products and services.

In sum, prior literature outlines various frictions, which may delay or limit the benefits of adopting new technology. Hence, whether the new digital technologies provide net value to firms when adopted is an open empirical question.

Challenges in Empirical Research on Technology Adoption

A key empirical challenge in many studies on technology adoption is the difficulty in identifying investments in new technologies. Measures of R&D or CapEx do not suffice, as these capture the firms' total investment and not just in the new technologies. Therefore, scholars have had to rely on alternative methods of identifying new technology investment. For example, several studies on IT adoption have relied on survey data – One source was Computer Intelligence Infocorp, which tracked the stock of computer hardware across Fortune 1000 firms (Hitt 1999). Another source is the Census Bureau; however, census survey data are limited to only the industry level.

Firm-level data on digital and AI-related technologies are even more sparse. This has led to calls for new measures of digital technology adoption (Seamans and Raj 2018). Recent studies have developed measure of AI adoption through the technology skillsets of workers (Babina et al. 2020; Rock 2019; Tambe et al. 2019; Abis and Veldkamp 2020). However, these labor-based measures cover only a subset of public firms and are not widely available to researchers. In contrast, we develop a new measure of digital technology based on the firm's disclosure of digital activities, which is easily replicable for a large sample of publicly listed firms.

2.1.2 Valuation and Non-Financial Information

The Growing Wedge Between Book Value and Equity Values

Following the rapid growth of the technology industry in the 1990s, several studies examined the failure of accounting systems in measuring the technology investment by firms. Specifically, scholars expressed concern that the rules on accounting for R&D expenditures reduced the value-relevance of accounting numbers because under FAS No. 2, R&D must be immediately expensed. Thus, the accounting for R&D does not capture the underlying economics of the investment. To illustrate that accounting rules obscured a key source of information from markets, Lev and Sougiannis (1996) showed that R&D capitalization is value-relevant to capital markets.

A key point in Lev and Sougiannis (1996) is that the standard accounting of firm performance is unsuited to firms that engage in high levels of R&D. This fact is especially concerning in today's economy, with increasing investment in intangibles through R&D and other forms of expenditures. Indeed, Lev and Zarowin (1999) and Core et al. (2003) find that the value-relevance of accounting measures have decreased over time as a result of the greater importance of intangible investments. This trend suggests that there is a growing wedge between accounting and economic value, which highlight a need for more research into value-relevant, non-financial information.

Value-Relevance of Non-Financial Information

One of the first studies to investigate the value-relevance properties of non-financial information was Amir and Lev (1996). The authors found that for cell phone firms, non-financial metrics, such as the population of the service area, were value-relevant to investors. Similarly, Trueman et al. (2000) and Trueman et al. (2001) showed that internet usage metrics provided value-relevant information about IT companies to investors, above and beyond accounting numbers.

Furthermore, studies have conducted textual analysis of corporate disclosures to examine relationships between non-financial variables and prices, much like we do in this paper (Li 2008, 2010; F. Li et al. 2013; You and Zhang 2009; Feldman et al. 2010; Campbell et al. 2014; Hope et al. 2016). For example, Li (2010) shows that some linguistic features of the qualitative disclosures in the MD&A section of the 10-K are associated with future performance and returns. In sum, these studies emphasize that disclosure of non-accounting/financial information is relevant to markets.

The findings of our study build on the above literature that uses textual disclosures or non-financial metrics to uncover information that is useful for valuation and predicting future performance. In particular, we find that our textual-based proxy for digital activities can help predict future improvements in asset turnover, as well as return performance over a 3-year horizon.

2.1.3 Innovation and Return Predictability

Our study is also related to several studies that have explored the return predictability of innovation activity in firms. A key insight from this literature is that innovation is hard to value for investors and the valuation uncertainty can lead to higher returns. For instance, studies have shown that innovation links (Lee et al. 2019), innovative efficiency (Hirshleifer et al. 2013) and the originality of innovation (Hirshleifer et al. 2018) are undervalued by investors and thus predicts future returns.

On the other hand, some accounting studies have questioned whether innovation activity, as measured by R&D, continues to drive higher future profitability and returns for firms. For instance, Resutek (2021) argues that R&D behaves more like a fixed cost, and thus future returns and profitability of R&D expenditure is due to the operating leverage effect of R&D. Moreover, Curtis et al. (2020) show that the profitability of R&D is declining due to shifts towards investment in safer innovations. Thus, there is also reason to expect that the innovation payoffs, as measured by R&D, is becoming more certain and should thus be associated with less return predictability.

We contribute to these studies in the innovation literature by examining the return predictability of digital activity. The setting of digital activities merits further study for two reasons.

First, the value of digital technologies has increased substantially over the past few decades, due to structural changes in the US economy. Digital technologies have become more valuable partly because macro-economic trends (such as globalization) have shifted the US economy away from manual routine labor and have pushed companies to automate some of their operations (Autor 2015). Furthermore, patent protections for software patents have also increased substantively over the past few decades and have thus increased the value of software-based intellectual property (Jaffe and Lerner 2004). Hence, there is some reason to believe that there is increasing value in the adoption of digital technologies, relative to other forms of investment.

Second, in contrast to existing studies in the innovation literature, our setting studies innovation activity through the deployment of existing digital technologies in non-IT firms. Thus the principal bottleneck in the digital innovation process is arguably the organizational co-invention costs (Bresnahan and Greenstein 1996), rather than the technology itself (Lakhani and Iansiti 2020). In particular, the uncertainty arising from these costs is primarily strategic and are thus quite different from the technological uncertainty observed in typical R&D and patenting activity. Moreover, the uncertainty concerning co-invention costs are also difficult to measure with tangible metrics, which thus portend high uncertainty for investors, and hence greater return predictability.

2.2 Data

We construct our sample from several sources. We begin with all US incorporated or headquartered firms from the intersection of COMPUSTAT and CRSP from fiscal years 2010 to 2020 with share codes 10, 11 and 12 in CRSP. We also include earnings/sales forecasts from IBES and 10-K filings from the SEC Edgar Database.

Our analysis focuses on the digital activities of non-IT firms, so we construct a sample of non-IT firms from our initial sample of firms. We first define IT firms as firms that operate in industries relating to computers, electronics, communications, data processing and internet services. Subsequently, we develop a parsimonious industry-based classification for IT firms by searching through the SIC, NAICS and GICS industries definitions[13] (presented in Appendix A.2) and we remove all firms that are classified as IT firms from our analysis.

The main subject of our study is digital activities, and we proxy for these activities by identifying digital terms in the firms' disclosures. Specifically, we construct a dictionary of digital terms, revolving around 7 topics – analytics, automation, artificial intelligence (AI), big data, cloud (-computing), digitization and machine learning (ML)[14] – to count digital terms in the firms' disclosures. These terms were identified from numerous articles on the digital phenomena as well as glossaries of digital terms provided by consulting firms that specialize in digital transformation.

We analyze the 10-K report to count mentions of digital terms. We focus on the business description section of the 10-K, and we identify the beginning and end of this section by searching for the line with either "Item1" or "Business." and the lines with either "Item1A" or "Risk Factors".

To address concerns that the raw count of words is a noisy measure of digital activity, we quantize the raw counts into terciles that are coded as follows: 0 if no digital activity is disclosed, 1, 2 and 3 if digital mentions fall in the bottom, middle and top tercile of digital mentions in the year, respectively. In the subsequent tests, we use this score as our main proxy for digital activity.

2.2.1 Sample Statistics

We report the sample statistics for the main variables in our study in Table [15] and describe several key characteristics of the sample of non-IT firms below. First, the market-to-book ratio of non-IT firms in our sample, tends to be lower at a mean (median) market-to-book of approximately 3.1 (1.8),

13. For NAICS industries, we use the 2017 industry classification, and convert industries defined in past versions via the NAICS crosswalks. We also drop firms that do not have at least a 4-digit NAICS code. For GICS industries, we define the technology industries based on current (2018 version) and historical GICS codes.
14. We outline the specific words within these topic groups in Appendix A.1
15. The construction of these variables are detailed in Appendix A.3.

compared to 4.2 (2.5) for IT firms. Additionally, the sample firms are older, with a mean (median) age of 25 (21) years compared to 20 (18) years for IT firms[16].

Table 2.1: Summary Statistics

	Mean	Std Dev	Median	25%	75%	N
Market Cap. (Millions)	4597	10114	799	171	3386	21046
Market-to-Book	3.05	3.72	1.79	1.1	3.33	21046
Firm Age	25	16	21	13	34	21046
Leverage Ratio	1.01	1.47	0.56	0.18	1.16	20963
β	1.02	0.54	1.02	0.7	1.35	21019
β_{Tech}	0.04	0.39	0.05	-0.18	0.28	17084
β_{NTech}	1.03	0.63	0.96	0.61	1.38	17084
Earnings Persistence	0.22	0.4	0.13	-0.06	0.46	20445
Return Volatility	0.03	0.02	0.02	0.02	0.03	21021
Days to EA	32	12	32	22	41	17388
Days to 10-K Filing	45	11	43	38	52	21046
Days Between 10-K & EA	10	11	7	0	17	17388
EA CAR(-1,40)	0	0.19	0	-0.08	0.08	15004
Unexpected Earnings	-0.01	0.87	0	-0	0	15101
Unexpected Sales	-0	0.14	0	-0	0.01	14910
Market-Adj. Annual Returns	0.02	0.85	-0.04	-0.25	0.17	21014
Return-on-Assets	0.03	0.16	0.05	0.01	0.1	21045
Return-on-Net Operating Assets	0.04	0.57	0.11	0.03	0.21	12911
Profit Margins	0.13	0.21	0.11	0.04	0.25	19933
Asset Turnover	0.78	0.72	0.61	0.17	1.16	21045
Net Operating Asset Turnover	2.18	2.02	1.57	0.85	2.75	12911
Sales Growth$_{t,t-3}$	0.06	0.16	0.04	-0.02	0.12	21046
SG&A Expense	0.15	0.19	0.07	0.02	0.21	21046
R&D Expense	0.07	0.12	0.02	0	0.06	8935
Loss (Indicator)	0.18	0.39	0	0	0	21046
Tech Manager	0.04	0.19	0	0	0	21046
Total Digital Words	0.6	2.69	0	0	0	21046
Quantized Digital Score	0.31	0.77	0	0	0	21046
Initial Digital Disclosure	0.03	0.17	0	0	0	21046
Digital Patents (Indicator)	0.05	0.21	0	0	0	17396
IT Workers	2.73	3.51	1.7	0.77	3.38	11735

We report the summary statistics of the main control variables in this table for the sample of non-IT firms in fiscal years 2010-2020 (summary statistics of IT firms are reported in Table A.1 in Appendix A.5.1). We examine the statistics of the following variables: market capitalization, market-to-book, firm age, leverage ratio, market beta, beta with respect to the IT and non-IT portfolios, earnings persistence, return volatility, no. of days from fiscal year end to earnings announcement, no. of days from fiscal year end to 10-K filing, no. of days between earnings announcement and 10-K filing, 40-day cumulative abnormal returns after the earnings announcement, unexpected earnings, unexpected sales, market-adjusted annual returns, return-on-assets, return-on-net operating assets, profit margins, asset turnover, net operating asset turnover, past 3-year sales growth (annualized), SG&A expense, R&D expense, an indicator for loss firms and an indicator for firms with technology-related top executives, digital patents (indicator), percentage of IT workers. Descriptions of the variables are outlined in detail in Appendix A.3.

Second, the non-IT firms do not significantly co-move with the IT portfolio, as the average beta on this portfolio is 0.04. By contrast, non-IT firms co-move strongly with the non-IT portfolio, as the

16. We report the sample statistics of the IT firms in Table A.1 in Appendix A.5.1.

average beta on this portfolio is 1.03. Taken together, these statistics suggests that there are substantial characteristic differences between non-IT and IT firms.

2.3 Non-IT Firms and Digital Activity

Our first key finding is that non-IT firms are increasingly engaging in digital activities. To illustrate this, we aggregate the number of digital terms in the 10-K and plot the distribution over time. Figure 2.1 shows that the disclosure of digital activity is steadily increasing over time[17]. This trend speaks to the increasing relevance of the phenomenon and motivates our study.

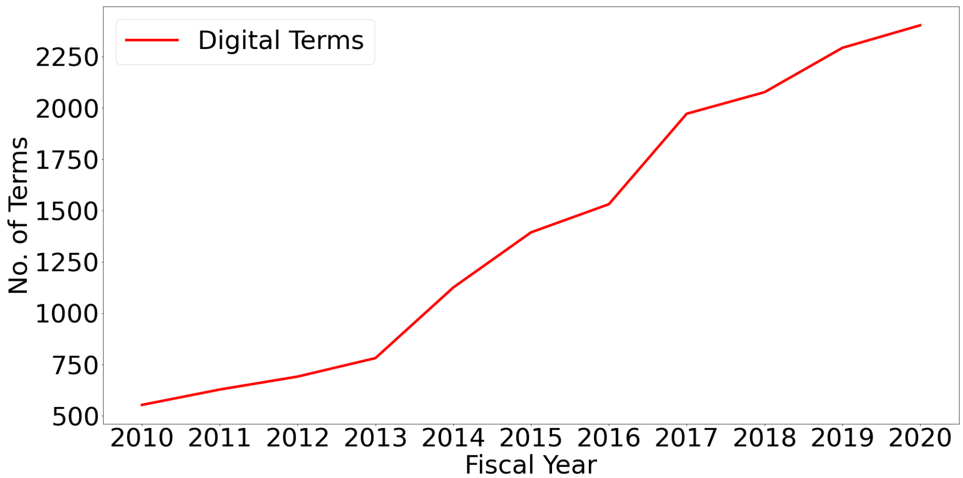
Next, we break down the aggregate digital terms by topic group in Panel A of Table 2.2 and find that the increasing trend exists across all topics. Notably, digital terms are most concentrated in "analytics", which has 1140 mentions in 10-Ks across 338 firms in 2020. The disclosure of "digitization" is also quite frequent, with 415 mentions across 232 firms in 2020. In addition, we report the digital word counts across industries in Panel B of Table 2.2 and find that it is highest in the manufacturing, financial, and services industries, but is also growing across other industries.

2.3.1 Validation of the Digital Activities Proxy

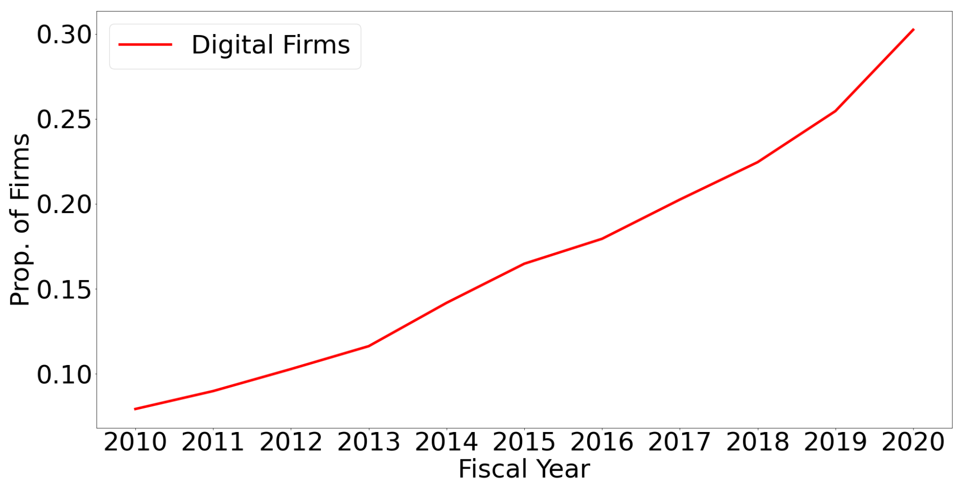
We validate our textual-based proxy of digital activities by examining whether our measure is also associated with other measures that also track the extent of digital activities in firms. We implement this analysis by regressing alternative measures of digital activities on our textual-based proxy. Specifically, we implement the following regression model:

$$Var_{i,t} = \alpha_j + \alpha_t + \beta_1 Digital_{i,t} + \sum_s \gamma_s X_{s,i,t} + \varepsilon_{i,t} \qquad (2.1)$$

17. In Appendix A.4, we provide some examples of how these digital terms are used in the firms' disclosures.



(a) Number of Digital Terms



(b) Proportion of Firms with Digital Terms

Figure 2.1: Number of Digital Terms over Years (a) and Proportion of Firms (b) Disclosing Digital Terms in the Business Description of the 10-Ks

Table 2.2: Distribution of Digital Words

Panel A: Word Group-Year Distribution

	Analytics	Automation	AI	Big Data	Cloud	Digitization	ML
2010	308	34	23	12	12	108	56
2011	326	44	23	18	31	129	57
2012	359	43	19	24	55	158	33
2013	443	29	14	17	82	147	49
2014	646	30	14	39	109	232	55
2015	810	40	15	68	145	257	59
2016	876	29	35	66	166	289	70
2017	1081	55	114	112	172	314	124
2018	1067	66	126	103	219	359	137
2019	1171	70	175	100	243	356	178
2020	1140	79	202	124	237	415	206

Panel B: Industry-Year Distribution

	0100-0999	1000-1499	1500-1799	2000-3999	4000-4999	5000-5199	5200-5999	6000-6799	7000-8999	Total Firms	Prop. Firms
2010	1	7	0	148	6	16	16	178	182	189	0.08
2011	0	6	0	167	11	13	27	198	206	196	0.09
2012	0	3	3	178	15	15	29	189	259	223	0.103
2013	0	3	3	232	21	21	36	191	274	265	0.117
2014	1	5	5	298	34	24	65	314	377	344	0.141
2015	0	6	4	318	58	27	82	349	549	393	0.165
2016	0	10	8	382	66	25	101	378	560	418	0.18
2017	0	15	10	497	70	33	92	523	719	462	0.201
2018	0	17	22	594	94	49	84	450	775	512	0.225
2019	0	23	42	673	133	49	89	473	800	542	0.255
2020	0	31	19	707	134	62	156	550	746	592	0.302

In Panel A, we report the distribution of individual digital words in the business description section of 10-Ks of non-IT firms by fiscal year from 2010 to 2020. The regex expressions used to identify these words are described in the Appendix A.1. In Panel B. we report the distribution of digital words in 10-Ks by SIC divisions-years for non-IT firms from 2010 to 2020. The industry divisions reported are Agriculture, Forestry and Fishing (0100-0999), Mining (1000-1499), Construction (1500-1799), Manufacturing (2000-3999), Transportation, Communications, Electric, Gas and Sanitary service (4000-4999), Wholesale Trade (5000-5199), Retail Trade (5200-5999), Finance, Insurance and Real Estate (6000-6799) and Services (7000-8999). The second-to-last column reports the number of firms that disclose at least one digital term in the year. The last column reports the proportion of firms that disclose at least one digital term in the year.

where we regress the dependent variable on the digital activity proxy and a set of control variables ($\sum_s \gamma_s X_{s,i,t}$), namely, size (log of market cap) and other firm characteristics, namely, firm age, leverage ratio, market-to-book, ROA, sales growth, annual market-adjusted return and the number of words in the business description section of the 10-K. We also control for year and industry (Fama-French 48-industry) fixed effects and cluster standard errors at the firm level.

The first variable that we examine is the probability of filing digital-related patents. These patents are identified using search terms relating to AI, cloud computing, machine learning, neural networks, robots, self-driving cars and software (taken from Bloom et al. 2018; Webb 2020)[18]. The first column of Panel A in Table 2.3 examines the relationship between our digital activities proxy and digital patents, and we show that firms with digital activities exhibit a 1-3% higher probability of filing such patents (or 20-60% higher relative to the sample average of 5%). The second variable measures the proportion of IT workers in the firm[19]. Our analysis in the second column of Panel A show that firms with digital activities tend to have 1-3% higher more IT workers (or 37-111% higher relative to the sample average of 2.7%). Thus overall, our digital activity proxy correlates well with alternative measures of digital-related activities.

We continue our validation analysis by examining whether our digital activity proxy identifies non-IT firms that are economically more (less) similar to other IT (non-IT) firms. To implement this analysis, we first estimate the return co-movement with the IT and non-IT portfolios, in the following regression:

$$R_{i,t} = \alpha + \beta_{Tech} R_{Tech,t} + \beta_{NTech} R_{NTech,t} + \varepsilon_{i,t} \tag{2.2}$$

where daily returns, $R_{i,t}$, is regressed on the value-weighted returns of the IT ($R_{Tech,t}$) and the non-IT

18. We link the patent data to our dataset with the crsp-patent link table provided in Kogan et al. (2017), which covers patent data from calendar years 1925-2019 (September). Thus, we examine the patent data for a sub-sample of firm-year observations from fiscal years 2010-2018: 16,390.

19. We measure the proportion of IT workers from Revelio Labs, which covers a sub-sample of firms in our sample (11,681 firm-year observations).

Table 2.3: Validation Analysis

Panel A: Patent and Employment-Based Validation		
	Digital Patents	IT Workers
Dependent Variable	Indicator	Percentage
$\text{Digital}_{i,t}$	0.010***	1.065***
	(0.003)	(0.157)
Controls	Yes	Yes
Time FE	Yes	Yes
Industry FE	Yes	Yes
Observations	16,390	11,681
Pseudo/Adj. R^2	0.3300	0.2517

Panel B: IT Portfolio Co-Movement (β_{Tech})		
	Levels	Past 3 Year Change
Dependent Variable	$\beta_{Tech,t}$	$\beta_{Tech,t}$ - $\beta_{Tech,t-3}$
$\text{Digital}_{i,t}$	0.018***	0.017***
	(0.004)	(0.005)
Controls	Yes	Yes
Time FE	Yes	Yes
Industry FE	Yes	Yes
Observations	17,024	12,616
Adj. R^2	0.3395	0.1281

Panel C: Non-IT Portfolio Co-Movement (β_{NTech})		
	Levels	Past 3 Year Change
Dependent Variable	$\beta_{NTech,t}$	$\beta_{NTech,t}$ - $\beta_{NTech,t-3}$
$\text{Digital}_{i,t}$	-0.051***	-0.013*
	(0.008)	(0.007)
Controls	Yes	Yes
Time FE	Yes	Yes
Industry FE	Yes	Yes
Observations	17,024	12,616
Adj. R^2	0.4337	0.1616

We validate the digital proxy in this table. In Panel A, we perform the validation analysis by studying the relationship between the digital proxy and digital-related patents (following the classification in Bloom et al. 2018; Webb 2020), a measure of IT workers for the sample of non-IT firms in fiscal years 2010-2020. For the patent variable, we examine the incidence of filing patents with the probit model (average margins are reported) from 2010-2018. For the IT worker variable, we regress on the proportion of IT workers relative to the total number of employees in the firm. In all regressions, we proxy for digital by using a quantized score of the number of digital mentions in the business description of the 10-K (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). All regressions control for size, firm age, leverage, market-to-book, return-on-assets, past 3-year sales growth, market-adjusted annual returns, return volatility, and the number of words in the business description section as well as industry (Fama-French 48-industry) and year fixed effects. In Panels B and C, we report the coefficients of the regressions of IT and non-IT portfolio betas on the proxy for digital activities and controls (we include an additional control for share turnover). β_{Tech} and β_{NTech} are estimated for each fiscal year, by regressing the firm's daily returns on the IT and non-IT portfolio returns. We perform regressions using the levels specification in column 1. In column 2, we perform regressions on the past 3-year changes. Panel A reports the estimates from the IT portfolio co-movement (β_{Tech}), and panel B reports the estimates from the non-IT portfolio co-movement (β_{NTech}). In the changes specification, control variables are lagged by 3-years. Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

portfolio ($R_{NTech,t}$) over the fiscal period for each firm-year[20]. The estimates of interest are β_{Tech} and β_{NTech}, which are the co-movement to the IT and non-IT portfolio, respectively.

Using these estimated β's, we apply the regression model specified in equation 1, with an additional control for share turnover to address measurement errors arising from bid-ask bounce in thinly traded and low visibility firms (Piotroski and Roulstone 2004; Crawford et al. 2012). Panel B in Table 2.3 presents our results on the association between β_{Tech} and digital activity using the levels specification and 3-year changes, respectively. In Column 1 in Panel B, we report the levels specification and find that digital activity is strongly associated with greater co-movement with the IT portfolio, as firms with digital activities exhibit 45-135% higher co-movement with this portfolio (i.e., a firm in the top tercile of digital disclosure has a β_{Tech} that is 0.054 higher than the sample average of 0.04, or 135%). Column 2 report the changes from 3 years-prior to current β_{Tech}, regressed on the digital activity proxy and the lagged controls. Our results show that firms with digital activities have increased co-movement by 42-127% over 3 years (that is, a firm in the top tercile of digital disclosure increases β_{Tech} by 0.051 relative to the sample average of 0.04, or 127%), which is roughly 88% of the contemporaneous difference between the β_{Tech} of firms with digital activities and industry peers. Combined, these results suggest that our digital activity proxy is measuring activities within non-IT firms that lead these firms to become more tech-like.

Next, we examine whether firms with digital activities co-move less with the non-IT firms, in Panel C. We find that digital activity is associated with less co-movement with the non-IT portfolio, as firms with digital activities exhibit 5-15% less co-movement with this portfolio. We also find that firms that engage in digital activities become less like non-IT firms over prior years. Column 2 report

20. The IT portfolio consists of all IT firms classified based on Appendix A.2. The returns within the portfolio are value-weighted, and we re-balance portfolio weights at the daily-level. The non-IT portfolio is defined similarly but consists of firms that are classified as non-IT. To reduce the effects of low liquidity stocks from inducing measurement error in the return regressions, we drop penny stock entries with less than $5 in price. Also, to reduce measurement error, betas estimated with less than 200 observations are dropped from the analysis. Due to these sample restrictions, the analysis is based on a sub-sample of 17,024 firm-year observations.

the changes from 3 years-prior to current β_{NTech}, regressed on the digital activity proxy and the lagged controls, and shows that firms with digital activities experience declines of 1.31-3.93% in β_{NTech} over 3 years, which is roughly 26% of the contemporaneous difference between the β_{NTech} of firms with digital activities and industry peers. Combined, the results in this panel complement our findings for β_{Tech} and suggest that our digital activity proxy is also measuring activities within non-IT firms that lead these firms to be less like their peers.

2.3.2 Determinants of Digital Activity

Next, we examine various determinants of firm-level digital activity in the following regression model, which regresses our proxies for digital activity on lagged determinant variables:

$$\begin{gathered} Digital_{i,t} = \alpha_j + \alpha_t + \beta_1 Digital_{i,t-1} + \beta_2 SIZE_{i,t-1} + \beta_3 MB_{i,t-1} + \beta_4 LEV_{i,t-1} \qquad (2.3) \\ +\beta_6 AGE_{i,t-1} + \beta_7 SALES_{i,t-1,t-4} + \beta_8 CASH_{i,t-1} + \beta_9 R\&D_{i,t-1} + \beta_{10} Missing\ R\&D_{i,t-1} \\ \beta_{11} SG\&A_{i,t-1} + \beta_{12} CapEx_{i,t-1} + \beta_{13} Returns_{i,t-1} + \beta_{14} TechManager_{i,t-1} + \varepsilon_{i,t} \end{gathered}$$

where our proxy for digital activity, is regressed on lagged digital activity and several determinant variables, which we describe in Appendix A.3. We control for year and industry (Fama-French 48-industry) fixed effects, to perform cross-sectional and within industry comparisons, and cluster standard errors at the firm level.

Table 2.4 presents regression results on the determinants of digital activity. We first run a determinant analysis of equation 3 in columns 1 and 2, we drop the industry fixed effects and include the lagged industry-level digital activity as a determinant. We perform a similar analysis on the sub-sample of observations with initial disclosures and controls (i.e. we drop observations with subsequent digital disclosures) in columns 3 and 4.

Across all columns, we find that several variables significantly explain digital activities, namely, lagged digital activity, size, return volatility, age, SG&A expenditure and the lagged industry-level of digital activity. The coefficients on size, age and return volatility indicate that firms that are larger, younger and face more uncertainty tend to engage in digital activities. Finally, the results also suggest that more SG&A intensive firms engage in digital activity, while more CapEx intensive firms are less likely to engage in digital activities.

Notably, lagged digital activity explains a significant amount of variation in the determinants model as this variable increases the adjusted R^2 from 0.20 to 0.73. Furthermore, a firm that has disclosed at least 1 digital term has a 88% unconditional probability of doing the same the next year, which suggests that digital activity is a persistent process. Lagged industry-level of digital activity also determines firm-level digital activity, which suggests that there are industry-wide complementarities that increase the benefits and thus likelihood of engagement in digital activities.

Next, we study the determinants of initial digital activity digital activity in Columns 3-4. Notably, we find that sales growth, market-to-book and stock returns are either negatively associated or not statistically significantly related with initial activity[21][22]. We therefore find little evidence of more successful firms engaging in digital activities which alleviates concerns that the valuation effects of these technologies are due to selection bias. Somewhat surprisingly, we also find that firms with less R&D tend to engage in digital activity more. As Koh and Reeb (2015) show that R&D disclosure is affected by voluntary non-disclosure, we also examine the indicator of missing R&D as a determinant. We find that the coefficient on this indicator is negative, which suggests that non-disclosure is not a likely explanation for the negative relationship between R&D and digital activities.

21. For the initial activity sample, we drop all observations of subsequent digital activity, which leads the sample size to drop from 20,527 to 16,563. Note that this regression compares first disclosers to non-disclosers who form the majority of the sample (there are only 549 first disclosers). This suggest that the majority of the disclosed digital activity is subsequent disclosure, which aligns with the fact that digital disclosure is highly persistent.

22. We do find that lagged ROA is positively associated with higher digital score in the full sample. But we do not find a statistically significant effect in the initial activity sample with industry fixed effects, suggesting that this finding is driven by the increase in ROA after digital disclosure is initiated (and this is corroborated in Panel

Table 2.4: Determinants of Digital Activity

Dependent Variable	Full Sample	Full Sample	Initial Activity	Initial Activity
$\text{Digital}_{i,t-1}$	0.865^{***}	0.885^{***}		
	(0.008)	(0.007)		
$\text{SIZE}_{i,t-1}$	0.018^{***}	0.018^{***}	0.013^{***}	0.014^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
$\text{Market-to-Book}_{i,t-1}$	0.001	0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
$\beta_{i,t-1}$	-0.005	-0.002	-0.005	-0.002
	(0.006)	(0.005)	(0.004)	(0.004)
$\text{Return Volatility}_{i,t-1}$	0.589^{**}	0.715^{***}	0.717^{***}	0.822^{***}
	(0.249)	(0.234)	(0.199)	(0.187)
$\text{Leverage}_{i,t-1}$	-0.000	0.000	0.004^{*}	0.004^{*}
	(0.002)	(0.002)	(0.002)	(0.002)
$\text{Return-on-Assets}_{i,t-1}$	0.051^{*}	0.092^{***}	0.010	0.043^{*}
	(0.029)	(0.028)	(0.025)	(0.024)
$\text{AGE}_{i,t-1}$	-0.027^{***}	-0.022^{***}	-0.017^{***}	-0.014^{***}
	(0.005)	(0.005)	(0.004)	(0.004)
$\text{Sales Growth}_{i,t-1,t-4}$	-0.043^{**}	-0.053^{***}	-0.028^{*}	-0.032^{**}
	(0.018)	(0.018)	(0.015)	(0.014)
$\text{CASH}_{i,t-1}$	0.010	0.033^{*}	0.026	0.046^{***}
	(0.022)	(0.020)	(0.018)	(0.016)
$\text{R\&D}_{i,t-1}$	-0.004	-0.082	-0.109^{**}	-0.117^{**}
	(0.065)	(0.056)	(0.051)	(0.047)
$\text{Missing R\&D}_{i,t-1}$	-0.009	-0.005	-0.012	-0.014^{**}
	(0.010)	(0.007)	(0.008)	(0.006)
$\text{SG\&A}_{i,t-1}$	0.057^{**}	0.094^{***}	0.023	0.067^{***}
	(0.024)	(0.019)	(0.019)	(0.015)
$\text{CAPEX}_{i,t-1}$	-0.192^{***}	-0.119^{**}	-0.115^{**}	-0.040
	(0.073)	(0.051)	(0.056)	(0.039)
$\text{Stock Returns}_{i,t-1}$	-0.008^{*}	-0.008^{*}	-0.007^{***}	-0.006^{**}
	(0.004)	(0.005)	(0.002)	(0.002)
$\text{Tech Manager}_{i,t-1}$	0.014	0.017	0.015	0.018
	(0.017)	(0.017)	(0.015)	(0.015)
$\text{Industry Digital}_{j,t-1}$		0.001^{***}		0.001^{***}
		(0.000)		(0.000)
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	No	Yes	No
Observations	20,527	20,527	16,563	16,563
Adj. R^2	0.7297	0.7266	0.0245	0.0166

We report the determinants of digital activity in this table for the sample of non-IT firms in fiscal years 2010-2020. The dependent variable in all regression model is the quantized score of digital mentions in the business description of 10-Ks (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). In Columns 1 and 2, we study the full sample. In Columns 3 and 4, we study only the first disclosure of digital terms, by removing observations where the firm makes subsequent disclosure of digital terms. Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level respectively.

2.4 Valuation and Performance Implications

2.4.1 Market Valuations and Digital Activity

We begin by examining whether the market-to-book ratio reflects digital activity. In these tests, we regress market-to-book on our digital activity proxy and controls, in the following model:

$$MB_{i,t} = \alpha_j + \alpha_t + \beta Digital_{i,t} + \sum_s \gamma_s X_{s,i,t} + \varepsilon_{i,t} \tag{2.4}$$

where $MB_{i,t}$ is the ratio of the market value on the day after the 10-K filing date to stockholder'books equity[23]. The independent variables consist of the quantized scores for digital activities ($Digital_{i,t}$), and a set of control variables ($\sum_s \gamma_s X_{s,i,t}$), namely, size, age, leverage ratio, ROA, past 3-year sales growth, market-adjusted annual returns, return volatility, and the number of words in the business description section. Additionally, we control for year and industry fixed effects, to perform cross-sectional and within industry comparisons[24], and cluster standard errors at the firm level.

Table 2.5 presents our market-to-book results. In Panel A, we report the associations between digital activity and market-to-book. In column 1, we find that firms with digital activities are associated with a market-to-book that is 8-26% higher than peers[25] (i.e., a firm in the top tercile of digital disclosure has a market-to-book that is 0.798 higher than the sample average of 3.05, or 26%). The second column indicates that part of the differences in market-to-book is due to capitalization restrictions of intangibles, as the magnitudes range from 5-14% when intangible investment (SG&A, R&D

A of Table 9).

23. We use stockholder's equity in Compustat to measure book equity (following, Abdel-Meguid et al. 2021; Jennings et al. 2020)

24. We compare ratios relative to industry peers, as prior work shows that investors benchmark accounting and valuation ratios relative to industry peers (Jennings et al. 2020).

25. We also examine the valuation effects of digital activity in the cross-section in Table A.5 in Appendix A.5.5. We find firms that are larger, expend more on SG&A, CapEx and in high digital adoption industries, receive higher valuations.

and an missing R&D indicator to control for voluntary non-disclosure of R&D following, Koh and Reeb (2015)) are added as a control[26]. Notably, the economically significant magnitude in this specification indicates that the differences in market-to-book are not fully driven by capitalization restrictions. To address concerns of selection bias, we also implement a regression with the lagged dependent variable to assess the valuation effects of digital activities between firms with similar market-to-book. In this specification, we find that digital activities are associated with 2-5% higher market-to-book. While magnitudes are substantially smaller, we note that this specification yields highly conservative estimates as it removes the valuation effects from past years of digital activity.

We further study the effects of capitalization restrictions on the market-to-book differences and changes, by examining the valuation effects on conservatism corrected market-to-book (McNichols et al. 2014). Specifically, we estimate the conservatism correction factor for a sub-sample of firms with sufficient length of investment histories and compute a corrected-version of market-to-book that explicitly adjusts for the missing capitalization of intangibles[27]. We present the results with this sub-sample in Panel B, Table 2.5. In columns 1-2, we report the results using the unadjusted market-to-book to provide a baseline for this sub-sample of firms[28]. Our estimates show that firms with digital activities in this sub-sample exhibit a 13-39% higher market-to-book relative to industry peers (i.e., a firm in the top tercile of digital disclosure has a market-to-book ratio that is 1.296 higher than the sample average of 3.34[29], or 39%) and this effect is robust to a specification with the lagged depen-

26. Note that this estimate could be viewed as conservative as the controls for intangibles also absorbs the effects that digital investments may have on future investment opportunities and growth.

27. This approach corrects for accounting conservatism by first estimating the conservatism correction factor, which is the ratio of the capitalized tangible and intangible assets (via the cost accounting method over the estimated useful life of assets) to capitalized tangible assets (via the straight-line depreciation method over the estimated useful life of assets). Market-to-book is then adjusted by dividing by this ratio. See Appendix A.5.2 for more details on the methodology and theory behind the computation of this conservatism correction factor.

28. As the conservatism-adjustment drops firms with insufficient investment histories, we conduct the analysis on a small sub-sample of firms. We also follow McNichols et al. (2014) in dropping financial firms (SIC 6000-6779), firms with assets < 4 million and net PPE-to-asset ratio < 0.1 as the conservatism correction is less suited for these firms. Consequently, our sample for this analysis consists of 7,257 firm-year observations.

29. This is based on the average market-to-book of this sub-sample reported in Table A.2 in Appendix A.5.2

Table 2.5: Market-to-Book

Panel A: Market-to-Book			
Dep. Var.	$MB_{i,t}$	$MB_{i,t}$	$MB_{i,t}$
$Digital_{i,t}$	0.266***	0.138**	0.055*
	(0.073)	(0.066)	(0.032)
Baseline Controls	Yes	Yes	Yes
Intangibles Controls	No	Yes	Yes
Lagged Dep. Var.	No	No	Yes
Time FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Observations	20,930	20,930	20,254
Adj. R^2	0.3774	0.4485	0.7042

Panel B: Conservatism-Corrected Market-to-Book				
	Unadjusted		Adjusted	
Dep. Var.	$MB_{i,t}$	$MB_{i,t}$	$MB_{i,t}$	$MB_{i,t}$
$Digital_{i,t}$	0.432***	0.119**	0.149***	0.026*
	(0.129)	(0.051)	(0.048)	(0.015)
Controls	Yes	Yes	Yes	Yes
Lagged Dep. Var.	No	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	7,257	6,342	7,257	6,342
Adj. R^2	0.4509	0.8060	0.2663	0.7992

Panel C: IV First Stage				
Dep. Var.	$Digital_{i,t}$	$Digital_{i,t}$	$Digital_{i,t}$	$Digital_{i,t}$
AI Technology $Exposure_{i,t-1}$	0.068***	0.067***	0.103***	0.085***
	(0.013)	(0.013)	(0.020)	(0.022)
Baseline Controls	Yes	Yes	Yes	Yes
Intangibles Controls	Yes	Yes	No	No
Lagged Dep. Var.	No	Yes	No	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Sample	Full	Full	CC-Sample	CC-Sample
Observations	20,254	20,254	6,948	6,342
Adj. R^2	0.2144	0.2151	0.1924	0.1814

Table 2.5: Market-to-Book (Continued)

Panel D: IV Regression				
	Unadjusted Market-to-Book		CC-Adjusted Market-to-Book	
Dep. Var.	$MB_{i,t}$	$MB_{i,t}$	$MB_{i,t}$	$MB_{i,t}$
$\widehat{Digital}_{i,t}$	2.686***	0.955**	0.737**	0.343*
	(0.826)	(0.430)	(0.343)	(0.193)
Baseline Controls	Yes	Yes	Yes	Yes
Intangibles Controls	Yes	Yes	No	No
Lagged Dep. Var.	No	Yes	No	Yes
Sample	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Full	Full	CC-Sample	CC-Sample
Observations	20,254	20,254	6,948	6,342
Adj. R^2	0.0414	0.6096	0.1309	0.7578
Kleibergen-Paap F-Stat	28.8919	27.8955	25.5953	14.9241

We report the coefficients of the regressions of market-to-book on the proxy for digital activities for the sample of non-IT firms in fiscal years 2010-2020. In Panel A, we report the associations between market-to-book and digital activity. In Panel B, we report the associations between conservatism-corrected market-to-book and digital activity. In the regressions, we proxy for digital by using a quantized score of the number of digital mentions in the business description of the 10-K (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). All regressions control for size, firm age, leverage, return-on-assets, past 3-year sales growth, market-adjusted annual returns, return volatility, number of words in the business description section and industry (Fama-French 48-industry) and year fixed effects. In Panel A, we also run an additional specification with controls for intangible investment, namely SG&A, R&D and an indicator for missing R&D, as well as a specification that controls for the lagged dependent variable. In Panel B, we focus on a sub-sample of observations with valid consertaivism-corrected market-to-book, and we present the unadjusted multiple as the dependent variable in the first two columns and the conservatism-corrected multiple in the next two columns. In Panels C and D, We report the coefficients of the 2SLS regressions of market-to-book on the proxy for digital activities for the sample of non-IT firms in fiscal years 2010-2020, with one-year lagged AI technology exposure as an instrument for digital activities. In Panel C, we report the first stage regression of digital activities on the AI exposure score. In Panel D, we report the relationship between market-to-book and the instrumented digital activity score. In columns 1-2, we report the results for the unadjusted market-to-book for the full sample of firms. In columns 3-4, we report the results for the adjusted market-to-book. For columns 1-2, we also include controls for intangible investment, namely SG&A, R&D and an indicator for missing R&D. In columns 2 and 4, we include the lagged dependent variable. Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

dent variable. We control for the capitalization effects in next two columns, by using conservatism-corrected market-to-book as the dependent variable. In this specification, we find that firms with digital activities exhibit 10-29% higher market-to-book (i.e., a firm in the top tercile of digital disclosure has a market-to-book ratio that is 0.447 higher than the sample average of 1.53[30], or 29%) and again, this estimate is robust to the specification with the lagged dependent variable. We compare the estimates from both versions of market-to-book, and the comparison suggests that roughly 25% of the level differences in market-to-book can be attributed to capitalization restrictions.

To further address concerns of selection bias, we implement an IV analysis. We instrument for the digital activity score by the extent to which industries are exposed to AI technology. This exposure variable is estimated by the GloVe cosine similarity between the title/abstracts in AI-related patents (which are identified using search terms in Webb 2020) and the NAICS industry description (following the approach in Kogan et al. 2020)[31]. The relevance criterion of this instrument is likely satisfied as the exposure variable identifies industries that are more likely to benefit from, and hence, adopt AI technologies. Panel C of Table 2.5 confirms this view, as the instrument exhibits a strong first stage. For the exclusivity criterion, we argue that the instrument likely satisfies this condition as most patents are filed in universities and thus it's unlikely that the patent-based measure is linked to selection biases relating to the firm. With this instrument, we also find that digital activity is associated with higher market-to-book in Panel D. For the unadjusted specification, we find that digital activities are associated with a 88-264% higher market-to-book relative to peers. And for the adjusted specification, we find that digital activities are associated with a 48-144% higher conservatism-corrected market-to-book relative to peers.

Next, we supplement our market-to-book tests by examining whether the market values earnings more following digital activity. If digital activities increase firm valuations, we should also observe

30. See Table A.2 in Appendix A.5.2 for more details on the sample statistics of the conservatism-corrected market-to-book.

31. For more details on the methodology please see Appendix A.5.3

increases in the earnings response coefficients (ERC) as investors foresee higher future growth opportunities for the firm and consequently value current earnings more[32]. We measure the changes in investors' valuation of earnings using the following ERC regression:

$$CAR_{i,t} = \alpha_j + \alpha_t + \beta_1 UE_{i,t} + \beta_2 Digital_{i,t} + \beta_3 UE_{i,t} \times Digital_{i,t} + \sum_s \gamma_s X_{s,i,t} \qquad (2.5)$$
$$+ \sum_s \partial_s UE_{i,t} \times X_{s,i,t} + \sum_s \partial_s UE_{i,t} \times \alpha_j + \sum_s \partial_s UE_{i,t} \times \alpha_t + \varepsilon_{i,t}$$

where $CAR_{i,t}$ represents the (-1,40) window[33] cumulative abnormal returns[34] around the earnings announcement and is regressed on the unexpected earnings (UE), which are estimated by the actual EPS minus the most recent median IBES consensus[35], and several controls and interactions that incrementally explain the baseline returns-earnings relationship, which is measured by β_1, the earnings response coefficient (ERC). Our primary coefficient of interest is β_3, which measures the incremental impact of digital activity ($Digital_{i,t}$) on the ERC. $\sum_s \gamma_s X_{s,i,t}$ represent the list of controls in the ERC regression. Following prior literature (Easton and Zmijewski 1989; Collins and Kothari 1989; Dehaan et al. 2017; Gipper et al. 2019), we control for several variables (and their interactions with UE) that explain variation in the ERC: market cap., market-to-book, leverage ratio, market beta, loss (indicator), persistence, return volatility, earnings announcement and 10-K filing lag. In addition, we also control for intangibles by including SG&A, R&D and the indicator for missing R&D (following Koh and Reeb 2015) in the set of controls. To ensure that industry- or time-based trends do not in-

32. The logic underlying the valuation interpretation of the ERC stems from an accounting literature that views the ERC coefficient as capturing the market's expectation of the capitalization rate of earnings (Easton and Zmijewski 1989; Collins and Kothari 1989; Dechow et al. 2014)

33. We chose this return window as the 99th percentile of the lag between earnings announcement and 10-K filing date is 39 days. We drop observations where the lag is greater than 40 days.

34. Abnormal daily returns are calculated by taking the raw return minus the Carhart, Fama-French four-factor expected returns (Carhart 1997), where the expected returns are estimated with the β's of the four-factor model that are estimated in a (-280,-60) window.

35. We remove consensus forecasts that are more than 100 days old at the time of the announcement and remove forecasts in which the price at the end of the fiscal period is less than $1 and UE greater than the price.

fluence our findings, we also add industry and time fixed effects (and their interactions with UE), and cluster standard errors at the firm level[36].

Panel A of Table 2.6 reports the results of ERC tests. Column 1 presents the baseline (the regression model with only controls and UE interactions) ERC coefficient, and we report an ERC coefficient of 2.710. Column 2 explores the interactive effect of digital activity, proxied by the quantized score of digital terms, on the ERC model. Consistent with our expectations, we find that the coefficient on $UE_{i,t} \times Digital_{i,t}$ is statistically significant and suggests that firms with digital activities exhibit ERCs that are 16-50% higher than industry peers (i.e., a firm in the top tercile of digital disclosure has an ERC that is 1.35 higher than the baseline ERC of 2.710, or 50%)[37].

As the earnings-returns relationship is characterized by non-linearities (Freeman and Tse 1992), one might be concerned that the increase in ERC for firms with digital activities could be an artifact of these non-linearities. To address this potential confound, we implement the same regressions, but with UE deciles which are ranked each year in columns 3-4. Like our previous results, we find that firms with digital activities tend to have higher ERCs as the estimates suggest that a firm with digital activities exhibit a decile-based ERC that is 22-66% higher than the baseline[38].

We recognize that the same accounting rules regarding expensing of R&D affect the ERC tests as they did in the market-to-book tests. Therefore, we also examine the market response to unexpected sales or the sales response coefficient (SRC). Unlike book equity and earnings, sales are not affected

36. To further control for firm-level heterogeneity in the UE and returns relationship, we also examine an alternative specification with grouped firm fixed effects based on 10 by 10 size and beta portfolios in Table A.3 in Appendix A.5.4. We find similar results under this specification.

37. To further control for firm-level heterogeneity in the UE and returns relationship, we also examine an alternative specification with grouped firm fixed effects based on 10 by 10 size and beta portfolios in A.3 in Appendix A.5.4. We find similar results under this specification.

38. In addition, following the approach in, we also examine the fitted ERC curves for digital and non-digital firms using fractional polynomials to model ERC non-linearities. Our results, presented in Figure A.1 in Appendix A.5.4 shows that digital firms tend to exhibit greater return reactions to both positive and negative unexpected earnings (albeit at the more extreme end for negative earnings), consistent with these firms exhibiting a higher ERC coefficient.

Table 2.6: Market Response to Earnings and Sales

	Raw Values		Yearly Deciles	
	Baseline	With Digital	Baseline	With Digital
Panel A: Unexpected Earnings				
Dependent Variable	CAR (-1,40)	CAR (-1,40)	CAR (-1,40)	CAR (-1,40)
Unexpected Earnings$_{i,t}$	2.710***	4.365***	0.009***	0.013**
	(0.411)	(1.374)	(0.001)	(0.006)
Digital$_{i,t}$		0.003**		-0.007
		(0.002)		(0.005)
Digital$_{i,t}\times$Unexpected Earnings$_{i,t}$		0.529**		0.002**
		(0.268)		(0.001)
Controls	Yes	Yes	Yes	Yes
Time FE	No	Yes	No	Yes
Industry FE	No	Yes	No	Yes
Unexpected Earnings $\times$ Controls	Yes	Yes	Yes	Yes
Unexpected Earnings $\times$ Time FE	No	Yes	No	Yes
Unexpected Earnings $\times$ Industry FE	No	Yes	No	Yes
Observations	14,308	14,308	14,308	14,308
Adj. R^2	0.0244	0.0355	0.0339	0.0375
Panel B: Unexpected Sales				
Dependent Variable	CAR (-1,40)	CAR (-1,40)	CAR (-1,40)	CAR (-1,40)
Unexpected Sales$_{i,t}$	0.415***	-1.080**	0.004***	-0.002
	(0.153)	(0.456)	(0.001)	(0.005)
Digital$_{i,t}$		0.002		-0.006
		(0.002)		(0.004)
Digital$_{i,t}\times$Unexpected Sales$_{i,t}$		0.257**		0.001*
		(0.117)		(0.001)
Controls	Yes	Yes	Yes	Yes
Time FE	No	Yes	No	Yes
Industry FE	No	Yes	No	Yes
Unexpected Sales $\times$ Controls	Yes	Yes	Yes	Yes
Unexpected Sales $\times$ Time FE	No	Yes	No	Yes
Unexpected Sales $\times$ Industry FE	No	Yes	No	Yes
Observations	14,006	14,006	14,006	14,006
Adj. R^2	0.0234	0.0299	0.0251	0.0305

We report the coefficients to the ERC (Earnings Response Coefficient)/SRC (Sales Response Coefficient) regression with the proxy for digital activities in this table for the sample of non-IT firms in fiscal years 2010-2020. In Columns 1 and 2, we report the ERC/SRC regression at the using raw unexpected values, where CAR(-1,40) is regressed on unexpected earnings/sales, controls, industry and year fixed effects as well as their interactions with unexpected earnings/sales. In Columns 3 and 4, we report the ERC/SRC regression at the using yearly decile rankings of unexpected earnings/sales, where CAR(-1,40) is regressed on unexpected earnings/sales, controls, industry and year fixed effects, as well as their interactions with unexpected earnings/sales. Columns 2 and 4 include our proxy for digital activities as an interaction variable. We proxy for digital activity in the regression models by the quantized score of the number of digital mentions in the business description of the 10-K (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). All regressions control for size, leverage ratio, loss (ind.), persistence, return volatility, past 3-year sales growth, SG&A expenditure, R&D expenditure, indicator for missing R&D, capital expenditures, the number of days to EA, the number of days to 10-K filing and the number of words in the business description section. For ease of interpretation of the unexpected earnings/sales coefficient, we mean-center all continuous control variables. Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

by the capitalization restrictions and thus the SRC analysis serves as a useful robustness analysis that addresses the capitalization restriction concern. To examine the SRC, we run the same regression model as before but with unexpected sales, which are estimated by the actual sales per share minus the most recent median IBES consensus[39]. Panel B of Table 2.6 report the results for the SRC tests. As before, we report the baseline SRC model in Column 1, which is 0.415. In Column 2, we explore the interactive effects of digital activities, and our findings mirror the results obtained in the ERC tests, as our estimates suggest that firms with digital activities exhibit an SRC that is 62-186% higher than industry peers. Columns 3 and 4 implements the same regression model with yearly unexpected sales deciles and we find also find that firms with digital activities exhibit a decile-based SRC that is 25-75% higher than the baseline[40].

To address this question, we examine return predictability based on our digital proxy. We first construct portfolios in June of each year, starting from 2011, by holding firms in the long position if they are in the top tercile of firms that disclose digital terms in the business description section of the 10-K[41], and holding firms in the short position if they have not disclosed digital terms.

We then track the performance of these long-short digital portfolios over 3 years using DGTW-adjusted returns following the methodology in Daniel and Titman (1997). These risk-adjusted returns are first calculated at the firm level by deducting the corresponding size, book-to-market and momentum quintile portfolios from the raw returns. We then aggregate to the digital portfolio returns by taking the weighted average of these returns based on the market capitalization of the firms at the portfolio formation date. To account for the changing risk profile of firms, we allow the benchmark portfolio to change every year in June. Furthermore, to help address survivorship bias, if a firm delists

39. Similarly, we remove consensus forecasts that are more than 100 days old at the time of earnings announcement and remove forecasts in which the price at the end of the fiscal period is less than $1 and unexpected sales greater than the price.

40. We also examine the robustness of the SRC results by implementing SRC regressions with grouped fixed effects and by examining the fitted SRC using fractional polynomials in Table A.3 and Figure A.2 of Appendix A.5.4. The inferences from both sets of analysis corroborate the main results presented above.

41. We assume 10-K information to be publicly available by four months after the fiscal year end.

before the end of the returns horizon period we include the delisting return[42] and reinvest the proceeds in a value-weighted market portfolio, while also reinvesting the benchmark return to the end of the horizon period[43].

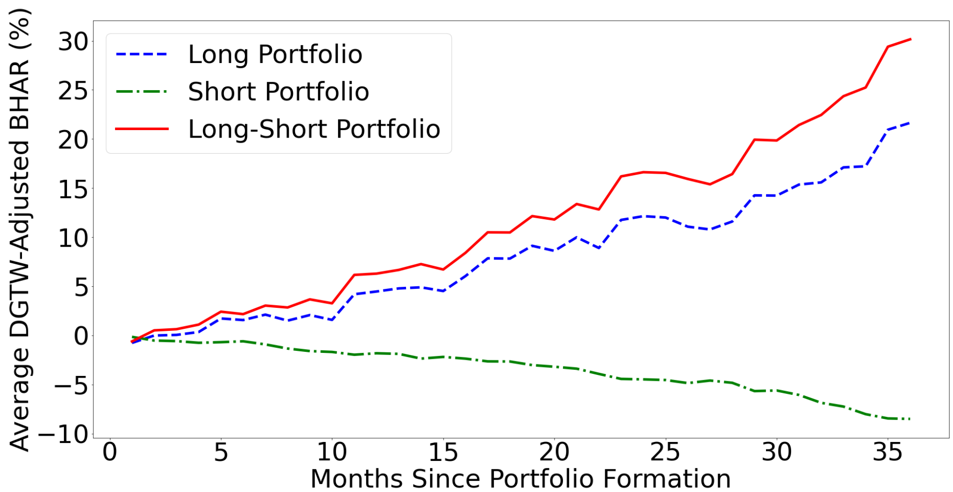


Figure 2.2: Average Size, Book-to-Market and Momentum (DGTW) Adjusted Returns to Value-Weighted Portfolios Formed on Digital Disclosure

Our results reported in Figure 2.2 shows that value-weighted portfolios formed on digital disclosure consistently predict positive returns. Notably, by the end of the third year, our results indicate that an investor can earn a 30% risk-adjusted return. We tabulate the time-series average return performance at the full 3-year horizon in column 1 of Panel A in Table 2.7 and find that the long-short portfolio formed on digital disclosure exhibits statistically significant returns over 3 years. In addition, we examine the portfolio returns in the 1-3 years of the holding period separately in columns 2-4 and observe significant long-short returns for all three intervals. One caveat to our return results is that a portion of the long-short returns comes from the short side. While this may be puzzling because firms

42. Following Shumway (1997) and Shumway and Warther (1999) we code the delisting return as -30% and -55% if the firm delists for performance reasons from NYSE and NASDAQ respectively.

43. Additionally, to further account for low liquidity and high transactions costs in penny stocks, we also remove stocks with prices below $5 at the portfolio formation date.

with digital activities form a small proportion of our sample, we note that we are considering only the universe of non-IT firms. Thus, our results also suggest that non-IT firms that fail to engage in digital activities have performed poorly relative to size, book-to-market and momentum matched portfolios in our sample period.

To further address concerns that other forms of risk may be driving our results, we run panel regressions of raw returns on an indicator ($Digital_{i,t}$) proxying for the digital long-short strategy (that is, an indicator that is coded 1 for top tercile digital disclosure, -1 for no digital disclosure and 0 otherwise), and controls for risks, namely, size ($SIZE_{i,t}$), book-to-market ($BM_{i,t}$), operating profit ($OP_{i,t}$), investment ($INV_{i,t}$), momentum ($MOM_{i,t}$), SG&A and R&D, as well as year fixed effects. Specifically, we implement the following regression model:

$$R_{i,t} = \alpha_t + \beta_1 Digital_{i,t} + \beta_2 SIZE_{i,t} + \beta_3 BM_{i,t} + \beta_4 OP_{i,t} + \beta_5 INV_{i,t} + \beta_6 MOM_{i,t} \quad (2.6)$$
$$+\beta_7 SG\&A_{i,t} + \beta_8 R\&D_{i,t} + \varepsilon_{i,t}$$

We implement the regression at the portfolio year level. To address serial correlation, we cluster standard errors at the firm level. Panel B of Table 2.7 reports these results and confirming our prior portfolio-level results, we find that the long-short strategy yields significant risk-adjusted returns of 0.2%, 2.9%, 5.8% and 8.0% at the monthly and 1-3-year horizons[44].

We also assess whether the digital long-short strategy yields positive returns in calendar time, by turning to calendar-time portfolio regressions. We implement this by evaluating the alpha from a regression of the long-short portfolio returns on the Fama and French (2015) five factors and the momentum factor as described below:

$$R_{p,t} = \alpha_p + \beta_1 MKT_{i,t} + \beta_2 SMB_{i,t} + \beta_3 HML_{i,t} + \beta_4 RMW_{i,t} + \beta_5 CMA_{i,t} + \beta_6 MOM_{i,t} + \varepsilon_{i,t} \quad (2.7)$$

44. We also examine the returns without controls for risk factors and find that the results are relatively unchanged.

Table 2.7: Portfolio Returns

Panel A: Time-Series Average Long-Run Portfolio Returns				
Portfolio	RET(1,36)	RET(1,12)	RET(13,24)	RET(25,36)
Long	0.216**	0.034	0.060*	0.131*
	(0.081)	(0.022)	(0.032)	(0.058)
Short	-0.085***	-0.027***	-0.033**	-0.060***
	(0.021)	(0.008)	(0.010)	(0.015)
Long - Short	0.302***	0.062**	0.091**	0.187***
	(0.068)	(0.022)	(0.031)	(0.051)

Panel B: Returns Panel Regressions				
Dep. Var.	Monthly Returns	1-Year Ahead Buy-Hold Returns	2-Year Ahead Buy-Hold Returns	3-Year Ahead Buy-Hold Returns
$\text{Digital}_{i,t}$	0.002***	0.029***	0.058***	0.080***
	(0.000)	(0.007)	(0.014)	(0.022)
$\text{Log(Market Cap.)}_{i,t}$	-0.000	-0.004*	-0.006	-0.011*
	(0.000)	(0.002)	(0.004)	(0.006)
$\text{Book-to-Market}_{i,t}$	0.002**	0.026*	0.044**	0.076***
	(0.001)	(0.014)	(0.019)	(0.028)
$\text{Operating Profit}_{i,t}$	0.003**	0.042**	0.074**	0.151***
	(0.001)	(0.019)	(0.031)	(0.043)
$\text{Investment}_{i,t}$	-0.000	0.019	0.011	-0.086
	(0.001)	(0.019)	(0.032)	(0.056)
$\text{Momentum}_{i,t}$	-0.001	-0.023	0.018	0.077*
	(0.001)	(0.014)	(0.020)	(0.040)
$\text{SG\&A}_{i,t}$	0.003	0.044	0.078	0.191**
	(0.002)	(0.027)	(0.051)	(0.087)
$\text{R\&D}_{i,t}$	0.007	0.029	0.495***	1.212***
	(0.007)	(0.094)	(0.179)	(0.468)
$\text{Missing R\&D}_{i,t}$	-0.001**	-0.022***	-0.038**	-0.022
	(0.001)	(0.008)	(0.016)	(0.029)
Time FE	Yes	Yes	Yes	Yes
Observations	210,759	17,433	15,816	14,065
Adj. R^2	0.1923	0.2009	0.1092	0.0565

Table 2.7: Portfolio Returns (Continued)

Panel C: Calendar Portfolio Regressions			
	Long-Short	Long	Short
α	0.511***	0.331*	-0.180**
	(0.177)	(0.171)	(0.072)
MKT - Rf	-0.042	0.934***	0.976***
	(0.047)	(0.041)	(0.019)
SMB	0.007	0.089	0.082**
	(0.086)	(0.087)	(0.034)
HML	-0.098	0.076	0.174***
	(0.070)	(0.072)	(0.034)
RMW	-0.097	-0.105	-0.008
	(0.118)	(0.122)	(0.040)
CMA	-0.094	0.049	0.143**
	(0.116)	(0.123)	(0.068)
MOM	0.115*	0.096*	-0.019
	(0.058)	(0.055)	(0.023)
Observations	120	120	120
Adj. R^2	0.1048	0.8167	0.9734

We report the risk-adjusted returns for portfolios formed on digital disclosure. In Panel A, we report the time-series average size, book-to-market and momentum-adjusted portfolio returns, which are computed using the methodology in Daniel et al. (1997). Each portfolio is formed at the end of June of each year, starting from 2011, and firms in the top tercile of digital disclosures are placed in the long portfolio, while firms with no digital disclosures are placed in the short portfolio. To address liquidity issues in penny stocks, we drop firms with less than $5 in share price. All portfolios are value-weighted, and if a firm delists during the holding period, the proceeds from the delisting returns are reinvested in the CRSP value-weighted portfolio. The benchmark portfolios are also allowed to change each year in June. Time-series standard errors are reported in parentheses. In Panel B, we report panel regressions of monthly, 1 to 3-year ahead returns on size, book-to-market, operating profit, investment, momentum, SG&A, R&D, an indicator for missing R&D as well as a "Digital" long-short strategy indicator that is coded 1 for top tercile digital disclosure, -1 for no digital disclosure and 0 otherwise. We control for time fixed effects and cluster at the firm-level to address serial correlation due to overlapping returns. *, **, *** denote 10%, 5% and 1% significance level, respectively. In Panel C, we report the α from regressing monthly portfolio returns from July 2011 to June 2020 on 6 risk factors: market (MKT-RF), size (SMB), value (HML), profitability (RMW), investment (CMA) and momentum (MOM). The monthly returns for the risk factors are taken from Ken French's website. The portfolios formed on digital disclosures are rebalanced monthly and are value-weighted. Portfolio allocations based on digital disclosure (top tercile digital disclosure are allocated to the long portfolio, while firms with no digital disclosures are allocated to the short portfolio) are revised at the end of June in each year. To address liquidity issues in penny stocks, we drop firms with less than $5 in share price before the portfolio formation date. Robust standard errors are reported in parentheses. *, **, *** denote 10%, 5% and 1% significance level, respectively.

where $R_{p,t}$ is the monthly long-short portfolio return in excess of the risk-free rate. The allocations to the long and the short side of the portfolio are based on the previously described portfolio construction methodology and are revised in June of each year. The monthly portfolio return is estimated by value-weighting the firm-level raw returns, and the weights are re-balanced monthly. $MKT_{i,t}$ is the monthly market return in excess of the risk-free rate. $SMB_{i,t}$, $HML_{i,t}$, $RMW_{i,t}$ and $CMA_{i,t}$ are the size, value, profitability and investment risk-factor-mimicking monthly portfolio returns Fama and French (2015). The coefficient of interest is α_p, the excess return on the portfolio, after controlling for exposure to the five risk factors in the regression model.

Panel C of Table 2.7 reports our calendar portfolio regression results for the long-short, long-side and short-side portfolio returns. In the first column, we report the results for the long-short portfolios, and our results indicate that the portfolio returns a 51-basis-point alpha[45], which on an annualized basis, is roughly 6%. We examine the long- and short-side portfolios in column 2 and find that these portfolios return a positive 33-basis-point alpha and a negative 18-basis-point alpha.

Taken together, our results suggest that markets are slowly incorporating the value implications of the disclosure of digital activity. In particular, we find that trading strategies formed on the digital disclosure tend to perform well and can deliver significant risk-adjusted returns.

Finally, we investigate the role that disclosure plays in driving the return patterns that we observe in Table 2.7. One conjecture for why we find returns over a long-horizon, is that the returns are due to markets are slowly incorporating the value implications of digital activities through continual disclosure of these activities. Our results in Table 2.8, suggest that much of the long-run returns are due to continuous disclosure. In this panel, we split the digital disclosure variable into non-continuous and continuous high disclosure (i.e. continuous top tercile disclosure over the return window)[46] of

45. The monthly returns on an unadjusted basis is 49 basis points and is also statistically significant.

46. To be clear, we code the continuous and non-continuous digital disclosure in the following way. For continuous disclosures, we recode the $Digital_{i,t}$ variable in equation 6 as 0 for firms that do not make top-tercile disclosure continuously in the return window, and vice versa for non-continuous disclosures.

digital terms, for the 2-year and 3-year buy-and-hold returns. Our results show that non-continuous disclosers of high digital activity (those that do not disclose continuously in the top tercile over the return window), are not associated with abnormal returns. And we also show that firms that continuously disclose high levels of digital activity, exhibit higher magnitudes of positive returns, relative to the full sample results in Table 8 (8.7% compared to 5.8%, and 10.3% compared to 8.0% respectively). Thus, our analysis suggests that markets are slowly incorporating value in digital activities through continual disclosure of these activities.

Table 2.8: Return Patterns of Continuous and Non-Continuous Digital Activity Disclosure

Panel A: Sample Breakdown of One-Time and Continuous High Digital Activity				
Return Window	2-Year		3-Year	
Sample	Non-Continuous	Continuous	Non-Continuous	Continuous
	246	508	339	289

Panel B: Return Patterns of One-Time and Continuous High Digital Activity		
Dep. Var.	2-Year Ahead Buy-Hold Returns	3-Year Ahead Buy-Hold Returns
Non-Continuous Digital$_{i,t}$	-0.028	-0.011
	(0.029)	(0.042)
Continuous Digital$_{i,t}$	0.087***	0.103**
	(0.027)	(0.044)
Controls	Yes	Yes
Time FE	Yes	Yes
Observations	15,816	14,065
Adj. R^2	0.1093	0.0565

We report the return patterns following continuous and non-continuous digital activity disclosure in this table. To measure continuous digital disclosure, we code the "Digital" long-short strategy indicator as 1 if the firm discloses in the top tercile of digital disclosure at the beginning *and* in the subsequent years in the return window. We then code the variable as -1 for firms that do not disclose digital activities at the beginning of the return window, and 0 otherwise. To measure non-continuous digital disclosure, we code the "Digital" long-short strategy indicator as 1 if the firm discloses in the top tercile of digital disclosure at the beginning of the return window *and does not* disclose continuously in the top tercile in the subsequent part of the return window. We then code the variable as -1 for firms that do not disclose digital activities at the beginning of the return window, and 0 otherwise. In Panel A, we report the number of observations that are classified as non-continuous and continuous in the 2-year and 3-year return windows. In Panel B, we report the 2 and 3-year returns to firms that do not continuously disclose high amounts of digital terms in the return window versus firms that disclose high amounts of digital terms continuously. The regression design of both panels follows Panel B of Table 7, with the same set of controls for risk factors. *, **, *** denote 10%, 5% and 1% significance level, respectively.

2.4.2 Digital Activity and Fundamental Performance

In this subsection, we report the changes to fundamental performance due to digital activity. The previous sections have revealed a link between digital activity and higher valuations and returns. We investigate whether the increased valuations are validated by better fundamental performance.

The framework of our tests in this subsection is similar to the design of our market-to-book tests (equation 4). We regress performance ratios on the digital activity proxy and a set of controls ($\sum_s \gamma_s X_{s,i,t}$): size, age, leverage ratio, ROA, past three-year sales growth, annual market-adjusted returns, return volatility, SG&A, R&D, missing R&D indicator and the number of words in the business description section of the 10-K, as well as industry and time fixed effects.

Additionally, we examine future changes in performance ratios by implementing a changes analysis for the sub-sample of firms that initiate digital disclosure, with the following design:

$$VAR_{i,t+s} - VAR_{i,t-1} = \alpha_j + \alpha_t + \beta_1 First_{i,t} + \beta_2 Var_{j,t} + \beta_3 Var_{j-i,t} + \sum_s \gamma_s X_{s,i,t} + \varepsilon_{i,t} \quad (2.8)$$

where, $VAR_{i,t+s} - VAR_{i,t-1}$ is the changes in the performance ratio relative to the year before the initial disclosure, and $First_{i,t}$ is the tercile score in the initial year of disclosure. Subsequent observations of digital disclosures are dropped, to ensure that the control group consists of only non-disclosers. Moreover, to ensure that firms are continually disclosing in the changes interval, we drop firms that do not continuously disclose over that interval. We also control for the industry median ($Var_{j,t}$) and the industry-adjusted level ($Var_{j-i,t}$) of $Var_{i,t}$ to address mean-reversion in the accounting ratios (Dechow et al. 2001; Soliman 2005; Healy et al. 2014). Finally, controls in this analysis follows the set of controls in previous analyses and standard errors are clustered at the firm-level.

We first examine the various accounting metrics of productivity in Table 2.9. In Panels A and B, we study ROA and asset turnover. The analysis in the first column shows that firms with digital activities yield an additional 0.4-1.2% of ROA[47] and 2-6% of asset turnover compared to peers (or 13-40% and 2.5-7.5% relative to the sample average of 3% and 78%). In the following columns, we study the 1-3 year changes in ROA after initial digital disclosure, and show that ROA increases by 1.1-3.3% over the three year period after initial disclosure (or 36-110% on a relative basis).

Table 2.9: Analysis of Accounting Metrics of Productivity

Treatment Sample	Full Sample	Initial Digital Activity		
	Levels	One Year Ahead Change	Two Year Ahead Change	Three Year Ahead Change
Panel A: Return-on-Assets				
Dependent Variable	ROA_t	ROA_{t+1} - ROA_{t-1}	ROA_{t+2} - ROA_{t-1}	ROA_{t+3} - ROA_{t-1}
$Digital_{i,t}$	0.004*	-0.001	0.003	0.011**
	(0.002)	(0.003)	(0.005)	(0.005)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	20,779	12,368	9,981	8,082
Adj. R^2	0.6152	0.0508	0.0752	0.0851
Panel B: Asset Turnover				
Dependent Variable	ATO_t	ATO_{t+1} - ATO_{t-1}	ATO_{t+2} - ATO_{t-1}	ATO_{t+3} - ATO_{t-1}
$Digital_{i,t}$	0.020*	-0.005	-0.011	0.020
	(0.010)	(0.008)	(0.012)	(0.015)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	20,779	12,323	9,941	8,050
Adj. R^2	0.6355	0.0716	0.1028	0.1507
Panel C: Return on Net Operating Assets				
Dependent Variable	$RNOA_t$	$RNOA_{t+1}$ - $RNOA_{t-1}$	$RNOA_{t+2}$ - $RNOA_{t-1}$	$RNOA_{t+3}$ - $RNOA_{t-1}$
$Digital_{i,t}$	0.016*	0.009	0.011	0.028
	(0.010)	(0.010)	(0.019)	(0.025)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,822	6,982	5,433	4,325
Adj. R^2	0.4855	0.0304	0.0685	0.0882

47. Consistent with managerial expertise playing a key role in digital adoption, in Table A.6 in Appendix A.5.5, we also show that firms with digital activity and tech-saavy managers yield higher ROA to peers with digital activity.

Table 2.9: Analysis of Accounting Metrics of Productivity (Continued)

Panel D: Net Operating Asset Turnover				
Dependent Variable	$NOATO_t$	$NOATO_{t+1}$ - $NOATO_{t-1}$	$NOATO_{t+2}$ - $NOATO_{t-1}$	$NOATO_{t+3}$ - $NOATO_{t-1}$
$Digital_{i,t}$	0.013	0.005	0.100	0.150*
	(0.043)	(0.040)	(0.068)	(0.081)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	12,849	6,984	5,434	4,330
Adj. R^2	0.3651	0.0698	0.1432	0.2194

We report the coefficients of regressions of return-on-assets (ROA), asset turnover (ATO), return-on-net operating assets (RNOA) and net operating asset turnover (NOATO) on the proxy for digital activities and controls in this table for the sample of non-IT firms in fiscal years 2010-2020. For each accounting performance measure, we report the level differences of firms with digital activity versus peers, as well as the one-, two- and three-year-ahead change in the performance measure for firms that are initiating digital activities (relative to firms with no digital activity), in columns 1-4, respectively. Panel A reports the results for return-on-assets. Panel B reports the results for asset turnover. Panel C reports the results for return-on-net operating assets. Panel D reports the results for net operating asset turnover. In all regression models, we proxy for digital by using a quantized score of the number of digital mentions in the business description of the 10-K (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). All regressions control for size, firm age, leverage, past 3-year sales growth, SG&A, R&D, an indicator for missing R&D, market-adjusted annual returns, return volatility and industry (Fama-French 48-industry) and year fixed effects. Additionally in the changes specification, we control for the industry median and the industry median-adjustment of the dependent variable. Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

Next, we first examine return-on-net operating assets (RNOA) and net operating asset turnover (NOATO) (following, Soliman 2008; Nissim and Penman 2001) in Panels C and D. We find that RNOA is 1.6-4.8% higher in firms with digital activities compared to peers (or 40-120% relative to the sample average of 4%). While NOATO in firms with digital activities is not statistically different compared to peers, we find that NOATO increases by 15-45% (or 7-20% relative to the sample average of 2.18) over the 3 years after initial digital disclosure. Thus, consistent with prior work (Tambe 2014)[48], the weight of the evidence suggest that digital activity improves productivity.

In Table 2.10, we continue our investigation into the accounting performance consequences of digital activities by examining accounting metrics of market competition, namely, profit margins and sales growth. For profit margins, which is reported in Panel A, we find no significant differences be-

48. Notably, we also find that the 3-year changes in RNOA and asset turnover conditional on initial digital disclosure are on average positive. Statistical power could explain why we find weak robustness in our results, as our sample is only 11 years. Future research could explore whether the productivity findings are more robust with a longer sample.

tween firms with digital activity and peers, and we find no significant changes in margins for firms that are initiating digital activity. Surprisingly, we find that firms with digital activities are associated with declines in sales growth. In Panel B, we find that firms with digital activities exhibit 0.8-2.4% lower sales growth relative to industry peers (or 13-40% relative to the sample average of 6%). However, the declines in sales growth do not persist over time as we find that firms initiating digital activity do not exhibit statistically significant changes in sales growth.

Table 2.10: Analysis of Accounting Metrics of Market Competition

Treatment Sample	Full Sample	Initial Digital Activity		
	Levels	One Year Ahead Change	Two Year Ahead Change	Three Year Ahead Change
Panel A: Profit Margins				
Dependent Variable	MARGINS_t	MARGINS_{t+1} - MARGINS_{t-1}	MARGINS_{t+2} - MARGINS_{t-1}	MARGINS_{t+3} - MARGINS_{t-1}
$\text{Digital}_{i,t}$	0.033	-0.031	-0.069	-0.176
	(0.031)	(0.038)	(0.063)	(0.117)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	20,854	12,903	10,565	8,518
Adj. R^2	0.3608	0.0242	0.0590	0.0828
Panel B: Sales Growth				
Dependent Variable	SALES $\text{GROWTH}_{t,t-1}$	SALES $\text{GROWTH}_{t+1,t-1}$	SALES $\text{GROWTH}_{t+2,t-1}$	SALES $\text{GROWTH}_{t+3,t-1}$
$\text{Digital}_{i,t}$	-0.008***	-0.002	-0.002	0.003
	(0.003)	(0.004)	(0.010)	(0.020)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	20,544	15,145	11,873	9,255
Adj. R^2	0.0861	0.5889	0.4412	0.3567

We report the coefficients of regressions of profit margins (MARGINS) and sales growth (SALES GROWTH) on the proxy for digital activities and controls in this table for the sample of non-IT firms in fiscal years 2010-2020. For each accounting performance measure, we report the level differences of firms with digital activity versus peers, as well as the one-, two- and three-year-ahead change in the performance measure for firms that are initiating digital activities (relative to firms with no digital activity), in columns 1-4, respectively. Panel A reports the results for profit margins. Panel B reports the results for sales growth. In all regression models, we proxy for digital by using a quantized score of the number of digital mentions in the business description of the 10-K (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). All regressions control for size, firm age, leverage, past 3-year sales growth, SG&A, R&D, an indicator for missing R&D, market-adjusted annual returns, return volatility and industry (Fama-French 48-industry) and year fixed effects. Additionally in the changes specification, we control for the industry median and the industry median-adjustment of the dependent variable. Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

2.4.3 Discussion

Reconciling the Valuation and Fundamental Performance Results

Previously, we reported weak evidence that digital activity improves fundamental performance. Although our findings suggest that productivity improves for firms with digital activity, we find no significant changes in profit margins and decreases in sales growth for these firms. These results are puzzling given our earlier findings on a positive association between digital activity and valuations. We offer several explanations to help reconcile this apparent puzzle.

First, the increases in valuation could be driven by increases in the market expectation of growth opportunities and that are not reflected in immediate changes in performance. Although these growth opportunities should eventually be realized in improvements to future performance, it is unclear when these improvements would occur. Notably, prior evidence on the payoffs to IT investment, show that the payoffs to these take up to 7 years to fully realize (Brynjolfsson and Hitt 2003). Thus, our limited results on the changes in future financial ratios possibly reflect the fact that investment in digital technologies also take a long time to bear fruit.

Second, some of the gains from digital activities could be eroded by market competition. For profit margins and sales growth, there may be little improvement in these performance measures if competitors are also making similar investments in digital technologies[49]. Moreover, under the market competition story, one should still observe gains in productivity-based metrics because productivity is unlikely to be affected by market pressures on price. And indeed, we find some consistent associations between digital activity and return-on-assets/asset turnover, like prior work (Tambe 2014). Under this interpretation, the pricing effects of digital activities, would therefore be driven by the potential future productivity gains rather than the margin gains of digital technologies.

49. In support of this conjecture, we also find that gross margins (defined as revenues minus cost of goods sold, scaled by sales) is lower in firms with digital activity compared to peers (see Table **??** in Appendix A.5.5), which is suggestive of competitive price pressures that are eroding margins.

Third, the performance benefits of digital technologies could well be limited, but the pricing effects reflect a demand from investors for firms that are involved with digital technologies. This interpretation is motivated by anecdotes from the dot-com era, when firms that were marginally linked to the internet received higher valuation, despite limited performance gains. For instance, Cooper et al. (2001) document that merely changing a firm's name to one with a "dot-com" is sufficient to drive large and persistent valuation changes, without commensurate changes in the firm's fundamentals. Given the weak accounting performance results that we document, we acknowledge that such a phenomenon could be a reason that reconciles our accounting performance and valuation results.

Potential Selection Bias

A concern in interpreting our results is that they may be driven by selection bias. One such concern is that our results may be driven by better performing firms that also adopt digital technologies. The higher valuations of firms with digital activities would thus be an artifact of the higher valuations of better performing firms. We argue that this form of selection bias is unlikely, as our determinant results (Table 4) show that lagged performance is generally unrelated to digital disclosure. Moreover, we also address this form of selection bias in Table 5 by showing robustness of our main results in the lagged market-to-book specification, which compares firms with digital activities, with similar market-to-book peers. Finally, we also address this concern by implementing an IV analysis, and we find that our results are also robust to this specification.

Second, another concern related to selection bias is that our results may be driven by the selective disclosure. We argue that this is unlikely to be a contributing factor to our results because we document mixed evidence on the accounting performance benefits for digital activities and a long-run market response to the value implications of these activities. Both of these results suggest that the success of digital activities is difficult to assess ex ante, and thus casts doubt on whether firms are able to selectively disclose only successful digital activities. Moreover, we also argue that the proprietary cost channel for selective disclosure is also an unlikely explanation for our findings. Specifically, for the proprietary cost argument to hold, firms that withhold information should generate long-run value as they possess proprietary technologies. In contrast, our results show that firms that do not disclose digital activities, earn significant long-run *negative* returns.

Finally, another concern is that the non-IT firms with digital activities in our study could be misclassified IT firms. We address this concern in two ways. First, we drop firms that exhibit a median β_{Tech} that is in the top 2.5 percentile of the median β_{Tech} for all non-IT firms. With this sample, we find that our main inferences are unchanged. Second, we drop firms that made digital disclosures in the years before 2010, which is the year when we observe an upswing in non-IT firms starting to go digital. We find that our main inferences are unchanged with this sample[50].

2.5 Conclusion

In recent years, a growing number of non-IT firms have made investments in the new wave of digital technologies that have the potential to transform businesses and create greater firm value (Brynjolfsson et al. 2017). Motivated by this growing phenomenon, our objective in this study is to characterize the firms that adopt these technologies and to evaluate the valuation and performance benefits of adopting these technologies.

To that end, we develop a textual-based measure of digital activity to create a large sample of firms that are going digital. We show that this measure captures the growing trend of going digital amongst non-IT firms. We find that these non-IT firms that go digital tend to be firms that are large and young, invest more in SG&A and are in industries with higher digital activity.

We find that going digital improves valuations as the market-to-book of firms that engage in digital activities is 8-26% higher than their industry peers. While part of these level differences is attributable

50. We report these analyses in Tables A.7 and A.8 in Appendix A.5.6.

to capitalization restrictions on digital activity expenditures, we still observe economically and statistically significant level differences after controlling for the capitalization restrictions effect. Furthermore, we also find that digital activities increase the valuations of earnings and sales, which further corroborates the claim that digital activities increase valuations. Moreover, portfolios formed on digital disclosure significantly predict returns and deliver a 51-basis point alpha in a Fama-French 5 factor plus momentum model[51].

However, we find weak results when examining the implication of digital activities on accounting performance measures. ROA and asset turnover improve suggesting that digital activities offer gains in firm productivity. While profit margins and sales growth are either insignificantly, or negatively associated with digital activities. We reconcile the weak accounting performance gains of digital activities with our valuation findings as possibly due to (1) the long payoffs in digital investment, (2) competitive pressures and (3) market euphoria on the digital phenomena (Cooper et al. 2001). Understanding which of these factors explains our results is an intriguing question, and we leave a detailed study of this for future research.

51. However, we caution the reader that real-time profits are likely to be lower than the returns reported in this study, as trading frictions could impose additional costs for the investors. Moreover, the adoption of digital technologies by later adopters could receive different valuations compared to the early adopters that we study.

3

Financial Analyst Interest in Digitization and AI Investment in Non-IT Firms

Over the past decade, it has been widely reported that many non-IT firms have been engaging in digitization[1] (or digital transformation) by deploying digital technologies, such as big data, analytics and

1. Digitization is defined as the process of adopting digital technologies, a complementary set of technologies comprising of big data, analytics, artificial intelligence, and machine learning (Goldfarb et al. 2019).

artificial intelligence (AI), in their business operations (Babina et al. 2020; Acemoglu et al. 2020). At the same time, key market intermediaries, such as financial analysts are also starting to take notice of this phenomena (see Figure 3.1) and have been opining on non-IT firms' digitization activities. Motivated by this recent development, I study the drivers of analysts' interest in non-IT firms' digitization activities and the relationship between their interest in these activities and future investment in advanced digital technologies – artificial intelligence (AI).

One reason for studying analysts' interest in the digitization of non-IT firms is the uncertainty and tension in the trade-off between the potential benefits and costs in digitization. While prior academic work has illustrated how these technologies could substantively benefit (Brynjolfsson et al. 2017) non-IT firms[2], the realized net benefits of digitization varies quite widely[3]. Complementary co-invention costs (Bresnahan and Greenstein 1996), for instance, are idiosyncratic and significant frictions for non-IT firms, owing to their legacy structures and processes (Lakhani and Iansiti 2020). These frictions are therefore likely to drive high tension and uncertainty in the net benefits of digitization. This in turn, leaves open questions on how key market institutions, such as financial analysts, are assessing this process.

Notably, the uncertainty from digital investments is likely quite different compared to those from other types of investments studied in the accounting literature. AI and other digital investments are typically expensed immediately, with payoffs that are likely to occur only after a span of several years[4]. Thus, the high and immediate costs of these investments are more likely to concern key market players, compared to general capital expenditures.

2. The amount of capital spent on digitization perhaps further illustrates the importance of this process to firms. Statistics from the World Economic Forum estimate that corporations have spent a total of 2.5 T. USD on digital investments from 2016-2020. See: http://reports.weforum.org/digital-transformation/files/2018/05/201805-DTI-Maximizing-the-Return-on-Digital-Investments.pdf

3. In fact, practitioner surveys on digitization suggest that most firms fail at this process: https://www.mckinsey.com/business-functions/organization/our-insights/unlocking-success-in-digital-transformations

4. For instance, prior work in the ICT literature, shows that for ICT investments, the full benefits of the technology only emerged in the 5-7 years after the investment (Brynjolfsson and Hitt 1996).

Moreover, compared to uncertainty in intangible investments that are immediately expensed such as R&D (Lev and Sougiannis 1996), uncertainty in digital investment is also quite distinct. Specifically, investment uncertainty in R&D stems from developing technologies, whereas for digital technologies, the uncertainty arises from *applying* new technologies[5]. This distinction has implications for how market participants evaluate these investments – For R&D, market participants can use past experience and expertise to help assess the uncertainty in the investment, especially in the recent period, where R&D is used to develop mostly incremental and safe innovations (Curtis et al. 2020). However, for digital technologies, analysts and other market institutions, may be less able to draw from past experiences and expertise to assess these technologies. As digital technologies are a form of general purpose technology (Brynjolfsson et al. 2017; Goldfarb et al. 2019), these technologies have unique firm-level applications (Bresnahan 2010) and co-invention costs, which limits learning from prior examples of investment. Consequently, for market participants, investment uncertainty in these technologies is quite different and perhaps more challenging to assess. In addition, digital and AI technologies are often viewed as productivity-enhancing technologies, that can improve productivity in innovation (Cockburn et al. 2017) and business operations. Thus, these investments also have a countervailing, decreasing effect on R&D and other intangible expenditures[6]. This potential for reducing expenditures also distinguishes digital investments from product-based R&D investments and could therefore also lead to differences in how market participants are assessing these investments.

5. Another potential source of distinction lies in the market uncertainty of the new products & services that are developed through R&D and digital investments. While new products & services from both forms of investment are likely to drive some degree of market uncertainty, there is some reason to believe that the uncertainty is higher for digital investments. As suggested in Curtis et al. (2020), payoffs of more recent R&D investments tend to be less risky, which should lead to greater market certainty over new products & services. In contrast, the challenges in getting consumers to use early versions of digital technologies, such as e-commerce (Brynjolfsson and Smith 2000), suggest that the success of new digital products & services could be quite uncertain from a market perspective.

6. Similarly, for other intangible investments such as SG&A, these investments are also likely to have an ambiguous effect. On the one hand, investments in these technologies should increase SG&A costs through human capital investment in these technologies. But on the other hand, greater automation could also drive down these expenditures as automation replaces routine labor (Autor et al. 2003).

Another reason for studying analysts' interest in the digitization, is that an investigation of its relationship with future AI or digital investments could also contribute to the broader debate on how analysts' views influences investment decisions. Some scholars have raised concerns that analysts and other capital market institutions could increase short-term pressures on managers, leading firms to decrease investment in long-term investment projects (Graham et al. 2005; He and Tian 2013). Others contend that analysts and other capital market institutions play a positive governance role in shaping investment decisions (Derrien and Kecskés 2013). Still others argue that analysts could drive over-investment in long-term investment projects (Bebchuk and Stole 1993), particularly for long-term investments with limited capitalization such as digital investments. The implications of this debate is especially salient given the importance of digital technologies has rapidly increased for non-IT firms in recent years (Lakhani and Iansiti 2020; Bai et al. 2020).

Motivated by the above discussion, this paper studies financial analysts' interest in the digitization of non-IT firms, by examining two research questions. First, I examine the factors that drive analysts' interest in digitization. To measure analysts' interest, the empirical analysis uses digital questions from quarterly earnings conference calls[7], a disclosure medium that covers a wide range of analyst views[8]. Second, I examine the relationship between analysts' interest in digitization and future investment in digital technologies in non-IT firms, by studying the associations between digital questions and non-IT firms' future investments in advanced digital technologies (AI). To measure AI investments, the

7. These questions are identified using a digital dictionary, which is developed using words relating to digitalization (following the word list in Section A.1 and an expanded dictionary of digital-related technologies, generated by cosine similarities in word-embeddings from a trained Word2Vec neural network model (following the approach in K. Li et al. 2021; Zhang 2021; Mikolov et al. 2013). For more details, see Appendix B.2.1.

8. While other studies have used textual data in the *entire* conference call transcripts as a proxy for a firms' exposure to specific technologies (Bloom et al. 2021), its perhaps worth emphasizing that this study takes a different approach to distinguish between two key sources of text/views within the conference calls. (1) The analysts' questions on digitization is used as a proxy for their views and interest in these technologies. (2) The managers' disclosure of these technologies in the presentation portion of the conference call, as a proxy for the firms' exposure or intention to invest in these technologies.

empirical analysis uses AI job postings[9], following recent studies that measure AI adoption through job postings data (Alekseeva et al. 2021; Acemoglu et al. 2020; Babina et al. 2020; Law and Shen 2020).

I study the first research question, by examining the drivers of analysts' interest in digitization. The literature on the drivers of corporate investment and technology adoption suggests that five factors should play an important role — (1) the suitability of digital/AI technologies (Brynjolfsson et al. 2018; Felten et al. 2021), (2) financing and economic risks (Whited and Wu 2006; Hayashi 1982), (3) execution expertise (Chemmanur et al. 2019; Matta et al. 2019), (4) product market competition (Arrow 1962; Aghion et al. 2005) and (5) organizational frictions (Henderson and Clark 1990; Bresnahan and Greenstein 1996; Bresnahan et al. 2012)[10]. Moreover, a literature in accounting and finance highlights the importance of managerial disclosures (Bowen et al. 2002) and peer analyst activity (Welch 2000; Merkley et al. 2017) in shaping individual analysts' decisions. My empirical analysis therefore studies whether digital questions from analysts are associated with these factors.

Consistent with expectations, the empirical evidence show that analysts' interest in digitization is positively associated with (1) AI's suitability across industries, (2) the governance attributes as measured by management team/board's experience with digital technologies, and (3) managerial and industry-wide analyst discussion of digital topics. In a industry and quarter-fixed effect regression, the estimates show that analysts pose more questions when the firm is in an industry that has a high AI suitability score, and when the firm has managers or directors with experience in digital technologies[11]. Ques-

9. These postings are identified by the set of AI-related skills in Alekseeva et al. (2021).

10. As a validation test of the theorized investment and technology adoption factors in the AI investment setting, these factors are shown to drive AI job posting intensity. The suitability of AI technologies, as measured by the textual overlap between AI patents and NAICS industry descriptions correlates positively with the AI job postings. Consistent with financing, economic risks and execution ability playing a role in driving investment decisions, firms with lower capital constraints, higher Tobin's Q and with board experience in digital technologies tend to invest more in AI. Furthermore, consistent with product market competition and organizational factors shaping AI investment decisions, industry-level AI job postings, firm age and the lifecycle stage of the firm are associated with firms' AI investment decisions.

11. These managers and directors are identified using digital-related search terms in the biographies reported in the *Capital IQ Professionals* database. See Appendix **??** for more details on the search terms.

tions are also higher when the firm's management team discusses digital topics in the presentation portion of the conference call, and when other analysts of peer firms pose digital questions in conference calls. Moreover, the sentiment of analysts' interest in digitization[12] is also associated with various factors of digital investments, as the sentiment of digital questions tend to be positive when firms are facing (1) lower financing and economic risks and possess (2) execution expertise (as measured by board experience in digital technologies).

Somewhat surprisingly, there is some evidence that product market competition, as measured by the adoption of AI at the industry-level is negatively associated with both the sentiment and proportion of digital questions. One interpretation of this result is that analysts are pressing firms to invest more in AI when there is an opportunity to be a first-mover in the adoption of AI technologies. Thus, this result opens the possibility that the levels and types of digital questions could vary depending on the extent of AI investment.

To further unpack how the levels and types of digital questions varies according to the level of AI investment, the dynamics of digital disclosures (i.e. the managerial disclosure and analysts' questions on the digital topics) are examined across sub-samples of observations where firms investments in AI are abnormally low, high and normal[13]. Analyst questions and management discussions on digital topics are more pronounced when firms' investments are at abnormally low or high levels. Topic analysis across the sub-samples further show that the types of managerial disclosure and analysts' questions differ depending on the levels of AI investments. Managers tend to be more positive when firms invest in AI at abnormally high-levels. In contrast, analysts tend to be more negative in these situa-

12. Measured by applying the FinBert model (A. Huang et al. 2020), a deep-learning BERT model for measuring sentiment in a financial context, on the digital questions, and aggregated at the firm-quarter level.

13. These samples are created by first bench-marking the AI job postings relative to the expected level of postings as predicted by the five aforementioned factors of AI investment. Observations are then defined as normal investment if the residual from a quarter-fixed effect regression of the five factors is within 1 σ from the mean in each quarter, and as abnormally low (high) investment if the residual is below (above) 0.5 σ from the mean. Additional robustness analysis shows that the main sub-sample results are robust to alternative specifications — (1) regressions based on panel OLS or based on industry-quarter FE model, (2) and bandwidth cut-offs defined at 0.5 or 1.5 σs from the mean.

tions and focus on the firms' competitive position (measured following F. Li et al. 2013). Thus, these analyses suggest that analysts apply monitoring pressure on highly-optimistic managers by taking a more negative view on their digitization activities when there are signs of abnormally high investment in these technologies.

Taken together, the analysis of the drivers of analysts' interest in digitization, suggest that analysts play a role in analyzing and monitoring these types of firm investment activities. In particular, analysts take into account the suitability of the technology and whether the firm has the governance attributes to engage in digitization, when forming their views on the process. Moreover, analysts appear to monitor firms that are highly optimistic and are investing in AI at abnormally high-levels.

In the second part of the study, I examine whether analyst interest is associated with increases or decreases in digitization activities through future investment in AI technologies. Notably, prior work suggest that the direction of the relationship between analysts and investment is ex ante unclear, as some argue that analysts can drive lower investment (Graham et al. 2005; He and Tian 2013; Bebchuk and Stole 1993), while others contend that analysts can drive higher investment (Derrien and Kecskés 2013). Thus, this debate motivates an analysis of the relationship between analysts and AI investment in this setting.

The main empirical test is implemented through a firm and quarter-fixed effect regression that regresses future investment in AI (as measured by AI job postings[14]) on the proportion of digital questions from analysts and controls[15]. The estimates of this regression shows that firms with more digital questions invest more in AI. Specifically, a 1 σ increase in the proportion of digital questions increases the 1-year ahead probability and intensity of AI investment by roughly 1.1% and 0.05% respectively (or 5% and 10% relative to the average probability and intensity of AI postings respectively). More-

14. To be clear, AI job postings is measured in two ways — (1) an indicator for whether the firm posts an AI posting, and (2) a measure of AI job posting intensity (defined as AI job postings scaled by total job postings, following, Acemoglu et al. 2020; Babina et al. 2020)

15. Controls are the six determinant factors that were examined in the previous analysis.

over, digital questions on performance, competition, specific technologies and questions that exhibit positive sentiment, are also positively associated with future investment in AI.

To further probe whether analysts are playing an *active* role in encouraging companies to increase future AI investments, several additional cross-sectional tests are studied. The first test examines whether analyst interest is associated with greater AI investment when the firm is not engaging in any current AI investment. The results of this test suggest that analysts plays a role in encouraging firms to engage in new AI investment, as the positive association between analyst interest in digitization and future AI investment is significant only for firms that are not investing in AI. The second test studies whether analyst interest is associated with greater AI investment when there is no firm-level disclosure on digital topics[16]. Consistent with analysts playing an active role in facilitating more AI investment, the results of this test show that the positive relationship between analysts interest in digitization and future AI investment is higher when there is no firm disclosure on digital topics. Thus, the overall evidence suggest that analyst interest plays an active role in encouraging firms to increase future AI investments[17].

Moreover, to examine whether analysts are encouraging firms to invest in AI at more normal levels, the last set of analyses studies sub-samples where AI investment is at abnormally low, high or at normal-levels. These sub-sample analyses show that digital questions are positively associated with more future AI investments when AI investment is abnormally low. The results therefore suggest that

16. To search for digital-related firm disclosure, I examine the managerial disclosure on the presentation portion of the conference call, product announcements in Key Development database in Capital IQ and the business description & MD&A section of the 10-K.

17. Its perhaps worth noting that the positive association between digital questions and future AI investments, is interpreted as analysts *encouraging* firms, rather than pushing firms, to invest more in AI. This is because managers are the primary decision-maker in investment decisions, and so its unlikely that individual analysts are independently pushing managers to invest in AI. On the other hand, analysts can encourage managers to make AI investment decisions in several ways, by (1) providing managers with insights on the activities in peer firms and (2) by serving as a conduit for the broader market and investor community's view on digitization. These two factors are key considerations in a manager's decision-making process, and thus, analysts could encourage more AI investment decisions by providing additional information on these factors.

analysts play a positive governance role by encouraging companies to increase future AI investments when they are currently making too little investment in these technologies.

Finally, I examine the various economic channels that could explain why analysts' questions are associated with more AI investment in non-IT firms. I study two potential channels. First, one interpretation of the positive association between the digital questions and AI investment is that it could reflect firm reactions to market-wide sentiment or pressures on investment in digital technologies. To examine this potential explanation, I implement the main regression model with analyst fixed effects that controls for heterogeneity in analysts' views on digital technologies and isolates the aggregate analysts' views on the technology. With this fixed effect structure, I find that the digital questions are also positively associated with future AI investments, which provides support for the hypothesis that aggregate market views on digital technologies could be driving the positive association between the digital questions and future AI investments. In addition, I find further evidence that markets tend to reward companies that act on market interest in digital technologies, as firms that invest in AI technologies after facing analysts' questions on digitization tend to exhibit significantly positive changes in industry-adjusted Tobin's Q.

Another interpretation of the main results, is that analysts could be providing some expertise on digital technologies that are helping managers invest in newer technologies such as AI. To examine this potential interpretation, I examine whether questions from analysts' that cover technology firms (the technology analysts) are associated with more AI investments. And I find some evidence that digital questions from technology analysts are associated with more AI investment, as measured by the intensity in AI job postings. To further probe the analyst expertise channel, I also examine cross-sections in the firm-quarter panel, where firms could benefit from technology expertise from analysts due to a lack of such expertise in the management team. Consistent with analysts providing some expertise on digital technologies in firms with less managerial expertise, I find that the digital questions are positively associated with AI investment only in cross-sections with no executive expertise in digital technologies. Thus, these set of analyses also provide some evidence that analysts could be providing some expertise to firms in their investment in advanced digital technologies.

While the findings presented in this study are associative, the findings are insightful due to several reasons. First of all, the findings are addressing an unanswered but important question on the relationship between capital market institutions and the digitization of non-IT firms. While the lack of exogenous variation sources in this setting prevents one from implementing a natural experiment or instrumental variable analysis, the research design includes firm, quarter fixed effects and key covariates as controls to address a novel question in an empirically rigorous way. Secondly, the estimated bounds on one of the main endogeneity concerns, the potential effects of within-firm unobserved selection (following the approach in, Oster 2019), suggest that these selection effects may play less of a role in explaining the main association between digital questions and future AI investment. Third, this association is also unlikely due to spurious fluctuations in analyst activity, as a placebo analysis of non-digital questions and future AI investment shows no statistically significant relationship between the two variables.

Overall, this study contributes to the literature in two ways. First, the findings of this study contribute to the accounting and finance literature that studies the role of financial analysts in financial markets. From an ex ante perspective, these findings provide novel evidence on the factors that analysts take into account when responding to the emergence of new technologies, such as digital/AI technologies. Thus, these results complement existing studies on the anatomy of the decision-making process of analysts (Joos et al. 2016; Brown et al. 2015; A. H. Huang et al. 2017; Green et al. 2016), and shows that analysts, on average, make decisions that are consistent with their roles as information analyzers and monitors. From an ex post perspective, the findings of this study contributes to a group of studies that has studied the investment consequences of analyst following (He and Tian 2013; Derrien and Kecskés 2013). In contrast to recent practitioner and academic concerns that short-term pressures from analysts might force firms to reduce investments (Graham et al. 2005; Fuller and

Jensen 2010), the results of the paper show that analysts' interest in digitization is positively associated with more investment in advanced digital technologies (AI).

While causality cannot be established with the research setting of this paper, the findings of this study are nonetheless interesting considering the debate on the relationship between the presence of financial analysts and investment behavior.

Second, this study contributes to the economics and strategy literature on technology adoption in established firms. By studying the role of financial analysts in the adoption of digital technologies in non-IT firms, this paper touches on a novel angle — the capital market factors that are related to the investment in new technologies. This angle is new as the existing literature on incumbent adoption of new technologies has so far focused mainly on the organizational (Henderson and Clark 1990; Bresnahan and Greenstein 1996), customer dynamics (Christensen 1997), and product market factors that shape new technology adoption (Arrow 1962). Thus, this paper potentially contributes to this literature by providing some insight into how capital markets may play a role in technology adoption.

3.1 Literature Review

3.1.1 Digitization in Non-IT Firms — Benefits and Costs

I refer to digitization (or digital transformation), as the process of adopting digital technologies to help improve business processes, increase value for customers and to spur innovation. Digital technologies refers to a broad set of technologies that have been commonly referenced by companies in their digital transformation initiatives. Specifically, these are a complementary set of technologies (Goldfarb et al. 2019), that range from the big data and analytics technologies which have been popular since the early 2010s, to the more advanced artificial intelligence technologies that have become more popular in recent years.

What motivates a study of these technologies is the large potential benefits and costs to these technologies for a broad set of firms. On the benefits, recent scholarship on the topic have argued that these technologies can lead to substantial productivity (Brynjolfsson et al. 2017) and innovation gains across multiple industries (Cockburn et al. 2017), much like how information technologies (IT) did for many firms in the late 90s and early 00s. Moreover, historical trends in the patent protections of software-related patents (Jaffe and Lerner 2004) and the overall movement towards greater automation in the US economy (Autor et al. 2015), has further added to the potential value that digital technologies can provide to firms.

A core reason for why digital technologies create value for such a *wide* range of firms (including the non-IT firms in this study) is that these technologies are potentially general-purpose technologies (GPT) (Bresnahan and Trajtenberg 1995). They create value through two channels — (1) by complementing existing assets, and (2) by developing follow-on innovation from the GPT. Scholars have argued that many digital technologies are GPTs. For instance, there are several studies that have argued that the more advanced digital technologies, artificial intelligence, and machine learning technologies, are GPTs (Brynjolfsson et al. 2017; Cockburn et al. 2017), while other have also argued that several digital technologies that are complements to AI and machine learning, such as analytics and big data, should also be viewed collectively as GPTs as well (Goldfarb et al. 2019).

Empirical evidence on the value of digital technologies generally supports the view that these technologies can create value for firms across many non-IT industries. Tambe et al. (2020) show that a substantial share of firms' market value lies in digital capital. Chapter 2 focuses specifically on non-IT firms and shows that these firms benefit from higher valuations and returns when making disclosures on the adoption of digital technologies. Prior work has also shown that value created by digital technologies is likely to come from two main channels. First, these technologies generate substantive productivity gains by automating business processes (Hitt 1999; Tambe 2014). Second, they create opportunities for firms to develop new product and services (Brynjolfsson and Smith 2000), which helps firms to scale and to compete for more market share (Babina et al. 2020).

Although there is consensus that digital technologies are valuable opportunities for non-IT companies, the process of adopting these technologies is highly costly and risky. Notably, the key bottleneck or cost in digitization tends not to be technological but is most often organizational (Lakhani and Iansiti 2020). Specifically, the adoption of digital technologies requires complementary investment in organizational co-invention costs (Bresnahan and Greenstein 1996), such as redeveloping business processes, new organizational structures, and establishing new relationships between suppliers and customers. For non-IT firms, restructuring the organization to incorporate the new digital technology is particularly challenging — as these firms operate through legacy hierarchical structures and processes that are less suited for these technologies (Lakhani and Iansiti 2020; Cao and Iansiti 2021).

There are several anecdotes of established companies that failed to adopt new technologies, because these companies faced high diseconomies of scope (Bresnahan et al. 2012) and were thus unable to change the organization to meet the demands of the technology. One such example, is the case of Encyclopaedia Britannica, a company, that faced stiff challenges in shifting its print encyclopedias online (Greenstein 2017), and eventually lost to new entrants offering online alternatives to its product. Moreover, as digital technologies are GPTs, with unique applications to each firm (Bresnahan 2010), there is limited potential for firms to learn from each other's experiences. Thus, companies are likely to face significant challenges when adopting new digital technologies.

3.1.2 Market Perspectives On The Digitization of Non-IT Firms

Due to the high potential benefits and costs of digitization, there is much room for market institutions to play a role in this process. In fact, key activist investors, have engaged several different activist campaigns to encourage or deter digitization efforts in non-IT firms. One example is Starboard Value's activist campaign at Darden Restaurants, where Starboard Value actively pushed for more investment in digital marketing to cut advertising costs[18]. Activist investors have also pushed to curtail digitization activities. For instance, Carl Icahn's activist campaign at Blockbuster, was a key force in preventing the firm from transitioning its operations into the online space[19].

Financial analysts, as key information intermediaries, have also been active in opining on the implications of digitization for non-IT firms. For instance, in the retail space, several analysts from Barclays and Cowen have written detailed analyst reports discussing the digital investment of key non-IT retailers such as Walmart and Target (McClintock et al. 2016; Chen, Rakhlenko, Origenes, et al. 2017). The reports delve into how these traditional retailers can complement digital technologies with their existing organization and follows metrics that track the digitization progress of these firms. Moreover, in the banking space, there have also been much analyst interest in the rise of FinTech and how large banks are leveraging these technologies to improve their services (see for example, a recent UBS report surveying the state of digitization across banks Martinez et al. 2019).

These analysts and activist investors examples, suggest that markets take notice of non-IT firms' digitization initiatives. Motivated by this phenomenon, my study provides some insight into the determinants and consequences of market views on these activities, by focusing on the digital questions posed by financial analysts in quarterly conference calls.

However, one important caveat to keep in mind is that the questions and opinions from analysts is not a *full* representation of market views. Nonetheless, there is some appeal in using analysts' questions as a proxy for those views. Firstly, analysts' opinions play an important role in shaping how markets are reacting to value-relevant activity in firms. Within financial markets, the role of financial analysts is to spotlight and to provide research analysis of investment opportunities for investors.

18. See Starboard Value's presentation to investors of Darden Restaurant, where the investor argued for greater investment in digital technologies to cut costs: https://www.sec.gov/Archives/edgar/data/940944/000092189514002031/ex991dfan14a06297125_091114.pdf

19. In a HBR article, Blockbuster's former CEO, John Antioco, suggested that the activist investor, Carl Icahn, delayed the transition attempt, which arguably led to Blockbuster's eventual demise: https://hbr.org/2011/04/how-i-did-it-blockbusters-former-ceo-on-sparring-with-an-activist-shareholder

Consequently, their views can move markets, as the information provided by analysts are often used by investors in investment decisions (see for example, the price movement effects of analyst recommendations, Beneish 1991).

Secondly, analyst questions in conference call represent a broad set of market views. Recent evidence suggests that buy-side analysts (representing institutional investors) also participate in the conference calls (Jung et al. 2018) and so questions from these analysts represent the direct views of certain institutional investors. Moreover, survey evidence suggest that sell-side analysts ask questions on behalf of certain institutional investors during these conference calls as well (Brown et al. 2015). Thus, there is some reason to believe that analysts' questions in conference calls represent a broad set of views from market participants. Thirdly, the dataset of analysts' questions in conference calls covers a wide range of firms, and so focusing on these questions enables a study of a large sample of non-IT firms that are potentially engaging in digitization.

3.1.3 Drivers of Analysts' Interest in the Digitization of Non-IT Firms

To better understand how analysts form their views on the digitization of non-IT firms, my first research question studies the drivers of these views. As a starting point, I discuss several potential drivers that are based on the investment and technology adoption literatures.

The first factor that should drive analyst interest in digitization in non-IT firms is the suitability of the technology in the firms' industry. This hypothesis is motivated by recent studies that have attempted to categorize occupations, industries, geographies on the likelihood of exposure to AI-related technologies (see for example, Felten et al. 2021; Webb 2020; Brynjolfsson et al. 2018). These studies argue that variation across occupation in AI suitability occurs as AI and other software technologies are particularly suited for certain abstract and routine tasks (Autor et al. 2003). Thus, industries with occupations and processes that are more easily automated are more likely to benefit from these technologies, which should in turn, drive higher adoption of these technologies. Consistent with this view, Acemoglu et al. (2020) show that some of the AI suitability indices are correlated with industry-level AI-related job postings. Hence, if analysts are aware of the variation in AI suitability across industries, this variation should lead to the following hypothesis:

H1a: Analysts' interest in digitization is associated with AI's suitability in the firm's industry.

The second factor that should drive analysts' interest in digitization in non-IT firms is the financing and economic risk that the firm faces. In theories of corporate investment, two financing and economic variables have been proposed to drive investment decisions. The first variable is Tobin's Q, based on Q-theory which argues that the marginal value of investment relative to marginal costs (or marginal q) should drive investment decisions (Hayashi 1982; Blundell et al. 1999). The second variable is capital constraints, as finance studies have also argued that capital constraints can be binding and prevent companies from investing effectively (Whited and Wu 2006; Rauh 2006). Thus, given the importance of these variables in explaining investment behavior, I expect the following hypothesis:

H1b: Analysts' interest in digitization is associated with financing and economic risks.

The third factor that should also drive analyst interest in the digitization of non-IT firms, is the execution expertise in firms. One such attribute is managerial skills. There is increasing evidence that managerial skills are important for innovative success (Chemmanur et al. 2019; Custódio et al. 2019; Cho et al. 2016). Recent studies that show that managers with specific technical experience in technologies is important for innovation (Islam and Zein 2020). Thus, firms that have executives with experience in digital technologies should be more successful at driving digital technology adoption. In a similar vein, board members with technology experience should also play a role in guiding firms to engage in successful digitization. Specifically, there is also an emerging literature that studies the skill sets of directors (Adams et al. 2018), and some have shown that director experiences can help firms increase innovative activity (Matta et al. 2019). Thus, the insight from these studies suggest that execution expertise should play an important role in driving digital investment, which should therefore inform analysts' interest and motivates the following hypothesis:

H1c: Analyst interest in digitization is associated with execution expertise.

The fourth potential determinant of analyst interest in digitization is organizational frictions. Prior work in strategy and economics have argued that established firms (which is another characterization of the non-IT firms in this study) have legacy structures and value networks which creates additional frictions in the adoption of new technologies. Specifically, established firms may have existing innovation architectures (Henderson and Clark 1990) or value networks (Christensen 1997) that prevent firms from bringing in new technologies. Moreover, the co-invention costs of implementing new business processes could be large and difficult to navigate (Bresnahan and Greenstein 1996; Bresnahan 2010). Thus overall, these studies suggest that organizational friction should play a role in digital investment decisions.

While the theory is clear on how organizational frictions affect technology adoption, how analysts react to these frictions is somewhat of an open question. Analysts could press firms with legacy organization structures and processes to curtail digitization activities, as these firms are unlikely to succeed in this process. On the flip side, analysts could engage with management to address some of the organizational challenges to facilitate more effective digitization. Nonetheless, both types of reactions predict greater analysts' interest in digitization when there are organizational frictions, which therefore motivates the following hypothesis:

H1d: Analyst' interest in digitization is associated with organizational frictions.

The fifth potential determinant of analyst interest in digitization is product market competition. Insights from the economics and strategy literature suggest that product market competition should play an important role in driving digital technology adoption due to two reasons. First, in more competitive industries, there are greater incentives to invest and adopt new technologies (Arrow 1962). In particular, the empirical evidence suggests that competitive forces between rival firms spurs companies engage in more innovation (Blundell et al. 1999; Aghion et al. 2005). Given the highly competitive nature of digital adoption across industries (Lakhani and Iansiti 2020), this general prediction for innovation behavior should generalize well to the setting of non-IT firms adopting digital-related technologies. Hence, there is reason to expect the following hypothesis:

H1e: Analyst interest in digitization is associated with product market competition.

In addition to the technology adoption-related reasons for analyst interest in digitization, the accounting and finance literature on analysts, suggest that the milieu of disclosure should also play a role in shaping analyst interest as well. For one, managerial disclosures could be a key information source for analysts as studies have shown that analysts forecasts improve after voluntary disclosure from managers (Bowen et al. 2002). In fact, survey evidence suggest that analyst value information from managers (albeit private information) more than they do their own research on publicly obtained sources (Brown et al. 2015). Moreover, evidence on forecast attributes after private access to management was cut for analysts (following Reg FD) showed significant decreases in forecasting accuracy (Agrawal et al. 2006), which further suggests that information from managers is important in shaping analyst behavior. Thus, I expect managerial disclosure to play an important role in shaping analysts' interest in the digitization of non-IT firms.

Another source of disclosure that could shape analysts' interest in digitization is the industry-level analyst discussion of digital topics. Recent studies show that industry-level analyst activity can generate positive spillover effects, that increases individual analyst effort (Merkley et al. 2017). Thus, analysts in industries where there is more active discussion on digital-related topics could be more incentivized to perform research on their firms' digitization efforts and to engage with management on those efforts. On the other hand, herding effects from industry-level activity, could also compel analysts to engage in discussion of digital-related topics. Prior work has shown that analysts engage in herding behavior (Trueman 1994; Welch 2000; Hong et al. 2000) by mimicking forecasts of other analysts. In the digitization setting, analysts could exhibit similar herding behavior by mimicking the digital discussion of peer analysts. Thus, industry-wide analyst activity could drive greater analyst interest in the digitization of non-IT firms. And therefore, I expect the following hypothesis:

H1f: Analyst interest in digitization is positively associated with managerial and industry-wide analyst interest on these activities.

3.1.4 How Do Analysts' Interest in Digitization Shape Investment in AI?

The key tension of this study, is whether analysts' interest in the digitization of non-IT firms encourages or deters investment in advanced digital technologies (AI)? Broadly, the accounting and finance literatures makes three different predictions on the relationship between analysts and investments. Thus, in this subsection, I discuss these predictions in turn.

One view, is that the short-term market pressures from analysts can deter investment in the new technology. Studies in accounting and finance have argued that pressures from investors and analysts can drive companies to under-invest in positive NPV projects (Stein 1989; Bebchuk and Stole 1993; Graham et al. 2005), if markets are sufficiently impatient (Gigler et al. 2014). In support of this view, several studies show that analysts are associated with greater pressures to meet short-term benchmarks. Bhojraj et al. (2009) show that firms cut discretionary expenditure to meet analysts expectations of near-term earnings. More specific to investments, Irani and Oesch (2016) and He and Tian (2013) show that exogenous decreases in analyst coverage are associated with less real earnings management and greater patenting activity. Thus, the evidence presented in these studies, suggest that the presence of analysts does play some role in cutting investment in long-term projects.

Notably, the short-termism concerns of analyst coverage, generalizes well to the setting of digital or AI investments. These technologies involve high short-run costs in SG&A and R&D that are not always capitalizable. Moreover, prior research on the payoffs of the related investment in ICTs, suggest that it could take several years for the full benefits of new digital technologies to be fully realized (Brynjolfsson and Hitt 2003; Brynjolfsson et al. 2017). Thus, the investment in these technologies are exceptionally prone to the influence of short-term pressures from analysts. And hence, I expect the following hypothesis:

H2a: Analysts' interest in digitization is associated with under-investment in AI.

Another view, is that analysts can facilitate effective investment in AI technologies. This view is motivated by a group of studies that argue that analysts and other market intermediaries facilitate investments by monitoring corporate decisions. From a broad perspective, there is some recent evidence that short-term oriented market intermediaries are associated with more effective investment decisions. For instance, Giannetti and Yu (2020) show that short-term investors help firms invest more in periods of radical change, while studies on shareholder activism, have also shown that activist hedge funds are associated with more effective long-term investment (Brav et al. 2018).

For analysts, there is also some evidence in favor of the claim that these intermediaries facilitate investment decisions. Derrien and Kecskés (2013) show that analysts help reduce information asymmetries and cost of capital, which allows companies to invest more in R&D and CapEx. Similarly, To et al. (2018) show that analysts are associated with better investment efficiency in firms. And several studies show that the quality of analysts, as measure by the forecasting attributes, is associated with greater investment efficiency in firms (T. Chen et al. 2017; Choi et al. 2020). Thus, there is also some reason to expect the following:

H2b: Analysts' interest in digitization is associated with efficient-investment in AI.

Finally, studies from the accounting and finance literature argue that analysts could also play a role in driving over-investment in AI technologies. Specifically, in situations where markets are uncertain over the productivity of investment, the short-term expectations from analysts could drive firms to over-invest in new technologies to signal their type (Bebchuk and Stole 1993). Notably, the digital adoption setting likely fits well with this model: The payoffs to digital investment are highly uncertain (see the discussion in Section 2.1), and the full productivity benefits of these technologies is still unknown (Brynjolfsson et al. 2017).

Evidence from the dot-com bubble, provides some supporting empirical evidence for the signaling effect of short-term market expectations. Specifically, studies have shown that during the market frenzy over IT investments, investors anchored on actions that signalled higher IT investments, rather than on metrics of current profitability or productivity. For example, Demers and Lev (2001), Bartov et al. (2002) and Rajgopal et al. (2002) showed that investors of internet firms focused mainly on cash burn and managerial actions rather than on other metrics of profitability. And analysts potentially played some role in driving this market frenzy, as studies have shown that analysts forecast were highly optimistic (Healy and Palepu 2003; O'Brien and Tian 2006).

The current trend of non-IT firms investing in digital technologies draws some parallel with the dot-com era, as both trends are based on emergent GPTs. For these technologies, the full extent of the benefits and costs of these technologies have yet to be proven, which drives high uncertainty over its productivity. This uncertainty could in turn drive signaling incentives (Bebchuk and Stole 1993) to over-invest, which leads to the following hypothesis:

H2c: Analysts' interest in digitization is associated with over-investment in AI.

3.2 Data

In this section, I begin with the sample selection. The second sub-section, outlines the measurement methodology of the two key variables in this study — analysts' interest in digitization and AI investment. And the final sub-section discusses the key summary statistics.

3.2.1 Sample Selection

The sample of this study consists of all US-listed non-IT firms in the Compustat/CRSP universe from 2010Q1-2020Q4. The choice of focusing on only non-IT firms is motivated by the fact that these firms are at the early stages of digitization, and so the concerns relating to co-invention costs and technology suitability is perhaps more salient. To limit the sample to non-IT firms, firms in technology industries identified in Section A.2 or with names containing the words "Technologies" and "Analytics", are dropped from the sample.

Table 3.1: Sample Selection

Sampling Criteria	Firm-Quarters	Firms
Compustat-CRSP Merge from Q12010-Q42020	230,889	8,752
Share Codes 10, 11, 12 and US Headquartered and Incorporated Firms	157,275	6,231
Non-IT Firms	113,832	4,378
Conference Call Transcripts with Analyst Questions	69,608	2,925
BurningGlass Observations	43,353	2,218
Observations with non-missing β	39,478	1,963
6-Digit GICS Industry-Quarters with at least 3 observations	39,143	1,956

This table reports the sample selection for the study, which is based on a subset of US-listed, non-IT firms from 2010Q1-2020Q4.

In addition, several other sample restrictions are also implemented. To enter the sample, each firm-quarter observation requires valid conference call data (from *Thomson Reuter StreetEvents*) with analyst questions and job postings data from *BurningGlass*[20]. The data is also restricted to 6-digit

20. The conference call data was matched to the main dataset by the stock tickers and name, while the job posting data was matched to the main dataset by matching names using fuzzy matching and manual validation. For more details on the matching procedure, see Appendix B.4.

GICS industry-quarters with more than 3 firms, and for firm-quarter observations with sufficient history of weekly returns for β estimation purposes[21]. In sum, the final sample has a total of 39,135 observations, across 1,956 firms (see Table 3.1, for the full breakdown of observations).

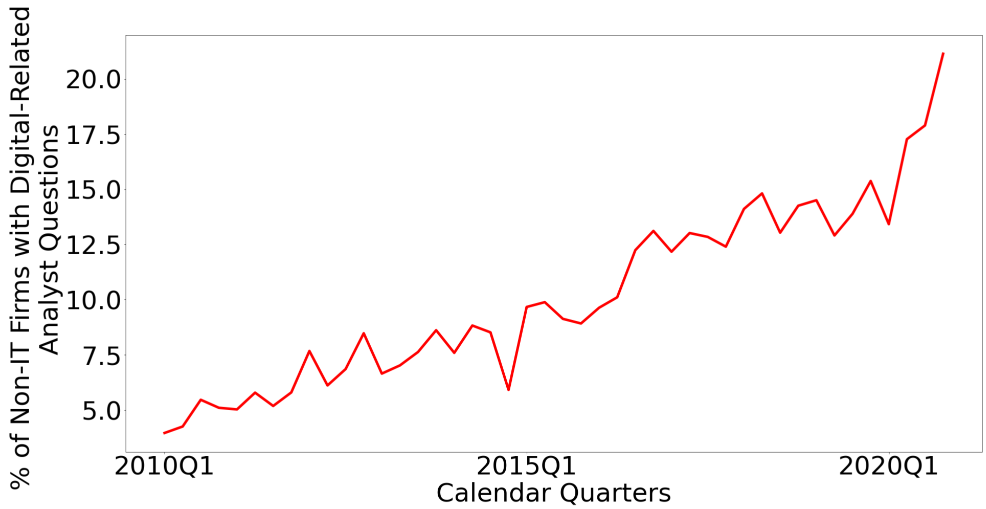
3.2.2 Measuring Analyst Interest in Digitization

The main independent variable in this study is analyst interest in the digitization of non-IT firms. This construct is measured using the proportion of digital analyst questions in quarterly earnings conference calls. The appeal of using analysts' questions in earnings conference to measure analyst interest in digitization is two-fold — (1) conference calls cover a wide-set of publicly listed firms and (2) prior work has shown that information obtained from the Q&A in conference calls is important to financial markets (Matsumoto et al. 2011).

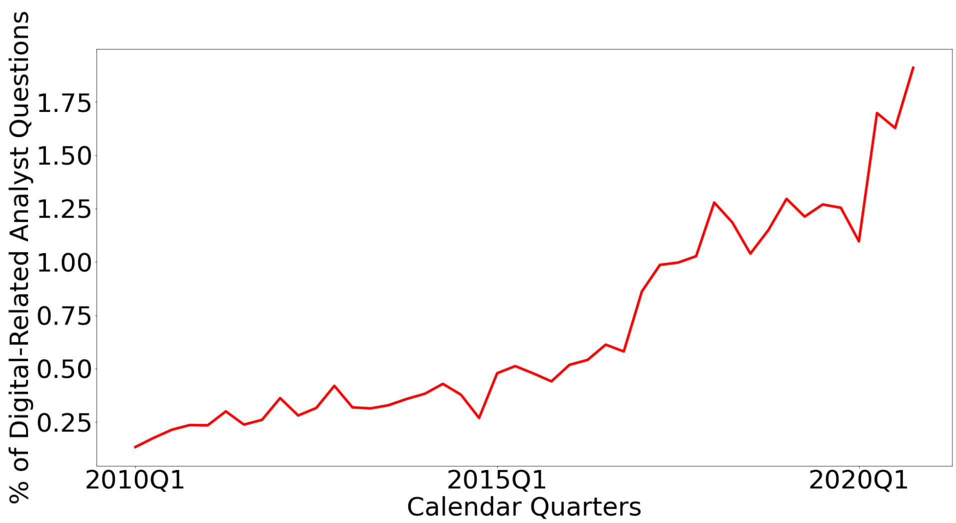
To parse digital questions from the conference call transcripts, analyst questions are first extracted from the transcripts. Each of these transcripts, provide corporate participants lists, and I use the names in that list to identify question segments of analysts in the Q&A portion of the conference call. These questions are then coded as digital questions if the question mentions terms that are relating to digital strategy, analytics, big data or artificial intelligence technologies. For the digital strategy topic, I use regex expressions with the lemmatized version of the word "digital" to search for terms relating to digital strategy. For the topics relating to the three specific groups of digital technologies (analytics, big data and AI), I use the terms from the word list in Section A.1 as seed words to form an expanded dictionary, using cosine similarity scores between word-embeddings (by training and applying a neural network Word2Vec model, following, K. Li et al. 2021)[22].

Panel A, Figure 3.1 presents the distribution of firms in non-IT industries that are covered by analysts that are asking digital questions in the conference calls (Panel B presents the proportion of

21. I use β based on 3-year weekly returns as a control for risk in all the main regressions
22. See Appendix B.2.1 for more details on the methodology and the specific word lists



(a) Proportion of Firms with Digital Questions From Analysts



(b) Proportion of Analyst Questions with Digital Terms

Figure 3.1: Proportion of Non-IT Firms with Digital Questions From Analysts over Calendar Quarters (a) and (b) Proportion of Analyst Questions with Digital Terms

digital questions in all non-IT firms over time). From the figure, one can clearly see a marked increase in analyst discussion of digital topics over time. In first quarter of 2010, only 3% had analysts who were discussing digital-related topics, and in the last quarter of 2020, this proportion has risen to about 20%.

Across industries, there is substantial variation in the intensity of digital questions. Figure 3.2 presents the heat map of industry-years of non-IT firms with digital questions, and shows that the intensity is highest amongst the media, retailing, services, and financial industries. Furthermore, across time, the figure shows a growth in intensity in most industries.

To shed some light into the types of questions that analysts ask regarding digitization, the digital questions are split into various topic groups. Following prior work, digital questions on performance and competition topics are identified based on the word lists in Matsumoto et al. (2011) and F. Li et al. (2013) respectively. Digital questions relating to strategy are identified using the regex expression in Appendix B.2.2, and the questions relating to digital technologies are identified based on the terms in the digital dictionary in Appendix B.2.3[23]. Finally, the sentiment of the questions is measured following A. Huang et al. (2020).

Overall, analysts appear to be more interested in the performance aspect of digitization than competition or technology-related topics. Statistics reported in Table 3.2 suggest that roughly 20% of digital questions are devoted to performance topics, while digital questions on competition and technology topics comprise of roughly 5% and 50% of all questions.

3.2.3 Measuring AI Investment

The next key variable of interest is AI investment, which is measured through the human capital investment in AI technologies, reported in online job posting data (Babina et al. 2020; Goldfarb et al. 2019; Acemoglu et al. 2020). To measure AI job postings, I identify job postings with AI skills

23. For some examples of digital questions on the various topics, see Appendix B.3.1

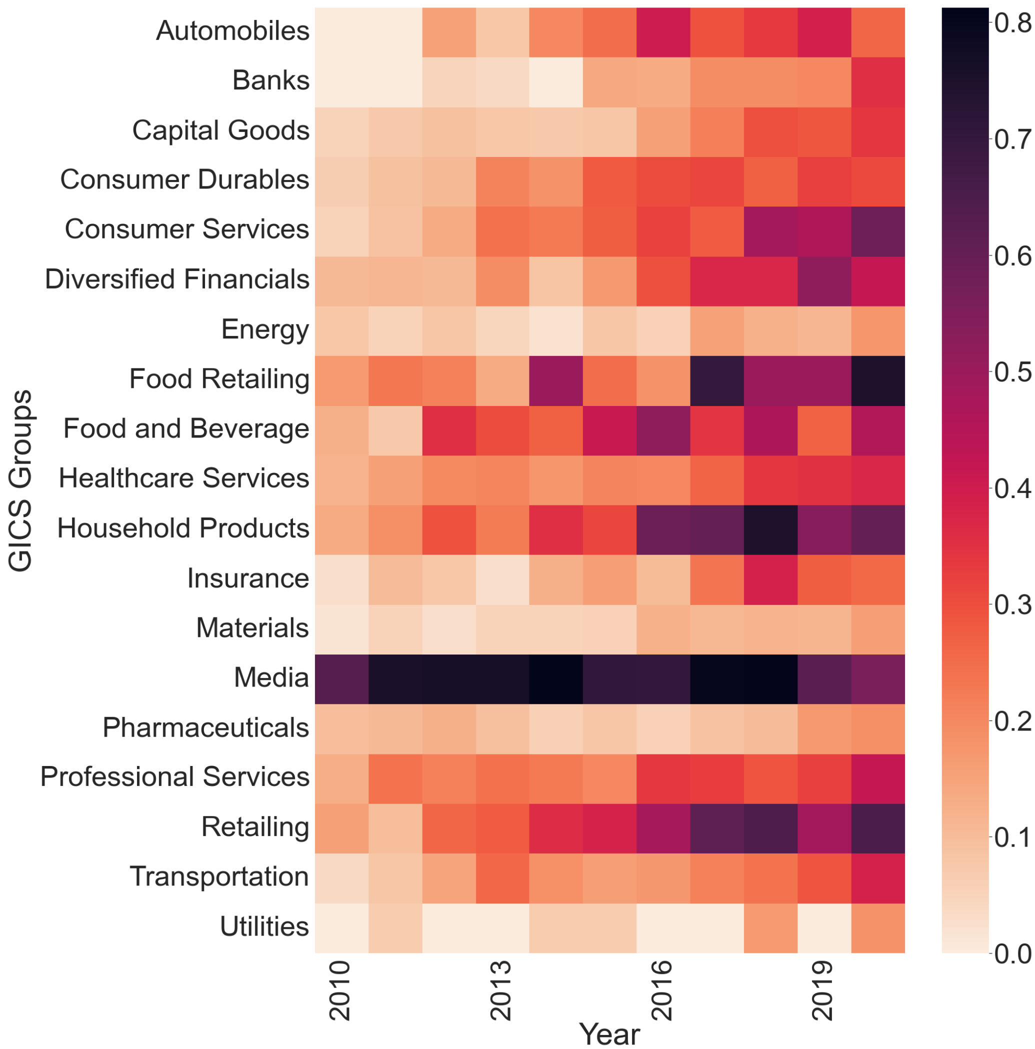


Figure 3.2: Proportion of Non-IT Firms with Digital Questions From Analysts Over 4-Digit GICS Groups.

Table 3.2: What Types of Digital Questions Do Analysts Ask?

Panel A: Topics by Year									
Year	Total	Perf.	Comp.	Digital Strat.	Big Data	Analytics	AI	Tech.	Sentiment
2010	224	18.3	4.9	50.9	2.7	25	21.9	49.1	28.7
2011	325	17.8	0.9	53.2	1.5	22.2	23.1	46.8	24.4
2012	431	17.6	2.6	52.9	2.3	26.2	19.7	47.3	21.6
2013	410	18.8	4.1	52.2	2	26.6	20	47.8	29.2
2014	469	20.3	3.2	51.8	2.6	29	17.3	48.6	39.8
2015	558	18.3	3	51.1	2.5	28.7	18.1	49.1	29.6
2016	607	26.2	4.8	45.3	3.1	33.1	20.3	55.5	30
2017	687	26.3	4.1	47.6	7.6	24.5	24.5	54.6	28.8
2018	855	24.6	5.8	47.7	5.4	23.2	26.3	53.9	30.7
2019	885	25.8	5	44.5	5.1	24.5	27.9	56.9	31.8
2020	1079	19.8	2.3	61.2	3.2	17.1	21.5	41.1	39.1
Panel B: Topics by GICS Groups									
GICS	Total	Perf.	Comp.	Digital Strat.	Big Data	Analytics	AI	Tech.	Sentiment
Automobiles	88	21.6	9.1	26.1	17	26.1	37.5	77.3	33.1
Banks	209	11.5	1.9	86.6	3.3	6.2	9.1	17.2	39.9
Capital Goods	525	37.3	4.6	17.3	8.4	13	63.4	83.8	29.1
Consumer Durables	321	19	2.2	63.2	3.1	22.4	13.1	38.3	36.1
Consumer Services	463	13.8	0.6	73.7	0.6	21.4	6.3	28.3	34.2
Diversified Financials	357	46.8	7.8	21.6	5	54.9	21.6	79.6	33
Energy	171	35.7	2.3	28.1	8.2	14	55.6	74.3	33.3
Food Retailing	120	24.2	5.8	59.2	3.3	35	3.3	41.7	25.6
Food and Beverage	190	46.3	3.2	13.2	0	61.1	26.3	87.4	12
Healthcare Services	835	30.9	9.8	11.1	2.8	40.1	47.4	89.5	29.9
Household Products	134	35.8	4.5	44.8	1.5	52.2	3.7	56	16.4
Insurance	103	26.2	3.9	39.8	1.9	38.8	21.4	62.1	33.6
Materials	106	25.5	0.9	49.1	2.8	25.5	24.5	51.9	24
Media	1349	3.6	0.7	91.5	2.6	5	1.6	8.9	29.5
Pharmaceuticals	282	26.6	6	17	10.6	26.6	47.9	84.4	24.4
Professional Services	384	28.4	4.2	45.1	7.8	33.9	16.1	56.8	40.7
Retailing	771	12.8	1.6	69.4	1	25.6	5.8	31.6	39.2
Transportation	115	33	9.6	21.7	2.6	18.3	59.1	79.1	27.8
Utilities	7	28.6	0	0	14.3	0	85.7	100	28.6

This table reports the distribution of topics related to the analyst questions on digital strategy. The table reports the cross-year and -industry distributions for three different topics, ranging from digital strategy-related (see Appendix B.2.2 for definitions), technology-related (see Appendix B.2.3 for definitions), performance-related (see Matsumoto et al. (2011) and Appendix B.2.4 for definitions), competition-related topics (see F. Li et al. (2013) and Appendix B.2.5 for definitions). For the technology-related topics, the topic is further split into 3 subtopics — namely, big data-related, analytics-related, and artificial intelligence (AI-related). Furthermore, the last column, reports the distribution of sentiment (A. Huang et al. 2020) across industries and years. Panel A presents the distribution of questions-relating to each topic over years. Panel B reports the distribution of questions-relating to 4-digit GICS groups. The GICS group studied are the following: Energy (1010), Materials (1510), Capital Goods (2010), Commercial & Professional Services (2020), Transportation (2030), Automobiles & Components (2510), Consumer Durables & Apparel (2520), Consumer Services (2530), Retailing (2550), Food & Staples Retailing (3010), Food, Beverage & Tobacco (3020), Household & Personal Products (3030), Health Care Equipment & Services (3010), Pharmaceuticals, Biotechnology & Life Sciences (3020), Banks (4010), Diversified Financials (4020), Insurance (4030), Media (5020), Utilities (5510), Real Estate (6010).

using the set of AI skills in Alekseeva et al. (2021)[24]. With this measure, two measures of AI investment, measured at the quarterly level are used in the empirical analysis — (1) an indicator variable for whether the firm posts AI-related job postings, and (2) the intensity of AI-related job postings relative to all job posting issued at the firm (following, Babina et al. 2020; Acemoglu et al. 2020).

Panel A, Figure 3.3 presents the distribution of firms in non-IT industries that make AI-related job postings. There is a marked increase in posting of AI-related jobs over time, growing from 12% in first quarter of 2010 to about 30% in last quarter of 2020[25]. The distribution AI job posting as a fraction of all job-postings in non-IT firms is also rising as well. Moreover, there is also a marked increase from 2015, which is consistent with the findings in Acemoglu et al. (2020) and aligns with the technological breakthroughs in neural network algorithms during the mid-2010s.

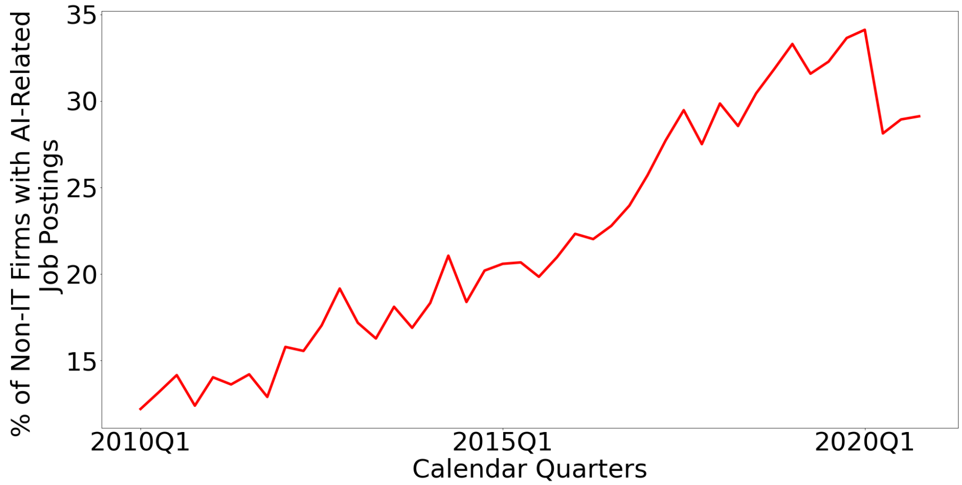
Across industries, there is also variation in the intensity of AI job postings. Figure 3.4 presents the heat map of industry-years of non-IT firms with these postings, and shows that the intensity is highest amongst the media, retailing, transportation, services, and financial industries. Across time, the figure also shows a growth in intensity in most industries.

3.2.4 Summary Statistics

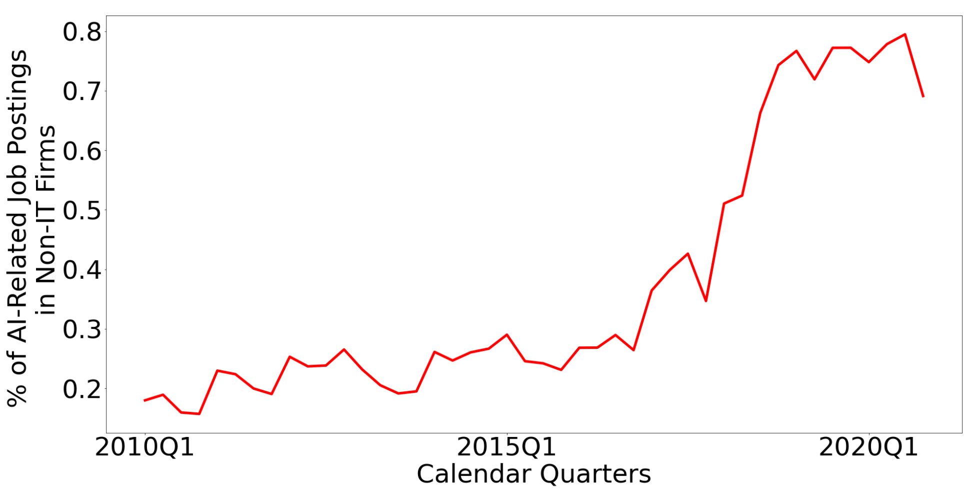
To provide a brief overview of the sample, I present the sample statistics of the operating characteristics of the non-IT firms, as well as the sample statistics of the key variables. These statistics are reported in Table 3.3, and continuous variables are winsorized at the 1% and 99% level. The average (median) size of the firm in the 17 (2) Billion USD in assets, and the average (median) performance of the firm is 1.1% (1.8%) in return-on-assets and 11% (5%) in sales growth. These firms tend to be more established, traditional firms, as the average age of non-IT firms is around 27 years. Moreover, these firms

24. For an example, see Figure B.1 in Appendix B.3.2

25. In the figure, there is a noticeable drop in the proportion of firms posting AI job postings in 2020. One potential explanation for this pattern could be the labor market shocks that occurred due to the COVID-19 crisis, which prevented companies from engaging in their regular hiring practices.



(a) Proportion of Firms with AI-Skills



(b) Proportion of AI-related Job Postings to All Job Postings

Figure 3.3: Proportion of Non-IT Firms with (a) AI-Related Job Postings and (b) AI-Related Job Postings Relative to Total Job Postings

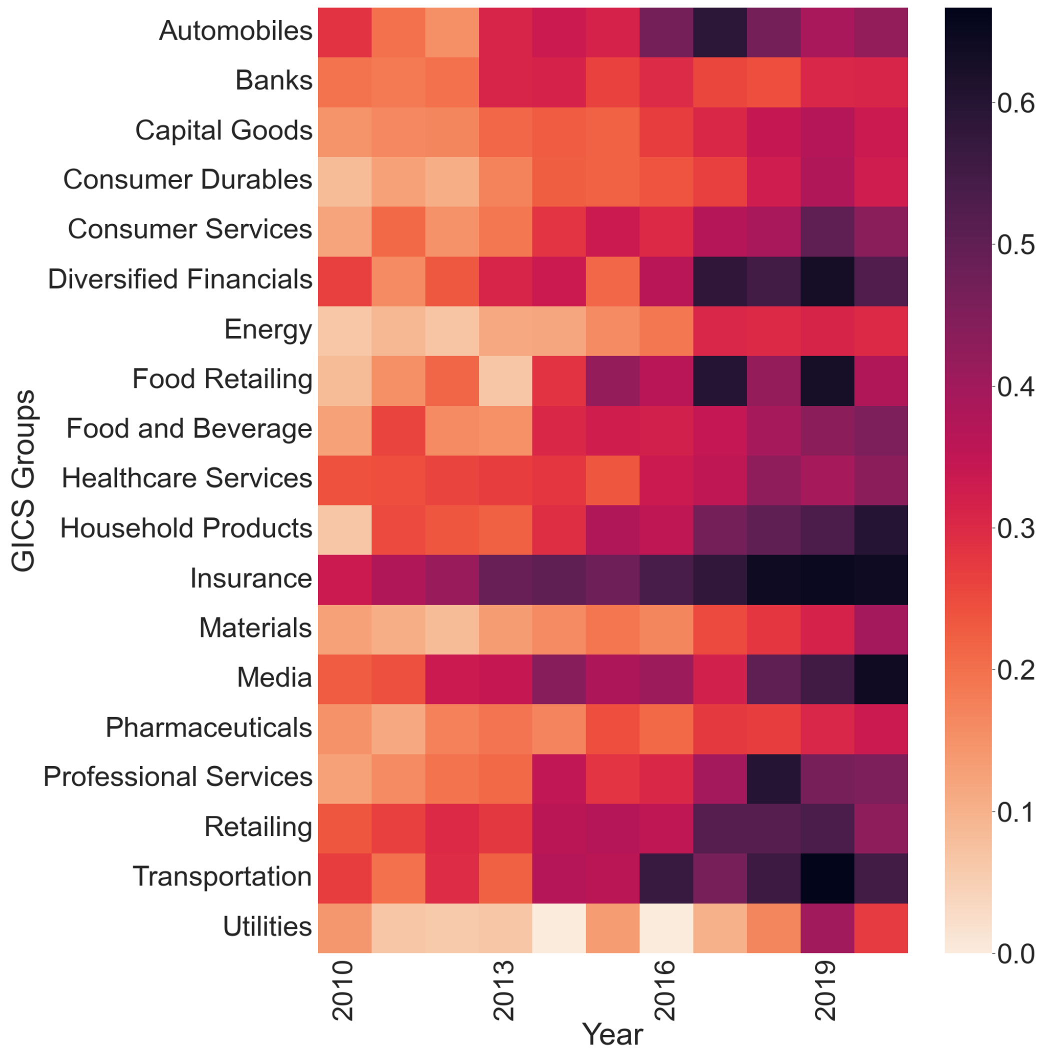


Figure 3.4: Proportion of Non-IT Firms with AI-Related Job Postings Over 4-Digit GICS Groups.

tend to exhibit more complex operating structures as most firms in the sample tend to have more than 1 segment. And they also tend to pay dividends, a further indication that they are more established.

Table 3.3: Summary Statistics

Dependent Variable	Mean	SD	Median	25%	75%	N
Total Assets (Billions)	17.092	120.266	2.098	0.613	7.595	39135
Cash Flow to Assets	0.014	0.047	0.018	0.002	0.036	39115
Dividend Firm	0.53	0.499	1	0	1	39143
Debt to Assets	0.272	0.225	0.239	0.088	0.391	38774
Sales Growth	0.114	0.45	0.053	-0.027	0.154	38432
Industry-Adjusted Sales Growth	0.065	0.434	0.003	-0.063	0.091	38432
Return-on-Assets	0.011	0.046	0.018	0.005	0.031	38897
Industry-Adjusted Return-on-Assets	0.007	0.039	0.002	-0.006	0.017	38897
Herfindahl Index	0.114	0.072	0.092	0.061	0.145	39143
SG&A-to-Sales	0.212	0.221	0.161	0.068	0.301	38721
R&D-to-Sales	0.273	1.654	0	0	0.025	38721
Missing R&D (Indicator)	0.566	0.496	1	0	1	39143
Proportion of Major Customer Sales	0.203	0.3	0	0	0.34	39143
Segment Number	2.038	1.326	2	1	3	39143
Firm Age	27.87	17.576	23.014	14.258	39.151	39143
Tobins Q	1.701	1.615	1.208	0.805	1.972	38773
Industry-Adjusted Tobin's Q	0.425	1.336	0.036	-0.209	0.584	38773
Market Beta	1.35	0.657	1.242	0.904	1.666	39143
Intensity of AI Job Postings	0.005	0.025	0	0	0	39143
AI Job Postings (Indicator)	0.226	0.418	0	0	0	39143
Proportion of Digital Questions	0.008	0.031	0	0	0	39143
Managerial Digital Disclosure (Indicator)	0.336	0.472	0	0	1	39143
Managerial Digital Experience (Indicator)	0.076	0.266	0	0	0	38956
Board Digital Experience (Indicator)	0.234	0.423	0	0	0	38947

This table presents the summary statistics of the sample of Non-IT firm-quarter observations from 2010Q1-2020Q4. Statistics of 24 variables are reported, namely, the size of the firm (total assets), cash flows to assets, indicator for a dividend-paying firm, debt to assets, sales growth, industry-adjusted sales growth, return-on-assets, industry-adjusted return-on-assets, herfindahl index, SG&A-to-sales, R&D-to-sales, indicator for missing R&D, proportion of major customer sales, number of segments, firm age, Tobin's Q, industry-adjusted Tobin's Q, market beta, intensity of AI job postings, indicator for whether a firm makes an AI job postings, proportion of digital questions, indicator for whether the management team discloses digital terms in the presentation portion of the conference call, indicator for whether the firm has a top-10 executive with digital experience and indicator for whether the firm has a board member with digital experience.

Yet, there are also indications that a sizeable portion of these non-IT firms are engaging in some form of digitization. Around 33% of firm-quarter observations exhibit some degree of managerial disclosure on digital-related topics. These firms also tend to have executives and board members with experience in digital technologies. The sample statistics suggest that 7% of firm-quarter observations exhibit a top-10 executive with experience in digital technologies, while 23% of observations exhibit a board member with experience in the same technologies. Thus, these statistics speak to the relevance of digitization to these non-IT firms and motivates an empirical study of the phenomenon.

3.3 Empirical Analysis

In this section, I present the empirical results in two parts. The first studies the factors that drive analysts' interest in the digitization of non-IT firms. The second examines the association between analysts' interest in this process and AI investment.

3.3.1 Firm and Analyst Focus on Digitization in Non-IT Firms

The goal of this subsection, is to unpack the drivers of analysts' interest in the digitization of non-IT firms. While the discussion in Section 2.3 outlined several key digital investment factors that should also drive analysts' interest in digitization, it is a priori, unclear whether these factors hold in the AI investment setting. Therefore, to benchmark the drivers of analysts' interest in digitization, I first evaluate the drivers of AI investment.

Firm Determinants of AI Investments

As argued in Section 2.3, five investment factors should drive AI investment and these are, namely, the suitability of AI across industries, the financing/economic risks, governance attributes, product market competition and organizational frictions within the firm. To evaluate these drivers, I rely on a fixed effects design that is implemented as follows:

$$\begin{aligned} Posting_{i,q} &= \alpha_q + \alpha_j + \beta_1 AI\ Index_{l,t-1} + \beta_2 Constraints_{i,q} + \beta_3 Q_{i,q} \\ &+\beta_4 Managerial\ Experience_{i,t-1} + \beta_5 Board\ Experience_{i,t-1} + \beta_6 Age_{i,q} \\ &+\beta_7 Late_{i,t-1} + \beta_8 Segments_{i,t-1} + \beta_9 Posting_{j,q} + \beta_{10} Herfindahl_{j,t-1} \\ &+\beta_{11} Sales\ Growth_{i,q,q-4} + \textstyle\sum_k \gamma_k X_{i,k,t} + \varepsilon_{i,t} \end{aligned} \tag{3.1}$$

where the dependent variable, $Posting_{i,q}$, is the intensity of AI job postings (following the methodology in, Babina et al. 2020; Acemoglu et al. 2020). To measure the different sources of variation, two types of fixed effects models are used. One, a quarter fixed effects model (α_q) that estimates effects of the predictors relative to peer firms in the same quarter. Two, a 6-digit GICS industry and quarter fixed effects model (α_q and α_j), that holds time invariant industry factors fixed and studies the effects of the predictors as within-industry changes.

The set of determinants used in the analysis is based on the five investment factors. The textual-based industry-year index of AI suitability ($AI\ Suitability_{l,t-1}$)[26] proxies for AI technology suitability (Brynjolfsson et al. 2018; Felten et al. 2021). The capital constraints index ($Constraints_{i,t-1}$) and industry-adjusted Tobin's Q ($Q_{i,q}$) measures financing and economic risks (Hayashi 1982; Whited and Wu 2006). Indicators for whether the firm has a top-10 executive with digital experience ($Managerial\ Experience_{i,t}$) or board members with digital experience ($Board\ Experience_{i,t}$) proxies for the execution expertise of the firm (Matta et al. 2019). The firm's age ($Age_{i,t}$), an indicator for late lifecycle firms[27] ($Late_{i,t}$) and the number of segments ($Segments_{i,t-1}$), measures organizational frictions (Bresnahan and Greenstein 1996; Henderson and Clark 1990; Bresnahan 2010; Christensen 1997). The industry-level AI adoption rates ($AI\ Posting_{j,q}$), industry-adjusted sales growth ($Sales\ Growth_{i,q,q-4}$), and the herfindahl index ($Herfindahl_{j,t-1}$)[28], proxy for product market competition (Arrow 1962; Aghion et al. 2005). Finally, to control for the operating profile of the firm, the regression model includes a set of controls, namely, industry-adjusted ROA, market beta, SG&A, R&D (and a missing R&D indicator following Koh and Reeb 2015) and the proportion of major customers.

Table 3.4 presents the results of this determinants analysis for the regression without fixed effects,

26. This is measured using textual overlap between titles/abstracts in filed US patents on AI technologies and the 2017 6-digit NAICS industry descriptions (following methodologies in Kogan et al. 2020; Webb 2020). For more details, see Appendix B.5.

27. The "mature","shake-out" and "decline" stages in the cash-flow based classification in Dickinson (2011).

28. Note that these industries are indexed by j, which is at the 6-digit GICS level, as prior work has suggested that the GICS classification is better at grouping peer firms in a financial context (Bhojraj et al. 2003)

with quarter fixed effects and with quarter plus firm fixed effects respectively. Overall, the results shows that all five investment factors are significantly associated with AI investments. Across all models, there is consistent evidence that industries that are more suitable for AI as measured by the textual-based AI index is associated with more AI investment. Specifically, the estimates from the industry fixed effects model suggests that a 1 σ increase in the AI suitability score is associated with a 0.06% increase in the intensity of AI investment (or a 12% increase relative to the average AI investment intensity).

For financing and economic risks, the results suggests that firms with higher industry-adjusted Tobin's Q tend to exhibit higher AI investment. Estimates from the industry fixed effects model show that a 1 σ increase in Tobin's Q is also associated with roughly 0.06% in AI investment intensity (or a 13% increase relative to the average AI investment intensity). On the flip side, the results show that firms with greater capital constraints exhibit lower AI posting intensity in the quarter fixed effects model. Estimates from this regression indicate that a 1 σ increase in the capital constraints index is associated with a 0.07% decrease in AI posting intensity (or a 14% decrease relative to the average AI investment intensity).

Next, I examine the variables relating to the execution expertise of the firm and organization frictions. For execution expertise, whether the board has experience digital technologies appears to be an important indicator of greater AI investments. Such firms exhibit a 0.25% higher AI investment intensity compared to firms with no such board members (or a 50% increase relative to the average AI investment intensity), in the quarter fixed effects model. While the indicator for managerial experience is statistically insignificant, it is positively related to AI investment intensity, and untabulated analysis reveals that when the board experience variable is dropped, managerial experience is significantly associated with a 0.2% increase in AI investment intensity[29]. Moreover, consistent with organizational frictions in technology adoption, the results also suggest that late life-cycle firms also tend to exhibit

29. Moreover an indicator that is coded as 1 for firms that have either board or managerial experience in digital technologies also exhibit a significantly positive correlation with AI investment intensity.

Table 3.4: What Drives Firm-level Investment in AI?

Dependent Variable	% AI Posting Intensity	% AI Posting Intensity	% AI Posting Intensity
Technology Suitability			
AI Suitability$_{l,t-1}$	0.137***	0.082***	0.065***
	(0.025)	(0.024)	(0.023)
Financing and Economic Risks			
Capital Constraints Index$_{i,q}$	-0.336***	-0.341***	-0.264***
	(0.070)	(0.070)	(0.066)
Industry Adjusted Tobin's Q$_{i,q}$	0.063***	0.063***	0.058***
	(0.015)	(0.015)	(0.014)
Execution Expertise			
Managerial Experience$_{i,t-1}$	0.122	0.088	0.122
	(0.082)	(0.082)	(0.082)
Board Experience$_{i,t-1}$	0.262***	0.256***	0.250***
	(0.055)	(0.054)	(0.053)
Organizational Frictions			
Number of Segments$_{i,t-1}$	0.029**	0.035**	0.012
	(0.013)	(0.013)	(0.013)
Logarithm of Age$_{i,q}$	-0.004	-0.013	0.009
	(0.023)	(0.023)	(0.024)
Proportion of Major Customer Sales$_{i,t-1}$	0.041	0.019	-0.048
	(0.056)	(0.056)	(0.065)
Late Lifecycle$_{i,t-1}$	-0.107**	-0.110**	-0.062
	(0.044)	(0.043)	(0.040)
Product Market Competition			
AI Postings$_{j,q}$	1.060***	0.740***	-0.381**
	(0.164)	(0.168)	(0.145)
Herfindahl Index$_{j,t-1}$	-0.073	-0.001	-1.314**
	(0.240)	(0.236)	(0.641)
Industry Adjusted Sales Growth$_{i,q,q-4}$	0.023	0.018	0.005
	(0.018)	(0.018)	(0.018)
Calendar Quarter FE	No	Yes	Yes
6-Digit GICS FE	No	No	Yes
Observations	37,659	37,659	37,659
R^2	0.065	0.075	0.107

This table reports the determinants of AI investment as measured by the intensity of AI job postings. This table examines five potential drivers of AI investments. The first factor, technology suitability, is measured by the extent of textual overlap between AI patent text and the textual description of the firm's NAICS industry. The second factor, financing and economic risks, is proxied by industry-adjusted Tobin's Q, and the Whited-Wu capital constraints index. The third factor, governance attributes, is proxied by managerial and board-level experience in digital technologies. The fourth factor, organizational factors, is proxied by the number of segments, firm age and an indicator for late lifecycle firms. The fifth factor, product market competition, is measured with the adoption rate of AI technology at the 6-digit GICS industry-level, industry-adjusted sales growth and the herfindahl index of the same industry. Finally, several variables measuring the operating profile of the firms are added as controls, namely, market beta, industry-adjusted ROA, SG&A expenditure, R&D expenditure (also includes an indicator for missing R&D) and the proportion of major customer sales. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

lower AI investment. Specifically, the estimates from the quarter fixed effects model show that a 1 σ increase in life-cycle is associated with 0.11% decrease in AI investment intensity relative to peers (or a 22% decrease relative to the average AI investment intensity).

Finally, I study the product market competition factors that may shape AI investment. I find that the intensity of AI postings at the industry-level predicts greater AI investment relative to peers in the quarter fixed effects model[30]. Specifically, a 1 σ increase in AI posting intensity at the industry-level is associated with a 0.04% increase in AI investment intensity (or a 8% decrease relative to the average AI investment intensity). In addition, the industry-fixed effects model shows that a 1 σ decrease in the herfindahl index is associated with a 0.09% increase in AI investment intensity (or a 18% decrease relative to the average AI investment intensity), which suggest that as industries become more competitive, firms tend to invest more in AI-based technologies.

Thus, taken together, the above analysis is suggestive that the five theorized drivers of investment, predict AI investment decisions. Thus, in the next set of analyses, I use these drivers as a benchmark, to assess the factors that predict analysts' interest in digitization.

Analysts' Interest in Digitization

Broadly, this determinants analysis relies on the same variables used in the previous regression model, but adds managerial and peer analyst disclosure variables. The regression analysis follows a similar fixed effects design, that is implemented as follows:

$$Questions_{i,q} = \alpha_q + \alpha_j + \beta_1 AI\ Index_{l,t-1} + \beta_2 Constraints_{i,t-1} + \beta_3 Q_{i,q} \tag{3.2}$$

30. One caveat is that the industry fixed effect specification shows a negative relationship between the industry-level AI postings and firm-level AI postings. This is somewhat of a mechanical result as the industry-level posting variable is defined as the proportion of peer firms that have AI posting at the industry-quarter level (i.e. the numerator is the total number of firms in the industry that have AI postings minus one if the firm itself has such a posting). Thus, with the industry-fixed effects included, the variation in the industry-level posting variable is mostly driven by whether the firm has an AI posting, and consequently drives a mechanically negative relationship between the industry-level AI posting variable and the firm-level AI intensity variable.

$$+\beta_4 Managerial\ Experience_{i,t-1} + \beta_5 Board\ Experience_{i,t-1} + \beta_6 Age_{i,t-1}$$

$$+\beta_7 Late_{i,t-1} + \beta_8 Segments_{i,t-1} + \beta_9 Posting_{j,q} + \beta_{10} Herfindahl_{j,t-1}$$

$$+\beta_{11} Sales\ Growth_{i,q,q-4} + \beta_{12} Managerial\ Disclosure_{i,q}$$

$$+\beta_{13} Managerial\ Disclosure_{j,q} + \beta_{14} Questions_{j,q} + \sum_k \gamma_k X_{i,k,t} + \varepsilon_{i,t}$$

where the dependent variable, $Questions_{i,q}$, is the ratio of the total number of digital questions to the total number of questions posed by analysts of firm i in conference call in quarter t. Like the previous analyses, two variants of the model are studied, one with quarter-fixed effects, and one with quarter plus industry fixed effects. As disclosures from alternative sources arguably play an important role in shaping analysts disclosures (Bowen et al. 2002; Merkley et al. 2017), several disclosure variables are also included in the determinant analysis. These are an indicator for whether there digital-related topics in management's presentation during the conference call[31] ($Managerial\ Disclosure_{i,q}$), the proportion of firms with managerial disclosure on digital topics ($Managerial\ Disclosure_{j,q}$) and the proportion of firms with digital questions from analysts at the industry level ($Questions_{j,q}$).

Table 3.5 presents the determinants analysis of the digital questions for the OLS, quarter fixed effects and industry fixed effects regression models, respectively. In the first row, the AI suitability index is positively associated with digital questions. Specifically, the estimates from the industry fixed effects model suggest that a 1 σ increase in the AI index score, is associated with a 0.12% increase in digital questions (or a 15% increase relative to the average proportion of questions)[32].

On the other hand, the results shows that financing and economic risks plays a limited role in driving analysts' questions on digitization. Across the various regression models capital constraints

31. I measure managerial disclosure through disclosure in the conference call for two reasons. One, disclosure through this channel is timelier compared to other sources such as the disclosures in the 10-K. Two, the disclosure in the conference call is more targeted to financial analysts; in contrast, disclosures in the 10-K, for instance, is consumed by a wider range of stake holders.

32. As a further robustness check, I also show in Appendix Table B.2 that digital questions are also positively associated with the Brynjolfsson et al. (2018) and Felten et al. (2021) cross-sectional indices of AI suitability.

Table 3.5: What Drives Analyst Interest in Digitization?

Dependent Variable	% Digital Questions	% Digital Questions	% Digital Questions
Technology Suitability			
AI Suitability$_{l,t-1}$	0.186***	0.165**	0.126*
	(0.065)	(0.067)	(0.073)
Financing and Economic Risks			
Capital Constraints Index$_{i,q}$	-0.037	-0.029	-0.107
	(0.095)	(0.095)	(0.100)
Industry Adjusted Tobin's Q$_{i,q}$	0.039	0.039	0.045
	(0.028)	(0.028)	(0.028)
Execution Expertise			
Managerial Experience$_{i,t-1}$	0.641***	0.631***	0.534***
	(0.172)	(0.174)	(0.160)
Board Experience$_{i,t-1}$	0.222***	0.226***	0.184***
	(0.069)	(0.069)	(0.065)
Organizational Frictions			
Number of Segments$_{i,t-1}$	-0.030*	-0.026	-0.031*
	(0.016)	(0.017)	(0.018)
Logarithm of Age$_{i,q}$	0.054	0.051	0.033
	(0.052)	(0.053)	(0.055)
Proportion of Major Customer Sales$_{i,t-1}$	-0.101	-0.118	-0.053
	(0.070)	(0.070)	(0.091)
Late Lifecycle$_{i,t-1}$	0.018	0.014	-0.001
	(0.067)	(0.067)	(0.054)
Product Market Competition			
AI Postings$_{j,q}$	-0.357*	-0.531**	0.138
	(0.192)	(0.198)	(0.225)
Herfindahl Index$_{j,t-1}$	0.045	0.079	3.554**
	(0.383)	(0.382)	(1.398)
Industry Adjusted Sales Growth$_{i,q,q-4}$	-0.005	-0.013	-0.001
	(0.036)	(0.036)	(0.036)

Table 3.5: What Drives Analyst Interest in Digitization? (Continued)

Managerial and Peer Analyst Digital Disclosure			
Managerial Disclosure$_{i,q}$	1.042***	1.036***	0.998***
	(0.088)	(0.087)	(0.084)
Digital Questions$_{j,q}$	2.962***	2.913***	1.052***
	(0.412)	(0.408)	(0.383)
Managerial Disclosure$_{j,q}$	0.814***	0.766***	0.311*
	(0.135)	(0.146)	(0.162)
Calendar Quarter FE	No	Yes	Yes
4-Digit GICS FE	No	No	Yes
Observations	37,659	37,659	37,659
R^2	0.145	0.148	0.163

This table reports the determinants of digital questions. The first factor, technology suitability, is measured by the extent of textual overlap between AI patent text and the textual description of the firm's NAICS industry. The second factor, financing and economic risks, is proxied by industry-adjusted Tobin's Q and the Whited-Wu capital constraints index. The third factor, governance attributes, is proxied by managerial and board-level experience in digital technologies. The fourth factor, organizational factors, is proxied by the number of segments, firm age and an indicator for late lifecycle firms. The fifth factor, product market competition, is measured with the adoption rate of AI technology at the 6-digit GICS industry-level, industry-adjusted sales growth and the herfindahl index of the same industry. Lastly, the sixth factor, managerial and peer analysts' digital disclosures, is measured with an indicator for managerial disclosure of digital topics, industry-level proportion of digital questions from analysts of peer firms and digital disclosures from managers of peer firms. Finally, several variables measuring the operating profile of the firms are added as controls, namely, market beta, industry-adjusted ROA, SG&A expenditure, R&D expenditure (also includes an indicator for missing R&D) and the proportion of major customer sales. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

and industry-adjusted Tobin's Q exhibit statistically insignificant relationship with digital questions posed by analysts.

Turning to the execution expertise variables in the following two rows, there is some evidence that analysts ask more digital questions when firms have managers or board members with digital experience. Estimates from the industry fixed effects model suggest that firms with such managers exhibit a 0.53% higher proportion of digital questions (or a 66% increase relative to the average proportion of questions). Furthermore, firms with board members that have digital-related experience exhibit a 0.18% higher proportion of digital questions (or a 22% increase relative to the average proportion of questions). Thus, these results suggest that the extent of digital-related experience within the firm is an important determinant of whether an analyst takes interest in a firm's digitization activity.

On the other hand, analysts do not appear to respond to the organizational frictions within the firm by asking more digital questions on the conference calls. Across the variables measuring organizational complexity (number of segments) and legacy structures (age and the indicator for late lifecycle), there is limited evidence (with the exception of the number of segments) that analyst questions are driven by these factors.

There is also limited evidence that product market competition drives more digital questions. In fact, for the industry-level AI postings, this variable is negatively related to digital questions in the quarter fixed effect model. One interpretation of this result, is that analysts are pushing firms to be a first-mover in AI adoption when adoption rates are low in the industry[33]. Due to the rapid scalability of digital and AI technologies (Babina et al. 2020), firms can quickly build-up market share if they are the first to adopt these technologies. Thus, analysts could be aware of the strong first-mover advantages and are therefore posing more questions when the firm has an opportunity to be an early AI adopter in its industry.

The last three rows of Table 3.5, examines the role of managerial and industry-level analyst disclosure on digital topics in shaping individual analysts' interest in the digitization of non-IT firms. The results show that both sources of disclosure are significantly associated with greater digital questions. In the industry fixed effects model, firms with managerial disclosure on digital topics, exhibit a 1% increase in digital questions (or a 140% increase relative to the average proportion of questions). Moreover, a 1 σ increase in the proportion of peer firms with digital questions from analysts is associated with a 0.18% increase in the proportion of digital questions at an individual firm's conference call (or a 36% increase relative to the average proportion of questions). And a 1 σ increase in the proportion of peer firms with managerial digital disclosure is associated with a 0.09% increase in the proportion of digital questions at an individual firm's conference call (or a 18% increase relative to the average proportion of questions).

Next, I examine the determinants of the sentiment of the digital questions in Table 3.6, with the same set of determinants and fixed effects structure. I find that the sentiment of the digital question

33. This implies that the negative relationship should emerge when firms in industry are under-investing in AI, and untabulated analysis show that the negative association is indeed only statistically significant in the sub-sample with abnormally-low AI investment (See Section 4.1.3 for details on the sub-sample formation).

exhibits a positive but generally, statistically insignificant relationship with the AI suitability score. On the other hand, the sentiment of the digital questions exhibits a significant positive association with lower financing and economic risks, as lower values of the capital constraint index and higher Tobin's Q are associated with more positive sentiment in the digital questions. In addition, execution expertise is also an important determinant of positive sentiment in the digital questions. Specifically, firms with some degree of managerial and board experience in digital technologies are associated with significantly more positive sentiment in the digital questions.

Table 3.6: What Drives Analyst Sentiment on Digitization?

Dependent Variable	% Digital Sentiment	% Digital Sentiment	% Digital Sentiment
Technology Suitability			
AI Suitability$_{l,t-1}$	0.008*	0.005	0.005
	(0.004)	(0.004)	(0.005)
Financing and Economic Risks			
Capital Constraints Index$_{i,q}$	0.004	0.004	-0.012*
	(0.005)	(0.005)	(0.006)
Industry Adjusted Tobin's Q$_{i,q}$	0.006***	0.006***	0.007***
	(0.002)	(0.002)	(0.002)
Execution Expertise			
Managerial Experience$_{i,t-1}$	0.016*	0.015*	0.004
	(0.009)	(0.009)	(0.009)
Board Experience$_{i,t-1}$	0.017***	0.017***	0.009***
	(0.004)	(0.004)	(0.003)
Organizational Frictions			
Number of Segments$_{i,t-1}$	-0.003**	-0.003**	-0.002
	(0.001)	(0.001)	(0.001)
Logarithm of Age$_{i,q}$	-0.000	-0.001	-0.002
	(0.003)	(0.003)	(0.003)
Proportion of Major Customer Sales$_{i,t-1}$	-0.023***	-0.023***	-0.008
	(0.005)	(0.005)	(0.006)
Late Lifecycle$_{i,t-1}$	0.012**	0.012**	0.010*
	(0.005)	(0.005)	(0.005)
Product Market Competition			
AI Postings$_{j,q}$	0.020*	0.015	-0.004
	(0.010)	(0.011)	(0.016)
Herfindahl Index$_{j,t-1}$	-0.049*	-0.053**	-0.132
	(0.026)	(0.026)	(0.089)
Industry Adjusted Sales Growth$_{i,q,q-4}$	-0.000	-0.000	0.003
	(0.002)	(0.002)	(0.003)

Table 3.6: What Drives Analyst Sentiment on Digitization? (Continued)

Managerial and Peer Analyst Digital Disclosure			
Managerial Sentiment$_{i,q}$	0.066***	0.065***	0.056***
	(0.005)	(0.005)	(0.005)
Digital Sentiment$_{j,q}$	0.006	0.004	-0.002
	(0.004)	(0.004)	(0.004)
Managerial Sentiment$_{j,q}$	-0.000	-0.002	-0.005
	(0.003)	(0.004)	(0.004)
Calendar Quarter FE	No	Yes	Yes
6-Digit GICS FE	No	No	Yes
Observations	37,659	37,659	37,659
R^2	0.031	0.032	0.042

This table reports the determinants of the sentiment of digital questions, as measured using the FinBert-derived sentiment scores (A. Huang et al. 2020). The first factor, technology suitability, is measured by the extent of textual overlap between AI patent text and the textual description of the firm's NAICS industry. The second factor, financing and economic risks, is proxied by industry-adjusted Tobin's Q and the Whited-Wu capital constraints index. The third factor, governance attributes, is proxied by managerial and board-level experience in digital technologies. The fourth factor, organizational factors, is proxied by the number of segments, firm age and an indicator for late lifecycle firms. The fifth factor, product market competition, is measured with the adoption rate of AI technology at the 6-digit GICS industry-level, industry-adjusted sales growth and the herfindahl index of the same industry. Lastly, the sixth factor, managerial and peer analysts' digital sentiment, is measured with the sentiment score of managerial disclosure of digital topics, industry-level proportion of digital questions from analysts of peer firms and digital disclosures from managers of peer firms. Finally, several variables measuring the operating profile of the firms are added as controls, namely, market beta, industry-adjusted ROA, SG&A expenditure, R&D expenditure (also includes an indicator for missing R&D) and the proportion of major customer sales. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

For the organizational friction variables, the sentiment of the digital questions appears to be associated with firms with lower organizational frictions, due to lower number of segment and lower proportion of major customer sales. On the other hand, I find somewhat surprisingly, that firms in the later stage of their lifecycle are more likely to face digital questions with more positive sentiment. One reason that could explain this finding is that analysts do not view organizational frictions as a particular salient issue in the adoption of digital technologies. In fact, the discussion in analyst reports, suggests that analysts view legacy assets-in-place as a position of strength for non-IT firms seeking to adopt these technologies. For instance, analyst reports on Walmart's digital investments suggest that Walmart can utilize digital technologies more effectively due to its existing assets-in-place (Chen, Rakhlenko, and Kim 2017). These reports, while highlighting the potential complementarities between digital technologies and existing assets such as physical stores and employees, do not discuss the costs of organizational frictions. Thus, the absence of these types of discussions in analyst reports adds further support to the claim that analysts pay limited attention to organizational frictions when deciding to opine on digital-related topics.

Turning to the product market competition and disclosure variables in the following last few rows, I find some evidence that the herfindahl index is negatively associated with the digital question sentiment. This result suggests that analysts are more positive about digital technologies, when the firm is operating in a more competitive industry.

Lastly, I examine the relationship between the sentiment of alternative sources of disclosure on the sentiment of the digital questions. My analysis show that the sentiment of analysts' questions appears to partly respond to managerial sentiment on the conference call as a 10% increase in managerial sentiment is associated with a 0.6% increase in sentiment. On the other hand, the sentiment of the digital questions does not appear to respond to the digital sentiment of either the peer analysts or managers.

Abnormal Levels of AI Investments and Disclosure Dynamics

As indicated in the preceding analysis, the level of digital questions varies by whether the industry is exhibiting low or high levels of AI investment. Arguably, this result suggests that how analysts pose questions could vary when firms/industries are exhibiting low or high levels of investment. For instance, the negative relationship between AI adoption at the industry-level and digital questions suggest that when firms are investing in AI at abnormally low levels, analysts could be asking questions that help nudge firms towards greater investments. On the other hand, when firms are investing in AI at abnormally high levels, analysts could either be joining in the euphoria or they could start becoming more negative in an attempt to curtail excessive investments.

The following analyses, unpacks how analysts' questions vary when firms investing at low or high-levels, by studying the variation in disclosure dynamics (i.e manager's discussion of digital topics and analysts' question responses) when firms are either investing at abnormally high, low or at normal levels. To partition the sample into abnormally low, high and normal investment samples, residuals of AI investment are first computed relative to the benchmark model of AI investment in column two of Table 3.4. This benchmark model assumes that normal investment in AI is based on the five investment factors, controls, and quarter fixed effects[34][35]. Next, residuals within 1 σ of the mean in each quarter are classified as normal investment, while residuals that are above (below) 0.5 σs from the mean are labelled as abnormally high (low) investment[36].

Analysis on the various topics of managerial and analyst disclosure when firms are investing at abnormally high or low investment, reveals some intriguing empirical patterns. Panel A of Table 3.7 reports the differences in managerial disclosure across topics and sentiment (measured using sentences in the presentation portion of the conference call with digital terms; for more details see variable definitions in Appendix B.1). The results show that managers tend to discuss more digital topics when firms are investing in AI at abnormally low or high levels. Notably, the comparison of the probability of disclosure between sub-samples of low and high levels of investment suggests that there is a 50% greater likelihood of discussion of digital topics when firms are exhibiting abnormally low or high investment.

Managers also tend to be more positive on digital topics when firms are investing at abnormally high levels. In particular, the comparisons of the average percentage of positive sentences (following the FinBert sentiment model in, A. Huang et al. 2020) show that mangers are 8% more positive in the discussion of digital topics when firms are investing at abnormally high levels. Moreover, managers also mention more words relating to specific digital technologies, which suggests that managers are optimistic about the prospect of specific technologies that they are investing in the firm.

34. Quarter fixed effect are used to address the overall increase in AI adoption over time (see Figure 3.3).

35. Additional robustness analysis, in Appendix Tables B.3 and B.4, show that results are robust to residuals from a panel OLS model and a industry-quarter fixed effects model.

36. Under normality assumptions, the choice of a 1 σ bandwidth assigns two-thirds of the observations as normal investment. Robustness analysis, reported in Appendix Tables B.5 and B.6 show that the inferences are not sensitive to this choice of bandwidth, as results are robust to bandwidths of 0.5 and 1.5 σs.

Table 3.7: What Do Analysts and Managers Say When Firms Are Investing in AI At Abnormally Low or High Levels?

Sub-Sample	Abn. Low	Normal	Abn. High	Abn. Low Minus Normal	Abn. High Minus Normal
Panel A: Managerial Disclosure of Digital Topics					
% Disclosure	43.326***	29.321***	46.228***	14.005***	16.907***
	(2.265)	(1.701)	(2.599)	(1.668)	(1.912)
% Sentiment	75.416***	72.016***	78.542***	3.400**	6.526***
	(1.436)	(1.111)	(1.317)	(1.287)	(1.404)
% Positive Sentences	79.692***	76.241***	82.769***	3.451***	6.528***
	(1.196)	(0.975)	(1.045)	(1.123)	(1.166)
% Negative Sentences	4.276***	4.225***	4.227***	0.051	0.002
	(0.468)	(0.326)	(0.526)	(0.469)	(0.498)
% Performance Words	0.999***	1.072***	0.914***	-0.073*	-0.158***
	(0.035)	(0.026)	(0.037)	(0.041)	(0.039)
% Competition Words	0.073***	0.054***	0.078***	0.019**	0.023*
	(0.007)	(0.005)	(0.011)	(0.009)	(0.012)
% Technology Words	1.972***	2.382***	2.363***	-0.410***	-0.019
	(0.088)	(0.076)	(0.112)	(0.091)	(0.098)
Panel B: Analysts' Digital Questions					
% Disclosure	15.387***	8.382***	15.765***	7.005***	7.383***
	(1.286)	(0.686)	(1.543)	(1.079)	(1.299)
% Sentiment	32.518***	31.494***	28.265***	1.024	-3.229
	(2.095)	(1.887)	(2.217)	(2.606)	(2.657)
% Positive Sentences	47.122***	45.595***	46.084***	1.527	0.488
	(1.726)	(1.269)	(1.950)	(2.061)	(2.179)
% Negative Sentences	14.604***	14.102***	17.819***	0.502	3.717***
	(0.892)	(1.028)	(1.192)	(1.168)	(1.320)
% Performance Questions	21.316***	25.218***	25.887***	-3.902*	0.669
	(1.709)	(1.504)	(2.481)	(2.048)	(2.645)
% Competition Questions	3.753***	3.601***	7.839***	0.152	4.238***
	(0.762)	(0.456)	(1.314)	(0.767)	(1.359)
% Technology Questions	46.715***	57.825***	64.086***	-11.110***	6.262
	(3.253)	(2.853)	(4.543)	(3.695)	(4.823)

This table reports the types of disclosures that analysts and managers make when firms are investing in AI at abnormally low or high levels. columns 1-3 reports the average percentage of disclosure, net sentiment, positive sentences, negative sentences, performance (words or questions), competition (words or questions) or technology (words or questions), for the sub-samples of abnormally low, normal and abnormally high AI investment levels. Columns 4 and 5, presents the comparison of the means between sub-samples of abnormally low and normal levels of AI investment, and sub-samples of abnormally high and normal levels of AI investments. Panel A reports the analysis for managerial disclosure of digital sentences (i.e. sentences with digital terms) in the presentation of the earnings conference calls. Panel B reports the analysis for digital questions in the Q&A section of the earnings conference calls. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

To study how analysts respond to firms' AI investment levels, Panel B present the differences in analysts' questions across topics and sentiment. The results show that analysts also respond to abnormal investment levels, as analysts are roughly 90% more likely to discuss of digital topics when firms are either investing at abnormally high or low levels.

On the other hand, analysts do not appear to share the more positive sentiment that managers show when firms are investing at abnormally high levels. In fact, there is some evidence of analysts monitoring managers through negative questioning in these situations. In particular, the differences in means suggest that analysts are 26% more negative when firms are investing in AI at abnormally high levels. Furthermore, consistent with analysts' concerns over the levels of investments relative to peers, analysts also appear to be asking questions related to benchmarks in these situations. Specifically, the comparison of the average competition-related questions across sub-samples show that these questions are 117% more likely to be posed when firms are investing in AI at abnormally high levels.

Taken together, the topic analysis across sub-samples yields the following insights. First, both managers and analysts are more active in discussing digital topics when firms are either investing at abnormally low or high levels. Second, managers tend to be more optimistic about the technology when firms are investing at abnormally high levels. On the other hand, consistent with analysts playing a monitoring role, analysts are more negative and engage in more bench-marking questions when firms are investing at abnormally high-levels.

3.3.2 Analysts' Interest in Digitization and Future AI Investment

Motivated by the discussion in Section 2.4, the next set of analyses examines whether analysts' interest in non-IT firms' digitization activities is related to more or less future investment in advanced digital technologies (AI).

The research design of these tests follows a similar firm-fixed effects framework as shown in previous regression models. Specifically, this model regresses 1-year ahead AI job posting, as measured by an indicator or the intensity, on the proportion of digital questions:

$$AI\ Posting_{i,q+4} = \alpha_i + \alpha_q + \beta_1 Questions_{i,q} + \sum_s \gamma_s Y_{i,s,q} + \varepsilon_{i,q} \tag{3.3}$$

where the coefficient of interest is β_1 which estimates the association between digital questions ($Questions_{i,q}$) and future investment in AI technologies ($AI\ Posting_{i,q+4}$). Also included in the regressions are the variables outlined in equation 2 ($\sum_s Y_{i,s,q}$), as well as firm and quarter fixed effects. In addition, standard errors are clustered at the firm and quarter-level.

Table 3.8 reports the results of the analysis from the above regression model, for the future AI job posting indicator in Panel A and the AI job posting intensity in Panel B. Across columns 1-6, the regressions are implemented for the full set of digital questions, and the questions on performance, competition, and technology, as well as positive and negative digital questions. Column 1 of Panel A shows that analysts' interest in the digitization of non-IT firms, is positively associated with greater future investment in advanced digital technologies (AI). Estimates from this regression suggest that a 1 σ increase in the proportion of digital questions is associated with a 1.1% increase in the probability of a future AI job posting (or 5% relative to the average probability of AI postings)[37]. The results for the AI job posting intensity measure paints a similar picture. Estimates from this regression show that a 1 σ increase in the proportion of digital questions is associated with a 0.05% increase in the intensity of AI job postings (or 10% relative to the average intensity of AI postings).

The following three columns performs the same analysis on the sub-group of digital questions. In both panels, the results in these columns continue to show that digital questions are positively

37. The reader might be concerned with the number of observations in this analysis — 29,518, compared to 37,659 in the determinants analysis. The low number of observations is due to periodic gaps in the *Burning-Glass* data at the quarterly level. To address concerns that this feature of the data might confound interpretation of the results, I examine the robustness of the main findings to an alternative specification that assumes missing future investment observations as 0. This analysis presented in Appendix Table B.7. shows that results are robust to this specification as well.

Table 3.8: Is Analyst Interest in Digitization Associated with Greater AI Investment?

Panel A: Probability of Future AI Job Postings					
Dependent Variable	1-Year Ahead AI Posting (Indicator)	1-Year Ahead AI Posting (Indicator)	1-Year Ahead AI Posting (Indicator)	1-Year Ahead AI Posting (Indicator)	1-Year Ahead AI Posting (Indicator)
Digital Questions$_{i,q}$	0.351***				
	(0.122)				
Performance Questions$_{i,q}$		1.071***			
		(0.374)			
Competition Questions$_{i,q}$			2.139*		
			(1.087)		
Technology Questions$_{i,q}$				0.430*	
				(0.225)	
Sentiment$_{i,q}$					0.020**
					(0.010)
Firm FE	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	29,518	29,518	29,518	29,518	29,518
R^2	0.514	0.514	0.513	0.514	0.513

Panel B: Intensity of Future AI Job Postings					
Dependent Variable	1-Year Ahead AI Posting Intensity	1-Year Ahead AI Posting Intensity	1-Year Ahead AI Posting Intensity	1-Year Ahead AI Posting Intensity	1-Year Ahead AI Posting Intensity
Digital Questions$_{i,q}$	0.016**				
	(0.007)				
Performance Questions$_{i,q}$		0.052***			
		(0.018)			
Competition Questions$_{i,q}$			0.119***		
			(0.044)		
Technology Questions$_{i,q}$				0.029**	
				(0.012)	
Sentiment$_{i,q}$					0.001***
					(0.000)
Firm FE	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	29,518	29,518	29,518	29,518	29,518
R^2	0.409	0.409	0.408	0.409	0.408

This table reports the relationship between analysts' digital questions and future job postings of AI conditional on analysts' digital questions. The regressions in Panel A, regresses an indicator for 1-year ahead AI-related job postings, on either the proportion of digital questions posed by analysts, the proportion of performance-related questions, the proportion of competition related questions, the proportion of technology-related questions, the proportion of positive and negative sentiment in the digital questions. Regressions in Panel B performs the same analysis but with the intensity of AI job postings as the dependent variable. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Firm and calendar quarter fixed effects are also included. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

associated with future AI investment. For digital questions on performance, the estimates indicate that a 1 σ increase in the proportion of these questions increase the probability of future AI postings by 0.9% (or 4% relative to the average probability of AI postings) and the intensity of AI posting by 0.04% (or 8% relative to the average intensity of AI postings). Similarly, for digital questions on competition, the estimates indicate that a 1 σ increase in the proportion of these questions increase the probability of future AI postings by 0.32% (or 1% relative to the average probability of AI postings) and the intensity of AI posting by 0.02% (or 4% relative to the average intensity of AI postings). For digital questions on technologies, the estimates indicate that a 1 σ increase in the proportion of these questions increases the probability of future AI postings by 0.62% (or 3% relative to the average probability of AI postings) and the intensity of AI posting by 0.04% (or 8% relative to the average intensity of AI postings).

Lastly, I examine whether the sentiment of digital questions from analysts are also associated with greater AI investments. The results in the last column of Panel A of Table 3.8 suggest that greater positive sentiment is associated with more future AI investments. Estimates suggest that a 1 σ increase in the sentiment of question sis associated with a 0.4% increase in 1-year ahead AI posting probability (or 2% relative to the average probability of AI postings) and a 0.02% increase in 1-year ahead AI posting intensity (or 4% relative to the average intensity of AI postings).

Are Analysts Actively Encouraging Firms to Invest More in AI?

The previous set of analyses presented broad evidence of a positive association between analysts' digital questions and future AI job postings. However, the direction of this relationship is unclear as there are two competing interpretations. One, that analysts are actively encouraging companies to increase AI investments. Two, that analysts discuss more about digital topics when managers are signaling their intent to invest in AI technologies.

To help disentangle the two interpretations, I study the association between digital questions and future AI job-postings in three sub-samples, split by — (1) whether the firm is actively engaging in AI investments and (2) whether the firm has made disclosures relating to digitization. Panel A of Table 3.9 shows that digital questions tend to significantly predict future AI job postings only when the firm does not engage in current AI investment. Moreover, the positive association is significant mostly in sub-samples where there are no firm disclosures on digital topics. Specifically, Panel B of Table 3.9, shows that the positive association between the digital questions and the probability of future AI investment is significant and larger in the sub-sample with no firm disclosure of digital topics.

Taken together, these findings show that the positive relationship between analysts' interest in digitization and future AI investment is stronger when there are limited indications that the firm is initiating such investment on its own. Thus, these analyses suggest that analysts are playing an active role in encouraging firms to invest more in AI.

When Do Analysts Encourage Firms to Invest More in AI?

The discussion in Section 2.4, argues that the positive association between analysts' interest in digitization and future AI investment is consistent with two hypotheses. One, that analysts are encouraging firms to invest more efficiently in these technologies (H2b). Two, that analysts are overly-optimistic about the new technologies and are driving a frenzy that compels firms to over-invest in AI (H2c). In the following analyses, I present some cross-sectional analyses that aims to disentangle these two hypotheses.

These cross-sectional analyses study the association between analysts' interest in digitization and AI investment in sub-samples of abnormally low, high and normal investment. Following the approach outlined in Section 4.2.3, samples where firms are investing at abnormally low or high levels are created by splitting the main sample with the residuals from a benchmark AI investment model[38].

38. Additional robustness analysis, presented in Appendix Tables B.3-B.6, show that the results are robust to different regression model specifications and bandwidth cut-offs for sub-sample formation.

Table 3.9: Are Analysts Actively Encouraging Firms to Invest More in AI? Cross-Sectional Evidence

Panel A: Current AI Job Postings				
Sample	No AI Job Postings		AI Job Postings	
Dependent Variable (1-Year Ahead)	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity
Digital Questions$_{i,q}$	0.343**	0.012*	0.253	0.013
	(0.154)	(0.007)	(0.178)	(0.009)
Firm FE	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	22,254	22,254	7,128	7,128
R^2	0.315	0.248	0.413	0.529

Panel B: Firm Disclosure				
Sample	No Firm Disclosure		Firm Disclosure	
Dependent Variable (1-Year Ahead)	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity
Digital Questions$_{i,q}$	0.533*	0.024*	0.219*	0.011
	(0.304)	(0.013)	(0.123)	(0.007)
Firm FE	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	15,938	15,938	11,290	11,290
R^2	0.490	0.356	0.544	0.467

This table reports the sub-sample analysis of the relationship between analysts' digital questions and future job postings of AI conditional on analysts' digital questions. All regressions use the proportion of digital questions as the main independent variable. Regressions in Panel A regresses either the indicator or intensity of 1-year ahead AI-related job postings on the proportion of digital questions in sub-samples of observations with current AI-related job postings and no current AI-related job postings. Regressions in Panel B perform the same analysis but with partitions along whether the firm discloses digital topics, either in the presentation of the conference call, the business description, MD&A of the 10-K or through product-announcements. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

Table 3.10: When Is Analyst Interest in Digitization Associated with Greater AI Investment?

Sub-Sample	Abn. Low-Investment		Normal-Investment		Abn. High-Investment	
Dependent Variable	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity
Digital Questions$_{i,q}$	0.450**	0.025**	0.294**	-0.000	-0.053	0.024
	(0.206)	(0.009)	(0.134)	(0.004)	(0.181)	(0.016)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,407	5,407	20,623	20,623	2,995	2,995
R^2	0.464	0.315	0.514	0.304	0.503	0.532

This table reports the sub-sample analysis of the relationship between analysts' digital questions and future job postings of AI conditional on the current level of AI investment. The sample is split by whether the firm is investing in AI at abnormally low, high or normal-levels. Regressions regresses either the indicator or intensity of 1-year ahead AI-related job postings on the proportion of digital questions in the three sub-samples. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Panel C studies the interaction effect of the current abnormal investment level (coded as -1 for abnormally low investment, 1 for abnormally high investment and 0 otherwise) and digital questions or positive/negative sentiment, on future AI job posting probability in the main sample. In addition to the baseline set of controls, regressions in this panel also include controls for the abnormal investment level, and the digital questions, or positive/negative sentiment respectively. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

With these sub-samples, Table 3.10 shows consistent evidence that analysts' digital questions tend to be associated with higher future probability and intensity of AI investment when firms are investing in AI at abnormally low levels. Thus, the evidence in this analysis suggests that analysts are encouraging firms to increase future AI investment when the current level of AI investment is too low.

Overall, these analyses further suggests that analysts are playing a positive governance role in shaping AI investments, as analysts encourage companies to increase future investment in AI, mostly when they exhibit lower investment relative to peers. Thus, these analyses also suggest that analysts are not driving over-investment in AI (H2c).

3.3.3 Discussion: The Relationship Between Digital Questions and Future AI Investments

The main result in the preceding section is this positive association between the digital questions and future AI investments. In this section, I discuss the various economic mechanisms that might explain this relationship, as well as the various endogeneity concerns that may confound the interpretation of this main result.

Potential Economic Mechanisms

To flesh out them claim that analysts are encouraging managers to invest more in advanced digital technologies, I propose two economic channels that ties analysts' views on digital technologies and managers investing in advanced digital technologies.

First, analysts could also encourage managers to invest more in AI by providing insights on the broader market views on digitization. Analysts serve as a conduit for market views as they aggregate information from other market participants, and write opinions that are widely followed in capital markets. Recent surveys, for example, finds that analysts occasionally ask questions on conference calls on behalf of investors (Brown et al. 2015). Moreover, studies also show that investors act upon the views of analysts (Beneish 1991; Womack 1996). Thus, one might expect that the digital questions measured in this study, to some extent, proxies for the views of the broader set of market participants. This in turn suggests that analysts could also be encouraging managers to invest more in AI, by providing some insight on the markets' view on digitization.

To examine this economic channel, I perform two set of analyses. First, I extend the main sample into an analyst-firm-quarter panel, to study the relationship between digital questions and future AI investments, while holding analyst characteristics fixed. This analysis controls for heterogeneity in analysts' views on the new technologies and allows the researcher to isolate the effects of the aggregate

Table 3.11: Aggregate Market Views and AI Investments

Panel A: AI Posting Probability			
Dependent Variable (1-Year Ahead)	Job Posting Probability	Job Posting Probability	Job Posting Probability
Digital Questions$_{i,t}$	0.033**	0.030*	0.028*
	(0.016)	(0.016)	(0.015)
Firm FE	Yes	Yes	Yes
Analyst FE	No	Yes	No
Analyst-Broker FE	No	No	Yes
Calendar Quarter FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	196,875	194,245	187,183
R^2	0.524	0.548	0.562

Panel B: AI Posting Intensity			
Dependent Variable (1-Year Ahead)	Job Posting Intensity	Job Posting Intensity	Job Posting Intensity
Digital Questions$_{i,t}$	0.002*	0.002*	0.001*
	(0.001)	(0.001)	(0.001)
Firm FE	Yes	Yes	Yes
Analyst FE	No	Yes	No
Analyst-Broker FE	No	No	Yes
Calendar Quarter FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	196,875	194,245	187,183
R^2	0.450	0.509	0.529

This table reports the relationship between analysts' digital questions and future job postings of AI conditional on analysts' digital questions, for the analysis on the analyst-firm-quarter panel. The regressions in Panel A, regresses an indicator for 1-year ahead AI-related job postings, on the proportion of digital questions posed by analysts. Regressions in Panel B performs the same analysis but with the intensity of AI job postings as the dependent variable. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Firm and calendar quarter fixed effects are also included across all regression models. Analyst and Analyst-Broker FE are also included in models 2 and 3. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

analyst or market views on the technology and future AI investments. These results are presented in Table 3.11, and shows consistent results that after controlling for analyst or analyst-broker fixed effects, the digital questions continue to exhibit a positive and significant association with future AI job postings. Thus, this result suggests that aggregate market views or sentiment on digital technologies could be driving companies to invest more in AI technologies.

Second, to further examine the hypothesis that aggregate market views drive future AI investments, I examine whether markets reward companies for investing in new AI technologies when they face more digital-related questions from the analyst community. This analysis is implemented by studying the one-year ahead changes in industry-adjusted Tobin's Q in the following interactions model:

$$\begin{aligned} Q_{i,q+4} = \alpha_i + \alpha_q + \beta_0 AI\,Posting_{i,q+4} \times Questions_{i,q} + \beta_1 AI\,Posting_{i,q+4} \\ + \beta_2 Questions_{i,q} + \textstyle\sum_s \gamma_s Y_{i,s,q} + \varepsilon_{i,t} \end{aligned} \tag{3.4}$$

where the coefficient of interest is β_0 which measures the incremental effect on changes in one-year ahead Tobin's Q for firms that face digital questions in time q and invest in AI technologies in the year ahead $(q + 4)$.

The analysis with this regression model is presented in Table 3.12. I find that for both the indicator and intensity of AI job postings, companies that invest in one-year ahead AI technologies and had received digital questions from analysts tend to exhibit higher changes in 1-year ahead Tobin's Q. Specifically, firms that exhibit a 1 σ increase in digital questions and posts an AI job posting in the year after, exhibits a 0.05 increase in Tobin's Q. On the other hand, firms that faced digital questions and had not invested in AI technologies, did not experience significant changes in Tobin's Q in the year ahead. Thus, this set of analysis suggests that markets reward firms for engaging in AI investments that are also linked with prior digital-related questions from analysts.

Table 3.12: Market Responses to AI Investment and Analyst Interest in Digitization

Dependent Variable (1-Year Ahead)	Industry-Adjusted Tobin's Q	Industry-Adjusted Tobin's Q
AI Investment Variable (1-Year Ahead)	Job Posting Indicator	Job Posting Intensity
AI Investment$_{i,1+4}$ × Digital Questions$_{i,q}$	1.532**	46.120***
	(0.609)	(16.664)
Digital Questions$_{i,q}$	0.080	0.316
	(0.341)	(0.348)
AI Investment$_{i,q+4}$	-0.008	0.201
	(0.020)	(0.691)
Firm FE	Yes	Yes
Calendar Quarter FE	Yes	Yes
Controls	Yes	Yes
Observations	29,410	29,410
R^2	0.773	0.773

This table reports the relationship between 1-year ahead industry-adjusted Tobin's Q and the interaction between 1-year ahead AI investment and current analysts' questions on digitization. The regressions regresses industry-adjusted Tobin's Q on the proportion of digital questions posed by analysts and an interaction with 1-year ahead AI investment as measured by either the intensity or proportion of AI job postings. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Firm and calendar quarter fixed effects are also included. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

Another potential economic channel that could tie analyst questions on digitization with higher AI investments, is information transfer from analysts to managers on the conference call. In the Q&A discussion on the earnings call, analysts could provide information relating to the competitive environment relating to these technologies, or could provide some expertise relating to these technologies. To investigate this channel, I first turn to the analyst-firm-quarter panel, and examine whether the main results are driven by analysts' with some degree of technology expertise. Using the names of the analysts, I identify analysts with technology expertise (or technology analysts) as those that cover technology companies in the same calendar quarter. I then examine the main regression specification but with separate independent variables for digital questions posed by technology and non-technology analysts respectively[39]. These results presented in Table 3.13 show some evidence that technology analysts ask digital questions that are associated with higher future AI investments. Specifically, in the regression model with AI job posting intensity as the dependent variable, digital questions posed by technology analysts are associated with higher future AI investments, while those posed by non-technology analysts are not associated with changes in future AI investments.

In addition, to further highlight the possibility that analysts could be providing expertise that help firms invest more in AI technologies, I examine cross-sections of the main sample based on whether the firm's executive team has some experience in digital technologies. This analysis presented in Panel A of Table 3.14, shows that digital questions tend to be positively associated with future AI investment, only for firms that do not have top executives with digital technology experience. This result suggests that analysts play a bigger role in encouraging companies to invest more in AI technologies, when management lack the expertise in digital technologies. Moreover, I find further supportive evidence that this result holds true particularly in cases where the board has some expertise in digital technologies. As shown in Panel B of Table 3.14, digital questions are positively associated with future AI investment, mostly in situations where the management team lacks digital technology expertise, while the

39. These technology analysts form roughly 14% of all analyst-firm-quarter observations

Table 3.13: Technology Analysts and AI Investments

Dependent Variable (1-Year Ahead)	Job Posting Probability	Job Posting Intensity
Tech Analyst Digital Sentiment$_{i,t}$	0.035	0.004**
	(0.028)	(0.002)
Non-Tech Analyst Digital Sentiment$_{i,t}$	0.030	0.001
	(0.021)	(0.001)
Firm FE	Yes	Yes
Calendar Quarter FE	Yes	Yes
Controls	Yes	Yes
Observations	183,175	183,175
R^2	0.528	0.454

This table reports the sub-sample analysis of the relationship between analysts' digital questions and future job postings of AI conditional on analysts' digital questions. All regressions use the proportion of digital questions from technology analysts and non-technology analysts as the main independent variables. Technology analysts are defined as analysts that also cover a technology firm in the same calendar quarter. The first (and second) regression model regresses the indicator (and intensity) of 1-year ahead AI-related job postings. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

board has some expertise in this area. this result suggests that analysts play a role in helping boards monitor AI investments, when management teams lack digital technology experience.

Endogeneity Concerns

One of the main endogeneity concerns is the measurement error in the proxy for AI investment. One source of error is in the timing of future AI investment. The horizon of one-year ahead investment was chosen as a sufficient time lag was needed to mitigate the concern that the digital questions were describing existing AI investments in firms. To address concerns that the time lag of one-year might be an empirical choice that drove a spurious correlation between the digital questions and future AI investments. I show that the main results are robust to time horizons of 1-4 quarters (See Table 3.15). Furthermore, there may also be concerns that the AI job posting measure, may not be fully capturing the concept of AI investment. Firms may choose to invest in these new technologies through acquisitions, and so in Table 3.16 I show that the positive association between digital questions and future

Table 3.14: When do Analysts Provide Digital Technology Expertise?

Panel A: Manager Technology Experience				
Sample	No Manager Tech. Exp.		Manager Tech. Exp.	
Dependent Variable (1-Year Ahead)	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity
Digital Questions$_{i,q}$	0.375**	0.016**	0.053	0.010
	(0.142)	(0.008)	(0.231)	(0.008)
Firm FE	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	27,479	27,479	2,005	2,005
R^2	0.510	0.399	0.567	0.570

Panel B: Manager and Board Technology Experience				
Sample	No Manager but Board Tech. Exp.		Manager and Board Tech. Exp.	
Dependent Variable (1-Year Ahead)	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity
Digital Questions$_{i,q}$	0.574***	0.026**	-0.086	0.010
	(0.208)	(0.013)	(0.280)	(0.010)
Firm FE	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	5,585	5,585	1,154	1,154
R^2	0.570	0.530	0.573	0.560

This table reports the sub-sample analysis of the relationship between analysts' digital questions and future job postings of AI conditional on analysts' digital questions. All regressions use the proportion of digital questions as the main independent variable. Regressions in Panel A regresses either the indicator or intensity of 1-year ahead AI-related job postings on the proportion of digital questions in sub-samples of observations with managerial experience in digital technologies and no managerial experience in digital technologies. Regressions in Panel B perform the same analysis but with partitions along whether there is managerial technology experience in the sub-sample of firms with board experience in digital technologies. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

AI investments is also robust to using technology M&As[40] as a dependent variable.

Table 3.15: Sensitivity Analysis: Varying Horizons of AI Investments

Panel A: AI Posting Probability				
Dependent Variable Horizon	1 Quarter Ahead	2 Quarter Ahead	3 Quarter Ahead	1 Year Ahead
Digital Questions$_{i,q}$	0.220**	0.189	0.356***	0.351***
	(0.101)	(0.114)	(0.118)	(0.122)
Firm FE	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	34,016	32,355	30,826	29,518
R^2	0.506	0.507	0.509	0.514

Panel A: AI Posting Intensity				
Dependent Variable Horizon	1 Quarter Ahead	2 Quarter Ahead	3 Quarter Ahead	1 Year Ahead
Digital Questions$_{i,q}$	0.010**	0.010*	0.010*	0.016**
	(0.005)	(0.006)	(0.005)	(0.007)
Firm FE	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	34,016	32,355	30,826	29,518
R^2	0.392	0.396	0.408	0.409

This table reports the relationship between analysts' digital questions and future job postings of AI conditional on analysts' digital questions, over varying horizons. All regressions use the proportion of digital questions as the main independent variable. Regressions in Panel A regresses the indicator of 1-year ahead AI-related job postings on the proportion of digital questions. Regressions in Panel B regresses the indicator of 1-year ahead AI-related job postings on the proportion of digital questions. Each panel examines the relationship between the digital question and the 1-4 quarter ahead AI postings. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

Another key endogeneity concerns is the potential presence of correlated omitted variables that might confound the reported relationship between digital questions and future AI investments. The inclusion of firm and quarter fixed effects, helps to rule out concerns relating to confounding time-invariant factors and aggregate time trends. The inclusion of the proxies for the five investment factors and the disclosure variables, helps to address some of the concerns relating to within-firm covari-

40. These M&A transactions are identified using the S&P's 451 database on technology acquisitions

Table 3.16: Sensitivity Analysis: Alternative Measures of Advanced Digital Investment

Dependent Variable (1-Year Ahead)	Technology M&A Counts	Technology M&A Indicator	Proportion of Software Workers	Salary-Weighted Proportion of Software Workers
Digital Questions$_{i,q}$	0.284**	0.194**	0.912**	0.954**
	(0.115)	(0.086)	(0.397)	(0.420)
Firm FE	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	29,518	29,518	20,369	20,369
R^2	0.164	0.163	0.948	0.956

This table reports the sub-sample analysis of the relationship between analysts' digital questions and future investment in digital technologies, as measured by technology M&As and software workers. All regressions use the proportion of digital questions as the main independent variable. From left to right, the estimates from regression with the 1-year ahead technology M&A counts, indicators, proportion of software workers and the salary-weighed proportion of software workers are reported. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

ates that might also confound the interpretation of the reported positive relationship. Moreover, the sub-sample analysis on measures of managerial intent on digitization activities (current investment in AI, and firms' digital disclosure) provides further suggestive evidence that the positive relationship is stronger when there is limited managerial intent. Thus, this helps to rule out one of the main potential omitted variable bias – that is, managerial intent on digitization activities could be driving both digital questions and future AI investments.

Nonetheless, there remain concerns over the possible role of within-firm unobservable selection effects in confounding the positive association reported in the preceding analysis. To further address these concerns, I turn to the approach proposed by Oster (2019) to help bound the potential confounding effects[41]. I implement this approach to estimate the lower bound of δ, which is the relative impact of within-firm the unobservable selection (compared to observable selection) that is re-

41. This approach has also been used in several recent accounting and finance papers to estimate bounds on unobservable selection effects (e.g. Ma et al. 2021; Akey et al. 2021).

quired to drive the coefficient on the main independent variable to zero[42]. Table 3.17 reports these estimates. For the regression with AI job posting probability (intensity), the estimates on ∂, given $R_{max} = 1.3 \times R^2$ and $\beta = 0$ is 1.18 (1.90). This suggests that the within-firm unobservable selection effects would have to be at least 1.2 (1.9) times the within-firm observable selection effects to yield a regression coefficient on digital questions that is zero. The magnitudes on ∂ is higher than the rule-of-thumb cutoff of 1 suggested by (Oster 2019) and further suggests that unobservable selection effects play a more limited role in explaining the reported positive association in the preceding analysis.

Table 3.17: Sensitivity Analysis: Estimating the Bounds on the Effect of Unobservable Omitted Variables in the Regressions of Future AI Investment on Digital Questions

	Uncontrolled		Controlled		∂	β
Dependent Variable	β	Within-Firm R^2	β	Within-Firm R^2	Given $R_{max} = 1.3 \times R^2$ and $\beta = 0$	Given $\partial = \pm 1$ and $R_{max} = 1.3 \times R^2$
AI Job Posting Probability	1.230	0.005	0.351	0.074	1.179	[0.054, 0.645]
AI Job Posting Intensity	0.039	0.005	0.016	0.058	1.905	[0.008, 0.024]

This table presents sensitivity analysis of the regression of future AI investment on digital questions (column 1 of Table 7). This analysis follows Oster (2019) and provides an estimate on the relative impact of within-firm unobservables that is needed to drive the regression coefficient on digital questions to zero. In columns 1 and 2, the coefficient on digital questions (β) and within-firm R^2 is presented for the "uncontrolled" which implements the baseline firm fixed effects (within firm changes) regression of future AI job posting probability/intensity on digital questions. In columns 3 and 4, the same estimates are presented for the "controlled" regression, which is the specification seen in column 1 of Table 7. The fifth column reports the ∂ given $R_{max} = 1.3 \times R^2$ (following the empirically-derived suggestion in, Oster 2019), or the relative importance of the unobservables, that is required to yield a β coefficient of zero. Finally, the last column reports the bounds on β based on the assumption that the unobservable selection effects are as important as the observed selection effects ($\partial = \pm 1$), given $R_{max} = 1.3 \times R^2$.

Another key endogeneity concern is that changes in overall analyst activity could lead to spurious correlations between digital questions and future AI investment. To address this concern, I perform a placebo test based on the premise that if overall analyst activity drives the positive relationship, then non-digital questions should also exhibit a positive association with future AI investment. Table

42. To be clear on the methodology and assumptions: (1) I use the firm fixed effects model as the "uncontrolled" specification and the regression model with all controls as the "controlled" specification. This allows me to estimate the relative impact (∂) of within-firm unobservable selection that is needed to drive the main independent variable coefficient to zero. (2) Following Oster (2019), I make an assumption that the maximum R^2 (R_{max}) of the model (including unobservables) is 1.3 times the controlled within-firm R^2.

3.18 investigates this possibility by regressing future AI job posting probability and intensity, on the total number of non-digital questions. The results show that the number of non-digital questions exhibits no statistically significant relationship with AI job posting probability and intensity. Thus, these findings suggest that the positive relationship between analysts' digital questions and future AI investment, is unlikely due to changes in overall analyst activity.

Table 3.18: Placebo Test: Non-Digital Questions and Future AI Investment

Panel A: Full Sample		
Dependent Variable	Job Posting Probability	Job Posting Intensity
Non-Digital Questions$_{i,q}$	0.000	-0.000
	(0.000)	(0.000)
Firm FE	Yes	Yes
Calendar Quarter FE	Yes	Yes
Controls	Yes	Yes
Observations	29,518	29,518
R^2	0.513	0.408
Panel B: Abnormally Low AI Investment Sub-Sample		
Dependent Variable	Job Posting Probability	Job Posting Intensity
Non-Digital Questions$_{i,q}$	0.000	0.000
	(0.001)	(0.000)
Firm FE	Yes	Yes
Calendar Quarter FE	Yes	Yes
Controls	Yes	Yes
Observations	5,407	5,407
R^2	0.463	0.312

This table reports the relationship between analysts' non-digital questions and future job postings of AI conditional on the current level of AI investment. Regressions in this table regresses either the indicator or intensity of 1-year ahead AI-related job postings on the non-digital questions (computed as the difference between the total number of questions in the Q&A of the conference call and the total number of digital questions). Panel A reports the analysis for the full sample, while Panel B and C reports the analysis for the sub-sample of abnormally-low investment and no current AI investment respectively. Regressions also control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

3.3.4 Robustness Analysis

In addition to the main analysis, I also report several robustness analyses to address concerns relating to certain specifications of the study.

One concern related to the analysis based on the sub-samples of abnormally low, high and normal-level AI investments, is the definition of the sub-samples and residuals (i.e. expected investment model). To examine the robustness of the results to alternative specifications of the investment model, I run the main sub-sample analyses with investment models based on a panel OLS and a industry-quarter FE specification. Results, presented in Tables B.3 and B.4 in the Appendix are robust to these specifications. In addition, the main sub-sample results are also robust to alternative bandwidth cut-offs. Specifically, Tables B.5 and B.6 show that the main sub-sample results are also robust to bandwidths of 0.5 and 1.5 σ's.

Next, I address two more concerns relating to the measurement of future AI job postings in this study, in additional robustness analysis. The first concern is due to periodic gaps in the *BurningGlass* coverage. Because of these gaps, the analysis on the future AI job postings is based on a sample that omits missing observations of 1-year ahead AI job postings. These gaps are unlikely to change main inferences, as Appendix Table B.7 show that the main results are robust to interpolating zeros for missing 1-year ahead AI job postings observations. The second concern is on the measurement of AI job posting intensity. This is measured by scaling the variable by the total number of job postings (following Babina et al. 2020; Acemoglu et al. 2020). But one might be concerned that the denominator is a flow, rather than a stock variable. To address this concern, I show in Appendix Table B.8 that the main results are broadly robust to an alternative definition of AI job posting intensity, that is defined with employee counts in *Compustat* as the denominator.

Finally, I conduct additional robustness analysis on the measurement of AI skills. In the main results, the definition of AI skills follows the AI list from Alekseeva et al. (2021), which is developed using BurningGlass classification of AI-related skills. Nonetheless, there is some debate over the list of skills that constitute AI skills. Thus, I examine the robustness of the analysis to another set of AI-skills proposed by Acemoglu et al. (2020), which is a narrower set of AI-related skills. Results presented in Appendix Table B.9 and B.10 show that the main inferences are unchanged with this specification.

3.4 Conclusion

Motivated by financial analysts' growing interest in the digitization of non-IT firms, I examine how analysts evaluate digital investments and how their views on digitization is related to future investments in advanced digital technologies (AI). Using analysts' digital questions in quarterly earnings conference calls as a proxy for their interest in the topic, I address the main research questions by studying the drivers of their interest in digitization and the associations between their interest in this process and future AI job postings.

My findings show that analysts develop interest in digitization based on two drivers of AI investment, namely, whether the firm is in an industry that is suited for AI investments, and whether the firm has managers or board members with experience in digital technologies. Moreover, analysts also show more positive interest in digital technologies when firms exhibit lower economic and financing risks, as measured by proxies of capital constraints and Tobin's Q. Further analysis of the topics in digital questions shows that analysts tend to ask more negative questions on the competitive position of firms relative to peers when firms are investing in AI at abnormally high-levels. Taken together, these results suggest that analysts play a role in both analyzing various sources of information and monitoring firm activities.

On the relationship between analysts' interest in digitization and future AI investment, my results suggest that analysts are encouraging firms to invest more in AI. Regressions of future AI investment on digital questions, show that analysts' questions are associated with more future AI investments.

Moreover, consistent with analysts playing an active role in encouraging firms to increase future AI investments, cross-sectional results show that the positive association is stronger in situations where the firm is not currently investing in AI investment or when managers are not discussing digital topics. Finally, consistent with analysts encouraging firms to make more future AI investments when firms are currently not investing enough in AI, cross-sectional results also show that the positive association is stronger in situations where the current levels of AI investment is abnormally low.

Overall, the evidence in this study suggests that financial analysts play a positive governance role in shaping AI investment decisions. Thus, from a broader perspective, my findings introduce the possibility that capital markets, as a whole, can play an important role in the digitization process. Notably, financial analysts are only one of many institutions in capital markets. Thus, there is room for more research in this area, as future work could further examine how other capital market institutions may play a role in the digitization process.

4

Capital Market Forces and Digital Investment in Non-IT Firms

In recent years, scholars have argued that the new wave of digital technologies, such as AI, big data analytics, and cloud-based technologies (Brynjolfsson et al. 2017; Cockburn et al. 2017; Goldfarb et al. 2019), have the potential to transform businesses across many industries. From a firm and investor's perspective, the adoption of these technologies is a rare opportunity to substantively improve the firm's technological capabilities and long-run viability. Yet, the adoption of these technologies is fraught with uncertainties, as the adoption process is quite unique to each firm/industry. It is therefore, ex ante unclear how capital markets perceive and influence the investment in these technologies. In this study, I provide new evidence on the relationship between capital markets and digital technology investment, by studying the potential capital market forces that shape the investment in the latest wave of GPTs, digital technologies.

Two reasons motivate an in-depth study of the role of capital markets in digital technology investments. First, despite an extensive literature on the market dynamics (Arrow 1962), customer dynamics (Christensen 1997) and organizational factors (Henderson and Clark 1990; Bresnahan 2010) factors that drive the adoption of new technologies, there has been little study of the role that capital markets play in new technologies such as digital technologies. Yet, capital markets arguably play an important role in facilitating these investments, due to the high costs and long-horizon payoffs associated with digital investments. Statistics from the World Economic Forum estimate that corporations have spent a total of 2.5 T. USD on digital investments from 2016-2020[1], and several large corporations are reportedly spending billions per year on these technologies[2]. These investments will likely take years to bear fruit as studies that examined the previous GPT revolution, computer technologies, suggest that the productivity benefits of the new technology only peaks in the 5-7 years after the initial investment (Brynjolfsson and Hitt 2003). Moreover, recent work has found limited evidence of current productivity gains from AI adoption as of yet (Brynjolfsson et al. 2017), which further suggests that the digital investments could also have long-horizon payoffs. Thus, these investments likely demand extensive engagement with capital markets, which in turn, motivates a study of the role of capital markets in facilitating these investments.

1. See: http://reports.weforum.org/digital-transformation/files/2018/05/201805-DTI-Maximizing-the-Return-on-Digital-Investments.pdf

2. See for instance, JP Morgan's 11 B. per year spending on technologies like AI and machine learning, https://www.jpmorganchase.com/news-stories/tech-investment-could-disrupt-banking

Second, from an accounting and finance perspective, the setting of digital investment is worthy of further study as these investments exhibit key differences compared to other forms of investment. As a form of general purpose technology (GPT), these new digital technologies likely require substantial complementary organizational changes to adjust the business process to the new technology (Bresnahan and Greenstein 1996; Tambe et al. 2020). Thus, unlike typical investments which builds on existing business structures, the investment in digital technologies could drive additional uncertainty from an investor's perspective, as firms will likely need to enact significant organizational changes for these forms of GPT investments. Another unique feature of digital technologies, is the fact that, for application sectors, these technologies are typically developed from technology spaces that are less familiar to investors of the application sector, and so the valuation implications of these technologies could also be more uncertain for markets. Thus, these factors suggest that investments in digital technologies are generally more drastic and uncertain compared to typical investment, which therefore, also motivates a study on the capital market factors that drive this particular form of investment.

To address the main research question, I follow the theoretical framework in Bebchuk and Stole (1993) and study how digital investment of non-IT firms is shaped by three capital market forces, namely, expected investment productivity, capital constraints and short-term market incentives. I begin my analysis by first developing measures of digital investment and expected digital investment productivity. Digital investment is measured by three different proxies, namely, the acquisition of technology companies, the proportion of software workers (from resume data following, Babina et al. 2020; Wu et al. 2020) and the posting of job vacancies with AI-related skills (Alekseeva et al. 2021; Acemoglu et al. 2020). To measure digital investment productivity, I develop a patent-value-weighted index of textual overlap between AI-related patents and the 6-digit NAICS industry descriptions (following methodologies in, Webb 2020; Kogan et al. 2020) as an industry-year proxy of expected digital investment productivity[3]. Cross-sectional validation analyses, show that this forward-looking score

3. At a conceptual level, the titles and abstract of AI patents discusses the applications of advanced digital

of digital investment productivity is significantly correlated with several alternative productivity measures that have been used in other studies, and is also correlated with alternative measures of the value of digital investment.

Following the prior literature on the drivers of corporate investment behavior, I use these measures to examine whether digital investment is associated with the expected productivity of these investments (Tobin 1969; Hayashi 1982). Consistent with this view, I find that a 1 standard deviation increase in the patent-weighted AI index is associated with a 0.8% increase in the number of technology acquisitions, 0.27% higher proportion of software workers, and a 0.1% increase in the intensity of AI worker hiring. Furthermore, motivated by prior studies that document the binding effects of capital constraints on firm-level investment (Fazzari et al. 1988; Rauh 2006), I also examine the effects of capital constraints on digital investments. My results suggest that capital constraints are binding, as I find that capital constraints, as measured by the Whited and Wu (2006) index, is negatively associated with digital investments. Estimates suggest that a 10% increase in the capital constraints index is associated with a 0.5% decrease in technology acquisitions and a 0.05% decrease in the hiring intensity of AI workers.

Next, I examine the effects of short-term market incentives on digital investment. Notably, prior work in accounting and finance have argued that the relationship between short-term market incentives and investment is ex ante unclear. On the one hand, several studies argue that the incentive to boost short-term share prices lead firms to underinvest in positive NPV projects (e.g Graham et al. 2005), as managers would shift future earnings for a short-term performance boost (Stein 1989). On the other hand, short-term competitive forces (Giannetti and Yu 2020) and signalling incentives

technologies (AI technologies) on specific processes and the NAICS industry descriptions outline the production processes used to produce goods and services at the industry-level. Thus, the textual overlap of these two sources of text, identifies processes that are relevant to a particular industry *and* can be improved by advanced digital technologies. Hence, an industry with high overall textual overlap with existing AI patents has more processes that can be improved by new digital technologies. In other words, industries with high textual overlap scores are likely to exhibit higher complementarities with these technologies.

(Bebchuk and Stole 1993) could also drive a positive relationship between short-term market incentives and investment.

To implement this analysis, I measure short-term market incentives by the delta of the equity and options that vest in a calendar quarter (defined as "vesting equity" following, Edmans et al. 2017). My results suggest that short-term market incentives is negatively associated with digital investment through acquisitions. Estimates suggest that for a 1 standard deviation increase in vesting equity, the probability of a technology acquistion falls by 0.2%. Thus, I find some evidence that the increase in vesting equity create incentives for firms to cut-back on digital-related investments.

The contribution of this study is three-fold. First, my study provides new evidence on how capital markets play a role in influencing digital investments. Given recent interest from academics and practitioners over the adoption of digital technologies, the findings of this study could be useful by introducing capital markets as an important player in the process. In particular, my study fills a gap in the academic literature by showing that three capital market forces, namely, the expected productivity of digital investments, capital constraints and short-term market incentives play a key role in explaining the variation in digital investment. And from a practitioner's point of view, the finding that these forces play a key role in explaining the variation in digital investment, could help practitioners formulate more effective digital adoption strategies.

Second, another potential contribution of this study is to the short-termism literature in accounting and finance. The evidence from this literature has been fairly mixed; with some studies providing evidence of short-term market pressures driving under-investment in positive NPV projects (Bushee 1998; Graham et al. 2005; Asker et al. 2014), while others report findings that suggests that these same pressures can encourage greater investment in certain settings (Giannetti and Yu 2020). My findings contribute to this literature by providing some evidence that short-term market incentives could deter digital investments in the form of acquisitions.

Third, I provide an incremental methodological contribution by adapting a measure of expected digital investment productivity, to a corporate investment context. The validation analyses that I examine suggests that this measure is consistent with measures developed in other contexts, and is also correlated with ex post measures of the value created by digital investment. Thus, future work could build on this measure to examine accounting and finance related questions relating to expected digital investment productivity.

The rest of this paper is organized as follows. In section 2, I discuss the relevant literature, and in section 3, I formulate the conceptual framework for the empirical analysis. Section 4 discusses the methodology for measuring the key variables as well as the empirical regression model. Section 5 presents the results and section 6 concludes.

4.1 Literature Review

4.1.1 The Importance of Digital Technologies

One of the more important technologies that have emerged in recent years, is arguably the digital technologies of AI, big data analytics and cloud-based technologies. In fact, some scholars have argued that these technologies are a form of general purpose technologies, a rare class of technologies that are important drivers of economic growth (Bresnahan and Trajtenberg 1995; Helpman 2003), that includes steam engines, electricity to computers. This is because digital technologies meet the three main criteria of GPTs (Bresnahan 2010). First, these technologies are widely used across many different sectors as shown in several academic studies (Goldfarb et al. 2019) and in several practitioner articles that examines the digital transformation phenomena. Two, digital technologies are capable of ongoing technical improvement, as a good portion of today's high-tech innovation activities are focused on AI and cloud-based technologies (Bloom et al. 2018). Three, digital technologies can spur innovation in multiple application sectors, as AI, big data analytics and cloud-based technologies not only have wide ranges of use, but can also automate the innovation process (Cockburn et al. 2017).

Thus, there is some consensus that digital technologies are the next-wave of GPTs, which suggests that these technologies are likely to drive significant changes in the ways that businesses operate.

4.1.2 Benefits of Digital Investment

As a form of GPT, the insights from the academic literature suggests that digital technologies have high potential benefits. In particular, due to the complementarities of GPT-type technologies, digital technologies have the potential to create value across multiple sectors, either by increasing the value of existing assets or by spurring new innovation in application sectors (industries that use the GPT). Evidence on ICT adoption, the previous wave of GPTs, show that GPT adoption can lead to significant benefits, as prior studies report that ICTs drive large productivity gains (Hitt 1999; Brynjolfsson and Hitt 2000), and open opportunities for firms to develop new products and services (Brynjolfsson and Smith 2000). Moreover, Chapter 2 finds that the latest wave of digital technologies is viewed positively by markets, as companies that disclose digital activities tend to receive higher valuations and returns.

Moreover, for existing firms, investment in new digital technologies, is in some ways, critical for the firm's long-run success. Chapter 2 show that firms that do not adopt or disclose digital technologies, yield significantly negative long-run risk-adjusted returns of roughly 9% over a 3 year horizon. And high-profile case examples also document significant negative consequences for companies that fail to adopt digital technologies. For instance, the discussion in Greenstein (2017) discusses shows that the demise of Encyclopedia Britannica was related to its failure to shift its physical product into the emerging digital medium in the 90s. Thus, from the perspective of existing firms, the investment in digital technologies is quite likely a critical component of their long-term success. And hence, understanding the drivers of digital investment is an important question for practitioners and academics.

4.1.3 Digital Technology Adoption Costs and Potential Frictions with Capital Markets

One key friction associated with digital technology adoption in existing firms, is the costly development of user co-invention when integrating new GPTs, such as digital technologies. Bresnahan and Greenstein (1996) illustrate this concept by showing that the transition from traditional mainframes to client/servers, required firms to address within-firm structural issues such as the conflicts between IT departments and their users. The authors further argue that co-invention is typically the bottleneck in adopting GPTs, as these inventions require changing organizations, which is likely a more difficult task relative to technical innovations. Alternatively, another way of framing this friction, is by casting GPTs and digital technologies as a form of radical or architectural innovation that requires substantial changes in the existing innovation process (Henderson and Clark 1990). Under this paradigm, prior work has shown that the introduction of new radical or architectural innovations can be quite difficult for firms, as existing innovation practices have to be revamped to accommodate the architectural/radical type innovation (Henderson 1993).

As addressing these costs involve changing entire business processes and practices (Brynjolfsson et al. 2018), the process of adopting new digital technologies can be quite complex. In addition, these costs are also difficult to quantify as the co-invention and other organizational investments are also highly intangible (Tambe et al. 2020). Thus, the organizational changes associated with digital investments likely drives high investor uncertainty, which in turn, drives potential frictions between markets and firms on digital investments.

Additionally, the adoption of these technologies is likely to face additional frictions from capital markets as investors are likely to be less familiar with the valuation implications of digital technologies. In particular, these technologies and other forms of GPTs are usually developed in technology spaces that are outside of the firm's own technology space (Helpman 2003; Bresnahan 2010). Thus, the introduction of digital technologies from a new technology space is likely to drive greater valuation uncertainty, which thus creates additional frictions with capital markets.

Due to the potential capital market frictions associated with digital investments, the role of capital markets in these investments is a question that is likely to generate interest from practitioners and regulators. Specifically, a study that elucidates how capital markets interact with digital investments could help practitioners formulate strategies on how to negotiate with capital markets when engaging in these investments. Furthermore, from the perspective of policymakers, such a study could also be useful, given growing interest in policies that facilitate the development of digital economies.

4.2 Conceptual Framework

4.2.1 Model

The conceptual framework builds on the investment model in Bebchuk and Stole (1993), and so in this section, I revisit the Bebchuk and Stole (1993) model and derive the key parameters that could determine investment in long-term projects such as digital technologies. The model is set-up over two periods as a choice between two investment projects — a short-term (ST) project that yields a return in the first period, and a long-term (LT) project that yields a return in the second period. The manager makes investment choices based on a utility function that is expressed as the market valuation of the firm in periods 1 and 2. In other words, the manager is concerned by both the ST and LT valuation of the firm. This utility function is expressed linearly:

$$U(V_1, V_2) = \delta + \alpha_1 V_1 + \alpha_2 V_2 \tag{4.1}$$

where α_1 and α_2 are the weights that the manager places on ST (V_1) and LT (V_2) valuation. The application of this model to my research question takes the following form — ST and LT investment correspond to the investment in existing and digital technologies respectively. The assumption underlying this framing, is that the digital technologies have longer payoffs (i.e. the returns only arrive in period 2). This is a valid assumption as prior work on GPTs has shown that the benefits of GPTs take several years to fully realize (Brynjolfsson and Hitt 2003; Brynjolfsson et al. 2017).

The production function of the investments are publicly known and are expressed as follows. For existing investments (ST projects), the payoffs are defined as: $\tilde{S} = S(k_1) + \varepsilon$, and $S'(.) > 0, S''(.) < 0$. For digital investments (LT projects), the payoffs are defined as: $\tilde{L} = \theta L(k_2) + \eta$, which varies by the firm's type θ, and $L'(.) > 0, L''(.) < 0$. Following Bebchuk and Stole (1993), I also further assume that there is a limited amount of capital K that can be allocated to both projects. And so the manager decides to allocate x to digital investments, and $K - x$ to existing investments.

Expected Investment Productivity and Capital Constraints

With these assumptions, the maximization problem, under the assumption of no information frictions, is the following:

$$Max\ W(x) = S(K - x) + \theta L(x) \tag{4.2}$$

which is characterized by the following FOC:

$$S'(k - x^*) = \theta L'(x^*) \tag{4.3}$$

To model the first best level of digital investment, I make an additional assumption of the investment/production function. I assume a log Cobb-Douglas function as the utility function is expressed linearly. And I also assume that both $S(.)$ and $L(.)$ have the same productivity parameter, normalized at 1, such that $\tilde{S}$ and $\tilde{L}$ differ only by θ. Thus, I assume that $S(k_1) = ln(k_1)$ and $L(k_2) = ln(k_2)$. Plugging this assumption into the FOC, I obtain:

$$x^* = \frac{K}{1 + \frac{1}{\theta}} \tag{4.4}$$

This equation indicates that digital investment is increasing with higher K, the capital budget, as well as θ, digital investment productivity. Thus, the interpretation of equation 4 is as follows: Digital investment is higher if the firm is endowed with higher expected digital investment productivity, but firms facing higher capital constraints will invest less in these technologies.

Unobservable Action: Under-Investment

The first part of Bebchuk and Stole (1993) examines the distortions created when there is market uncertainty over the investment action x. The authors derive a FOC, which characterizes the distortion:

$$\theta L'(x) - S'(K-x) = \frac{\alpha_1}{\alpha_1+\alpha_2}\theta L'(x) = \frac{\alpha\theta}{x} \tag{4.5}$$

Note that, $\frac{\alpha_1}{\alpha_1+\alpha_2} = \alpha_1 = \alpha$ as $\alpha_1+\alpha_2 = 1$. The right hand side of the equation characterizes the under-investment distortion relative to equation 4, and this expression indicates that the under-investment distortion is increasing in θ and α. Thus equation 5 indicates that the sensitivity to short-term market prices (or short-term market incentives) will drive down investment when there is uncertainty over investment actions.

Unobservable Productivity: Over-Investment

In the second part of Bebchuk and Stole (1993), the authors examine the distortions created by the market uncertainty over the uncertain component of digital investment productivity, θ. Specifically, the authors argue that more productive firms would over-invest to signal their type, in the spirit of Spence (1973). To illustrate this point, the authors characterize the over-investment distortion as the sensitivity of investment to the uncertain component of digital investment productivity:

$$\frac{dx}{d\theta} = \frac{\alpha_1}{\alpha_1+\alpha_2}\frac{L(x)}{S'(K-x)-\theta L'(x)} = \frac{\alpha ln(x)}{\frac{1}{K-x}-\frac{\theta}{x}} \tag{4.6}$$

The expression on the right hand side of this equation shows that the over-investment distortion relative to equation 4 is increasing in θ, α and in K. Thus, this equation indicates that the sensitivity to short-term market prices (or short-term market incentives) will increase investment when there is uncertainty over investment productivity.

4.2.2 Discussion of the Capital Market Forces

The theoretical model outlined above has highlighted three key capital market forces that are likely to play a role in driving digital investments. In the discussion below, I review the literature and develop ex ante predictions that are tested in the empirical analysis.

4.2.3 Q-Theory and Expected Investment Productivity

In neo-classical theories of investment, the *sufficient* summary metric for investment is Tobin's Q — the marginal value of investment relative to the marginal replacement cost (Tobin 1969; Tobin and Brainard 1977; Hayashi 1982), or the marginal profit of the investment. In support of this theory, prior work has shown that the market's expectation of the value of additional investment positively predicts investment behavior (Blundell et al. 1992; Erickson and Whited 2000), although these studies have also shown that the relationship is weaker than implied by theoretical models.

Specifically for digital investments, the key driver of expected marginal profit of these investments could be the expected productivity improvements. This is due to the fact that existing research on AI- and the previous GPT wave of computer-related technologies, have viewed these technologies as productivity-enhancing (Autor et al. 2003; Acemoglu and Restrepo 2019; Webb 2020) that drive value through automation and optimizing workflows. Thus, the expected productivity benefits of these investments should be a key driver of marginal profits. And therefore, firms with higher expected productivity of digital investment should investment more in these technologies.

On the other hand, one key feature of GPTs and digital technologies is that the value of these technologies are constantly evolving as complementary innovation and co-inventions are developed. Studies have argued that because of this feature, the true value of these technologies are typically only known with the benefit of multiple years of hindsight (Brynjolfsson and Hitt 2003; Bresnahan 2010). Due to this uncertainty, markets and firms may not fully invest based on ex-ante signals of expected investment productivity. Thus, the question of whether digital investment is associated with forward-looking measures of expected digital investment productivity is an open empirical question that I study in this paper.

4.2.4 Capital Constraints

Following Q-theory, a substantial literature on corporate investment has also explored several other capital market factors that incrementally explain investment behavior beyond marginal Q. The first capital market factor that I discuss is capital constraints. Early theoretical work has posited that capital constraints can play a role in driving up the cost of external financing, which in turn reduces the firm's ability to finance investment. For instance, from a moral hazard perspective, Holmstrom and Tirole (1997) and Leland and Pyle (1977) show that moral hazard drives imperfect pledgeability of assets, which thus drives down the amount of capital that can be raised from capital markets for more capital constrained firms. In addition, from an adverse selection perspective, Myers and Majluf (1984) argue that the choice to raise capital from public markets signals that the firm is overvalued, and consequently, markets charge a higher cost of capital for firms that rely more on external financing.

To examine whether capital constraints influence investment decisions, empirical studies have studied whether firms exhibit investment-cash flow sensitivities holding marginal q fixed, as a way of evaluating the capital constraints hypothesis. The conclusion from the prior literature, is however, somewhat mixed. Fazzari et al. (1988) show that firms that more financially constrained firms exhibit higher investment-cash flow sensitivities, while Kaplan and Zingales (1997) show that cash flow plays a limited role in explaining investment. More causal evidence in Rauh (2006) suggests that cash flow explains investment decision in a significant way. However, the evidence has been debated in the literature, and the results based on methodological refinements show that cash flow plays a limited role in explaining investment decisions (Bakke and Whited 2012).

Hence, given the divergent views on whether capital constraints are binding enough to influence investment decisions, I examine whether these factors plays a role in examining investment in digital technologies in this study.

4.2.5 Short-Term Market Incentives

Another capital market factor that I study in this paper, is the focus on short-term market incentives. Notably, there is a key tension in how short-term market incentives shape investment behavior, and in the following discussion, I outline the contrasting ways of how short-term pressures is related to digital investment.

Short-Term Market Incentives Deters Investment?

One potential consequence of capital market incentives is that short-term market pressures can deter investment in the new technology. Studies in accounting and finance have argued that pressures from investors and analysts can drive companies to under-invest in positive NPV projects (Stein 1989; Bebchuk and Stole 1993; Graham et al. 2005), if markets are sufficiently impatient (Gigler et al. 2014).

Empirical work in accounting and finance have found some evidence that short-termism pressures induce weaker investment in companies. Some find that firms reduce investment in certain scenarios where the pressures to meet earnings are high (Bushee 1998; Bhojraj et al. 2009; Kothari et al. 2016). Others show that capital market institutions, such as analysts and short-sellers, drive lower investment (He and Tian 2013; Grullon et al. 2015). Moreover, in comparisons between publicly-listed firms and private and government entities, studies have also shown that publicly-listed firms tend to under-invest in investment opportunities (Asker et al. 2014; Aggarwal and Hsu 2014; Budish et al. 2015; Bernstein 2015), which provides addition evidence that myopic forces tend to drive under-investment.

As digital technologies involve high short-run costs that are not always capitalizable and can only payoff in the long-run, the investment in these technologies is exceptionally prone to the influence of short-termism. Companies may have to cut shareholder payout to make cash available for these investment[4], and so short-term investor pressures on digital investment is likely to be high. In addition, successful digital adoption is difficult to measure and thus shareholders may not be able to effectively trade-off the long-term value for the short-term returns that they will give up when companies make digital investments. Thus, investors and boards often demand quick results for digital adoption[5], which adds significant short-term pressure on management in their digital transformation efforts. Hence, given the above discussion, there is much reason to expect that short-term market pressures could lead firms to under-invest in digital technologies.

Short-Term Market Incentives Encourages More Investment?

On the other hand, studies in the short-termism literature have argued that short-termism plays a limited role in shaping investment decisions. Studies that examine the aggregate trends of investment find limited evidence that expectations from capital markets lead to long-term adverse consequences for firms. Kaplan (2018) for instance, points out that the highly innovative biomedical and technology sectors thrive on capital market pressures and have shown success over the long-term. Roe (2018) suggests that an oft-cited statistic of myopic investment, cuts in capital expenditure investment, is a symptom of a changing economy rather than under-investment in the firm's future. Moreover, recent evidence on the consequences of the rise of shareholder activists, institutional investors who actively apply short-term pressures on managers, suggests that in net, these investors improve productivity (Brav et al. 2015), and the effectiveness of long-term investment (Brav et al. 2018). In addition, there

4. See for example, Mattel's dividend cut for digitalization: https://www.bloomberg.com/news/articles/2017-06-14/mattel-ceo-cuts-dividend-to-speed-up-transformation-of-toymaker

5. A UK study on CEOs in 2019, shows that 86% of CEOs are overwhelmed by their board's demands for quick digitalizaion results (https://home.kpmg/uk/en/home/insights/2018/05/uk-ceo-outlook-survey-2018.html). In addition, market commentators have also suggested that many bank investors are growing impatient over banks fintech adoption efforts (https://www.americanbanker.com/opinion/banks-big-tech-spending-is-testing-shareholders-patience)

is also some recent evidence that capital market expectations from short-term investors help firms to react more aggressively in periods of radical change (Giannetti and Yu 2020).

One view that could explain this phenomenon, is that markets anchor on performance metrics other than current earnings. Traditional models of short-termism make the assumption that incentives to under-invest stem from a focus on short-term prices, but short-term prices could be based on expectations on longer-term metrics such as sales on long-term earnings. Thus, short-term market incentives could also induce firms to invest more on projects that increase sales or long-run earnings, depending on the type of performance metrics that markets are focused on in the short-term.

Moreover, there are also other reasons that suggest that short-term market incentives could drive over-investment in digital technologies. From a behavioral finance perspective, capital market expectations can drive additional investment in digital technologies via inflated expectations. Specifically, prior work shows that inflated prices drives greater investment (Baker et al. 2003) and acquisition activity (Dong et al. 2006). In addition, evidence from the dot-com bubble, also shows that investor scrutiny and governance during periods of heightened optimism tends to be lax (Keating et al. 2003; Palepu and Elcock 2001; O'Brien and Tian 2006), which could also drive investment in inefficient projects. Thus, short-term market incentives in the form of market euphoria could also lead firms to over-invest in digital technologies.

Another view, is that signaling incentives, that arise from market uncertainty over the productivity of digital investment could also lead firms to over-invest in digital technologies. A model that outlines the conditions that drives this effect in a rational expectations framework is Bebchuk and Stole (1993). In this model, the authors show that when there is market uncertainty over the productivity of investments, short-term pressures drive firms to over-invest to signal their type to markets. Notably, the digital adoption setting is likely to fit well with this model — the payoffs to digital investment are highly uncertain, as scholars to date have yet to fully observe the productivity benefits of these technologies (Brynjolfsson et al. 2017).

Evidence from the dot-com bubble, provides some supporting evidence for the signalling effect of short-term market incentives. Specifically, during the market frenzy over IT investments, studies have shown that investors anchored on actions that signalled higher IT investments, rather than on metrics of current profitability or productivity. For example, Demers and Lev (2001), Bartov et al. (2002) and Rajgopal et al. (2002) showed that investors of internet firms focused mainly on cash burn and managerial actions rather than on other metrics of profitability. Moreover, the rich initial valuations of the latest technology start-ups such as Uber and SnapChat, both firms with significant cash burn and uncertain payoffs, suggests that the signaling dynamics from capital market expectations could still be a factor today.

Hence, given the above discussion, there is also much reason to expect that short-term market incentives could drive higher investment in digital technologies. And so in the paper, I examine the relationship between the extent of these incentives and digital investments.

4.3 Research Design

In this section, I discuss the data and empirical methodology to measure the key capital market forces, as well as the main regression model for this paper.

4.3.1 Sample Selection

The sample of this study consists of all US-listed non-IT firms in the Compustat/CRSP universe from 2010Q1-2020Q4. The choice of focusing only on non-IT firms is motivated by the fact that this study aims to examine the adoption of digital technologies in application sectors. For digital technologies, application sectors are the non-IT industries and thus the empirical analysis focuses on this subset of firms. In addition, to create a consistent sample for analysis, I retain firm-quarter observations with available observations from *BurningGlass* and *Revelio Labs*[6] (See Table 4.1 for the sample selection breakdown).

Table 4.1: Sample Selection

Sampling Criteria	Firm-Quarters	Firms
Compustat-CRSP Merge from Q12010-Q42020	230,889	8,752
Share Codes 10, 11, 12 and US Headquartered and Incorporated Firms	157,275	6,231
Non-IT Firms	113,832	4,378
ISS Incentive Lab Observations	26,438	815
BurningGlass Observations	17,918	643
Revelio Lab Observations	14,438	478

This table reports the sample selection for the study, which is based on a subset of US-listed, non-IT firms from 2010Q1-2020Q4, with valid *Burning Glass*, *Revelio Labs* and *ISS Incentive Lab* observations.

4.3.2 Measuring Digital Investment

The main dependent variable in the study, the investment in digital technologies, is proxied by three different measures. The first measure tracks digital investment by the number of technology acquisitions. This measure is motivated by anecdotes of non-IT firms acquiring technology start-ups for their advanced digital technologies[7]. These acquisitions are identified using the *S&P's 451 M&A database*, which tracks the acquisition of technology companies.

The next two measures are based on software and advanced digital technology employment. For the first measure, I obtain data on the number of software engineers and managers at the firm-level from a database provided by *Revelio Labs*. This database obtains firm-level employment data by extracting information from resumes and job postings. For the second measure, I obtain data on the hiring of workers with AI skills, by identifying job vacancies with AI-related skills (following the skill set list in, Alekseeva et al. 2021)[8].

6. These databases are used to compute the proportion of AI worker hiring and software workers respectively.

7. One example is Walmart's various acquisitions of AI tech firms:
https://www.cnbc.com/2019/05/16/from-advertising-to-ai-walmart-is-doing-a-lot-more-than-retail.html

8. Descriptions of these dependent variables, and other independent variables are described in Appendix C.1

4.3.3 Measuring Expected Digital Investment Productivity

I proxy for digital investment productivity at the industry-level by computing a patent-value weighted AI technology exposure scores following the methodology developed in Webb (2020) and Kogan et al. (2020). The basic idea of the methodology is to first compute scores based on the overlap of words in AI patents and the industry descriptions, to proxy for the complementarity between advanced digital technologies and industry-level applications. The second step, augments the score by adjusting for the value of software and AI patenting activity across industries.

The first step of the computation process is to identify a set of patents relating to advanced digital technologies. I start with a set of AI-related patents following the search procedure in Webb (2020)[9]. I search for these patents in the universe of US-based patents in the *Google Patents Database*[10][11]. To identify AI patents, I select patents that have at least one of the following terms in the title or abstract: "supervised learning", "reinforcement learning", "deep learning", "neural network" and "machine learning"[12].

In the second step, I use the Kogan et al. (2020) methodology to estimate the overlap between patents and the description of industries. The methodology, takes a GloVe approach (Pennington et al. 2014) in computing the textual similarity of the two sources of text through word embeddings. A key advantage of this approach is that it yields a cosine similarity metric that accounts for subtle differences in the textual meaning, and thus is particularly suited for situations where words are similar but are represented in different ways. In Appendix C.2, I discuss this approach in greater detail[13].

9. In addition to AI patents, the Webb (2020) paper also develops search terms for software and robotics technologies. I do not include both patents, as the AI patents arguably match better to the investment measure of recent technology worker hires and acquisitions.

10. https://console.cloud.google.com/marketplace/partners/patents-public-data

11. This databases includes patents that are filed globally. But to ensure that differences in the patenting process do not influence my results, I restrict the analysis to only patents filed in the US.

12. Webb (2020) does not include "machine learning" as a search term, but I include this term as it is a important component of AI technologies

13. For examples of patent-industry pairs with high similarity scores, see Appendix C.3.

The third step of the procedure adjusts for the value of patents across industries. To implement the adjustment, I compute the rolling 10-year average of the value[14] of software[15] and AI-related patents (identified as patents that satisfy the search critera, and follow-on patents that cite these patents), at the 3-digit NAICS-level. I then compute the yearly percentile rank of the average patent values, and multiply the similarity scores by the percentile rank. And to aggregate to the industry-year level, I sum up the adjusted scores.

Finally, to normalize the exposure scores, I deduct the raw scores by the pooled mean and scale by the pooled standard deviation.

Interpreting the Textual Overlap Scores

To help unpack the intuition behind the textual overlap scores, I first discuss the conceptual interpretation of the two key constituent components in the textual overlap scores, namely, the titles/abstracts of AI patents and the industry description.

The titles and abstract of patents provide a short summary of the essence of the patented invention, which could be in the form of either new processes, products or machines. For advanced digital technology patents, such as those of AI technologies, these patents are typically filed for new processes due to the nature of these technologies. Thus the titles/abstracts of AI patents describes what the new advanced digital technology is used for, and summarizes how the new technology can be adopted on specific production processes. Therefore, at a high-level, the corpus of text from AI patents can be interpreted as describing the full scope of how advanced digital technologies can be used to improve production processes.

The NAICS industry description provides an outline of the production process of specific indus-

14. The data on patent values is taken from the patent dataset in Kogan et al. (2017)
15. To identify software patents, I select patents that have at least one of the following terms in title or abstract: “software”, “program” and ”computer”, and none of the following terms: “chip”, “semiconductor”, “circuitry”, “circuit”.

tries. This is because the NAICS classification system was designed to group establish by production processes, and so the description of NAICS industries outline the processes that is used to produce goods and services of each industry.

Thus, the overlap of similar words between titles/abstracts of AI patents and NAICS industry descriptions, should identify production processes that are relevant to a particular industry *and* can be improved by certain advanced digital technologies[16][17]. Hence, higher textual overlap would indicate that there are more advanced digital technologies that could be used to improve the industry's production processes. In other words, higher textual overlap would indicate that an industry has higher *complementarities* with advanced digital technologies, and thus, indicate higher expected digital investment productivity.

Validation Analysis of the Textual Overlap Score

To provide some justification for the use of the patent-weighted AI index as a proxy for digital investment productivity, I compare this measure with ex post measures of digital productivity as well as other measures of expected digital technology productivity that have been used in other contexts.

16. Note that the overlap could also capture the overlap between words that are specific to AI technologies (i.e. machine learning or artificial intelligence), which would confound the production process interpretation of the textual overlap measure. However, for the non-IT industries (which is the main focus of this study), the corpus of text from the descriptions of these industries should be orthogonal to the set of words that are specific to the technology industry.

17. See Appendix C.3 for examples of patents identified by the algorithm, that outline new innovations that could improve the processes of the paired industries.

Table 4.2: External Validation of Expected Digital Technology Productivity Measures

	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index
Investment Correlation$_j$	0.255***					
	(0.039)					
Deal Value-to-Assets$_j$		0.278***				
		(0.043)				
Digital Intensity$_j$			0.312***			
			(0.052)			
Internet Adoption$_j$				-0.297***		
				(0.079)		
Occupation Exposure$_j$					0.321***	
					(0.049)	
Occupation Exposure (Kogan)$_j$						0.103**
						(0.041)
Observations	562	535	562	151	560	560
R^2	0.065	0.077	0.065	0.067	0.072	0.011

This table reports the validation analysis of the patent value-weighted AI index measure of digital technology productivity. In this table, I examine the associations between the cross-industry variation of the ranked weighted textual overlap score with other (ranked) metrics of digital investment productivity. I examine 6 measure of digital productivity — (1) the correlation in IT-related acquisitions in the industry-level, (2) the average deal value-to-assets for IT acquisitions at the industry-level, (3) the global taxonomy of digital intensive industries from the OECD, (4) the index of industries that have adopted the internet in 1998-2000, (5) the occupation exposure score from Webb (2020), (6) the occupation exposure score from Kogan et al. (2020). Robust standard errors are reported in parantheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

Table 4.2, Panel A presents the correlations between the alternative measures of digital investment productivity and percentile rank of the time-series average of the patent-value weighted AI index. In columns 1 and 2, I examine the association between the AI index score and metrics that proxy for technology acquisition success. These measures are namely, the serial-correlation of technology acquisitions at the 3-digit industry-level, as well as the average deal value-to-assets across 3-digit NAICS industry firms. Results of the analysis show that the association between these measures and the textual overlap score is statistically significant, which suggests that the textual overlap score is related to the ex post value in technology acquisitions.

In the following four columns, I compare the textual overlap scores with other measures of digital productivity that have been used in other settings. I find that the textual overlap score correlates strongly with the productivity measure from the OECD's global taxonomy of digital investment productivity across industries (Calvino et al. 2018), the occupation exposure score proposed in Webb (2020) and the occupation score proposed in Kogan et al. (2020). It is somewhat surprising that the textual overlap score is negatively correlated with the internet adoption score in Forman et al. (2003), and one potential explanation for this finding is that the internet technologies are quite different from the current digital phenomenon. Moreover, the industries examined in the study, were mostly the first-movers of internet adoption, while the subject of this study is the laggards of digital adoption.

Thus, taken together, these validation analyses thus provide some comfort that the textual overlap score is measuring variation in the expected digital investment productivity across industries[18].

Summary Statistics of the Textual Overlap Score

To provide some illustration of the AI index textual overlap scores, I present some summary statistics from the textual overlap scores. Table 4.3, reports the top and bottom 10 exposed 6-digit NAICS

18. In additional analysis, reported in Tables C.1 - C.3 in Appendix C.4, I also show that many of the validation analyses holds for alternative specification of the textual overlap score.

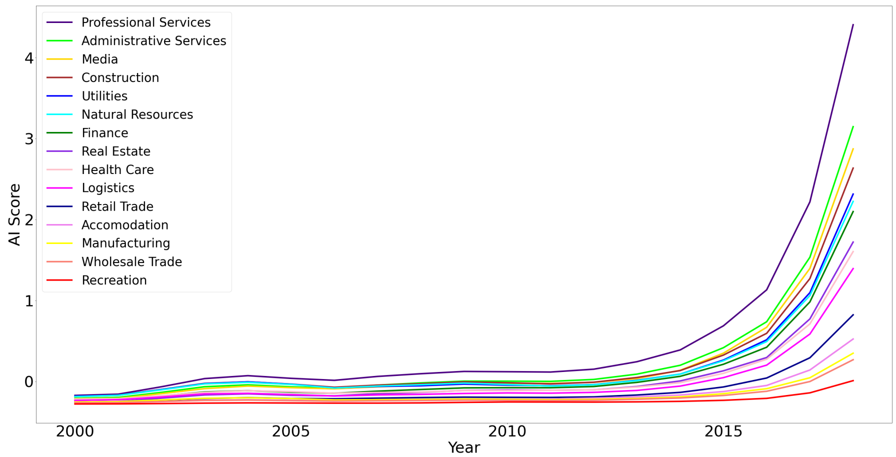
industries from the cross-sectional average of scores. Panel A, shows that the industries that are expected to benefit most from advanced digital investment are those from the media, finance and utilities/construction industries. On the other hand, Panel B suggests that light manufacturing and natural resource industries tend to benefit less from advanced digital investments.

Table 4.3: Top and Bottom AI Exposed Industries

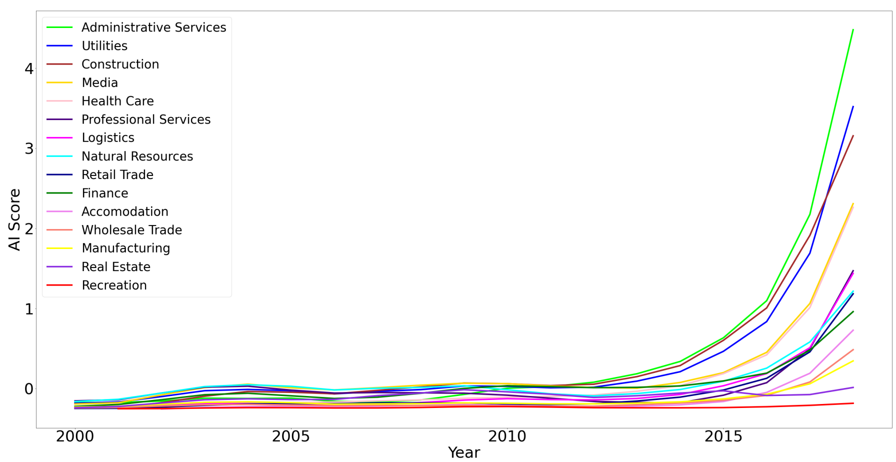
Panel A: Top Overlap Industries	
1	Cable and Other Subscription Programming
2	Investment Advice
3	Couriers and Express Delivery Services
4	Investment Banking and Securities Dealing
5	Securities and Commodity Exchanges
6	Portfolio Management
7	Financial Transactions Processing, Reserve, and Clearinghouse Activities
8	Television Broadcasting
9	Motion Picture and Video Distribution
10	Power and Communication Line and Related Structures Construction
Panel B: Bottom Overlap Industries	
1	Mattress Manufacturing
2	Roofing, Siding, and Insulation Material Merchant Wholesalers
3	Small Arms, Ordnance, and Ordnance Accessories Manufacturing
4	Gypsum Product Manufacturing
5	Breweries
6	Tobacco and Tobacco Product Merchant Wholesalers
7	Ethyl Alcohol Manufacturing
8	Metal Kitchen Cookware, Utensil, Cutlery, and Flatware (except Precious) Manufacturing
9	Distilleries
10	Nonchocolate Confectionery Manufacturing

This table reports the top 10 and bottom 10 AI-exposed 6-digit NAICS industries by textual overlap scores. Panel A reports the top 10 industries, while Panel B reports the bottom 10 industries for the textual overlap scores.

In Figure 4.1, I plot the time-series distribution of AI textual overlap scores for 2-digit NAICS industries. As the textual overlap is defined at the 6-digit level, I aggregate the exposure scores to the 2-digit level by taking a sales-weighted sum of the exposure over all firms. As seen in Figure 4.1, the time-series patterns of AI overlap across industries, matches with anecdotal reports of the path of AI technology development. Notably, since 2015, the figure show a steep rise in the textual overlap of AI patents with industry descriptions, which also aligns with the steep rise in the use of AI technologies across industries in that year (Acemoglu et al. 2020).



(a) Unweighted AI



(b) Weighted AI

Figure 4.1: Unweighted (a) and Patent Value-Weighted (b) AI Index Across 2-Digit NAICS Industries

4.3.4 Measuring Capital Constraints

To measure capital constraints, I compute an index of capital constraints proposed by Whited and Wu (2006). The Whited and Wu (WW) index is derived from a structural estimation exercise, that estimates the elasticity of the supply-of-capital curve. Based on the structural estimation, the authors develop the following index:

$$\begin{aligned} WW_{i,t} = 0.65 - 0.091 \times \frac{CF_{i,t}}{AT_{i,t}} - 0.062 \times 1_{DIV_{i,t}>0} + 0.021 \times \frac{DEBT_{i,t}}{AT_{i,t}} \\ -0.044 \times Ln(AT_{i,t}) + 0.102 \times SG_{j,t} + 0.035 \times SG_{i,t} \end{aligned} \tag{4.7}$$

where the WW index is computed as a function of cash flow-to-assets ($\frac{CF_{i,t}}{AT_{i,t}}$), an indicator for dividend payers ($1_{DIV_{i,t}>0}$), debt-to-assets ($\frac{DEBT_{i,t}}{AT_{i,t}}$), log of book assets ($Ln(AT_{i,t})$), industry-level sales growth ($SG_{j,t}$) and firm-level sales growth ($SG_{i,t}$). To facilitate the interpretation of this index, I convert the index into yearly percentiles.

4.3.5 Measuring Short-Term Market Incentives

The key driver of investment distortion in models like Stein (1989) and Bebchuk and Stole (1993) is short-term market incentives. In this paper I proxy for these incentives with the vesting equity of the CEO in the quarter. Prior work has argued that the variation in vesting equity is a proxy of managerial sensitivity to short-term stock prices (or short-term market incentives) (Edmans et al. 2017; Edmans et al. 2018), and so in this study, I follow Edmans et al. (2017) in computing the vesting equity at the quarterly-level.

The data on vesting equity schedules is taken from *ISS Incentive Lab*, and I follow the assumptions and methodology in Edmans et al. (2017) to compute the delta of vesting options and equity for CEO's in a quarter[19]. I then multiply the amount of vesting equity with the beginning fiscal quarter

19. See Appendix C.5, for a deeper discussion of the methodology.

stock price, such that the variable can be interpreted as the dollar change in the value of vesting equity for a 100% change in the stock price.

4.3.6 Regression Model

The baseline model studies the effect of the expected digital investment productivity on digital investment. I implement this in a regression model, outlined below:

$$Inv_{i,q} = \alpha_j + \alpha_t + \beta_1 AI\,Index_{j,t-1} + \beta_2 Constraints_{i,q} + \beta_3 Vesting\,Equity_{i,q} \\ + \textstyle\sum_s \beta_s X_{s,i,q} + \varepsilon_{i,q} \tag{4.8}$$

where $Inv_{i,t}$ is a measure of digital investment and is one of the three measures described in Section 5.1. $Tech_{j,t-1}$ is the textual-based estimates of expected digital investment productivity, and based on the discussion in Section 4, I expect β_1 to be either insignificant or positive. $Constraints_{i,q}$ is the Whited and Wu (2006) index for external financial constraints and based on the discussion in Section 4, I expect β_2 to be either insignificant or negative. Finally, $Vesting\,Equity_{i,q}$ is the amount of vesting equity in billions, and based on the discussion in Section 3, I hold no ex ante expectations on the direction of β_3. In addition, I include 6-digit GICS industry and quarter fixed effects, to control for time-invariant industry factors and time-trends. Standard errors are also clustered at the firm and quarter level.

Moreover, I also include a vector of controls for other time-varying confounds ($\sum_s X_{s,i,t}$). I follow Edmans et al. (2017) and first include CEO-level controls for the amount of vested and unvested equity, salary, bonus, tenure and an indicator of whether the CEO is a new appointment. For firm-level controls, I also include the current and lagged Tobin's Q, as well as lagged 3-month returns, leverage, retained earnings, return-on-assets and cash balance ratios.

In addition, I also include controls for industry herfindahl, industry-adjusted sales growth, as market dynamics could drive incentives to invest in new technologies (Arrow 1962). I also control for the proportion of sales to major customers, as customers could also play a role in shaping new technology adoption (Christensen 1997). In addition, I include controls for firm age, SG&A, R&D and segment number to control for organizational complexity and inertia, as prior studies show that these factors deter new technology adoption (Henderson 1993; Bresnahan et al. 2012). Finally, to control for firm risk, I control for market beta computed with weekly returns over a 3-year window[20].

4.4 Results

4.4.1 Summary Statistics

Table 4.4 presents the summary statistics of the key variables[21] used in this study. The average (median) size of the firms by assets is 36 B. USD (6 B. USD). These firms also tend to be older, with an average (median) age of 35 (30) years. On average, the return-on-assets of these firms tends to be around 2%, while the Tobin's Q of these firms tends to be around 1.9.

Turning to the digital investment measures in the last six rows. The number of technology acquisitions, on average is 0.035. As these acquisitions are infrequent, I also report the probability of technology acquisitions, which is 3% on average. I report the second measure of digital investment in the third-to-last row, and report that the average proportion of software worker salaries to total salaries in firms is 3%. Finally, the last row reports percentage of firm-quarter observations with AI worker hiring, and the proportion is around 38%.

20. I do not control for size explicitly, as size is constituent element of the capital constraint index. In particular, the logarithm of book assets exhibit a -84% correlation with constraints index. Nonetheless, in the IV analysis of the analyst coverage variable, I include lags of size in the regression and drop the capital constraints measure. And results are robust to that approach.

21. Description of the variable construction methodology is outlined in Appendix C.1

Table 4.4: Summary Statistics

	Mean	StDev	Median	25%	75%	N
Total Assets (Billions)	36.757	183.214	6.864	2.762	20.627	14438
Vesting Equity (Millions)	3.041	95.342	0	0	0.502	14438
Vested Equity (Millions)	254.272	1299.49	35.207	8.12	116.431	14438
Unvested Equity (Millions)	23.026	350.146	5.907	0.952	15.882	14438
Salary (Millions)	0.951	0.473	0.95	0.727	1.143	14438
Bonus (Millions)	0.182	0.82	0	0	0	14438
CEO Tenure	6.237	4.93	5	2	9	14024
Herfindahl Index	0.123	0.096	0.095	0.06	0.151	14438
Number of Segments	2.356	1.438	2	1	3	14438
Capital Constraint Index	0.287	0.206	0.242	0.13	0.406	14136
Firm Age	35.051	19.455	30.314	18.762	53.036	14438
SG&A-to-Sales	0.177	0.151	0.15	0.057	0.264	14351
R&D-to-Sales	0.061	0.339	0	0	0.015	14351
Major Customer Sales	0.151	0.252	0	0	0.227	14438
Market Beta	1.196	0.588	1.119	0.804	1.471	13901
Return-on-Assets	0.025	0.036	0.024	0.01	0.037	14349
Industry-Adjusted Sales Growth	0.04	0.263	0.001	-0.05	0.07	14186
Tobin's Q	1.945	1.8	1.445	0.936	2.306	14388
Stock Return Momentum	2.39	105.754	0.151	-0.039	0.405	13934
Leverage Ratio	0.285	0.201	0.262	0.134	0.393	14388
Retained Earnings	0.212	0.547	0.266	0.05	0.458	14294
Cash Balances	0.093	0.096	0.064	0.026	0.128	13613
Technology Acquisitions (Counts)	0.035	0.212	0	0	0	14438
Technology Acquisitions (Indicator)	0.03	0.17	0	0	0	14438
Salary-Weighted Proportion of Software Workers	3.753	3.559	2.736	1.667	4.771	14438
Proportion of Software Workers	3.276	3.315	2.328	1.313	4.264	14438
Proportion of AI Worker Hiring	0.005	0.016	0	0	0.004	14438
Indicator of AI Worker Hiring	0.385	0.487	0	0	1	14438

This table reports summary statistics of the main variables studied in the paper. The sample consists of US-listed non-IT firms over 2010Q1-2020Q4 with valid *Burning Glass*, *Revelio Labs* and *ISS Incentive Lab* observations. The variables reported are: Total assets (in Billions), vesting equity (millions), vested equity (millions), unvested equity (millions)., salary (millions), bonus (millions), CEO tenure, herfindahl index, number of segments, firm age, SG&A-to-sales, R&D-to-sales, proportion of sales to major customers, market beta, return-on-assets, industry-adjusted sales growth, Tobin's q, leverage ratio, retained earnings, cash balances, technology acquisitions (counts or indicator), salary-weighted proportion of software workers, proportion of software workers, proportion of AI worker hiring and indicator of AI worker hiring.

4.4.2 Expected Digital Investment Productivity, Capital Constraints and Short-Term Market Incentives

In the first part of my analysis, I examine the effects of expected digital investment productivity and capital constraints on digital investments. In Panel A of Table 4.5, I find that the textual overlap proxy of expected digital investment productivity is positively related to the extent of digital investments. Estimates show that a 1 standard deviation increase in the textual overlap score, is associated with a 0.8% increase in technology acquisitions (or a 26% increase relative to the average probability of technology acquisitions), a 0.27% increase in software workers (or a 8% increase relative to the average proportion of software workers) and a 0.1% increase in AI worker hiring (or a 20% increase relative to the average intensity of AI worker hiring).

On the other hand, I also find evidence that capital constraints are binding, as I find a negative relationship between the capital constraint index and several of the digital investment variables. A 10% increase in the percentile rank of the capital constraints index is associated with a 0.5% reduction in technology acquisitions (or a 16% decrease relative to the average probability of technology acquisitions) and a 0.05% decrease in the intensity of AI worker hiring (or a 10% decrease relative to average AI hiring intensity).

Lastly, I examine the associations between the vesting equity variable and the digital investment variables. While the relationship between vesting equity and the software worker/AI worker hiring variables is statistically insignificant, I find a significantly negative relationship between vesting equity and technology acquisitions. Specifically, the estimates suggest that for a 1 standard deviation increase in vesting equity, the probability of a technology acquisition falls by 0.2% (or a 6% decrease relative to the average probability of technology acquisitions).

Table 4.5: Expected Digital Investment Productivity, Capital Constraints and Short-Term Market Incentives

Dependent Variable	Technology Acquisitions (Indicator)	Technology Acquisitions (Count)	Proportion of Software Workers	Salary-Weighed Software Workers	AI Hiring Probability	AI Hiring Intensity
Patent Value-Weighted AI Index$_{j,t-1}$	0.008**	0.008*	0.270***	0.321***	0.009	0.001***
	(0.004)	(0.004)	(0.075)	(0.084)	(0.006)	(0.000)
Capital Constraints Index$_{i,q}$	-0.054***	-0.063***	-0.297	-0.343	-0.565***	-0.005**
	(0.016)	(0.021)	(0.556)	(0.624)	(0.074)	(0.002)
Vesting Equity$_{i,q}$	-0.020***	-0.023***	-0.047	-0.030	0.025	-0.000
	(0.006)	(0.006)	(0.146)	(0.160)	(0.028)	(0.001)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
6-Digit GICS FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,984	9,984	9,984	9,984	9,984	9,984
R^2	0.080	0.073	0.524	0.508	0.295	0.225

This table reports the effect of digital investment productivity, capital constraints and short-term market incentives on digital investment. The sample consists of US-listed non-IT firms over 2010Q1-2020Q4 with valid *Burning Glass*, *Revelio Labs* and *ISS Incentive Lab* observations. To proxy for digital investment productivity, I use the patent value-weighted AI index. To proxy for capital constraints, I use the Whited-Wu index of financial constraints (Whited and Wu 2006). To proxy for short-term market incentives, I use the vesting equity in a quarter, following the approach in Edmans et al. (2017). Regressions control for vested equity, unvested equity, salary, bonus, tenure, indicator for new CEO, firm age, lagged return-on-assets, lagged and current tobin's q, herfindahl index, market beta, industry-adjusted sales growth, SG&A-to-sales, proportion of major customer sales, number of segments, R&D-to-sales, missing R&D indicator, past 3-month stock return momentum, lagged leverage ratio, lagged retained earnings, lagged cash balances. 6-digit GICS group and calendar fixed effects are also included. Standard errors are clustered at the firm and year level. ***, **, * denote significance at the 1%, 5% and 10% level.

4.5 Conclusion

With a view to provide a better understanding of how capital markets shape digital technology investments, this study provides new evidence on the various capital market forces influence the investment in digital technologies. I examine three capital market forces, namely, the expected productivity of digital investment, as measured by the patent value-weighted AI index, capital constraints, as measured by the Whited and Wu (2006) index and short-term market incentives, as measured by vesting equity.

Consistent with the prior literature on corporate investment behavior, I find that companies with higher expected digital investment productivity and lower capital constraints tend to invest more in digital technologies. Furthermore, I find some evidence that short-term market incentives, is associated with lower digital investments, as measured by technology acquisitions.

Taken together, my study highlights the importance of capital markets in the investment in digital technologies, and shows that the capital market forces can play a significant role in shaping investment in these sets of technologies.

5
Conclusion

An emerging trend in financial markets in this digitization (or digital transformation) of the non-IT firms. With a view to both document this phenomenon and to understand the role that equity markets play in the digital technology investment in these firms, my dissertation examines (1) market valuations of digital technology investment, (2) the views of key market intermediaries and (3) the relationship between managerial stock price sensitivity and the investment in these technologies.

In the first chapter of the dissertation, I show in joint work with Suraj Srinivasan, that equity markets value digital investments (as measured by disclosure of digital terms in the 10-K) in non-IT firms. While a sizeable part of the valuation increases due to digital investment is coming from increasing intangibility of assets, digital investing non-IT firms still exhibit sizeable increases in valuation after adjustment for conservatism effects and these firms also yield significant future returns. On the other hand, we find limited evidence of future performance changes conditional on digital investment, which is suggestive that markets are pricing in potential benefits that will pay off only in the long-run.

In the second chapter of the dissertation, I measure the financial analysts' views of digitization in non-IT firms through their questions on the conference call, and I find two key results. First, the digital-related questions posed by analysts, and the sentiment of these questions correlate with several factors that determine digital technology adoption success, which thus suggests that analysts function as effective information analyzers. Second, the digital-related questions posted by financial analysts are strongly related with future investment in advanced digital technologies, in the form of AI-related job postings. Cross-sectional analyses suggest that analysts are playing an active role in encouraging firms to invest more in these new technologies, as the positive association is stronger in situations where the firm (1) not currently investing in AI or (2) not disclosing digital-related topics. Further economic mechanism analyses, suggest that the positive association could be due to two factors – (1) firms are viewing analysts' digital questions as market views on digital technologies and are investing to respond to those views, and (2) analysts are providing information/expertise to firms to encourage greater investment in AI technologies. In sum, the weight of the evidence suggests that analysts are playing a positive role in encouraging greater firm investment in AI technologies.

In the third chapter of the dissertation, I examine the effect of expected digital investment productivity, capital constraints and short-term market incentives, on digital technology investment. Specifically, for short-term market incentives, which is proxied by quarterly vesting equity, I examine the relationship between these incentives and three measures of digital investment – namely, technology acquisitions, proportion of software workers and AI job vacancies. My analysis yields mixed results, as I find no statistically significant relationship between vesting equity and the proportion of software workers and AI job vacancies. However, I do find a significantly relationship between vesting equity and technology acquisitions, which suggest that greater sensitivity to short-term stock prices reduces the investment in digital technology through acquisitions.

Overall, my research on the role of equity market views and digital technology investment shows that (1) equity markets plays an important role in assessing these investments in non-IT firms, and (2) provides some evidence that equity markets potentially plays a role in the investment of more advanced digital technologies in more traditional, non-IT firms.

A

Appendix to Chapter 2: Going Digital Implications for Firm Value and Performance

A.1 Digital Terms Regex Definitions

<table>
<tr><th>Digital Term</th><th>Regex Expression</th></tr>
<tr><td>Analytics:</td><td></td></tr>
<tr><td>analytics</td><td>(\banalytics\b)</td></tr>
<tr><td>proprietary algorithm</td><td>(\bproprietary algorithm)</td></tr>
<tr><td>business intelligence</td><td>(\bbusiness intelligence\b) | (\bcustomer intelligence\b) | (\boperating intelligence\b)</td></tr>
<tr><td>virtual reality</td><td>(\baugmented reality\b)|(\bvirtual realit)</td></tr>
<tr><td>Automation:</td><td></td></tr>
<tr><td>automation</td><td>(\bautomation solutions\b) | (\bintelligent automation\b) | (\bmarketing automation\b) | (\bprocess automation\b) | (\brobotic process automation\b)</td></tr>
<tr><td>autonomous technology</td><td>(\bautonomous ?[-]?tech)</td></tr>
<tr><td>AI:</td><td></td></tr>
<tr><td>artificial intelligence</td><td>(artificial ?[-]?intelligence) | (\bai ?[-]?tech) | (\bai ?[-]?related) | (\bconversational ai\b) | (\bevolutionary ai\b) | (\bevolutionary computing\b)</td></tr>
<tr><td>intelligence</td><td>(\bintelligent ?[-]?system)|(\bcomputer ?[-]?vision)</td></tr>
<tr><td>neural network</td><td>(\bneural ?[-]?network)</td></tr>
<tr><td>virtual assistant</td><td>(\bvirtual agent)|(\bvirtual ?[-]?assistant)</td></tr>
<tr><td>cognitive computing</td><td>(\bcognitive computing\b)</td></tr>
<tr><td>Big Data:</td><td></td></tr>
<tr><td>big data</td><td>(\bbig ?[-]?data)|(\bsmart ?[-]?data)</td></tr>
<tr><td>data science</td><td>(\bdata ?[-]?scien)</td></tr>
<tr><td>data mining</td><td>(\bdata ?[-]?mining)</td></tr>
<tr><td>data lake</td><td>(\bdata lake\b)</td></tr>
<tr><td>devops</td><td>(\bdevops\b)</td></tr>
<tr><td>digital twin</td><td>(\bdigital twin\b)</td></tr>
<tr><td>edge computing</td><td>(\bedge computing\b)</td></tr>
<tr><td>biometric</td><td>(\bbiometric)</td></tr>
</table>

Cloud:	
cloud platforms	(\bcloud ?[-]?platform) | (\bcloud ?[-]?based) | (\bcloud ?[-]?computing) | (\bcloud ?[-]?deployment)
cloud enablement	(\bcloud enablement\b) | (\bhybrid cloud\b)
virtual machines	(\bvirtual ?[-]?machine)
Digitization:	
digitization	(\bdigiti)|(\bdigital ?[-]?transformation)|(\bdigital ?[-]?revolution)
digital strategy	(\bdigital ?[-]?strateg)
digital marketing	(\bdigital ?[-]?marketing)
ML:	
deep learning	(\bdeep ?[-]?learning)
machine learning	(\bmachine ?[-]?learning)
NLP	(\bnatural ?[-]?language ?[-]?processing)
image recognition	(\bimage ?[-]?recognition)|(\bfacial ?[-]?recognition)
speech recognition	(\bspeech ?[-]?recognition)

A.2 IT Industry Classification Codes

Industry Codes	Industry Description
SIC Codes	
3570	Computer & Office Equipment
3571	Electronic Computers
3572	Computer Storage Devices
3575	Computer Terminals
3576	Computer Communications Equipment
3577	Computer Peripheral Equipment, NEC
3578	Calculating and Accounting Machines
3579	Office Machines, NEC
3661	Telephone & Telegraph Apparatus
3663	Radio & TV Broadcasting & Communications Equipment
3669	Communications Equipment, NEC
3670	Electronic Components & Accessories
3672	Printed Circuit Boards
3674	Semiconductors & Related Devices
3675	Electronic Capacitors
3675	Electronic Resistors
3677	Electronic Coils, Transformers & Other Inductors
3678	Electronic Connectors
3679	Electronic Components, NEC
4812	RadioTelephone Communications
4813	Telephone Communications (No Radiophone)
4899	Communications Services, NEC
7370	Services-Computer Programming, Data Processing, Etc.
7371	Services-Computer Programming Services
7372	Services-Prepackaged Software
7373	Services-Computer Integrated Systems Design
7374	Services-Computer Processing & Data Preparations
7375	Services-Information Retrieval
7376	Services-Computer Facilities Management Services
7377	Services-Computer Rental & Leasing

7378	Services-Computer Maintenance & Repair
7379	Services-Computer Related Services, NEC
NAICS Codes	
334100 (334110)	Computer and peripheral equipment manufacturing
334111	Electronic Computer Manufacturing
334112	Computer Storage Device Manufacturing
334118	Computer Terminal and Other Computer Peripheral Equipment Manufacturing
334200	Communications equipment manufacturing
334210	Telephone Apparatus Manufacturing
334220	Radio and Television Broadcasting and Wireless Communications Equipment Manufacturing
334290	Other Communications Equipment Manufacturing
334310	Audio and Video Equipment Manufacturing
334400 (334410)	Semiconductor and other electronic component manufacturing
334412	Bare Printed Circuit Board Manufacturing
334413	Semiconductor and Related Device Manufacturing
334416	Capacitor, Resistor, Coil, Transformer, & Other Inductor Manufacturing
334417	Electronic Connector Manufacturing
334418	Printed Circuit Assembly (Electronic Assembly) Manufacturing
334419	Other Electronic Component Manufacturing
334613	Blank Magnetic & Optical Recording Media Manufacturing
334614	Software & Other Prerecorded Compact Disc, Tape & Record Reproducing
335921	Fiber Optic Cable Manufacturing
511200	Software Publishers
511210	Software Publishers
517311	Wired Telecommunications Carriers
517312	Wireless Telecommunications Carriers (except Satellite)
517410	Satellite Telecommunications
517900 (517910)	Other Telecommunications
517911	Telecommunications Resellers
517919	All Other Telecommunications
518200 (518210)	Data Processing, Hosting & Related Services
519130	Internet Publishing & Broadcasting & Web Search Portals
541500 (541510)	Computer Systems Design and Related Services

541511	Custom Computer Programming Services
541512	Computer Systems Design Services
541513	Computer Facilities Management Services
541519	Other Computer Related Services
611420	Computer Training
GICS Codes	
25502010	Catalog Retail
25502020	Internet & Direct Marketing Retail
20201020	Data Processing Services
45101010	Internet Software & Services
45102010	IT Consulting & Other Services
45102020	Data Processing & Outsourced Services
45102030	Internet Services & Infrastructure
45103010	Application Software
45103020	Systems Software
45103030	Home Entertainment Software
45201010	Networking Equipment
45201020	Communications Equipment
45202010	Computer Hardware
45202020	Computer Storage & Peripherals
45202030	Technology Hardware, Storage & Peripherals
45203010	Electronic Equipment & Instruments
45203015	Electronic Components
45203020	Electronic Manufacturing Services
45203030	Technology Distributors
45205010	Semiconductor Equipment
45205020	Semiconductors
45301010	Semiconductor Equipment
45301020	Semiconductors
50101010	Alternative Carriers
50101020	Integrated Telecommunication Services
50102010	Wireless Telecommunication Services

A.3 Variable Definitions

Variable Name	Variable Description
Fundamental Variables:	
SIZE	Logarithm of Market Capitalization at the fiscal year end (*prc* × *shrout* in *CRSP*).
LEV	Leverage Ratio, defined as total debt divided by stockholder's equity. ($\frac{dlc+dltt}{seq}$ in *Compustat*)
LOSS	Coded as 1 where operating income after depreciation (*oiadp* in *Compustat*) is positive and 0 otherwise.
PERS	The AR(1) coefficient in quarterly EPS (defined as earnings per share before extraordinary items, *epspxq* in *Compustat*) regressed on EPS reported in the same quarter in the previous year, estimated over rolling 4-year windows (following, Piotroski et al, 2017).
MB	Market-to-Book Ratio, defined as the ratio of the market capitalization at the day after the 10-K filing date divided by stockholder's equity (*seq* in *Compustat*).
CC-MB	The Conservatism-Corrected Market-to-Book Ratio is defined as market-to-book adjusted for net financing assets, divided by the conservatism correction factor (following, McNichols et al. 2014). See the IA for more details on the computation of this factor.
ROA	Defined as operating income after depreciation scaled by average total assets from beginning to end of the fiscal period ($\frac{oiadp}{(at_t+at_{t-1})/2}$ in *Compustat*).
RNOA	Defined following Soliman (2008) as operating income after depreciation scaled by average net operating assets from beginning to end of the fiscal period ($\frac{oiadp}{(noa_t+noa_{t-1})/2}$ in *Compustat*). Net operating assets is defined as the difference in operating assets ($at - che - ivao$ in *Compustat*) and operating liabilties ($at - dltt - dlc - ceq - pstk - mib$ in *Compustat*). As net operating assets are not well-defined for finance firms (SIC code 6000-6799), we drop these firms from the estimation.

AGE	Logarithm of Firm Age. Age is determined by the number of years since the firm first appeared in *Compustat*.
MARGINS	Profit margins defined as operating income after depreciation scaled by sales ($\frac{oiadp}{sale}$ in *Compustat*).
ATO	Sales scaled by average total assets from beginning to end of the fiscal period. ($\frac{sale_t}{(at_t+at_{t-1})/2}$ in *Compustat*)
NOATO	Sales scaled by average net operating assets from beginning to end of the fiscal period.
SALES GROWTH$_{t+s,t}$	Sales Growth, annualized difference in future period $t + s$ sales minus current sales in period t scaled by the current period sales. ($\frac{sale_{t+s}-sale_t}{sale_t}$ in *Compustat*).
CapEx	Capital expenditure intensity, defined as capital expenditures scaled by assets ($\frac{capx}{at}$ in *Compustat*)
Missing R&D	Indicator variable coded as 1 if the R&D variable is missing and 0 otherwise.
SG&A	Selling, General and Administrative expenditure intensity, defined as SG&A expense minus R&D minus R&D in place scaled by assets ($\frac{xsga-xrd-rdip}{at}$ in *Compustat*). Following Peters and Taylor (2017) we keep the original *xsga* as SG&A expense if $xrd > xsga$ but $xrd < cogs$. If missing this variable is coded as 0.
R&D	Research and development expenditure intensity, defined as R&D scaled by assets ($\frac{xrd}{at}$ in *Compustat*). If missing, this variable is coded as 0.
Operating Profit	Defined as sales (*sale* in *Compustat*) minus cost of goods sold (*cogs* in *Compustat*) minus interest expense (*xint* in *Compustat*) minus sales, general and administrative expense (*xsga* in *Compustat*), divided by book equity defined above.
Investment	Defined as the difference in current period assets (*at* in *Compustat*) and lagged assets scaled by lagged assets.

Returns and Liquidity Variables:

β	The beta coefficient estimated from a regression of daily returns on *CRSP* value-weighted market returns over the fiscal year.

β_{Tech} or β_{NTech}	The IT or non-IT beta coefficient on the technology portfolio estimated from a regression of the following factor model: $R_{i,t} = \alpha_{i,t} + \beta_{Tech} R_{Tech,t} + \beta_{NTech} R_{NTech,t}$ that is estimated over the fiscal year. $R_{Tech,t}$ is the value-weighted portfolio returns of IT firms that are defined in Appendix B, and the portfolio is re-balanced daily. $R_{NTech,t}$ is the value-weighted portfolio returns of non-IT firms, and the portfolio is re-balanced daily. Daily returns with stock prices of less than \$5 are dropped from the estimation, and β_{Tech} or β_{NTech} estimated with less than 200 days of returns observations are also dropped.
Market-Adj. Returns	The buy-hold raw returns in the fiscal year minus the value-weighted *CRSP* market return distribution.
Share Turnover	Monthly share volume divided by the shares outstanding ($\frac{vol}{shrout}$ in *CRSP*), averaged over the fiscal year.
Return Volatility	Standard deviation of daily returns estimated over the fiscal year.
Momentum	Defined as the past 12 month to lagged 1 month buy-and-hold returns minus the buy-and-hold value-weighted market portfolio return over the same period.

Earnings Announcement Variables:

Unexpected Earnings or Sales	Unexpected earnings is actual minus median earnings (sales) per share forecasts scaled by the price at the end of the fiscal period from *IBES*. The median earnings (sales) forecasts is based on the most recent analyst consensus forecast, within 100 days before the earnings announcement. We remove observations where the price at the end of the fiscal period is less than \$1 and where the earnings surprise is in excess of price.
Days to EA	Number of business days between earnings announcement and fiscal year end.
Days to 10-K Filing	Number of business days between 10-K filing and fiscal year end.
Days Between 10-K & EA	Number of business days between 10-K filing and Earnings Announcement.
CAR (-1,40)	The cumulative adjusted returns from 1 day before the earnings announcement to 40 trading days after. Benchmark returns are estimated using the coefficients from the Carhart, Fama-French Four-Factor model (Carhart 1997) that are estimated based on a (-280, -60) window.

Digital Activity:	
Digital Score	A quantized score coded as 0 for no digital disclosure, 1 for yearly below tercile disclosure, 2 for yearly middle tercile disclosure and 3 for yearly top tercile disclosure. Counts of digital terms are taken from the business description section of the 10-K, and digital terms are taken from the dictionary of terms in Appendix A.
Digital Score (Returns Portfolio)	A score coded as -1 for no digital disclosure, 1 for yearly top tercile disclosure, and 0 otherwise.
Non-continuous/Continuous Digital Score (Returns Portfolio)	For non-continuous digital scores, we re-code the digital score as 0 for firms that continuously make top tercile disclosures of digital terms in the returns window (either 2-years or 3-years). For continuous digital scores, we re-code the digital score as 0 for all firms that do not make continuous top tercile disclosures of digital terms in the returns window.
Other Variables:	
Tech Manager	An indicator that is set to 1 and 0 otherwise, if one of the firm's top 5 executive has a technology-related managerial title. We define technology-related titles as either "VP Digital", "Chief Information Officer (CIO)" or "Chief Technology Officer (CTO)". Data on the top 5 executives is sourced from *Capital IQ's People Intelligence database*.
Total Words	Natural Logarithm of the total number of words in the business description section. Data taken from 10-K filings obtained through *SEC Edgar*.
Digital Patents	An indicator that is coded as 1 and 0 otherwise, if the firm files a patent relating to digital technologies. We identify such patents by the title and abstract search terms relating to AI, cloud computing, machine learning, neural network, robots, self-driving cars and software as defined in Bloom et al. (2018) and Webb (2020).
Proportion of IT workers	The proportion of IT workers relative to total number of employees in the firm. Data is from *Revelio Labs*.

A.4 Examples of Digital Disclosure

Mistras Group Inc, Fiscal Year: 2011
Historically, NDT solutions predominantly used qualitative testing methods aimed primarily at detecting defects in the tested materials. This methodology, which we categorize as traditional NDT, is typically labor intensive and, as a result, considerably dependent upon the availability and skill level of the certified technicians, engineers and scientists performing the inspection services. The traditional NDT market is highly fragmented, with a significant number of small vendors providing inspection services to divisions of companies or local governments situated in close proximity to the vendor s field inspection engineers and scientists. Today, we believe that customers are increasingly looking for a single vendor capable of providing a wider spectrum of asset protection solutions for their global infrastructure that we call one source . This shift in underlying demand, which began in the early 1990s, has contributed to a transition from traditional NDT solutions to more advanced solutions that employ automated digital sensor technologies and accompanying enterprise software, allowing for the effective capture, storage, analysis and reporting of inspection and engineering results electronically and in digital formats. These advanced techniques, taken together with advances in wired and wireless communication and information technologies, have further enabled the development of remote monitoring systems, asset-management and predictive maintenance capabilities and other data analytics and management. We believe that as advanced asset protection solutions continue to gain acceptance among asset-intensive organizations, only those vendors offering broad, complete and integrated solutions, scalable operations and a global footprint will have a distinct competitive advantage. Moreover, we believe that vendors that are able to effectively deliver both advanced solutions and data analytics, by virtue of their access to customers data, develop a significant barrier to entry for competitors, and so develop the capability to create significant recurring revenues.
Korn Ferry International, Fiscal Year: 2014
Talent Analytics

Companies are increasingly leveraging big data and analytics to measure the influence of activities across all aspects of their business, including HR. They expect their service providers to deliver superior metrics and measures and better ways of communicating results. Korn Ferry's go-to-market approach is increasingly focused on talent analytics we are injecting research-based intellectual property into all areas of our business, cascading innovation and new offerings up to our clients.
Insperity Inc., Fiscal Year: 2015
Our long-term strategy is to provide the best small and medium-sized businesses in the United States with our specialized human resources service offering and to leverage our buying power and expertise to provide additional valuable services to clients. Our most comprehensive HR services offerings are provided through our Workforce Optimization and Workforce Synchronization solutions (together, our PEO HR Outsourcing solutions), which encompass a broad range of human resources functions, including payroll and employment administration, employee benefits, workers compensation, government compliance, performance management and training and development services, along with our cloud-based human capital management platform, the Employee Service Center (ESC). Our Workforce Optimization solution is our most comprehensive HR outsourcing solution and is our

primary offering. Our Workforce Synchronization solution, which is generally offered only to our mid-market client segment, is a lower cost offering with a longer commitment that includes the same compliance and administrative services as our Workforce Optimization solution and makes available, for an additional fee, the strategic HR products and organizational development services that are included with our Workforce Optimization solution.

TransUnion, Fiscal Year: 2015

Our addressable market includes the big data and analytics market, which continues to grow as companies around the world recognize the benefits of building an analytical enterprise where decisions are made based on data and insights, and as consumers recognize the importance that data and analytics play in their ability to procure goods and services and protect their identities. International Data Corporation ("IDC") estimates worldwide spending on big data and analytics services to be approximately $52 billion in 2014, growing at a projected compounded annual growth rate (CAGR) of approximately 15% from 2014 through 2018. There are several underlying trends supporting this market growth, including the creation of large amounts of data, advances in technology and analytics that enable data to be processed more quickly and efficiently to provide business insights, and growing demand for these business insights across industries and geographies. Leveraging our 48-year operating history and our established position as a leading provider of risk and information solutions, we have evolved our business by investing in a number of strategic initiatives, such as transitioning to the latest big data and analytics technologies, expanding the breadth and depth of our data, strengthening our analytics capabilities and enhancing our business processes. As a result, we believe we are well positioned to expand our share within the markets we currently serve and capitalize on the larger big data and analytics opportunity.

Camping World Holdings, Inc., Fiscal Year: 2017

Customer Database. We have over 15.1 million unique RV contacts in our database of which approximately 3.6 million are Active Customers related to our RV products. We use a customized CRM system and database analytics to track customers and selectively market and cross-sell our offerings. We believe our customer database is a competitive advantage and significant barrier to entry.

A.5 Additional Methodology Details and Analysis

A.5.1 Summary Statistics of Technology Firms

In this Table A.1 of this section, we report the summary statistics of the IT sample of firms. The definition of IT firms is presented in Appendix B.

Table A.1: Summary Statistics of IT Firms

	Mean	Std Dev	Median	25%	75%	N
Market Cap. (Millions)	5218	15773	670	116	2954	5707
Market-to-Book	4.22	5.27	2.49	1.45	4.56	5707
Firm Age	20	12	18	12	26	5707
Leverage Ratio	0.56	1.04	0.17	0	0.65	5662
β	1.07	0.5	1.09	0.78	1.37	5700
β_{Tech}	0.68	0.53	0.61	0.3	0.99	4090
β_{NTech}	0.45	0.51	0.46	0.12	0.78	4090
Earnings Persistence	0.16	0.39	0.07	-0.1	0.36	5557
Return Volatility	0.03	0.01	0.03	0.02	0.04	5701
Days to EA	32	12	31	24	40	4847
Days to 10-K Filing	45	12	42	37	52	5707
Days Between 10-K & EA	10	9	9	1	16	4847
EA CAR(-1,40)	-0	0.18	-0	-0.1	0.09	4173
Unexpected Earnings	-0	0.04	0	0	0	4216
Unexpected Sales	0	0.05	0	-0	0	4216
Market-Adj. Annual Returns	0.06	0.69	-0.03	-0.26	0.24	5699
Return-on-Assets	0.02	0.14	0.05	-0.03	0.1	5706
Return-on-Net Operating Assets	0.05	0.57	0.1	-0.05	0.24	4681
Profit Margins	0.04	0.19	0.06	-0.02	0.14	5585
Asset Turnover	0.91	0.56	0.77	0.54	1.12	5706
Net Operating Asset Turnover	2.87	2.85	1.95	1.19	3.39	4681
Sales Growth$_{t,t-3}$	0.08	0.16	0.06	-0.02	0.16	5707
SG&A Expense	0.25	0.19	0.2	0.11	0.32	5707
R&D Expense	0.1	0.08	0.08	0.04	0.14	4553
Loss (Indicator)	0.31	0.46	0	0	1	5707
Tech Manager	0.11	0.31	0	0	0	5707
Total Digital Words	6.7	13.61	2	0	8	5707
Quantized Digital Score	1.22	1.13	1	0	2	5707
Initial Digital Disclosure	0.04	0.2	0	0	0	5707
Digital Patents (Indicator)	0.23	0.42	0	0	0	4768
IT Workers	19.68	12.47	17.47	9.53	28.62	3004

We report the summary statistics of the main control variables in this table for the sample of non-IT firms in fiscal years 2010-2020.We examine the statistics of the following variables: market capitalization, market-to-book, firm age, leverage ratio, market beta, beta with respect to the IT and non-IT portfolios, earnings persistence, return volatility, no. of days from fiscal year end to earnings announcement, no. of days from fiscal year end to 10-K filing, no. of days between earnings announcement and 10-K filing, 40-day cumulative abnormal returns after the earnings announcement, unexpected earnings, unexpected sales, market-adjusted annual returns, return-on-assets, return-on-net operating assets, profit margins, asset turnover, net operating asset turnover, past 3-year sales growth (annualized), SG&A expense, R&D expense, an indicator for loss firms and an indicator for firms with technology-related top executives, digital patents (indicator), percentage of IT workers. Descriptions of the variables are outlined in detail in Appendix C in the main text.

A.5.2 Conservatism-Corrected Market-to-Book Analysis

In this section, we describe the methodology and the assumptions behind the conservatism-corrected market-to-book proxy (McNichols et al. 2014) used in Panel B of Table 2.5. Broadly speaking, the conservative correction for market-to-book is an estimate of the ratio of the replacement cost of assets to book value of assets. Empirically, the McNichols et al. (2014) methodology estimates this ratio by using the firm's investment history, proportion of investment in intangibles and an estimate of the useful life of assets.

To help the reader better understand the basic mechanics of market-to-book conservatism correction factor, we provide a brief sketch of the theory behind the McNichols et al. (2014) conservatism correction method. First, we note that the motivation for the correction stems from the fact that market-to-book is the ratio of the market value of the history of investments divided by the book value of the history of investments, which may be understated due to capitalization restrictions on certain types of investments. Thus, to isolate the component of market-to-book that reflects future growth opportunities, one approach would be to apply a conservatism correction that converts the book value of assets in the denominator to the replacement cost of assets. Specifically, one could divide the standard market-to-book ratio by the following ratio:

$$CC_T = \frac{BV_T(I_T, d^*)}{BV_T(I_T, d^0)} = \frac{bv^*_{T-1} \cdot I_1 + bv^*_{T-2} \cdot I_2 + \ldots + bv^*_T \cdot I_T}{bv^0_{T-1} \cdot I_1 + bv^0_{T-2} \cdot I_2 + \ldots + bv^0_T \cdot I_T} \tag{A.1}$$

where CC_T is the conservatism correction factor, $BV_T(I_T, d^*)$ is the value of the history of investments I_T depreciated using replacement accounting rules (d^*) over the length of useful life of assets T. $BV_T(I_T, d^0)$ is the value of the history of investments I_T depreciated under unconditional conservatism accounting rules (d^0) over the length of useful life of assets T. Furthermore, with this definition of CC_T, we obtain an estimate of Tobin's q for an all-equity firm, by applying this factor to market-to-book: $q \equiv \frac{MB_T}{CC_T}$.

Empirically, we estimate the denominator of the CC_T by making the assumption that the firm depreciates its assets using straight-line depreciation. Thus, for a stream of investments over the useful life of assets T, which we estimate by dividing average (beginning and end fiscal year) net ppe and intangibles divided by depreciation, we compute the book value of assets as:

$$\begin{aligned} D_T = (1-\alpha_1)(1-\tfrac{T-1}{T}) + (1-\alpha_2)(1-\tfrac{T-2}{T})(1+\lambda_2) + \ldots \\ +(1-\alpha_{T-1})(1-\tfrac{1}{T})\prod_{i=2}^{T-1}(1+\lambda_i) + (1-\alpha_T)\prod_{i=2}^{T-1}(1+\lambda_i) \end{aligned} \quad \text{(A.2)}$$

where D_T is the aggregated book value of investments. α_t is the proportion of intangible investments (defined as advertising, R&D expense divided by advertising, R&D expense and capital expenditures) for $t \in 1 \ldots T$ history of investment periods. λ_t is the investment growth rate (defined as the sum of advertising, R&D and capital expenditures in the current period divided by the lagged sum minus 1) for $t \in 1 \ldots T$ history of investment periods. To build intuition, note that $1-\alpha_t$ is the normalized unit amount of investment that is capitalized in period t and these investments are straight-line depreciated to period T by $1-\frac{T-t}{T}$. These invested assets are growing to period T at a growth rate of $\prod_{i=2}^{T-t}(1+\lambda_i)$. Thus the product of these variables maps to $bv_t^0 \cdot I_t$ in equation 1.

For the numerator, we estimate the value of assets using the replacement cost accounting method. That is, we estimate the replacement cost of the total assets invested in each period and compute the total value of assets by summing the replacement costs. We first assume competitive markets for the assets, and so the per unit competitive price or replacement cost for the asset is given by:

$$c = \frac{1}{\sum_{t=1}^{T}\gamma_t} \quad \text{(A.3)}$$

where $\gamma_t = \frac{1}{(1+r)^t}$ is the discount factor and c the replacement cost is thus, defined as the reciprocal of the value of an annuity of \$1 paid over T years.

We apply an annuity depreciation for the book value of the asset in each investment period, and so $bv_t^* = c \cdot \sum_{i=t+1}^{T} \gamma_{i-t}$. Combined with the investment growth rates, the expression for the numerator of CC_T is thus the following:

$$N_T = \frac{1}{\Gamma^T} \cdot (\Gamma^1 + \Gamma^2(1+\lambda_2) + ... + \Gamma^T \prod_{i=2}^{T}(1+\lambda_i)) \tag{A.4}$$

where N_T is the aggregated replacement costs of investments. $\Gamma^t = \gamma + \gamma^2 + ...\gamma^t$, and $\gamma = \frac{1}{1+r}$ with r as the weighted average cost of capital. We estimate r by the following expression: $\frac{BV_T}{AT_T} \cdot r_e + \frac{AT_T - BV_T}{AT_T} r_d(1-\tau_T)$, where BV_T and AT_T are the book value of equity and assets respectively. r_e, the cost of equity capital is estimated at the end of each fiscal period for each firm with coefficients from the Fama and French (1992) two-factor model: $R_i - R_f = \delta_0 + \delta_1(R_m - R_f) + \delta_2(SMB) + \varepsilon$, which are estimated using monthly returns from the preceding five years. r_d, the cost of debt is estimated by dividing interest expense with the average of beginning and ending liabilities. τ_T is the statutory federal corporate tax rate. To map equation 4 back to the constructs outlined in the above discussion, note that $\frac{1}{\Gamma^T}$ is the per unit replacement cost of assets (c), and Γ^t is the annuity depreciation factor in period t. Combined with the compounded growth rate $\prod_{i=2}^{T-t}(1+\lambda_i)$, these variables map back to $bv_t^* \cdot I_t$ in equation 1.

Table A.2 presents the summary statistics of the key variables in our estimation of the conservatism corrected market-to-book. We first report the statistics of the market-to-book of the firms in this sub-sample in the first row, and these firms have an average market-to-book of around 3.34. Next, we adjust the raw market-to-book for net financial assets following McNichols et al. (2014). That is, we deduct the numerator and denominator of the market-to-book ratio by the sum of net ppe, intangibles minus liabilities (see Table A.2 for the summary statistics of this variable).

We then compute the conservatism-corrected factor following the methodology outlined above. To preserve sample size, we limit the estimated useful life of assets to at most 25 years (Livdan and

Nezlobin 2017)[1], and we interpolate any missing investment growth and percentage of intangible investment observations from the last reported value. Following McNichols et al. (2014) we also limit the sample to firms with greater than 4 million in assets, as well as firms that have a net ppe-to-asset ratio of above 0.1. The conservatism correction factor, reported in the second-to-last row, is greater than 1 reflecting the fact that the replacement cost of assets is greater than its reported book value. Finally, we report the conservatism-corrected market-to-book which has an average of 1.53 and a median of 1.14.

Table A.2: Summary Statistics of Conservatism-Corrected Market-to-Book Sub-Sample of Non-IT Firms

	Mean	Std Dev	Median	25%	75%	N
Market-to-Book	3.34	3.73	2.19	1.37	3.74	7377
Adjusted Market-to-Book	3.36	5.65	1.9	1.27	3.35	7377
Estimated Useful Life of Assets	18.69	5.59	19	14	25	7377
Conservatism Correction Factor	2.07	1.2	1.59	1.27	2.4	7377
Conservatism Corrected Market-to-Book	1.53	1.32	1.14	0.77	1.77	7377

We report the summary statistics of the main variables for the conservatism-corrected market-to-book analysis in this table for the sample of non-IT firms in fiscal years 2010-2020. Details of the variable construction methodology is outlined in the above text.

A.5.3 Instrument for Digital Adoption: AI Technology Exposure

In this section, we describe the methodology for constructing the AI technology exposure at the industry-level. The main intuition of the measure is to proxy for the level of the AI technology exposure, using the textual overlap between the title/abstract of AI-related patents and the industry descriptions in the NAICS classification.

We search for these patents from a database of global patents in the *Google Patents Database*[2] using *Google BigQuery*. To identify AI patents, we follow Webb (2020) and select patents that have at least one of the following terms in the title or abstract: "supervised learning", "reinforcement learning",

1. Livdan and Nezlobin (2017) use a maximum period of 20 years, but given the relatively high sample median of estimated useful life of assets (19 years relative to the typical length of 14 years reported in McNichols et al. (2014) and Livdan and Nezlobin (2017)) we use a longer maximum period of 25 years.

2. https://console.cloud.google.com/marketplace/partners/patents-public-data

“deep learning”, “neural network” and “machine learning”[3]. To ensure that patents identified in the search are not mechanically related to firm's patenting activity, we remove patents that are granted to US-listed firms using the *CRSP*-patent linked dataset in Kogan et al. (2017)[4].

In the second step, we use the Kogan et al. (2020) methodology to estimate the overlap between patents and the description of occupations and industries. The methodology first computes the word embedding of each document (either patent (abstract and title) and the 4-digit 2017 NAICS industry description (over all sub-industry description)), by characterising each word by GloVe vectors (Pennington et al. 2014)[5]. This method uses the co-occurrences of words to develop a mapping to a vector space which aligns words that are different but have similar meaning. To aggregate to the document level, we take the weighted average of the word vectors in the whole document by using the following formula:

$$X_i = \sum_{x_k \in A_i} w_{i,k} x_k \tag{A.5}$$

where X_i is the document vector and x_k is a word-vector in the set of vectors in document A_i. $w_{i,k}$ is the word TF-IDF weight which is determined as follows:

$$w_{i,k} = TF_{i,k} \times IDF_k \tag{A.6}$$

where $TF_{i,k}$ is term frequency of word k in document i and IDF_k is the inverse document frequency

3. Webb (2020) does not include “machine learning” as a search term, but we include this term as it is a important component of AI technologies

4. Taken from Noah Stoffman's website: https://github.com/KPSS2017/Technological-Innovation-Resource-Allocation-and-Growth-Extended-Data

5. We use the largest set of pre-trained vector representations provided on the StanfordNLP website: https://nlp.stanford.edu/projects/glove/. The dataset is “glove.840B.300d”, which is a dictionary trained on 840 B tokens from Common Crawl and maps words to 300-dimensional clusters.

of word k, which is computed as

$$IDF_k = log(\frac{\#\ documents}{\#\ documents\ with\ word\ k}) \tag{A.7}$$

The similarity scores are then estimated by:

$$\rho_{i,j} = \frac{X_i}{|X_i|} \cdot \frac{X_j}{|X_j|} \tag{A.8}$$

To estimate the industry-year level exposure score, we follow Kogan et al. (2020) and apply the following aggregation formula on cosine similarity scores:

$$\eta_{i,t} = \sum_{j \in \Gamma_t} \tilde{\rho}_{i,j} \tag{A.9}$$

where $\eta_{i,t}$ is the occupation or industry i AI exposure score in year t. $\tilde{\rho}_{i,j}$ is the adjusted similarity scores for occupation i and patent j. The similarity score adjustment takes two steps: (1) year fixed effects are removed from the similarity scores and (2) sparsity is imposed by setting scores below the 80th percentile to 0[6]. Finally, to normalize the exposure scores, we deduct the raw scores by the pooled mean and scale by the pooled standard deviation.

A.5.4 ERC/SRC Robustness Analysis

In this section, we examine the robustness of the ERC/SRC results presented in Table 2.6 in the main text. We first control for further heterogeneity in the ERC and SRC at the firm-level by including grouped firm fixed effects following the approach outlined in Gipper et al. (2019). These fixed effects are coded based on the 10 by 10 yearly decile ranks of size (market cap.) and market beta. As reported

6. Kogan et al. (2020) also scale the non-zero values to run from 0-1, but we do not do so as it gives undue weight to patent exposure in the early part of the sample.

in the below Table A.3, our inferences with these grouped firm fixed effects are similar to the results from our main specification.

Table A.3: ERC/SRC Analysis with Grouped Firm Fixed Effects

	Unexpected Earnings		Unexpected Sales	
	Raw Values	Yearly Deciles	Raw Values	Yearly Deciles
Dependent Variable	CAR(-1,40)	CAR(-1,40)	CAR(-1,40)	CAR(-1,40)
Unexpected Earnings or Sales$_{i,t}$	15.780***	0.020*	2.117*	0.016**
	(2.673)	(0.012)	(1.087)	(0.007)
Digital$_{i,t}$	0.001	-0.009**	0.001	-0.008*
	(0.002)	(0.005)	(0.002)	(0.004)
Digital$_{i,t}$ $\times$ Unexpected Earnings or Sales$_{i,t}$	0.497*	0.002**	0.179*	0.001**
	(0.254)	(0.001)	(0.108)	(0.001)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Grouped Firm FE	Yes	Yes	Yes	Yes
Unexpected Earnings $\times$ Controls	Yes	Yes	Yes	Yes
Unexpected Earnings $\times$ Time FE	Yes	Yes	Yes	Yes
Unexpected Earnings $\times$ Grouped Firm FE	Yes	Yes	Yes	Yes
Observations	14,309	14,309	14,006	14,006
Adj. R^2	0.0373	0.0431	0.0342	0.0370

We report the coefficients to the ERC (Earnings Response Coefficient)/SRC (Sales Response Coefficient) regression with the proxy for digital activities in this table for the sample of non-IT firms in fiscal years 2010-2020. Column 1 reports the ERC regression using raw unexpected earnings, where CAR(-1,40) is regressed on unexpected earnings, controls, grouped firm (indicators for yearly 10 by 10 decile groupings of firms with similar size and beta characteristics) and year fixed effects, as well as their interactions with unexpected earnings. Column 2 reports the same regression but with yearly unexpected earnings deciles. Column 3 reports the SRC regression using raw unexpected sales, where CAR(-1,40) is regressed on unexpected sales, controls, grouped firm and year fixed effects, as well as their interactions with unexpected sales. Column 4 reports the same regression but with yearly unexpected sales deciles. All regressions control for size, leverage ratio, loss (ind.), persistence, return volatility, past 3-year sales growth, market-to-book, SG&A expenditure, R&D expenditure, indicator for missing R&D, capital expenditures, the number of days to EA, the number of days to 10-K filing and the number of words in the business description section. In all regression models, we proxy for digital by using a quantized score of the number of digital mentions in the business description of the 10-K (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). For ease of interpretation of the unexpected earnings/sales coefficient, we mean-center all continuous control variables. Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

We next examine the ERC/SRC results with an analysis that further addresses the non-linearities in the earnings/sales-returns relationship (Gipper et al. 2019). We fit the earnings/sales-returns relationship using fractional polynomials of unexpected earnings and sales[7], and observe differences in the ERC/SRC curves for digital and non-digital firms. If digital firms exhibit greater ERC/SRCs then one would expect that ERC/SRC curves for digital firms to exhibit to be steeper than non-digital

7. We use the function *fp* in stata. As the functional polynomials include natural logarithms, we scale the unexpected earnings/sales by 0.02 and re-scale the fitted relationship in the plots presented in Figures A.1 and A.2 We also only keep unexpected earnings/sales observations within $\pm 2\%$ of price

firms. In Figure A.1 presented below, we plot the ERC curves and find that digital firms exhibit stronger return sensitivity to both positive (albeit only for the more extreme positive) and negative unexpected earnings. This is consistent with digital firms exhibiting a higher ERC coefficient, as the figure suggests that digital firms exhibit stronger returns reaction to earnings surprises.

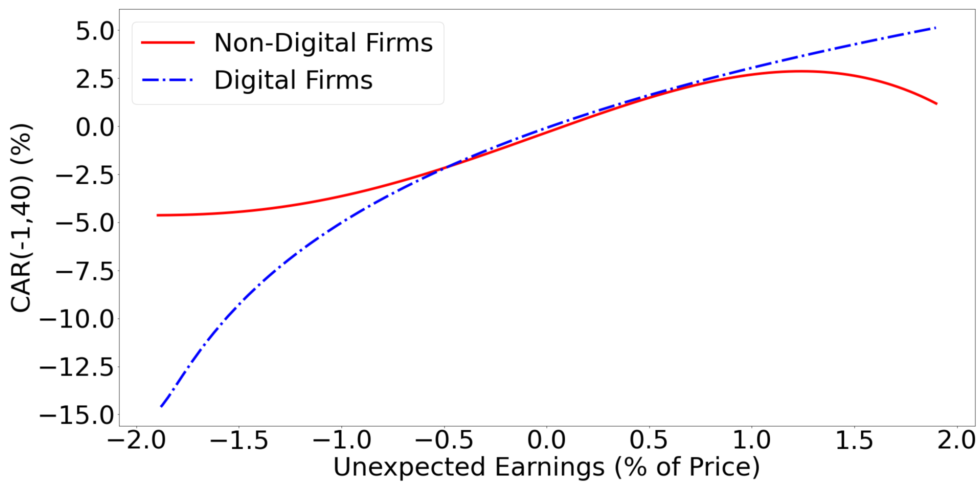


Figure A.1: Fitted Returns-Unexpected Earnings Relationship Using Fractional Polynomial Powers of Unexpected Earnings, For Digital and Non-Digital Firms

In Figure A.2 presented below, we plot the SRC curves using the same methodology as the ERC curves in Figure A.1 and find that digital firms exhibit stronger return sensitivity to both positive (albeit only for the less extreme positive) and negative unexpected earnings. Thus, this analysis also provides some support that digital firms exhibit higher SRC coefficients, as it suggests that digital firms experience higher return reactions for positive sales surprises.

A.5.5 Additional Analysis of Valuation and Accounting Performance

In this section, we present additional analysis of valuation and accounting metrics that are studied in our paper.

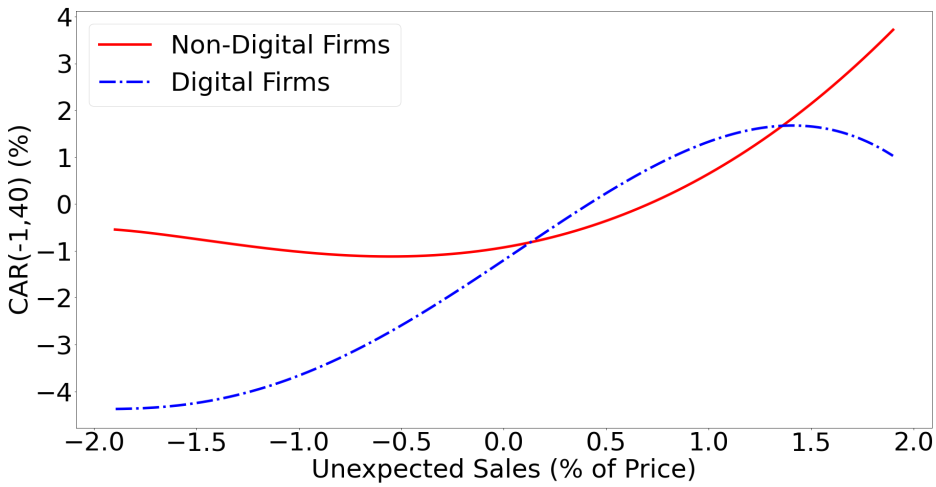


Figure A.2: Fitted Returns-Unexpected Sales Relationship Using Fractional Polynomial Powers of Unexpected Sales, For Digital and Non-Digital Firms

We begin with an additional analysis of the effect of digital activities on margin performance. We examine gross margins in Table A.4. We find that gross margins is lower for digital firms by 1.1-3.3% relative to peers (or 2.6-7.9% relative to the sample average of 42%). However, we do not find statistically significant changes in gross margins in the subsequent years after the initial digital disclosure.

To examine which digital firms receive higher valuations in the cross-section, we perform a regression of market-to-book on our digital proxy, interacted with various cross-sectional variables, namely: size, age, R&D, SG&A, CapEx, ROA, sales growth, leverage, cash balances and the industry-level of digital adoption. Results in Table A.5 show that firms that are larger, expend more on SG&A, CapEx and are in industries with high digital adoption, tend to receive higher valuations when they go digital. In particular, the positive relationship between size and digital activity suggests that digital technologies help larger companies increase the net benefits of scale, which is consistent with prior work that argues that IT technologies help companies expand and integrate more effectively (Hitt 1999; Baker and Hubbard 2004). Additionally, the finding that firms with higher SG&A and CapEx tend to re-

Table A.4: Gross Margins

Treatment Sample	Full Sample	Initial Digital Activity		
	Levels	One Year Ahead Change	Two Year Ahead Change	Three Year Ahead Change
Dependent Variable	$\text{MARGINS}_{i,t}$	MARGINS_{t+1}- $\text{MARGINS}_{i,t}$	MARGINS_{t+2} - $\text{MARGINS}_{i,t}$	MARGINS_{t+3} - $\text{MARGINS}_{i,t}$
$\text{Digital}_{i,t}$	-0.011**	-0.007	-0.007	0.002
	(0.004)	(0.005)	(0.006)	(0.006)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	20,108	11,874	9,604	7,793
Adj. R^2	0.5532	0.0844	0.1405	0.1850

We report the coefficients of regressions of gross margins (MARGINS) on the proxy for digital activities and controls in this table for the sample of non-IT firms in fiscal years 2010-2020. For the gross margins measure, we report the level differences of firms with digital activity versus peers, as well as the one-, two- and three-year-ahead change in the performance measure for firms that are initiating digital activities (relative to firms with no digital activity), in columns 1-4, respectively. In all regression models, we proxy for digital by using a quantized score of the number of digital mentions in the business description of the 10-K (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). All regressions control for size, firm age, leverage, past 3-year sales growth, SG&A, R&D, an indicator for missing R&D, market-adjusted annual returns, return volatility and industry (Fama-French 48-industry) and year fixed effects. Additionally in the changes specification, we control for the industry median and the industry median-adjustment of the dependent variable. Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

ceive higher valuations for digital investment suggests that digital technologies are general purpose technologies (Cockburn et al. 2017), which complement other investments. Moreover, consistent with industry-wide complementarities in digital technologies, we find that firms in industry that have adopted digital technologies to a large extent, tend to receive higher valuation for digital activities.

Motivated by the idea that management plays a key role in technology adoption, our last set of analyses investigate how firms with tech-saavy managers can improve performance through digital technologies. In Table A.6, we re-run the ROA analysis in Panel A, Table 2.9 for the sub-sample of non-IT digital firms, and include a proxy for tech-saavy managers, based on executive titles in *Capital IQ's People Intelligence* database. Our analysis show the presence of a tech-saavy manager improves performance through digital activities. In Table A.6, we find that non-IT digital firms with these managers earn 3.1% higher ROA compared to other digital firms. Thus, our results in this table suggest that managerial expertise within the firm is important for integrating and generating value from new

Table A.5: Cross-Sectional Analysis of Market-to-Book

Dependent Variable	$MB_{i,t}$	$MB_{i,t}$
$Digital_{i,t}$	-1.792***	-2.004***
	(0.374)	(0.394)
$Digital_{i,t} \times SIZE_{i,t}$	0.162***	0.154***
	(0.036)	(0.036)
$Digital_{i,t} \times AGE_{i,t}$	0.005	0.003
	(0.089)	(0.091)
$Digital_{i,t} \times R\&D_{i,t}$	0.744	1.003
	(1.202)	(1.244)
$Digital_{i,t} \times SG\&A_{i,t}$	0.973***	0.877***
	(0.335)	(0.339)
$Digital_{i,t} \times CAPEX_{i,t}$	4.476**	5.454***
	(1.764)	(1.866)
$Digital_{i,t} \times ROA_{i,t}$	0.551	0.659
	(0.666)	(0.690)
$Digital_{i,t} \times SG\ Growth_{i,t}$	0.374	0.300
	(0.378)	(0.379)
$Digital_{i,t} \times CASH_{i,t}$	1.183**	1.104**
	(0.516)	(0.504)
$Digital_{i,t} \times LEV_{i,t}$	0.124**	0.123**
	(0.052)	(0.053)
$Digital_{i,t} \times Industry\ Digital_{j,t}$		0.012***
		(0.004)
Controls	Yes	Yes
Time FE	Yes	Yes
Industry FE	Yes	No
Observations	20,877	20,877
Adj. R^2	0.4722	0.4529

We report cross-sectional associations between market-to-book and digital activity for the sample of non-IT firms in fiscal years 2010-2020. In the regressions, we proxy for digital by using a quantized score of the number of digital mentions in the business description of the 10-K (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). All regressions control for size, firm age, leverage, return-on-assets, past 3-year sales growth, market-adjusted annual returns, return volatility, number of words in the business description section and industry (Fama-French 48-industry) and year fixed effects. We also examine the digital activity proxy's interaction with the following variables: size, firm age, R&D expense, SG&A expense, capital expenditure, return-on-assets, past 3-year sales growth, cash balances, leverage and the industry-level of digital activity. Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

digital technologies.

Table A.6: Tech-Saavy Managers and Return-on-Assets

Dependent Variable	ROA_t
Tech $Manager_{i,t}$	0.031***
	(0.011)
$Digital_{i,t}$	0.004
	(0.004)
Controls	Yes
Time FE	Yes
Industry FE	Yes
Observations	3,488
Adj. R^2	0.6088

We report the coefficients of regression of the level of return-on-assets on the proxy for digital activities and the proxy for tech-saavy managers for the sub-sample of non-IT firms in fiscal years 2010-2020 that have made digital disclosures. For all regression models, we proxy for digital by using a quantized score of the number of digital mentions in the business description of the 10-K (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). All regressions control for size, firm age, leverage, past 3-year sales growth, SG&A, R&D, an indicator for missing R&D, market-adjusted annual returns, return volatility and industry (Fama-French 48-industry) and year fixed effects. Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

A.5.6 Robustness Analysis of Valuation Metrics

In the last section, we further examine the robustness of the valuation results reported in Tables 2.5 (Market-to-Book) and 2.6 (ERC/SRC), by dropping sub-sample of firms that could be mis-classified IT firms. We first drop non-IT firms that exhibit high IT co-movement. Specifically, we drop firms that exhibit median IT co-movement that is above the top 2.5 percentile of the distribution of median IT co-movement for non-IT firms. Our inferences of the valuation results under this specification, as reported in Table A.7, are mainly unchanged.

Next, we drop firms that disclose digital firms before 2010, the year before digital technologies were widely available to non-IT firms (due to the development of cloud services). These firms are potential mis-classified IT firms as they had access to the new digital technology before most non-IT firms. We re-examine our valuation results without this set of firms and again our inferences, as reported in Table A.8, are mainly unchanged.

Table A.7: Valuation Analysis Without Top IT Co-Movement Firms

Panel A: Market-to-Book

	Unadjusted Market-to-Book	Conservatism-Corrected Market-to-Book
Dep. Var.	$MB_{i,t}$	$MB_{i,t}$
$Digital_{i,t}$	0.133**	0.125***
	(0.066)	(0.047)
Controls	Yes	Yes
Time FE	Yes	Yes
Industry FE	Yes	Yes
Observations	20,773	7,366
Adj. R^2	0.4507	0.2914

Panel B: Earnings and Sales

	Unexpected Earnings		Unexpected Sales	
	Raw Values	Yearly Deciles	Raw Values	Yearly Deciles
Dependent Variable	CAR(-1,40)	CAR(-1,40)	CAR(-1,40)	CAR(-1,40)
Unexpected Earnings or $Sales_{i,t}$	4.111***	0.013**	-1.093**	-0.003
	(1.341)	(0.005)	(0.460)	(0.006)
$Digital_{i,t}$	0.003**	-0.008*	0.002	-0.006
	(0.002)	(0.005)	(0.002)	(0.004)
$Digital_{i,t} \times$ Unexpected Earnings or $Sales_{i,t}$	0.779***	0.002***	0.253**	0.001*
	(0.251)	(0.001)	(0.116)	(0.001)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Unexpected Earnings or Sales × Controls	Yes	Yes	Yes	Yes
Unexpected Earnings or Sales × Time FE	Yes	Yes	Yes	Yes
Unexpected Earnings or Sales × Industry FE	Yes	Yes	Yes	Yes
Observations	14,223	14,223	13,922	13,922
Adj. R^2	0.0373	0.0396	0.0310	0.0314

In Panel A, we report the coefficients of the regressions of market-to-book on the proxy for digital activities for the sample of non-IT firms in fiscal years 2010-2020. We drop non-IT firms that exhibit median co-movement with the IT portfolio that is greater than the top 2.5 percentile IT co-movement medians of all non-IT firms. We report the associations between market-to-book and digital activity in columns 1 and 2, in Panel A. All regressions control for size, firm age, leverage, market-to-book, past 3-year sales growth, SG&A, R&D, an indicator for missing R&D, market-adjusted annual returns, return volatility and industry (Fama-French 48-industry) and year fixed effects. In Panel B, we report the coefficients to the ERC (Earnings Response Coefficient)/SRC (Sales Response Coefficient) regression with the proxy for digital activities in this table for the same sample of non-IT firms. Column 1 reports the ERC regression using raw unexpected earnings, where CAR(-1,40) is regressed on unexpected earnings, controls, industry and year fixed effects, as well as their interactions with unexpected earnings. Column 2 reports the same regression but with yearly unexpected earnings deciles. Column 3 reports the SRC regression using raw unexpected sales, where CAR(-1,40) is regressed on unexpected sales, controls, industry and year fixed effects, as well as their interactions with unexpected sales. Column 4 reports the same regression but with yearly unexpected sales deciles. All regressions control for size, leverage ratio, loss (ind.), persistence, return volatility, past 3-year sales growth, market-to-book, SG&A expenditure, R&D expenditure, indicator for missing R&D, capital expenditures, the number of days to EA, the number of days to 10-K filing and the number of words in the business description section. For the ease of interpretation of the unexpected earnings/sales coefficient, we mean-center all continuous control variables. Standard errors are clustered at the firm level and are reported in parentheses. In all regressions across both panels, we proxy for digital by using a quantized score of the number of digital mentions in the business description of the 10-K (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

Table A.8: Valuation Analysis Without Firms that Disclose Digital Terms in 2001-2009

Panel A: Market-to-Book

	Unadjusted Market-to-Book	Conservatism-Corrected Market-to-Book
Dep. Var.	$MB_{i,t}$	$MB_{i,t}$
$Digital_{i,t}$	0.112*	0.141***
	(0.064)	(0.048)
Controls	Yes	Yes
Time FE	Yes	Yes
Industry FE	Yes	Yes
Observations	20,619	7,338
Adj. R^2	0.4525	0.2920

Panel B: Earnings and Sales

	Unexpected Earnings		Unexpected Sales	
	Raw Values	Yearly Deciles	Raw Values	Yearly Deciles
Dependent Variable	CAR(-1,40)	CAR(-1,40)	CAR(-1,40)	CAR(-1,40)
Unexpected Earnings or $Sales_{i,t}$	4.505***	0.013**	-1.027**	-0.002
	(1.392)	(0.006)	(0.447)	(0.005)
$Digital_{i,t}$	0.003	-0.010*	0.001	-0.009*
	(0.002)	(0.005)	(0.002)	(0.005)
$Digital_{i,t}\times$Unexpected Earnings or $Sales_{i,t}$	0.625**	0.002***	0.313**	0.002**
	(0.306)	(0.001)	(0.133)	(0.001)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Unexpected Earnings or Sales × Controls	Yes	Yes	Yes	Yes
Unexpected Earnings or Sales × Time FE	Yes	Yes	Yes	Yes
Unexpected Earnings or Sales × Industry FE	Yes	Yes	Yes	Yes
Observations	14,087	14,087	13,783	13,783
Adj. R^2	0.0361	0.0388	0.0302	0.0306

In Panel A, we report the coefficients of the regressions of market-to-book on the proxy for digital activities for the sample of non-IT firms in fiscal years 2010-2020. We drop non-IT firms that disclose digital terms in 2001-2009. We report the associations between market-to-book and digital activity in columns 1 and 2, in Panel A. All regressions control for size, firm age, leverage, market-to-book, past 3-year sales growth, SG&A, R&D, an indicator for missing R&D, market-adjusted annual returns, return volatility and industry (Fama-French 48-industry) and year fixed effects. In Panel B, we report the coefficients to the ERC (Earnings Response Coefficient)/SRC (Sales Response Coefficient) regression with the proxy for digital activities in this table for the same sample of non-IT firms. Column 1 reports the ERC regression using raw unexpected earnings, where CAR(-1,40) is regressed on unexpected earnings, controls, industry and year fixed effects, as well as their interactions with unexpected earnings. Column 2 reports the same regression but with yearly unexpected earnings deciles. Column 3 reports the SRC regression using raw unexpected sales, where CAR(-1,40) is regressed on unexpected sales, controls, industry and year fixed effects, as well as their interactions with unexpected sales. Column 4 reports the same regression but with yearly unexpected sales deciles. All regressions control for size, leverage ratio, loss (ind.), persistence, return volatility, past 3-year sales growth, market-to-book, SG&A expenditure, R&D expenditure, indicator for missing R&D, capital expenditures, the number of days to EA, the number of days to 10-K filing and the number of words in the business description section. For the ease of interpretation of the unexpected earnings/sales coefficient, we mean-center all continuous control variables. Standard errors are clustered at the firm level and are reported in parentheses. In all regressions across both panels, we proxy for digital by using a quantized score of the number of digital mentions in the business description of the 10-K (coded 0 for no disclosure, 1 for bottom tercile, 2 for middle tercile and 3 for top tercile disclosure). Standard errors are clustered at the firm level and are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

B

Appendix to Chapter 3: Financial Analyst Interest in Digitization and AI Investment in Non-IT Firms

B.1 Variable Definitions

Variable Name	Variable Description
Analyst Variables:	
Analyst Questions on Digital Technologies	Proportion of digital questions relative to the total number of questions posed in the conference call. Analyst questions are identified using the list of conference call participants, and digital questions are identified using the digital term-list in Section B.2.2 and B.2.3
Performance-Related Questions	Proportion of digital questions that are related to financial topics, relative to the total number of questions posed in the conference call. The list of performance-related terms is taken from Matsumoto et al. (2011), and is listed in Section B.2.4
Competition-Related Questions	Proportion of digital questions that are related to competition topics, relative to the total number of questions posed in the conference call. The list of competition terms is taken from F. Li et al. (2013), and is listed in Section B.2.5
Technology-Related Questions	Proportion of digital questions that are related to specific digital technologies, relative to the total number of questions posed in the conference call. The list of digital technologies is outlined in Section B.2.3
Question Sentiment	Proportion of positive minus negative sentences in digital question of the conference call divided by the total number of positive and negative sentences in the same set of questions (A. Huang et al. 2020).
Positive/Negative Sentiment	Proportion of positive or negative sentences in digital question of the conference call divided by the total number of positive and negative sentences in the same set of questions.
Technology Analysts	Coded as 1 if the analyst poses questions on the conference call of a firm that is classified as a technology firm.
Industry-Level Digital Questions	Proportion of peer firms in the 6-digit GICS industry and quarter that have digital questions from analysts.

Industry-Level Analyst Sentiment	The question sentiment variable defined using positive and negative sentences from analysts' digital questions at the 6-digit GICS industry and quarter level.

AI Skills and Suitability Measures:

Proportion of AI Job Posting	Total job postings with AI skills divided by total job postings posted by the firm. AI skills are identified using search terms in Acemoglu et al. (2020). See Section B.2.6 for full list of terms.
Incidence of AI Job Posting	Coded as 1 if the firm posts a job with AI skills and 0 otherwise.
AI Suitability Index	Sum of the textual overlap scores (following the approach in, Kogan et al. 2020) between AI-related (identified using search terms in, Webb 2020) patent abstract and titles, and 6-digit NAICS industry descriptions. Scores are standardized with the mean and standard deviation.

Financing and Economic Risk Variables:

Industry-Adjusted Q	Tobin's Q is defined as market capitalization plus long-term debt and debt in current liabilities divided by total assets. Industry adjustment is relative to the median value at the industry (6-digit GICS)-quarter level.
Capital Constraints	The Whited-Wu index of external capital constraints (Whited and Wu, 2006). The index computed from *Compustat* variables: $WW_{i,t} = 0.65 - 0.091 \times \frac{oancf_{i,t}}{at_{i,t}} - 0.062 \times 1_{div_{i,t}>0} + 0.021 \times \frac{dltt_{i,t}+dlc_{i,t}}{at_{i,t}} - 0.044 \times Ln(at_{i,t}) + 0.102 \times \frac{\Delta sale_{j,t}}{sale_{j,t-1}} + 0.035 \times \frac{\Delta sale_{i,t}}{sale_{i,t-1}}$.

Execution Expertise Variables:

Managerial Digital Experience	Coded as 1 if the firm has a top-10 executive with digital-related experience in the biography provided in *Capital IQ Professionals Database*, and 0 otherwise. List of digital-search terms is listed in Section B.2.7. Variable is measured at the beginning of the fiscal year.
Board Digital Experience	Coded as 1 if the firm has a board member with digital-related experience in the biography provided in *Capital IQ Professionals Database*, and 0 otherwise. List of digital-search terms is listed in Section B.2.7. Variable is measured at the beginning of the fiscal year.

Organizational Friction Variables:	
Age	Firm age, computed as the number of years since the firm's first appearance in compustat. Measured at the annual level in the beginning of the fiscal year.
Late Lifecycle	Coded as 1 if the firm is in the "mature", "shake-out" and "decline" part of its lifecycle, as defined based on the Dickinson (2011) cash flow-based lifecycle bucket. Variable is coded 0 otherwise. Measured at the annual level in the beginning of the fiscal year
Number of Segments	Total number of segments reported in *Compustat's* historical segment files. Measured at the annual level in the beginning of the fiscal year.
Product Market Competition Variables:	
Herfindahl	Herfindahl index, computed as the sum of squared sales-weight across all firms within each industry (6-digit GICS)-year cell. Measured at the annual-level in the beginning of the fiscal year.
Industry-Adjusted Sales Growth	Defined as the difference in current and 1-year prior sales divided by 1-year prior sales. Industry adjustment is relative to the median value at the industry (6-digit GICS)-quarter level.
Industry-Level AI Job Postings	Proportion of peer firms in the 6-digit GICS industry and quarter that has posted AI-related job postings.
Disclosure Variables:	
Managerial Disclosure	An indicator variable coded as 1 if management discloses digital-related terms in the presentation portion of the earnings conference call.
Industry-Level Managerial Disclosure	Proportion of peer firms in the 6-digit GICS industry and quarter that have managerial disclosure on digital topics.
Performance-Related Managerial Disclosure	Proportion of performance terms (See Section B.2.4) relative to the total words in sentences with digital terms in the presentation portion of the conference call.
Competition-Related Managerial Disclosure	Proportion of competition terms (See Section B.2.5) relative to the total words in sentences with digital terms in the presentation portion of the conference call.

Technology-Related Managerial Disclosure	Proportion of specific technology terms (See Section B.2.3) relative to the total words in sentences with digital terms in the presentation portion of the conference call.
Managerial Disclosure Sentiment	Proportion of positive minus negative sentences in sentences with digital terms in the presentation portion of the conference call divided by the total number of positive and negative sentences in the same set of sentences (A. Huang et al. 2020).
Positive/Negative Managerial Disclosure Sentiment	Proportion of positive or negative sentences in sentences with digital terms in the presentation portion of the conference call divided by the total number of positive and negative sentences in the same set of sentences.
Industry-Level Managerial Sentiment	The managerial sentiment variable defined using positive and negative sentences in in the presentation portion of conference calls in the 6-digit GICS industry and quarter level.
Firm Disclosure	An indicator variable coded as 1 if management discloses digital-related terms in the presentation portion of the earnings conference call, or if the firm discloses digital-related terms in the business description, MD&A portion of the 10-K (following the list of terms in Section A.1) or discloses a product-announcement with digital-related terms (see Section B.2.8 for the regex search terms).
Control Variables:	
Industry-Adjusted ROA	Return-on-assets, computed as operating income after depreciation divided by last-quarter assets. Industry adjustment is relative to the median value at the industry (6-digit GICS)-quarter level.
Market Beta	β from a market model regression of $R_{it} = \alpha + \beta R_{mt} + \varepsilon_{it}$, where R_{mt} is the weekly equal-weighted market returns, estimated over weekly returns over rolling 3 years, and a minimum of 52 weeks. Measured at the beginning of the fiscal year.
SG&A-to-Sales	Selling, General and Administrative expenditure intensity, defined as SG&A expense minus R&D minus R&D in place scaled by sales. Following Peters and Taylor (2017), SG&A as reported in *Compustat* is not adjusted if R&D expense is greater than SG&A expense and if R&D expense is less than Cost of Goods Sold. If missing this variable is coded as 0.

R&D-to-Sales	Research and development expenditure intensity, defined as research and development expenditures scaled by sales. If missing, this variable is coded as 0.
Missing R&D	Coded as 1 if research and development expenditure is missing and 0 otherwise.
Proportion of Major Customer Sales	Proportion of sales that are due to major customers, as reported in *Compustat's* customer segment files. Measured at the annual level in the beginning of the fiscal year.

B.2 Dictionary Lists for Analyst Questions and AI Skills

B.2.1 Methodology of the Digital Dictionary

To develop a dictionary of digital terms that are specific to the conference call transcripts, I create word list relating to digital strategy and the three groups of advanced digital technologies – big data, analytics and artificial intelligence.

For the first group, digital strategy, I use regex expressions that match with lemmatized versions of the word "digital". The specific regex expressions are denoted in Section B.2.2. As the bigrams of the word "digital" are also associated with older digital technologies, I also drop several "digital" word-pairs, which are outlined at the end of Section B.2.2.

To create the dictionary for the big data, analytics and artificial intelligence technologies, I implement a Word2Vec approach to create an expanded list of digital terms from a set of seed digital terms identified in Section A.1. The list of seed words are presented in Section B.2.3, and is organized into 4 groups, namely, those relating to digital strategy, big data, analytics and AI technologies.

To train the Word2Vec model, I use the entire corpus of earnings conference call text from 2001Q1-2020Q4. As a preprocessing step, I perform stemming and lemmatization, and drop stop words, using the Stanford Core NLP package. I then create bigrams and trigrams, by using CoreNLP's dependency parser and gensim's phraser module to detect bigrams and trigrams in the corpus of conference call text.

To implement the training of the Word2Vec algorithm, I follow the hyperparameters used in K. Li et al. (2021): (1) Maximum window size between focal word and neighbouring words = 5, (2) Dimension of word vector = 300, (3) Number of iteration of corpus = 20, (4) Minimum word count = 5, (5) Training method = negative sampling. I then use the Word2Vec model to convert the terms in the conference call text into vectors of 300-dimensions.

Next, I compute the average cosine similarity between the seed words in each group and other terms in the conference call transcripts. For each of the 3 groups, I then rank each word by their average cosine similarity score, and keep the top-200 words by cosine similarity in the expanded dictionary word list. For duplicate terms across groups, I assign the term to the group for which the term exhibits the highest cosine similarity.

This process yields new technology-specific digital terms by the degree of co-occurrence with the seed words, and so an additional verification step is required to check the terms in the expanded dictionary. Each term is manually reviewed and dropped if the term does not related to digital strategy, or any of the big data, analytics or AI technologies. In appendix B.1c, I present both the terms that are kept in the expanded dictionary and the terms that are dropped.

B.2.2 Regex Expressions for Digital Strategy Words

Digital Term	Regex Expression
digital*	(\bdigital\b)
digitization	(\bdigiti) | (\bdigitali)

* Note, in the counts, I drop the following "digital" word-pairs ending in: "imaging", "radiography", "television(tv)", "media", "cinema", "video", "imaging", "camera", "(bill)board", "music", "mammo(graphy)", "x-ray", "power", "printing","games", "photography", "projector", "boxes", "signage", "audio", "radio", "watermarking", "crt", "telephony","simulcast", "networking", "broadcast"

B.2.3 Expanded Dictionary of Terms for Big Data, Analytics and Artificial Intelligence Technologies

Seed Terms	Expanded Terms	Dropped Terms
Analytics:		
analytic, datum_science, natural_language _processing, nlp, proprietary_algorithm, sentiment_analysis, business_intelligence, customer_intelligence, operation_intelligence, augmented_reality, virtual_reality	datum_analytic, predictive_analytic, natural_language_processing, analytic_capability, datum_insight, datum_management, customer_analytic, datum_intelligence, business_analytic, datum_visualization, analytic_tool, rule_engine, analytic_engine, datum_aggregation, crm, decision_support, datum_analytic_capability, visualization_tool, analytic_platform, marketing_analytic, analytic_application, analytic_solution, business_intelligence_platform, master_datum_management, risk_analytic, datum_capability, time_analytic, marketing_technology, web_analytic, machine_learning_algorithm, business_intelligence_tool, performance_analytic, datum_technology, database_management, software_tool, machine_learning_model, analysis_tool, ai_engine, customer_insight, datum_warehouse, business_intelligence_capability, personalization, metadata, datum_tool, datum_management_capability, datum_science_team,datum_tool, datum_management_capability, datum_science_team, customer_relationship_management, threat_intelligence, datum_asset, leverage_datum, location_datum, precision_marketing, recommendation_engine, sale_automation, location_intelligence, datum_management_platform, business_insight, algorithm, analytic_area, bi_tool, business_intelligence_solution, datum_analytic_platform, analytic_service, analytic_offering, customer_datum, consumer_insight, datum_science_capability, algorithmic, datum_model, crm_application, technology_capability, data_warehouse, fraud_detection, analytic_product, customer_datum_platform, decision_support_tool, analytic_space, datum_analysis, technology_tool, datum_capture, web_content_management, modeling_capability, datum_resource, decision_engine, community_intelligence, business_performance_management, database_marketing, network_datum, fraud_prevention, enterprise_content_management, datum_hub, search_capability, crm_tool, modeling_tool, alteryx, datum_exchange, power_bi, crm_datum, dashboard, optimization_tool, data_lake, workflow_engine, platform_capability, datum_structure, client_datum, datum_basis	intelligence, conversational, contact_management, workflow_tool, reporting_tool, site_search, business_process, analytical, contextual, tool, campaign_management, knowledge_management, monitoring_tool, audience_targeting, predictive

Seed Terms	Expanded Terms	Dropped Terms
Big Data:		
big_datum, smart_datum, biometric, cognitive_computing, cloud_platform, cloud_based, cloud_computing, cloud_deployment, cloud_enablement, hybrid_cloud, virtual_machine, datum_mining, datum_lake, devop, edge_computing	cloud_technology, cloud_infrastructure, cloud_application, virtualization, cloud_service, cloud_solution, enterprise_application, multi-cloud, saa_application, cloud_architecture, web_service, datum_platform, cloud_environment, hybrid_cloud_environment, dev_op, application_delivery, application_platform, cloud_capability, hadoop, cloud_offering, hybrid_cloud_architecture, cloud_security, cloud_server, saas_application, kubernete, cloud_network, enterprise_platform, cloud_management, identity_management, virtualization_technology, sap_hana, soa, storage_virtualization, security_analytic, application_server, cloud_software, virtualize, serverless, enterprise_mobility, cloud_pak, datum_architecture, vdi, enterprise_cloud, application_infrastructure, enterprise_architecture, datum_governance, datum_cloud, data_integration, application_framework, multicloud, computing_capability, datum_movement, cloud_workload, microservice, cloud_networking, network_function_virtualization, google_cloud_platform, datum_fabric, hybrid_multicloud, web_application_firewall, datum_infrastructure, web_application, cloud_implementation, sase, multi-tenancy, ibm_cloud, ai_solution, cloud_storage, cloud_migration, application_layer, identity_platform, cloud_delivery, enterprise_app, application_environment, application_acceleration, legacy_modernization, application_service, software_platform, collaboration_tool, openstack, software_solution, cloud_product, server_virtualization, network_security, cloud_resource, nfv, application_integration, iot, cloud_connectivity, aw_cloud, cloud_provider, application_modernization, splunk, virtualization_solution, collaboration_platform, network_virtualization, microsoft_technology, cloud_native, server_environment, microsoft_azure, enterprise_datum_center, software_application, archiving_solution, cloud_world, sdn, saas_solution, webrtc, datum_store, collaboration_solution, information_infrastructure, security_solution, hybrid_environment, datum_security, core_application, network_automation, security_use_case, application_security, amazon_web_service, enterprise_technology, cloud_datum_center, iot_platform, utility_computing, cloud_area, cloud_-_native, storage_management, storage_platform, sharefile, application_software, server_technology, hybrid_cloud_strategy, de-duplication, network_application, deduplication, cybersecurity, google_cloud, web_security, cloudification, ai_application, datum_preparation, service_delivery_platform, datum_layer, datum_protection, mission_critical_application, endpoint_security, datum_solution, hybrid_architecture, application_development, software_infrastructure, azure, desktop_virtualization, azure_aw, enterprise_solution, cloud_strategy, ilm, microsoft_cloud, openshift, datum_management_solution, app_server, cloud_transformation, intrusion_prevention, datum_center_environment, hyperintelligence, datum_warehousing, insight_platform, amazon_cloud, data_platform, cloud_initiative, opentext_cloud, enterprise_analytic, xml, developer_tool, file_sharing, software_capability, datum_sharing	technology _platform, business_application, enterprise_content _management, policy_management, load_balancing, observability, composable, cloud, enterprise_scale, use_case, orchestration, content_management, disaster_recovery, legacy_modernization

Seed Terms	Expanded Terms	Dropped Terms
Artificial Intelligence:		
artificial _intelligence, chatbot, ai_related, ai_-_related, ai_technology, ai_-_technology, conversational_ai, intelligent_system, computer_vision, virtual_agent, virtual_assistant, autonomous _technology, neural_network, machine_learning, deep_learning, image_recognition, facial_recognition, speech_recognition, automation_solution, intelligent_automation, marketing_automation, process_automation, rpa, robotic_process_ automation	ai, artificial_intelligence_machine_learning, machine_learning_artificial_intelligence, ai_machine_learning, ai_ml, artificial_intelligence_technology, voice_recognition, ai_and_machine_learning, ai_capability, machine_learning_capability, artificial_intelligence_capability, speech_analytic, workflow_management, automation_tool, workflow_automation, machine_translation, artificial_intelligence_platform, voice_technology, automation, business_process_management, face_recognition, voice_biometric, ai_platform, chat_bot, ai_algorithm, automation_capability, productivity_tool, video_analytic, sensor_technology, image_analysis, automation_platform, visualization, video_technology, iot_technology, language_understanding, voice_search, voice_control, ai_model, intelligence_capability, quantum_computing, bot, videoconferencing, collaboration_capability, web_technology, core_technology, automate, software_technology, imaging_technology, automation_technology, automation_software, blockchain, learning_technology, ignio, intelligence_solution, vision_technology, business_process_automation, computational, communication_technology, work_flow, robotic, cmdb, image_processing, intelligence_platform, ar_vr, aiware, internet_technology, software_algorithm, technology_application, cutting_-_edge_technology, crowdsource, sensor_datum, biometric_authentication, technology_platform, pattern_recognition, user_interface, camera_technology, distribute_ledger_technology, simulation_tool, datum_application, smartsheet, machine_vision, bpm, voice_assistant, web_collaboration, automation_solution	telematic, quality_monitoring, workforce_management, design_thinking, incident_management, self_-_service, technology, e-signature, contact_center, workflow, intelligent

B.2.4 Performance-Related Words

The list of performance-related words is taken from Matsumoto et al. (2011). The full list of terms is the following:
"accounting", "derivatives", "accrual", "dividend", "accruals", "dividends", "accrued", "allowance", "dollar", "dollars", "allowances", "earnings", "amortization", "ebit", "amortize", "ebitda", "amortized", "eps", "asset", "equities", "assets", "equity", "bond", "borrowed", "euros", "euro", "borrowing", "expenditure", "borrowings", "expenditures", "budget", "expense", "budgeted", "expenses", "budgeting", "finance", "buybacks", "financed", "capex", "financial", "capital", "financially", "capitalization", "financials", "capitalize", "financing", "capitalized", "financings", "cash", "gain", "cent", "gains", "cents", "goodwill", "convertible", "hedge", "cost", "hedged", "costs", "hedges", "covenants", "hedging", "currencies", "impaired", "debentures", "impairment", "debt", "impairments", "debts", "income", "deferrals", "interest", "deposit", "investment", "deposits", "investments", "depreciation", "lease", "derivative", "leased", "leases", "revenues", "leasing", "roa", "lending", "roe", "leverage", "roi", "liabilities", "sales", "liability", "securities", "liquidity", "securitization", "loan", "security", "loans", "selling", "loss", "shares", "losses", "swaps", "margin", "tax", "margins", "taxable", "obligations", "taxes", "payable", "unamortized", "payables", "unleveraged", "payment", "warrants", "payments", "pound", "pounds", "prepaid", "prepayment", "prepayments", "pretax", "profit", "profitability", "profits", "receivable", "receivables", "redeemable", "refinance", "refinanced", "refinancing", "rent", "rental", "rentals", "repurchasing", "reserve", "reserves", "revenue".

B.2.5 Competition-Related Words

The list of competition-related words is taken from F. Li et al. (2013), and includes addition words such as "entrants", "players" and "peers". Full list of terms are the following:
"competition", "competitors", "competes", "competitor", "competitive", "compete", "competing", "peer", "peers", "entrants", "entrant", "players", "player".

B.2.6 AI Skills Search Terms for Job Postings

The list of terms are taken from Alekseeva et al. (2021).

Machine Learning, Computer Vision, Machine Vision, Deep Learning, Virtual Agents, Image Recognition, Natural Language Processing, Speech Recognition, Pattern Recognition, Object Recognition, Neural Networks, AI ChatBot, AI KIBIT, ANTLR, Apertium, Artificial Intelligence, Automatic Speech Recognition (ASR), Caffe Deep Learning Framework, Computational Linguistics, Decision Trees, Deeplearning4j, Distinguo, Google Cloud Machine Learning Platform, H2O (software), IBM Watson, IPSoft Amelia, Ithink, Lexalytics, Lexical Acquisition, Lexical Semantics, Machine Translation (MT), Madlib, Microsoft Cognitive Toolkit, MLPACK (C++ library), Mlpy, Modular Audio Recognition Framework (MARF), MoSes, MXNet, Natural Language Toolkit (NLTK), ND4J (software), Nearest Neighbor Algorithm, Object Tracking, OpenNLP, Semantic Driven Subtractive Clustering Method (SDSCM), Semi-Supervised Learning, Sentiment Classification, Supervised Learning (Machine Learning), TensorFlow, Text to Speech (TTS), Tokenization, Torch (Machine Learning), Vowpal, Wabbit, Text Mining, Unsupervised Learning, Image Processing, Mahout, Recommender Systems, Support Vector Machines (SVM), Random Forests, Latent Semantic Analysis, Sentiment Analysis / Opinion Mining, Latent Dirichlet Allocation, Predictive Models, Kernel Methods, Keras, Gradient boosting, OpenCV, Xgboost, Libsvm, Word2Vec, Chatbot, Machine Translation and Sentiment Classification

B.2.7 Digital Regex Search Expression for Managerial/Board Digital Experience

To search for digital-related terms in management and board biographies, I implement the following regex search expression:

(digital)|(artificial intelligence)|(\bai\b)|(cloud ?[-]?computing)|(machine learning) |(\bai-tech)|(\banalytics\b)|(big data)|(data science)

B.2.8 Digital Regex Search Expression for Product Announcement Relating to Digital Technologies

To search for digital-related terms in product announcement in the Key Development database, I search for the following terms (based on the word list in Section A.1 for announcement related to strategic alliances, client alliances, business expansions and product-related announcements:

(\banalytics\b)|(artificial ?[-]?intelligence)|(\bai ?[-]?tech)|(\bai ?[-]?related)|(\bautonomous ?[-]?tech)|(\bvirtual agent)|(\bvirtual ?[-]?assistant)|(\bbig ?[-]?data)|(\bbiometric)|(\bcloud ?[-]?platform)|(\bcloud ?[-]?based)|(\bcloud ?[-]?computing)|(\bcloud ?[-]?deployment) |(\bvirtual ?[-]?machine)|(\bdata ?[-]?scien)|(\bdata ?[-]?mining)|(\bdeep ?[-]?learning) |(\bdigital\b)|(\bdigiti)|(\bdigitali)|(\bmachine ?[-]?learning) |(\bnatural ?[-]?language ?[-]?processing)|(\bneural ?[-]?network)|(\bimage ?[-]?recognition) |(\bfacial ?[-]?recognition)|(\bintelligent ?[-]?system)|(\bcomputer ?[-]?vision)|(\bspeech ?[-]?recognition)|(\bsentiment ?[-]?analysis)|(\bvirtual realit)|(\bautomation solutions\b) |(\bintelligent automation\b)|(\bmarketing automation\b)|(\bprocess automation\b) |(\baugmented reality\b)|(\bedge computing\b)|(\bcognitive computing\b) |(\bbusiness intelligence\b)|(\bcustomer intelligence\b)|(\boperating intelligence\b) |(\bcloud enablement\b)|(\bhybrid cloud\b)|(\bconversational ai\b)|(\bevolutionary ai\b)

|(\bevolutionary computing\b)|(\bdata lake\b)|(\bdevops\b)|(\bproprietary algorithm)
|(\brobotic process automation\b)

B.3 Examples of Analysts Digital Questions and AI Skills

B.3.1 Examples of Analysts Digital Questions

Performance Questions

Discovery Inc, 2019Q3

Very interesting. And Gunnar, you were threatening to continue to delever, but there's only 10 basis points left to go. And you just talked about capital deployment a little bit. If you're doing close to $3 billion of free cash flow next year and you're basically at 3.0x as you end this year, it sounds like either that's going to M&A or capital returns. The only other thing that would be an input in the model is if you were investing so much in digital initiatives like the one David is talking about that EBITDA would be pressured. And that some of that cash flow would have to go to pay down debt. Am I sort of thinking about all that right? And is that last scenario in play?

Competition Questions

KeyCorp, 2018Q4

And then, I guess, Don or Beth, what is the impact that you're seeing in your markets from the other large banks expanding into your MSAs with digital banking? I guess, you're not really seeing anything yet, but there is that threat you'll see more competition, say, over the next 5 years. So are you worried about that? Are you preparing for that? Are you looking to strike back with more efforts of your own?

Technology Questions

Johnson & Johnson, 2019Q4

That's great. And then just a follow-up. I appreciate the update and overview on the robotic surgery strategy, Alex. But I guess, one of the things that we've noticed is just how quickly – and I realize it's sort of a buzz-worthy term, and we see it in a lot of magazines, but artificial intelligence in digital health. For whatever reason, implementation and innovation, but it seems to be moving perhaps the fastest of the categories of digital surgery and digital health. And just wondering, one of your competitors is partnered with a smaller innovator in that space around stroke. I'm sure you're watching it. I'm sure you're making investments and looking at it. I'm just curious if you could give us an update or a preview perhaps of what you'll talk about in the spring.

Positive Questions

Morgan Stanley, 2017Q4

Wanted to start off with a question on the investment strategy. And I know, Jon, you had just made some remarks on this topic. But Wealth Management in recent years, it's certainly been helped by your strong expense discipline. You spoke about your ability to make investments in digital technology while keeping that expense base flattish. And I'd say you've been a positive outlier relative to a lot of your peers in that regard within that business. I'm just curious, underlying the 26% to 28%, which James, you noted as conservative, what assumptions are you making on non-comp inflation? And then in addition to that, since it's been asked by a few investors, maybe you could speak to deposit beta assumptions in that segment as well.

Negative Questions

CarMax, 2019Q2

Just a question as we think about sort of the competitive dynamics in the market. And you've kind of talked about this a little bit, but I kind of wanted to put it all together. I mean, are there any specific markets where competition is being disruptive or any markets you can point to that you're being disruptive? And really, what I'm trying to get at here is there's an increasing focus by the new vehicle dealers. There's obviously some digital competitors that are out there. But at the same time, as you launch your omnichannel efforts and kind of really coalesce what you're doing online and within home delivery, you might be disruptive. Just trying to understand the competitive dynamics and if you can give us any examples in markets where you see market share shifts or opportunities. It just seems there's like – there's a lot going on right now. And you're probably going to be one of the big winners, but I think there's a lot of concern out there.

B.3.2 Example of AI Job Postings

Pure Michigan Talent Connect - Job Details

Job Title	**Senior Researcher**
Job Code Number	11380095
Organization Name	General Motors
Created Date	1/1/2021
Posted Date	1/1/2021
Expiration Date	2/1/2021
Job Description	

About GM

There's never been a more exciting time to work for General Motors.

To achieve our vision of a world with Zero Crashes, Zero Emissions and Zero Congestion, we need people to join us who are passionate about creating safer, better and more sustainable ways for people to get around. This bold vision won't happen overnight, but just as we transformed how the world moved in the last century, we are committed to transforming how we move today and in the future.

Why Work for Us

Our culture is focused on building inclusive teams, where differences and unique perspectives are embraced so you can contribute to your fullest potential as you pursue your career. Our locations feature a variety of work environments, including open work spaces and virtual connection platforms to inspire productivity and flexible collaboration. And we are proud to support our employees volunteer interests, and make it a priority to join together in efforts that give back to our communities.

Job Description

Job Summary:

The Analytics Research Department within the Advanced Analytics Center of Expertise (COE) is looking for highly motivated, collaborative individuals to join our growing team. Our mission is to identify and analyze strategic opportunities to improve GM's profitability, increase operational efficiency and provide customer insights. Application focus areas are enterprise-wide including customer analytics, finance, marketing, product development, revenue management, quality, and supply chain and logistics.

Position Details:

This position requires mathematical expertise, creative problem solving, cross-functional communication, teamwork, technical documentation, and presentation and project/technical leadership skills. Candidates should demonstrate an exceptional ability to adapt to ambiguous business objectives, and clearly communicate the value of business analytics to a non-technical audience.

Additional Job Description

Requirements:

- Ph.D. with 5-7 years of experience in a technical discipline, such as operations research, applied mathematics, statistics, decision science, econometrics, engineering, marketing, or artificial intelligence/machine learning
- Ability and willingness to contribute on multiple levels including: data preparation, algorithm development, documentation, programming, project management, conducting experiments, and interacting with clients
- Excellent collaboration skills with project teams, management, customers, and experts
- Excellent oral and written communication skills with a particular strength in explaining technical concepts to non-technical audiences

Figure B.1: AI-related Online Job Posting from General Motors

B.4 Database Matching Methodology

B.4.1 Matching Earnings Call Transcripts to Compustat/CRSP

The earnings conference call transcripts are taken from *Thomson Reuter*'s Streetevents from 2010Q1-2020Q4. The identifiers for these transcripts is the ticker and name (taken from the event title), so multiple approaches are taken to verify matches. In the first and second pass, the data is matched on tickers in *CRSP stocknames* and in the *Compustat* database, respectively. As the database backfills the ticker whenever the company changes the ticker or is acquired (K. Li et al. 2021), the ticker matches are verified with the names. This is done by first running a fuzzy match (with the Levenshtein algorithm), keeping matches with above 50% accuracy, and manually verifying each observations.

B.4.2 Matching BurningGlass Data to Compustat/CRSP

The data on job postings, is taken from *BurningGlass*. This dataset comprises of the near-universe of online job postings, and the only identifier at the employer-level is the company name. Thus, a combination fuzzy-matching and manual checks are conducted to ensure that the matches are correct.

As a first step, I run a fuzzy-matching procedure (with the levenshtein algorithm) that searches for the closest *CRSP* name (in stocknames) to each *BurningGlass* employer name. To ensure that suffixes are not changing the quality of matches, I remove common suffixes — "inc", "incorporated", "cos", "companies", "company", "co", "corp", "corporation", "llc", "lp", "limited", "ltd", "trust". Matches with below 90% matching accuracy are removed, and each match below 100% is verified manually.

To further validate the procedure, the same matching steps are implemented with the *Compustat* names. The alternative search yielded no differences in the matches, and so this gives some confidence that the procedure is identifying correct matches.

B.5 Methodology of the AI Suitability Score

B.5.1 Methodology Details

I develop an index for AI suitability at the industry-level following the methodology developed in Webb (2020) and Kogan et al. (2020). The basic idea of the methodology is to compute similarity scores based on the overlap of words in AI patents and the industry descriptions, to proxy for complementarity between AI technologies and industry-level applications.

The first step of the computation process is to identify a set of AI patents, following the search procedure in Webb (2020). These patents are sourced from the universe of US-based patents in the *Google Patents Database*[1 2]. To identify AI patents, I select patents that have at least one of the following terms in the title or abstract: "supervised learning", "reinforcement learning", "deep learning", "neural network" and "machine learning"[3].

In the second step, the Kogan et al. (2020) methodology is used to estimate the textual overlap between patents and the description of industries. The methodology takes a GloVe approach (Pennington et al. 2014) in computing the textual similarity of the two sources of text through word embeddings. A key advantage of this approach is that it yields a cosine similarity metric that accounts for subtle differences in the textual meaning, and thus is par-

1. https://console.cloud.google.com/marketplace/partners/patents-public-data
2. This database includes patents that are filed globally. But to ensure that differences in the patenting process do not influence my results, the analysis is restricted to only patents filed in the US.
3. Webb (2020) does not include "machine learning" as a search term, but I include this term as it is an important component of AI technologies

ticularly suited for situations where words are similar but are represented in different ways.

To compute the similarity scores based on Kogan et al. (2020), I first compute the word embedding of each document (either patent (abstract and title) and the 6-digit 2017 NAICS industry description)[4], by characterizing each word by GloVe vectors (Pennington et al. 2014)[5]. This method uses the co-occurrences of words to develop a mapping to a vector space which aligns words that are different but have similar meaning. For instance, a word like "dog" and "cat" will have a similar vector representation and thus the similarity between the two words will be assigned a high score, instead of the 0 computed in a traditional cosine similarity methodology. To aggregate to the document level, I take the weighted average of the word vectors in the whole document by the following formula:

$$X_i = \sum_{x_k \in A_i} w_{i,k} x_k \tag{B.1}$$

where X_i is the document vector and x_k is a word-vector in the set of vectors in document A_i. $w_{i,k}$ is the word TF-IDF weight which is determined as follows:

$$w_{i,k} = TF_{i,k} \times IDF_k \tag{B.2}$$

where $TF_{i,k}$ is term frequency of word k in document i and IDF_k is the inverse document

4. As a pre-processing step, I also remove stopwords, lemmatize the words and keep only nouns and verbs (based on the WordNet classification) in the patent and industry description text.

5. I use the largest set of pre-trained vector representations provided on the StanfordNLP website: https://nlp.stanford.edu/projects/glove/. The dataset is "glove.840B.300d", which is a dictionary trained on 840 B tokens from Common Crawl and maps words to 300-dimensional clusters.

frequency of word k, which is computed as:

$$IDF_k = log(\frac{\#\ documents}{\#\ documents\ with\ word\ k}) \tag{B.3}$$

The similarity scores are then estimated by:

$$\rho_{a,b} = \frac{X_a}{|X_a|} \cdot \frac{X_b}{|X_b|} \tag{B.4}$$

To estimate the industry-year level exposure score, I follow Kogan et al. (2020) and apply the following aggregation formula on cosine similarity scores:

$$\eta_{s,t} = \sum_{h \in t} \tilde{\rho}_{s,h} \tag{B.5}$$

where $\eta_{s,t}$ is the industry s exposure score in year t. $\tilde{\rho}_{s,h}$ is the adjusted similarity scores for industry s and patent h. The similarity score adjustment takes two steps: (1) year fixed effects are removed from the similarity scores and (2) sparsity is imposed by setting scores below the 80th percentile to 0[6].

Finally, to normalize the exposure scores, the raw scores are deducted by the pooled mean and scaled by the pooled standard deviation.

B.5.2 Descriptive Analysis of the AI Suitability Score

To provide some insight into the AI suitability scores, I discuss some of the key descriptive features of the data. Figure B.2 plots the distribution of scores, over the 2010-2020 period

6. Kogan et al. (2020) also scale the non-zero values to run from 0-1, but I do not do so as it gives undue weight to patent exposure in the early part of the sample.

for non-IT, 2-digit NAICS sectors. The figure indicates that there is wide variation in AI suitability across sectors. Moreover, there is a significant uptick in the suitability of AI to the various non-IT sectors from 2015, which aligns with anecdotes of the rapid advancement of neural network algorithms. Moreover, an inspection of the top and bottom 2-digit NAICS sectors shows that services sectors tend to benefit most from AI, while accommodation (hotels) and real estate, tend to benefit the least.

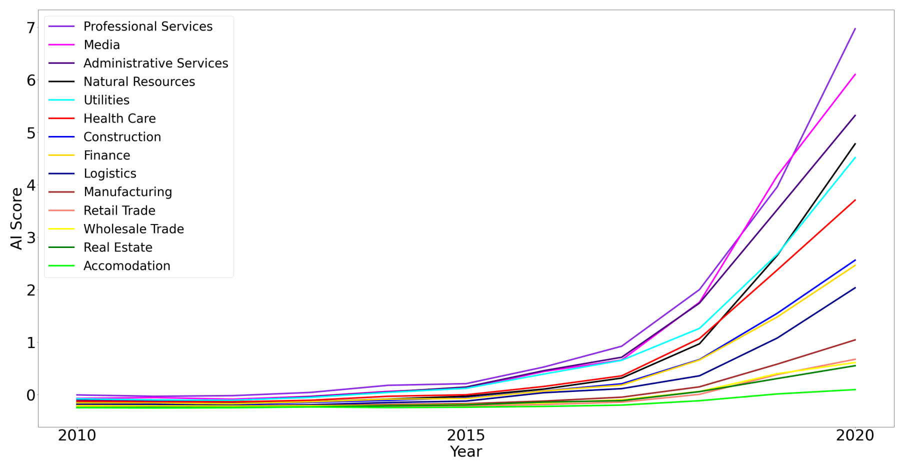


Figure B.2: AI Suitability Scores Across 2-Digit NAICS Industries

I further examine the suitability of AI across industries, using finer data in the Table B.1. The analysis reveals that the top industries by AI suitability are the finance, and media industries. On the other hand, the bottom industries by AI suitability comprise of industries such as, breweries and tobacco manufacturing. These descriptive results do align with the intuition that AI is more suited to more automatable tasks in the professional services industries but is less suited for tasks that require non-routine manual labor.

Table B.1: Top and Bottom Ranked Industries by AI Suitability Scores

Panel A: Top Overlap Industries	
1	Cable and Other Subscription Programming
2	Investment Advice
3	Couriers and Express Delivery Services
4	Investment Banking and Securities Dealing
5	Securities and Commodity Exchanges
6	Portfolio Management
7	Financial Transactions Processing, Reserve, and Clearinghouse Activities
8	Television Broadcasting
9	Motion Picture and Video Distribution
10	Power and Communication Line and Related Structures Construction
Panel B: Bottom Overlap Industries	
1	Mattress Manufacturing
2	Roofing, Siding, and Insulation Material Merchant Wholesalers
3	Small Arms, Ordnance, and Ordnance Accessories Manufacturing
4	Gypsum Product Manufacturing
5	Breweries
6	Tobacco and Tobacco Product Merchant Wholesalers
7	Ethyl Alcohol Manufacturing
8	Metal Kitchen Cookware, Utensil, Cutlery, and Flatware (except Precious) Manufacturing
9	Distilleries
10	Nonchocolate Confectionery Manufacturing

This table reports the top 10 and bottom 10 exposed 6-digit NAICS industries by textual overlap scores. Panel A reports the top 10 industries, while Panel B reports the bottom 10 industries for the textual overlap scores.

B.6 Additional Robustness Analysis

B.6.1 Robustness Analysis of the AI Suitability Score

This set of analysis examines the robustness of the positive relationship of the digital questions to the AI suitability scores. I use several alternative metrics of AI's suitability across suitability, namely, (1) the survey-based AI industry exposure index from Felten et al. (2021) and (2) the survey-based suitability for machine learning index from Brynjolfsson et al. (2018). Results are presented in Table B.2.

Table B.2: Are Analysts' Interest in Digitization Related to the Suitability of AI Across Industries?

Dependent Variable	Digital Questions	Digital Questions	Digital Questions	Digital Questions
AI Suitability$_{l,t-1}$	0.225***			0.229***
	(0.080)			(0.080)
AI Survey$_{b}$		0.121**		-0.001
		(0.049)		(0.052)
ML Survey$_{b}$			0.178***	0.188***
			(0.056)	(0.060)
Calendar Quarter FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	28,257	28,257	28,257	28,132
R^2	0.076	0.075	0.076	0.080

This table reports the associations between analysts' digital questions and indices that measure AI suitability over industries. Columns 1-4, examines three different indices of AI suitability — (1) the textual overlap-based index, (2) the survey-based AI industry exposure index from Felten et al. (2021) and (3) the survey-based suitability for machine learning index from Brynjolfsson et al. (2018). Regressions control for the industry-level of AI adoption, herfindahl index, the number of segments, the capital constraints index, the logarithm of age, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales, industry-adjusted sales growth and an indicator for late lifecycle firms. Calendar quarter fixed effects are also included, and standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

B.6.2 Robustness of the Abnormal-Level of AI Investments to Various Regression Models

This section presents the robustness of the empirical analyses with the abnormal-level of AI investment sub-samples. In Table B.3, I present the robustness of the analyses with sub-samples of AI investment residuals that are computed using an OLS model. In Table B.4, I present the robustness of the analyses with sub-samples of AI investment residuals that are computed using an industry-quarter FE model.

Table B.3: Robustness Analysis of Managerial/Analysts Disclosure and Future AI Investment Conditional on Sub-Samples of Abnormally Low, High and Normal AI Investments, Based on OLS Residuals

Sub-Sample	Abn. Low	Normal	Abn. High	Abn. Low Minus Normal	Abn. High Minus Normal
Panel A: Managerial Disclosure of Digital Topics					
% Disclosure	43.539***	29.159***	45.563***	14.380***	16.404***
	(2.262)	(1.686)	(2.612)	(1.624)	(1.893)
% Sentiment	75.275***	72.145***	77.968***	3.130**	5.823***
	(1.466)	(1.092)	(1.360)	(1.310)	(1.392)
% Positive Sentences	79.560***	76.358***	82.225***	3.202***	5.867***
	(1.214)	(0.965)	(1.089)	(1.149)	(1.160)
% Negative Sentences	4.285***	4.213***	4.257***	0.072	0.044
	(0.472)	(0.315)	(0.523)	(0.444)	(0.489)
% Performance Words	1.009***	1.067***	0.925***	-0.059	-0.143***
	(0.034)	(0.026)	(0.037)	(0.039)	(0.038)
% Competition Words	0.070***	0.055***	0.078***	0.015*	0.024*
	(0.007)	(0.005)	(0.011)	(0.008)	(0.012)
% Technology Words	1.996***	2.378***	2.375***	-0.382***	-0.003
	(0.085)	(0.077)	(0.116)	(0.090)	(0.099)
Panel B: Analysts' Digital Questions					
% Disclosure	15.288***	8.356***	15.446***	6.932***	7.090***
	(1.266)	(0.679)	(1.544)	(1.055)	(1.286)
% Sentiment	32.992***	31.414***	27.557***	1.578	-3.857
	(2.023)	(1.847)	(2.158)	(2.430)	(2.567)
% Positive Sentences	47.403***	45.526***	45.709***	1.878	0.184
	(1.638)	(1.267)	(1.890)	(1.940)	(2.123)
% Negative Sentences	14.412***	14.112***	18.152***	0.300	4.041***
	(0.892)	(1.029)	(1.170)	(1.169)	(1.303)
% Performance Questions	21.524***	25.190***	25.919***	-3.666*	0.729
	(1.658)	(1.540)	(2.504)	(2.028)	(2.710)
% Competition Questions	3.849***	3.534***	7.957***	0.315	4.423***
	(0.745)	(0.464)	(1.328)	(0.774)	(1.363)
% Technology Questions	47.687***	57.564***	64.206***	-9.878**	6.641
	(3.177)	(2.911)	(4.528)	(3.679)	(4.879)

Table B.3: Robustness Analysis of Managerial/Analysts Disclosure and Future AI Investment Conditional on Sub-Samples of Abnormally Low, High and Normal AI Investments, Based on OLS Residuals (Continued)

Panel C: Future AI Investment Conditional on Digital Questions						
Sub-Sample	Abn. Low-Investment		Normal-Investment		Abn. High-Investment	
Dependent Variable	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity
Digital Questions$_{i,q}$	0.422**	0.029***	0.348**	0.001	-0.131	0.024
	(0.175)	(0.011)	(0.148)	(0.004)	(0.171)	(0.016)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,667	5,667	20,331	20,331	3,018	3,018
R^2	0.469	0.316	0.514	0.299	0.513	0.535

This table reports the topics of managerial/analysts disclosure in sub-samples of abnormally low, high and normal AI investments, as well as the levels of future AI investment conditional on digital questions in the same sub-samples. Sub-samples are defined by residuals from an OLS regression of AI investment on the five investment factors. Panels A and B reports the average differences in the topics disclosed by management and analysts across the sub-samples. Panel C report the associations between future AI investment and digital questions across the sub-samples. Regressions in Panel C control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Standard errors in analyses reported in all panels are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

Table B.4: Robustness Analysis of Managerial/Analysts Disclosure and Future AI Investment Conditional on Sub-Samples of Abnormally Low, High and Normal AI Investments, Based on Industry-Quarter FE Residuals

Sub-Sample	Abn. Low	Normal	Abn. High	Abn. Low Minus Normal	Abn. High Minus Normal
Panel A: Managerial Disclosure of Digital Topics					
% Disclosure	39.390***	30.407***	42.367***	8.983***	11.960***
	(2.094)	(1.674)	(2.652)	(1.369)	(1.829)
% Sentiment	74.952***	72.517***	76.858***	2.435*	4.341***
	(1.458)	(1.063)	(1.491)	(1.215)	(1.480)
% Positive Sentences	79.060***	76.800***	81.109***	2.261**	4.309***
	(1.220)	(0.908)	(1.258)	(1.019)	(1.239)
% Negative Sentences	4.109***	4.283***	4.251***	-0.174	-0.032
	(0.430)	(0.336)	(0.515)	(0.425)	(0.519)
% Performance Words	1.007***	1.061***	0.958***	-0.055	-0.103**
	(0.035)	(0.025)	(0.039)	(0.037)	(0.040)
% Competition Words	0.065***	0.056***	0.078***	0.008	0.022*
	(0.007)	(0.004)	(0.011)	(0.008)	(0.011)
% Technology Words	2.139***	2.309***	2.385***	-0.170**	0.076
	(0.083)	(0.072)	(0.111)	(0.074)	(0.096)
Panel B: Analysts' Digital Questions					
% Disclosure	13.332***	8.964***	13.924***	4.368***	4.961***
	(1.101)	(0.690)	(1.425)	(0.825)	(1.135)
% Sentiment	31.921***	32.295***	26.576***	-0.374	-5.718**
	(2.027)	(1.784)	(2.231)	(2.362)	(2.532)
% Positive Sentences	47.229***	45.951***	44.801***	1.279	-1.149
	(1.519)	(1.116)	(1.889)	(1.529)	(1.888)
% Negative Sentences	15.308***	13.656***	18.225***	1.652	4.569***
	(1.029)	(1.052)	(1.281)	(1.355)	(1.572)
% Performance Questions	22.681***	24.613***	25.454***	-1.933	0.841
	(1.672)	(1.339)	(2.491)	(1.636)	(2.546)
% Competition Questions	3.574***	3.884***	6.938***	-0.310	3.055**
	(0.635)	(0.489)	(1.207)	(0.628)	(1.253)
% Technology Questions	49.956***	56.301***	62.963***	-6.345**	6.663
	(2.870)	(2.672)	(4.247)	(2.758)	(4.332)

Table B.4: Robustness Analysis of Managerial/Analysts Disclosure and Future AI Investment Conditional on Sub-Samples of Abnormally Low, High and Normal AI Investments, Based on Industry-Quarter FE Residuals (Continued)

Panel C: Future AI Investment Conditional on Digital Questions

Sub-Sample	Abn. Low-Investment		Normal-Investment		Abn. High-Investment	
Dependent Variable	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity
Digital Questions$_{i,q}$	0.485*	0.039**	0.364***	0.007	-0.150	0.016
	(0.277)	(0.017)	(0.130)	(0.008)	(0.167)	(0.014)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,056	5,056	20,602	20,602	3,226	3,226
R^2	0.502	0.349	0.504	0.302	0.509	0.543

This table reports the topics of managerial/analysts disclosure in sub-samples of abnormally low, high and normal AI investments, as well as the levels of future AI investment conditional on digital questions in the same sub-samples. Sub-samples are defined by residuals from an industry-quarter FE regression of AI investment on the five investment factors. Panels A and B reports the average differences in the topics disclosed by management and analysts across the sub-samples. Panel C reports the associations between future AI investment and digital questions across the sub-samples. Regressions in Panel C control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Standard errors in analyses reported in all panels are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

B.6.3 Robustness of the Abnormal-Level of AI Investments to Various Sub-Sample Cut-offs

In the next set of tables, I examine the robustness of the abnormal investment analyses, to changes to the abnormal investment-level cutoffs. Table B.5, reports the robustness of the analyses with sub-samples of AI investment residuals that are assigned based on a 0.25 σ cut-off. In Table B.6, I present the robustness of the analyses with sub-samples of AI investment residuals that are assigned based on a 0.75 σ cut-off.

Table B.5: Robustness Analysis of Managerial/Analysts Disclosure and Future AI Investment Conditional on Sub-Samples of Abnormally Low, High and Normal AI Investments, Based on 0.25 σ Cut-Off

Sub-Sample	Abn. Low	Normal	Abn. High	Abn. Low Minus Normal	Abn. High Minus Normal
Panel A: Managerial Disclosure of Digital Topics					
% Disclosure	38.575***	26.641***	40.518***	11.935***	13.878***
	(1.708)	(1.530)	(2.772)	(1.175)	(1.906)
% Sentiment	74.128***	71.198***	77.628***	2.930**	6.430***
	(1.180)	(1.214)	(1.363)	(1.292)	(1.397)
% Positive Sentences	78.356***	75.583***	81.609***	2.773**	6.026***
	(0.982)	(1.045)	(1.146)	(1.071)	(1.179)
% Negative Sentences	4.228***	4.385***	3.981***	-0.157	-0.404
	(0.359)	(0.425)	(0.455)	(0.469)	(0.497)
% Performance Words	1.024***	1.079***	0.965***	-0.055	-0.114***
	(0.029)	(0.030)	(0.035)	(0.039)	(0.038)
% Competition Words	0.061***	0.059***	0.070***	0.002	0.011
	(0.005)	(0.007)	(0.010)	(0.008)	(0.012)
% Technology Words	2.089***	2.483***	2.396***	-0.393***	-0.086
	(0.070)	(0.083)	(0.111)	(0.080)	(0.097)
Panel B: Analysts' Digital Questions					
% Disclosure	12.871***	7.112***	13.423***	5.759***	6.311***
	(0.875)	(0.569)	(1.412)	(0.680)	(1.152)
% Sentiment	33.332***	30.285***	27.381***	3.047	-2.904
	(1.757)	(2.125)	(2.212)	(2.329)	(2.662)
% Positive Sentences	47.213***	44.620***	45.390***	2.594	0.770
	(1.345)	(1.436)	(1.911)	(1.792)	(2.182)
% Negative Sentences	13.881***	14.335***	18.009***	-0.454	3.674***
	(0.806)	(1.281)	(0.956)	(1.214)	(1.231)
% Performance Questions	21.927***	27.268***	25.744***	-5.341***	-1.524
	(1.413)	(1.777)	(2.131)	(1.877)	(2.494)
% Competition Questions	3.436***	3.764***	7.394***	-0.328	3.631***
	(0.593)	(0.664)	(1.125)	(0.875)	(1.246)
% Technology Questions	49.504***	61.479***	63.419***	-11.975***	1.940
	(2.645)	(2.849)	(4.064)	(2.771)	(4.321)

Table B.5: Robustness Analysis of Managerial/Analysts Disclosure and Future AI Investment Conditional on Sub-Samples of Abnormally Low, High and Normal AI Investments, Based on 0.25 σ Cut-Off (Continued)

Panel C: Future AI Investment Conditional on Digital Questions						
Sub-Sample	Abn. Low-Investment		Normal-Investment		Abn. High-Investment	
Dependent Variable	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity
Digital Questions$_{i,t}$	0.282*	0.015*	0.535**	0.011*	-0.018	0.022
	(0.148)	(0.008)	(0.204)	(0.006)	(0.202)	(0.014)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,739	11,739	12,789	12,789	4,424	4,424
R^2	0.466	0.301	0.533	0.315	0.584	0.555

This table reports the topics of managerial/analysts disclosure in sub-samples of abnormally low, high and normal AI investments, as well as the levels of future AI investment conditional on digital questions in the same sub-samples. Sub-samples are defined by residuals from an quarter-fixed effects regressions of AI investment on the five investment factors, and abnormally low (high) investment sub-samples are split by whether the residual is below (above) 0.25 σs from the mean. Panels A and B report the average differences in the topics disclosed by management and analysts across the sub-samples. Panel C reports the associations between future AI investment and digital questions across the sub-samples. Regressions in Panel C control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Standard errors in analyses reported in all panels are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

Table B.6: Robustness Analysis of Managerial/Analysts Disclosure and Future AI Investment Conditional on Sub-Samples of Abnormally Low, High and Normal AI Investments, Based on 0.75 σ Cut-Off

Sub-Sample	Abn. Low	Normal	Abn. High	Abn. Low Minus Normal	Abn. High Minus Normal
Panel A: Managerial Disclosure of Digital Topics					
% Disclosure	44.782***	31.265***	47.516***	13.517***	16.251***
	(2.652)	(1.791)	(2.665)	(2.393)	(2.040)
% Sentiment	76.093***	72.804***	77.994***	3.289*	5.191***
	(1.819)	(1.117)	(1.398)	(1.776)	(1.480)
% Positive Sentences	80.003***	77.098***	82.139***	2.905*	5.041***
	(1.533)	(0.956)	(1.190)	(1.545)	(1.248)
% Negative Sentences	3.910***	4.294***	4.145***	-0.384	-0.150
	(0.610)	(0.320)	(0.520)	(0.602)	(0.502)
% Performance Words	0.991***	1.054***	0.927***	-0.063	-0.127***
	(0.048)	(0.024)	(0.039)	(0.052)	(0.040)
% Competition Words	0.095***	0.054***	0.085***	0.041***	0.031**
	(0.012)	(0.004)	(0.013)	(0.013)	(0.014)
% Technology Words	1.946***	2.315***	2.350***	-0.369***	0.035
	(0.120)	(0.074)	(0.116)	(0.128)	(0.101)
Panel B: Analysts' Digital Questions					
% Disclosure	18.080***	9.139***	16.516***	8.941***	7.377***
	(1.917)	(0.738)	(1.674)	(1.764)	(1.436)
% Sentiment	33.748***	31.314***	28.519***	2.434	-2.795
	(2.639)	(1.668)	(2.157)	(3.033)	(2.370)
% Positive Sentences	49.374***	45.494***	46.086***	3.880*	0.592
	(1.847)	(1.131)	(2.041)	(2.050)	(2.048)
% Negative Sentences	15.626***	14.180***	17.568***	1.446	3.387**
	(1.313)	(0.888)	(1.229)	(1.450)	(1.355)
% Performance Questions	20.296***	24.559***	26.583***	-4.263	2.025
	(2.584)	(1.305)	(2.790)	(2.684)	(2.853)
% Competition Questions	3.787***	3.732***	8.008***	0.055	4.276***
	(1.117)	(0.444)	(1.387)	(1.059)	(1.408)
% Technology Questions	48.515***	55.588***	64.123***	-7.074	8.534*
	(4.443)	(2.623)	(4.635)	(4.667)	(4.796)

Table B.6: Robustness Analysis of Managerial/Analysts Disclosure and Future AI Investment Conditional on Sub-Samples of Abnormally Low, High and Normal AI Investments, Based on 0.75 σ Cut-Off (Continued)

Panel C: Future AI Investment Conditional on Digital Questions						
Sub-Sample	Abn. Low-Investment		Normal-Investment		Abn. High-Investment	
Dependent Variable	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity	Job Posting Probability	Job Posting Intensity
Digital Questions$_{i,t}$	0.616**	0.021**	0.349***	0.010	-0.123	0.028
	(0.254)	(0.009)	(0.104)	(0.007)	(0.169)	(0.019)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,300	2,300	24,375	24,375	2,407	2,407
R^2	0.487	0.344	0.505	0.299	0.519	0.538

This table reports the topics of managerial/analysts disclosure in sub-samples of abnormally low, high and normal AI investments, as well as the levels of future AI investment conditional on digital questions in the same sub-samples. Sub-samples are defined by residuals from an quarter-fixed effects regressions of AI investment on the five investment factors, and abnormally low (high) investment sub-samples are split by whether the residual is below (above) 0.75 σs from the mean. Panels A and B report the average differences in the topics disclosed by management and analysts across the sub-samples. Panel C reports the associations between future AI investment and digital questions across the sub-samples. Regressions in Panel C control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Standard errors in analyses reported in all panels are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

B.6.4 Robustness Analysis Based on Various Measures of AI Investment

As there are periodic gaps in the quarterly frequency of the *BurningGlass* data, I examine the robustness of the main results with an interpolation of zero for missing observations. Results are presented in Table B.7. Additionally, I also present the analysis with an alternative way of measuring AI investment intensity. This approach scales the total number of AI job postings with the employee counts in the firm. In Table B.8 I report this analysis. Lastly, in tables B.9 and B.10, I examine the robustness of the main results to AI job postings that are measured using the "narrow" Acemoglu et al. (2020) AI skills list.

Table B.7: Robustness Analysis: Future AI Investment Probability with Interpolation for Missing *BurningGlass* Observations

Panel A: Probability of Future AI Job Postings and Sub-Sample Analysis

Dependent Variable	Posting Probability	Posting Probability	Posting Probability	Posting Probability	Posting Probability
Sub-Sample	All	No Job Posting	Job Posting	No Firm Disclosure	Firm Disclosure
Digital Questions$_{i,q}$	0.246**	0.312***	-0.075	0.511*	0.097
	(0.110)	(0.115)	(0.164)	(0.280)	(0.117)
Firm FE	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	33,855	26,077	7,618	18,402	12,826
R^2	0.526	0.258	0.468	0.478	0.579

Panel B: Abnormally Low/High Investment Analysis

Sub-Sample	Abn. Low-Investment	Normal Investment	Abn. High-Investment
Dependent Variable	Posting Probability	Posting Probability	Posting Probability
Digital Questions$_{i,q}$	0.599***	0.130	-0.219
	(0.206)	(0.134)	(0.215)
Firm FE	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	6,226	23,808	3,244
R^2	0.480	0.515	0.564

This table reports the relationship between analysts' digital questions and the probability of future job postings of AI conditional on digital questions. For the indicator of AI job postings, I interpolate observations where there are missing 1-year ahead *BurningGlass* observations and valid 1-year ahead sales observations with zero. Panel A reports the associations between future AI job posting probability and digital questions across the main sample, the sub-samples of no/current AI job postings and no/current firm digital disclosure. Panel B reports the same associations in sub-samples of abnormally low, high and normal-levels of AI investments. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Firm and calendar quarter fixed effects are also included. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

Table B.8: Robustness Analysis: Future AI Investment Intensity with Job Postings Scaled by Employee Counts

Panel A: Intensity of Future AI Job Postings and Sub-Sample Analysis

Dependent Variable	Posting Intensity	Posting Intensity	Posting Intensity	Posting Intensity	Posting Intensity
Sub-Sample	All	No Job Posting	Job Posting	No Firm Disclosure	Firm Disclosure
Digital Questions$_{i,q}$	0.026**	0.019**	0.026	0.059**	0.015
	(0.011)	(0.009)	(0.022)	(0.024)	(0.012)
Firm FE	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	28,041	21,658	6,240	15,416	10,392
R^2	0.400	0.215	0.518	0.343	0.474

Panel B: Abn. Low/High-Investment Sub-Sample Analysis

Sub-Sample	Abn. Low-Investment	Normal Investment	Abn. High-Investment
Dependent Variable	Posting Intensity	Posting Intensity	Posting Intensity
Digital Questions$_{i,q}$	0.006	0.019	0.047
	(0.012)	(0.012)	(0.035)
Firm FE	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	5,137	20,096	2,342
R^2	0.332	0.394	0.592

This table reports the relationship between analysts' digital questions and the intensity of future job postings of AI conditional on digital questions. For the intensity of AI job postings, I scale the total number of AI job postings with the total number of employee counts. To reduce the influence of outliers, which occurs due to small values in the denominator, I trim the job posting intensity variable at the 95% level. Panel A reports the associations between future AI job posting probability and digital questions across the main sample, the sub-samples of no/current AI job postings and no/current firm digital disclosure. Panel B reports the same associations in sub-samples of abnormally low, high and normal-levels of AI investments. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Firm and calendar quarter fixed effects are also included. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

Table B.9: Robustness Analysis: Future AI Investment Probability with AI Job Postings Measured by Acemoglu et al. (2020) Skills List

Panel A: Probability of Future AI Job Postings and Sub-Sample Analysis

Dependent Variable	Posting Probability	Posting Probability	Posting Probability	Posting Probability	Posting Probability
Sub-Sample	All	No Job Posting	Job Posting	No Firm Disclosure	Firm Disclosure
Digital Questions$_{i,q}$	0.345***	0.373**	0.217	0.605*	0.218*
	(0.110)	(0.142)	(0.169)	(0.300)	(0.125)
Firm FE	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	29,518	22,254	7,128	15,938	11,290
R^2	0.517	0.312	0.430	0.485	0.550

Panel B: Abnormally Low/High Investment Analysis

Sub-Sample	Abn. Low-Investment	Normal Investment	Abn. High-Investment
Dependent Variable	Posting Probability	Posting Probability	Posting Probability
Digital Questions$_{i,q}$	0.491*	0.309*	-0.058
	(0.251)	(0.179)	(0.217)
Firm FE	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	5,407	20,623	2,995
R^2	0.475	0.516	0.520

This table reports the relationship between analysts' digital questions and the probability of future job postings of AI conditional on digital questions. The set of skills used to measure AI-related job postings follows Acemoglu et al. (2020). Panel A reports the associations between future AI job posting probability and digital questions across the main sample, the sub-samples of no/current AI job postings and no/current firm digital disclosure. Panel B reports the same associations in sub-samples of abnormally low, high and normal-levels of AI investments. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Firm and calendar quarter fixed effects are also included. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

Table B.10: Robustness Analysis: Future AI Investment Intensity with AI Job Postings Measured by Acemoglu et al. (2020) Skills List

Panel A: Intensity of Future AI Job Postings and Sub-Sample Analysis					
Dependent Variable	Posting Intensity	Posting Intensity	Posting Intensity	Posting Intensity	Posting Intensity
Sub-Sample	All	No Job Posting	Job Posting	No Firm Disclosure	Firm Disclosure
Digital Questions$_{i,q}$	0.015**	0.013*	0.011	0.026**	0.009
	(0.006)	(0.007)	(0.009)	(0.013)	(0.006)
Firm FE	Yes	Yes	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	29,518	22,254	7,128	15,938	11,290
R^2	0.389	0.243	0.502	0.337	0.442

Panel B: Abn. Low/High-Investment Sub-Sample Analysis			
Sub-Sample	Abn. Low-Investment	Normal Investment	Abn. High-Investment
Dependent Variable	Posting Intensity	Posting Intensity	Posting Intensity
Digital Questions$_{i,q}$	0.024**	0.000	0.022
	(0.011)	(0.005)	(0.017)
Firm FE	Yes	Yes	Yes
Calendar Quarter FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	5,407	20,623	2,995
R^2	0.317	0.306	0.499

This table reports the relationship between analysts' digital questions and the intensity of future job postings of AI conditional on digital questions. The set of skills used to measure AI-related job postings follows Acemoglu et al. (2020). Panel A reports the associations between future AI job posting intensity and digital questions across the main sample, the sub-samples of no/current AI job postings and no/current firm digital disclosure. Panel B reports the same associations in sub-samples of abnormally low, high and normal-levels of AI investments. Regressions control for the industry-level AI suitability index, the industry-level proportion of peer firms with AI job postings, herfindahl index, an indicator for whether the firm has top-10 executives with experience in digital technologies, an indicator for whether the firm has board members with experience in digital technologies, an indicator for whether the management team discloses digital-related terms in the presentation portion of the conference call, the industry-level proportion of peer firms with digital questions from analysts, the industry-level proportion of peer firms with managerial disclosure with digital topics, the number of segments, the capital constraints index, the logarithm of age, an indicator for firms in the later part of their lifecycle, industry-adjusted return-on-assets, industry-adjusted Tobin's Q, market beta, proportion of sales to major customers, R&D-to-sales, indicator for missing R&D, SG&A-to-sales and industry-adjusted sales growth. Firm and calendar quarter fixed effects are also included. Standard errors are clustered at the firm and calendar quarter-level. ***, **, * denote significance at the 1%, 5% and 10% level.

C

Appendix to Chapter 4: Capital Market Forces and Digital Investment in Non-IT Firms

C.1 Variable Definitions

Variable Name	Variable Description
Dependent Variables:	
Technology Acquisitions	Either a raw count or an indicator for calendar quarters where the firm engages in an acquisition that is recorded in the S&P's 451 M&A database.
Proportion of Software Workers	Proportion of software-related workers relative to total number of employees in the firm. Data is taken from *RevelioLabs*.
Salary-weighted Proportion of Software Workers	Proportion of software-related workers's salary relative to total salary of employees in the firm. Data is taken from *RevelioLabs*.
Key Independent Variables:	
Patent Value-Weighted AI Index	Weighted sum of the textual overlap scores between AI-related (identified using search-terms in Webb (2020)) patent abstract and titles, and 6-digit NAICS industry descriptions. Weights are the percentile rank of average market value (taken from, Kogan et al. 2017) of AI-related patents (patents identified by the Webb (2020) search criteria as well as follow-on patents that cite these patents) in the NAICS 3-digit industry. Scores are standardized with the mean and standard deviation.
Capital Constraints	The Whited-Wu index of external capital constraints (Whited and Wu 2006). The index computed from *Compustat* variables: $WW_{i,t} = 0.65 - 0.091 \times \frac{oancf_{i,t}}{at_{i,t}} - 0.062 \times 1_{div_{i,t}>0} + 0.021 \times \frac{dltt_{i,t}+dlc_{i,t}}{at_{i,t}} - 0.044 \times Ln(at_{i,t}) + 0.102 \times \frac{\Delta sale_{j,t}}{sale_{j,t-1}} + 0.035 \times \frac{\Delta sale_{i,t}}{sale_{i,t-1}}$.
Vesting Equity	Total amount of vesting equity in billions, computed with *ISS Incentive Lab* data, using the approach in Edmans et al. (2017).

CEO Compensation Control Variables:

Vested Equity	Total amount of vested equity in billions, computed with *ISS Incentive Lab* data at the annual level, using the approach in Edmans et al. (2017).
Unvested Equity	Total amount of unvested equity in billions, computed with *ISS Incentive Lab* data at the annual level, using the approach in Edmans et al. (2017).
Salary	Total amount of cash-based salary in billions, measured with *ISS Incentive Lab* data at the annual level.
Bonus	Total amount of cash-based bonus in billions, measured with *ISS Incentive Lab* data at the annual level..
Tenure	CEO tenure at the firm-level, computed using *ISS Incentive Lab* data.

Firm-Level Control Variables:

Age	firm age, computed as the number of years since the firm's first appearance in compustat.
Return-on-Assets	Return-on-assets, computed as operating income after depreciation divided by last-quarter assets.
Q	Tobin's Q, defined as market capitalization plus long-term debt and debt in current liabilities divided by total assets
Industry-Adjusted Sales Growth	Defined as the difference in current and 1-year prior sales divided by 1-year prior sales. Industry adjustment is relative to the median value at the industry (6-digit GICS)-quarter level.
Market Beta	β from a market model regression of $R_{it} = \alpha + \beta R_{mt} + \varepsilon_{it}$, where R_{mt} is the weekly equal-weighted market returns, estimated over weekly returns over rolling 3 years, and a minimum of 52 weeks. Measured at the beginning of the fiscal year.
SG&A-to-Sales	Selling, General and Administrative expenditure intensity, defined as SG&A expense minus R&D minus R&D in place scaled by sales. Following Peters and Taylor (2017), SG&A as reported in *Compustat* is not adjusted if R&D expense is greater than SG&A expense and if R&D expense is less than Cost of Goods Sold. If missing this variable is coded as 0.
R&D-to-Sales	Research and development expenditure intensity, defined as research and development expenditures scaled by sales. If missing, this variable is coded as 0.

Cash Balances	Computed as total cash and cash equivalents scaled by total assets
Retained Earnings	Computed as total retained earnings scaled by total assets
Stock Return Momentum	Past 3-month returns computed with *CRSP* returns data.
Number of segments	Total number of segments reported in *Compustat's* historical segment files. Measured at the annual level in the beginning of the fiscal year.
Age	firm age, computed as the number of years since the firm's first appearance in *Compustat*.
Market Beta	β from a market model regression of $R_{it} = \alpha + \beta R_{mt} + \varepsilon_{it}$, where R_{mt} is the weekly value-weighted market returns, estimated over weekly returns over rolling 3 years, and a minimum of 52 weeks.
Herfindahl	Herfindahl index, computed as the sum of squared sales-weight across all firms within each industry-year cell.
Proportion of Major Customer Sales	Proportion of sales that are due to major customers, as reported in *Compustat's* customer segment files.

C.2 Methodology for Computing Similarity Scores

To compute the similarity scores, I first compute the word embedding of each document (either patent (abstract and title) and the 6-digit 2017 NAICS industry description)[1], by characterising each word by GloVe vectors (Pennington et al. 2014)[2]. This method uses the co-occurrences of words to develop a mapping to a vector space which aligns words that are different but have similar meaning. For instance, a word like "dog" and "cat" will have a similar vector representation and thus the similarity between the two words will be assigned a high score, instead of the 0 computed in a traditional cosine similarity methodology. To aggregate to the document level, I take the weighted average of the word vectors in the whole document by the following formula:

$$X_i = \sum_{x_k \in A_i} w_{i,k} x_k \tag{C.1}$$

where X_i is the document vector and x_k is a word-vector in the set of vectors in document A_i. $w_{i,k}$ is the word TF-IDF weight which is determined as follows:

$$w_{i,k} = TF_{i,k} \times IDF_k \tag{C.2}$$

1. As a pre-processing step, I also remove stopwords, lemmatize the words and keep only nouns and verbs (based on the WordNet classification) in the patent and industry description text.

2. I use the largest set of pre-trained vector representations provided on the StanfordNLP website: https://nlp.stanford.edu/projects/glove/. The dataset is "glove.840B.300d", which is a dictionary trained on 840 B tokens from Common Crawl and maps words to 300-dimensional clusters.

where $TF_{i,k}$ is term frequency of word k in document i and IDF_k is the inverse document frequency of word k, which is computed as:

$$IDF_k = log(\frac{\#\ documents}{\#\ documents\ with\ word\ k}) \tag{C.3}$$

The similarity scores are then estimated by:

$$\rho_{i,j} = \frac{X_i}{|X_i|} \cdot \frac{X_j}{|X_j|} \tag{C.4}$$

To estimate the industry-year level exposure score, I follow Kogan et al. (2020) and apply the following aggregation formula on cosine similarity scores:

$$\eta_{i,t,\Gamma} = \sum_{j \in \Gamma_t} \tilde{\rho}_{i,j} \tag{C.5}$$

where $\eta_{i,t,\Gamma}$ is the is occupation or industry i exposure score in year t for technology class Γ. $\tilde{\rho}_{i,j}$ is the adjusted similarity scores for occupation i and patent j. The similarity score adjustment takes two steps: (1) year fixed effects are removed from the similarity scores and (2) sparsity is imposed by setting scores below the 80th percentile to 0[3].

Finally, to normalize the scores, the raw scores are deducted by the pooled mean and scaled by the pooled standard deviation.

3. Kogan et al. (2020) also scale the non-zero values to run from 0-1, but I do not do so as it gives undue weight to patent exposure in the early part of the sample.

C.3 Patent Text – Industry Description

Example 1: Patent No. US-2002099645-A1 and NAICS 523110

Similarity Score: 0.490

Patent Title: Process of selecting portfolio managers based on automated artificial intelligence techniques

A method and a system applying data mining techniques and artificial intelligence algorithms, namely neural networks, operating via an Internet data exchange site, allowing portfolio management companies to access an on-line, standardized questionnaire (Request for Proposal) and present their capabilities. The method then analyses a large number of these questionnaires and classifies managers, categorizing them and ranking their capabilities. In addition, sponsors such as pension funds, endowments, and private clients can submit their offer for the management of their assets. The site provides a Request for Proposal that is modular and continuously adapted to new financial market conditions, legal considerations and Sponsors needs. Once completed by each manager, the questionnaire is maintained in a central database allowing each manager to have an up-to-date and on-line version of questionnaire. The site allows the manager to have access to offers for new mandates in an easier and more efficient way than the old approach that involves approaching each potential Sponsor individually in a time and money consuming manner. The site provides Sponsors the opportunity to request proposals for their mandate from a much higher number of management companies, thus increasing dramatically the efficiency and rationality of their final choice of managers.

Industry Title: Investment Advice

This industry comprises establishments primarily engaged in providing customized investment advice to clients on a fee basis, but do not have the authority to execute trades. Primary activities performed by establishments in this industry are providing financial planning advice and investment counseling to meet the goals and needs of specific clients.

Example 2: Patent No. US-2020073899-A1 and NAICS 541810

Similarity Score: 0.348

Patent Title: Content opportunity scoring and automation

This invention relates to marketing and creation of digital content (text, voice, video, imagery, etc.) and understanding what topics are most relevant for an intended audience, its size, the type of content that audience wants to consume, and the optimal distribution method (social media, email, podcasts, voice assistants, web pages, mobile apps, etc.), and leveraging machine learning to automatically create content with a high chance of success.

Industry Title: Advertising Agencies

This industry comprises establishments primarily engaged in creating advertising campaigns and placing such advertising in periodicals, newspapers, radio and television, or other media. These establishments are organized to provide a full range of services (i.e., through in-house capabilities or subcontracting), including advice, creative services, account management, production of advertising material, media planning, and buying (i.e., placing advertising).

Example 3: Patent No. US-2017286838-A1 and NAICS 221114

Similarity Score: 0.409

Patent Title: Predicting solar power generation using semi-supervised learning

A method for predicting solar power generation receives historical power profile data and historical weather micro-forecast data at a given location for a set of days. Based on power output features for the days, clusters are generated. A classification model that assigns a day to a generated cluster according to weather features is created. For each cluster, a regression model that takes as input weather features and outputs predicted solar power is built. A system includes a sensor for collecting meteorological data at a solar farm, a meter for measuring photovoltaic power output of the solar farm, and a computer processor for executing instructions to predict solar power generation at the solar farm according to the method disclosed, based on data from the sensor and the meter, for a predefined time period. Further instructions predict solar power generation at the solar farm based on a micro-forecast for the solar farm.

Industry Title: Solar Electric Power Generation

This U.S. industry comprises establishments primarily engaged in operating solar electric power generation facilities. These facilities use energy from the sun to produce electric energy. The electric energy produced in these establishments is provided to electric power transmission systems or to electric power distribution systems.

C.4 Additional Validation Analysis

In this section, I present various validation for the textual overlap scores. Table C.1 presents the validation analysis for the unweighted overlap score for expected digital technology productivity. Table C.2 presents the validation analysis for the overlap scores that are computed with global or US corporate patents. Table C.3 presents the validation analysis for the overlap scores that are computed with 4 or 5-digit NAICS industry descriptions.

Table C.1: External Validation of Unweighted Expected Digital Technology Productivity Measure

	Unweighted AI Index	Unweighted AI Index	Unweighted AI Index	Unweighted AI Index	Unweighted AI Index	Unweighted AI Index
Investment Correlation$_j$	0.376***					
	(0.038)					
Deal Value-to-Assets$_j$		0.277***				
		(0.040)				
Digital Intensity$_j$			0.508***			
			(0.050)			
Internet Adoption$_j$				-0.281***		
				(0.075)		
Occupation Exposure$_j$					0.370***	
					(0.047)	
Occupation Exposure (Kogan)$_j$						0.199***
						(0.041)
Observations	562	535	562	151	560	560
R^2	0.141	0.077	0.174	0.074	0.096	0.040

This table reports the validation analysis of the unweighted AI index measure of digital technology productivity. In this table, I examine the associations between the cross-industry variation of the ranked unweighted textual overlap score with other (ranked) metrics of digital investment productivity. I examine 6 measure of digital productivity — (1) the correlation in IT-related acquisitions in the industry-level, (2) the average deal value-to-assets for IT acquisitions at the industry-level, (3) the global taxonomy of digital intensive industries from the OECD, (4) the index of industries that have adopted the internet in 1998-2000, (5) the occupation exposure score from Webb (2020), (6) and from Kogan et al. (2020). Robust standard errors are reported in parantheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

Table C.2: External Validation of Expected Digital Technology Productivity Measure (Global and Corporate Patents)

Panel A: Textual-Based Digital Productivity Score (Global Patents) and Other Metrics of Digital Productivity

	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index
Investment Correlation$_j$	0.272***					
	(0.040)					
Deal Value-to-Assets$_j$		0.239***				
		(0.041)				
Digital Intensity$_j$			0.335***			
			(0.052)			
Internet Adoption$_j$				-0.214***		
				(0.079)		
Occupation Exposure$_j$					0.372***	
					(0.047)	
Occupation Exposure (Kogan)$_j$						0.218***
						(0.040)
Observations	562	535	562	151	560	560
R^2	0.074	0.057	0.076	0.040	0.097	0.047

Panel B: Textual-Based Digital Productivity Score (US Corporate Patents) and Other Metrics of Digital Productivity

	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index
Investment Correlation$_j$	0.295***					
	(0.040)					
Deal Value-to-Assets$_j$		0.256***				
		(0.041)				
Digital Intensity$_j$			0.385***			
			(0.054)			
Internet Adoption$_j$				-0.318***		
				(0.079)		
Occupation Exposure$_j$					0.381***	
					(0.049)	
Occupation Exposure (Kogan)$_j$						0.168***
						(0.041)
Observations	562	535	562	151	560	560
R^2	0.087	0.065	0.100	0.079	0.101	0.028

This table reports the validation analysis of the patent value-weighted AI index measure of digital technology productivity, based on either the set of global patents or US corporate patents. In Panel A, I examine the associations between the cross-industry variation of the ranked weighted textual overlap score with other (ranked) metrics of digital investment productivity. I examine 6 measure of digital productivity — (1) the correlation in IT-related acquisitions in the industry-level, (2) the average deal value-to-assets for IT acquisitions at the industry-level, (3) the global taxonomy of digital intensive industries from the OECD, (4) the index of industries that have adopted the internet in 1998-2000, (5) the occupation exposure score from Webb (2020), (6) and from Kogan et al. (2020). Robust standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

Table C.3: External Validation of Expected Digital Technology Productivity Measure (4 and 5 Digit NAICS Patents)

Panel A: Textual-Based Digital Productivity Score (4-digit NAICS) and Other Metrics of Digital Productivity

	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index
Investment Correlation$_j$	0.305***					
	(0.079)					
Deal Value-to-Assets$_j$		0.360***				
		(0.077)				
Digital Intensity$_j$			0.402***			
			(0.094)			
Internet Adoption$_j$				-0.503***		
				(0.126)		
Occupation Exposure$_j$					0.356***	
					(0.090)	
Occupation Exposure (Kogan)$_j$						0.206***
						(0.077)
Observations	153	145	153	44	152	152
R^2	0.093	0.131	0.119	0.165	0.108	0.043

Panel B: Textual-Based Digital Productivity Score (5-digit NAICS) and Other Metrics of Digital Productivity

	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index	Weighted AI Index
Investment Correlation$_j$	0.220***					
	(0.050)					
Deal Value-to-Assets$_j$		0.309***				
		(0.049)				
Digital Intensity$_j$			0.426***			
			(0.065)			
Internet Adoption$_j$				-0.414***		
				(0.089)		
Occupation Exposure$_j$					0.390***	
					(0.059)	
Occupation Exposure (Kogan)$_j$						0.193***
						(0.050)
Observations	362	347	362	103	360	360
R^2	0.048	0.096	0.117	0.143	0.110	0.037

This table reports the validation analysis of the patent value-weighted AI index measure of digital technology productivity, based on 4 or 5-digit NAICS industry descriptions. In Panel A, I examine the associations between the cross-industry variation of the ranked weighted textual overlap score with other (ranked) metrics of digital investment productivity. I examine 6 measure of digital productivity — (1) the correlation in IT-related acquisitions in the industry-level, (2) the average deal value-to-assets for IT acquisitions at the industry-level, (3) the global taxonomy of digital intensive industries from the OECD, (4) the index of industries that have adopted the internet in 1998-2000, (5) the occupation exposure score from Webb (2020), (6) and from Kogan et al. (2020). Robust standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

C.5 Methodology for Computing Vesting Equity

The data used to construct the vesting equity measure, comes from *ISS Incentive Lab*, which tracks the compensation packages of top executives in S&P 500 and S&P 400 firms. And this data is drawn from mandatory compensation disclosure in proxy statements of these firms.

With the compensation data from *ISS Incentive Lab*, I construct the vesting equity (from both stock and options) for CEOs at the quarterly frequency following the approach in Edmans et al. (2017). For options, I first compute the vesting amounts for the year, by calculating the changes in unexercisable options for each strike price-expiry date pair, and distribute across quarters by inferring from the anniversaries of the expiry date in each strike price-expiry date pair. I then compute the total vesting option, by computing the total delta for the amount of options that vest in a particular quarter.

For restricted stock, expiry dates are not available to estimate vesting dates and amounts. So I use the vesting schedules and the grant dates to infer the vesting dates and amounts. For CEOs there are two main types of restricted equity grants that are computed differently. First, for the cliff-vested grants, the vesting amounts are assigned to the quarter at the end of the vesting period. Second, for the graded-vesting grants, the vesting amounts are assigned annually on a straight-line basis (to the end of the vesting period) to quarters that align with the anniversaries of the grant-date. I then compute the total vesting equity, by multiplying the vesting amounts with the share price and adding the total vesting options, described in the previous paragraph.

For more details on the methodology and for specific examples, see the Appendix in Edmans et al. (2017).

Bibliography

Abdel-Meguid, A., J. N. Jennings, K. J. Olsen, and M. T. Soliman. 2021. "The Impact of the CEO's Personal Narcissism on Non-GAAP Earnings." *The Accounting Review* 96, no. 3 (May): 1–25.

Abis, S., and L. Veldkamp. 2020. "The Changing Economics of Knowledge Production." *Working Paper.*

Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo. 2020. "AI and Jobs: Evidence from Online Vacancies." *NBER Working Paper No. 28257.*

Acemoglu, D., and P. Restrepo. 2019. "Automation and new tasks: How technology displaces and reinstates labor." *Journal of Economic Perspectives* 33, no. 2 (March): 3–30.

Adams, R. B., A. C. Akyol, and P. Verwijmeren. 2018. "Director skill sets." *Journal of Financial Economics* 130, no. 3 (December): 641–662.

Aggarwal, V. A., and D. H. Hsu. 2014. "Entrepreneurial exits and innovation." *Management Science* 60, no. 4 (December): 867–887.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt. 2005. "Competition and Innovation: an Inverted-U Relationship." *The Quarterly Journal of Economics* 120, no. 2 (May): 701–728.

Agrawal, A., S. Chadha, and M. A. Chen. 2006. "Who is afraid of reg FD? the behavior and performance of sell-side analysts following the SEC's fair disclosure rules." *Journal of Business* 79, no. 6 (November): 2811–2834.

Akey, P., S. Lewellen, I. Liskovich, and C. Schiller. 2021. "Hacking Corporate Reputations." *Rotman School of Management Working Paper* 3143740.

Alekseeva, L., J. Azar, M. Giné, S. Samila, and B. Taska. 2021. "The demand for AI skills in the labor market." *Labour Economics* 71 (August): 102002.

Amir, E., and B. Lev. 1996. "Value-relevance of nonfinancial information: The wireless communications industry." *Journal of Accounting and Economics* 22 (1-3): 3–30.

Arrow, K. 1962. "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity: Economic and Social Factors,* edited by R. Nelson, 609–626. National Bureau of Economic Research, Inc.

Asker, J., J. Farre-Mensa, and A. Ljungqvist. 2014. "Corporate Investment and Stock Market Listing: A Puzzle?" *The Review of Financial Studies* 28, no. 2 (October): 342–390.

Autor, D. H., F. Levy, and R. J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118, no. 4 (November): 1279–1333.

Autor, D. H., D. Dorn, and G. H. Hanson. 2015. "Untangling Trade and Technology: Evidence from Local Labour Markets." *The Economic Journal* 125, no. 584 (May): 621–646.

Autor, D. H. 2015. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." *Journal of Economic Perspectives* 29 (3): 3–30.

Babina, T., A. Fedyk, A. X. He, and J. Hodson. 2020. "Artificial Intelligence, Firm Growth, and Industry Concentration." *SSRN Electronic Journal* (September).

Bai, J. (, W. Jin, S. Steffen, and C. Wan. 2020. "The Future of Work: Work from Home Preparedness and Firm Resilience During the COVID-19 Pandemic." *SSRN Electronic Journal* (June).

Baker, G. P., and T. N. Hubbard. 2004. "Contractibility and Asset Ownership: On-Board Computers and Governance in U. S. Trucking." *The Quarterly Journal of Economics* 119, no. 4 (November): 1443–1479.

Baker, M., J. C. Stein, and J. Wurgler. 2003. "When Does the Market Matter? Stock Prices and the Investment of Equity-Dependent Firms." *The Quarterly Journal of Economics* 118, no. 3 (August): 969–1005.

Bakke, T.-E., and T. M. Whited. 2012. "Threshold Events and Identification: A Study of Cash Shortfalls." *The Journal of Finance* 67, no. 3 (June): 1083–1111.

Bartov, E., P. Mohanram, and C. Seethamraju. 2002. "Valuation of Internet Stocks: An IPO Perspective." *Journal of Accounting Research* 40, no. 2 (December): 321–346.

Bass, A. S. 2018. "Non-tech businesses are beginning to use artificial intelligence at scale." *The Economist,* 3–5.

Bebchuk, L. A., and L. A. Stole. 1993. "Do Short-Term Objectives Lead to Under-or Overinvestment in Long-Term Projects?" *The Journal of Finance* 48, no. 2 (June): 719–730.

Beneish, M. D. 1991. "Stock Prices and the Dissemination of Analysts' Recommendation." *The Journal of Business* 64, no. 3 (August): 393–416.

Bernstein, S. 2015. "Does Going Public Affect Innovation?" *Journal of Finance* 70, no. 4 (August): 1365–1403.

Bhojraj, S., P. Hribar, M. Picconi, and J. McInnis. 2009. "Making sense of cents: An examination of firms that marginally miss or beat analyst forecasts." *Journal of Finance* 64, no. 5 (October): 2361–2388.

Bhojraj, S., C. M. C. Lee, and D. K. Oler. 2003. "What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research." *Journal of Accounting Research* 41, no. 5 (December): 745–774.

Bloom, N., T. A. Hassan, A. Kalyani, J. Lerner, and A. Tahoun. 2021. "The Diffusion of Disruptive Technologies." *NBER Working Paper 28999.*

Bloom, N., J. Lerner, N. Short, and M. Webb. 2018. "Some Facts of High-Tech Patenting." *HBS Working Paper* 19-014.

Bloom, N., R. Sadun, and J. V. Reenen. 2012. "Americans Do IT Better: US Multinationals and the Productivity Miracle." *American Economic Review* 102, no. 1 (February): 167–201.

Blundell, R., S. Bond, M. Devereux, and F. Schiantarelli. 1992. "Investment and Tobin's Q. Evidence from company panel data." *Journal of Econometrics* 51, nos. 1-2 (January): 233–257.

Blundell, R., R. Griffith, and J. Van Reenen. 1999. "Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms." *The Review of Economic Studies* 66, no. 3 (August): 529–554.

Bowen, R. M., A. K. Davis, and D. A. Matsumoto. 2002. "Do Conference Calls Affect Analysts' Forecasts?" *The Accounting Review* 77, no. 2 (February): 285–316.

Brav, A., W. Jiang, and H. Kim. 2015. "The Real Effects of Hedge Fund Activism: Productivity, Asset Allocation, and Labor Outcomes." *Review of Financial Studies* 28, no. 10 (October): 2723–2769.

Brav, A., W. Jiang, S. Ma, and X. Tian. 2018. "How does hedge fund activism reshape corporate innovation?" *Journal of Financial Economics* 130, no. 2 (November): 237–264.

Bresnahan, T. 2010. "General purpose technologies." In *Handbook of the Economics of Innovation,* 2:761–791. 1. Elsevier B.V., January.

Bresnahan, T. F., and S. Greenstein. 1996. "Technical Progress and Co-invention in Computing and in the Uses of Computers." *Brookings Papers on Economic Activity: Microeconomics,* 1–83.

Bresnahan, T. F., R. M. Henderson, and S. Greenstein. 2012. "Schumpeterian Competition and Diseconomies of Scope: Illustrations from the Histories of Microsoft and IBM." In *The Rate and Direction of Inventive Activity Revisted,* edited by J. Lerner and S. Stern. National Bureau of Economic Research, Inc.

Bresnahan, T. F., and M. Trajtenberg. 1995. "General purpose technologies 'Engines of growth'?" *Journal of Econometrics* 65, no. 1 (January): 83–108.

Brown, L. D., A. C. Call, M. Clement, and N. Y. Sharp. 2015. "Inside the "Black Box" of Sell-Side Financial Analysts." *Journal of Accounting Research* 53, no. 1 (March): 1–47.

Brynjolfsson, E., and L. Hitt. 1996. "Paradox Lost? Firm-Level Evidence on the Returns to Information Systems Spending." *Management Science* 42, no. 4 (April): 541–558.

———. 2000. "Beyond Computation: Information Technology, Organizational Transformation and Business Performance." *Journal of Economic Perspectives* 14, no. 4 (November): 23–48.

———. 2003. "Computing productivity: Firm-level evidence." *Review of Economics and Statistics* 85 (4): 793–808.

Brynjolfsson, E., T. Mitchell, and D. Rock. 2018. "What Can Machines Learn, and What Does It Mean for Occupations and the Economy?" *American Economic Association Papers and Proceedings* 108 (54): 43–47.

Brynjolfsson, E., D. Rock, and C. Syverson. 2017. "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics." *NBER Working Paper* 24001.

Brynjolfsson, E., and M. D. Smith. 2000. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." *Management Science* 46, no. 4 (April): 563–585.

Budish, E., B. N. Roin, and H. Williams. 2015. "Do firms underinvest in long-term research? Evidence from cancer clinical trials." *American Economic Review* 105, no. 7 (July): 2044–2085.

Bushee, B. J. 1998. "The Influence of Institutional Investors on Myopic R&D Investment Behavior." *The Accounting Review* 73 (3): 305–333.

Calvino, F., C. Criscuolo, L. Marcolin, and M. Squicciarini. 2018. "A taxonomy of digital intensive sectors." *OECD Science, Technology and Industry Working Papers.*

Campbell, J. L., H. Chen, D. S. Dhaliwal, H. m. Lu, and L. B. Steele. 2014. "The information content of mandatory risk factor disclosures in corporate filings." *Review of Accounting Studies* 19, no. 1 (March): 396–455.

Cao, R., and M. Iansiti. 2021. "Growth, Transformation and Digital Capital: The Importance of Technological and Organizational Architecture." *HBS Working Paper.*

Carhart, M. M. 1997. "On persistence in mutual fund performance." *Journal of Finance* 52, no. 1 (March): 57–82.

Chemmanur, T. J., L. Kong, K. Krishnan, and Q. Yu. 2019. "Top Management Human Capital, Inventor Mobility, and Corporate Innovation." *Journal of Financial and Quantitative Analysis* 54, no. 6 (December): 2383–2422.

Chen, M. A., Q. Wu, and B. Yang. 2018. "How Valuable is FinTech Innovation?" *Review of Financial Studies* Forthcomin (September).

Chen, O., M. Rakhlenko, and J. Kim. 2017. "WMT: Stores Are a Competitive Weapon To Reinvent Retail." *Cowen Equity Research.*

Chen, O., M. Rakhlenko, C. Origenes, and C. Wilson. 2017. "WMT vs. TGT Preview, the Digital Battle Intensifies." *Cowen Equity Research.*

Chen, T., L. Xie, and Y. Zhang. 2017. "How does analysts' forecast quality relate to corporate investment efficiency?" *Journal of Corporate Finance* 43 (April): 217–240.

Cho, C., J. T. Halford, S. Hsu, and L. Ng. 2016. "Do managers matter for corporate innovation?" *Journal of Corporate Finance* 36 (February): 206–229.

Choi, J. K., R. N. Hann, M. Subasi, and Y. Zheng. 2020. "An Empirical Analysis of Analysts' Capital Expenditure Forecasts: Evidence from Corporate Investment Efficiency*." *Contemporary Accounting Research* 37, no. 4 (December): 2615–2648.

Christensen, C. M. 1997. *The Innovator's Dilemma.* Boston: Harvard Business School Press.

Cockburn, I., R. Henderson, and S. Stern. 2017. "The Impact of Artificial Intelligence on Innovation." *NBER Working Paper* 24449.

Collins, D. W., and S. Kothari. 1989. "An analysis of intertemporal and cross-sectional determinants of earnings response coefficients." *Journal of Accounting and Economics* 11, nos. 2-3 (July): 143–181.

Cooper, M. J., O. Dimitrov, and P. R. Rau. 2001. "A Rose.com by any other name." *Journal of Finance* 56, no. 6 (December): 2371–2388.

Core, J. E., W. R. Guay, and A. Van Buskirk. 2003. "Market valuations in the New Economy: An investigation of what has changed." *Journal of Accounting and Economics* 34 (1-3): 43–67.

Crawford, S. S., D. T. Roulstone, and E. C. So. 2012. "Analyst initiations of coverage and stock return synchronicity." *Accounting Review* 87, no. 5 (September): 1527–1553.

Curtis, A., S. McVay, and S. Toynbee. 2020. "The changing implications of research and development expenditures for future profitability." *Review of Accounting Studies* 25, no. 2 (June): 405–437.

Custódio, C., M. A. Ferreira, and P. Matosc. 2019. "Do general managerial skills spur innovation?" *Management Science* 65, no. 2 (February): 459–476.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks." *The Journal of Finance* 52, no. 3 (July): 1035–1058.

Daniel, K., and S. Titman. 1997. "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns." *The Journal of Finance* 52, no. 1 (March): 1–33.

Dechow, P. M., A. P. Hutton, L. Meulbroek, and R. G. Sloan. 2001. "Short-sellers, fundamental analysis, and stock returns." *Journal of Financial Economics* 61, no. 1 (July): 77–106.

Dechow, P. M., R. G. Sloan, and J. Zha. 2014. *Stock prices and earnings: A history of research,* December.

Dehaan, E., J. Madsen, and J. D. Piotroski. 2017. "Do Weather-Induced Moods Affect the Processing of Earnings News?" *Journal of Accounting Research* 55 (3): 509–550.

Demers, E., and B. Lev. 2001. "A rude awakening: Internet shakeout in 2000." *Review of Accounting Studies* 6 (2-3): 331–359.

Derrien, F., and A. Kecskés. 2013. "The real effects of financial shocks: Evidence from exogenous changes in analyst coverage." *Journal of Finance* 68, no. 4 (August): 1407–1440.

Dickinson, V. 2011. "Cash flow patterns as a proxy for firm life cycle." *Accounting Review* 86, no. 6 (November): 1969–1994.

Dong, M., D. Hirshleifer, S. Richardson, and S. H. Teoh. 2006. "Does Investor Misvaluation Drive the Takeover Market?" *The Journal of Finance* 61, no. 2 (February): 725–762.

Dranove, D., C. Forman, A. Goldfarb, and S. Greenstein. 2014. "The Trillion Dollar Conundrum: Complementarities and Health Information Technology." *American Economic Journal: Economic Policy* 6, no. 4 (November): 239–270.

Easton, P. D., and M. E. Zmijewski. 1989. "Cross-sectional variation in the stock market response to accounting earnings announcements." *Journal of Accounting and Economics* 11, nos. 2-3 (July): 117–141.

Edmans, A., V. W. Fang, and A. Huang. 2018. "The Long-Term Consequences of Short-Term Incentives." *SSRN Electronic Journal* (April).

Edmans, A., V. W. Fang, and K. A. Lewellen. 2017. "Equity Vesting and Investment." *The Review of Financial Studies* 30, no. 7 (July): 2229–2271.

Erickson, T., and T. M. Whited. 2000. "Measurement error and the relationship between investment and q." *Journal of Political Economy* 108, no. 5 (July): 1027–1057.

Fama, E. F., and K. R. French. 1992. "The Cross-Section of Expected Stock Returns." *The Journal of Finance* 47, no. 2 (June): 427–465.

———. 2015. "A five-factor asset pricing model." *Journal of Financial Economics* 116, no. 1 (April): 1–22.

Fazzari, S. M., R. G. Hubbard, B. C. Petersen, A. S. Blinder, and J. M. Poterba. 1988. "Financing Constraints and Corporate Investment." *Brookings Papers on Economic Activity* 1988, no. 1 (May): 141–206.

Feldman, R., S. Govindaraj, J. Livnat, and B. Segal. 2010. "Management's tone change, post earnings announcement drift and accruals." *Review of Accounting Studies* 15, no. 4 (December): 915–953.

Felten, E., M. Raj, and R. Seamans. 2021. "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses." *Strategic Management Journal.*

Fitzgerald, T., B. Balsmeier, L. Fleming, and G. Manso. 2020. "Innovation Search Strategy and Predictable Returns." *Management Science* Forthcomin (August).

Forman, C., A. Goldfarb, and S. Greenstein. 2003. "Which Industries Use the Internet?" In *Advances in Applied Microeconomics Vol 12: Organizing the New Industrial Economy,* edited by M. R. Baye, 47–72. Elsevier.

Freeman, R. N., and S. Y. Tse. 1992. "A Nonlinear Model of Security Price Responses to Unexpected Earnings." *Journal of Accounting Research* 30, no. 2 (23): 185.

Fuller, J., and M. C. Jensen. 2010. "Just Say No to Wall Street: Putting a Stop to the Earnings Game." *Journal of Applied Corporate Finance* 22, no. 1 (January): 59–63.

Giannetti, M., and X. Yu. 2020. "Adapting to Radical Change: The Benefits of Short-Horizon Investors." *Management Science* (October).

Gigler, F., C. Kanodia, H. Sapra, and R. Venugopalan. 2014. "How Frequent Financial Reporting Can Cause Managerial Short-Termism: An Analysis of the Costs and Benefits of Increasing Reporting Frequency." *Journal of Accounting Research* 52, no. 2 (May): 357–387.

Gipper, B., C. Leuz, and M. Maffett. 2019. "Public Oversight and Reporting Credibility: Evidence from the PCAOB Audit Inspection Regime." *The Review of Financial Studies* (December).

Goldfarb, A., B. Taska, and F. Teodoridis. 2019. "Could Machine Learning Be a General-Purpose Technology? Evidence from Online Job Postings." *SSRN Electronic Journal* (October).

Graham, J. R., C. R. Harvey, and S. Rajgopal. 2005. "The economic implications of corporate financial reporting." *Journal of Accounting and Economics* 40, nos. 1-3 (December): 3–73.

Green, J., J. R. M. Hand, and X. F. Zhang. 2016. "Errors and questionable judgments in analysts' DCF models." *Review of Accounting Studies 2016 21:2* 21, no. 2 (February): 596–632.

Greenstein, S. 2017. "The reference wars: Encyclopaedia Britannica's decline and Encarta's emergence." *Strategic Management Journal* 38, no. 5 (May): 995–1017.

Grullon, G., S. Michenaud, and J. P. Weston. 2015. "The real effects of short-selling constraints." *Review of Financial Studies* 28, no. 6 (June): 1737–1767.

Hayashi, F. 1982. "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica* 50, no. 1 (May): 213–224.

He, J., and X. Tian. 2013. "The dark side of analyst coverage: The case of innovation." *Journal of Financial Economics* 109, no. 3 (September): 856–878.

Healy, P., G. Serafeim, S. Srinivasan, and G. Yu. 2014. "Market competition, earnings management, and persistence in accounting profitability around the world." *Review of Accounting Studies* 19, no. 4 (January): 1281–1308.

Healy, P. M., and K. G. Palepu. 2003. "The Fall of Enron." *Journal of Economic Perspectives* 17, no. 2 (May): 3–26.

Helpman, E. 2003. *General Purpose Technologies and Economic Growth.* Edited by E. Helpman. Cambrdige, MA: MIT Press.

Henderson, R. 1993. "Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry." *The RAND Journal of Economics* 24, no. 2 (22): 248.

Henderson, R. M., and K. B. Clark. 1990. "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms." *Administrative Science Quarterly* 35, no. 1 (March): 9.

Hirshleifer, D., P. H. Hsu, and D. Li. 2013. "Innovative efficiency and stock returns." *Journal of Financial Economics* 107, no. 3 (March): 632–654.

———. 2018. "Innovative originality, profitability, and stock returns." *Review of Financial Studies* 31, no. 7 (July): 2553–2605.

Hitt, L. M. 1999. "Information Technology and Firm Boundaries: Evidence from Panel Data." *Information Systems Research* 10 (2): 134–149.

Holmstrom, B., and J. Tirole. 1997. "Financial Intermediation, Loanable Funds, and the Real Sector." *The Quarterly Journal of Economics* 112, no. 3 (February): 663–691.

Hong, H., J. D. Kubik, and A. Solomon. 2000. "Security Analysts' Career Concerns and Herding of Earnings Forecasts." *The RAND Journal of Economics* 31, no. 1 (21): 121.

Hope, O. K., D. Hu, and H. Lu. 2016. "The benefits of specific risk-factor disclosures." *Review of Accounting Studies* 21, no. 4 (December): 1005–1045.

Huang, A., H. Wang, and Y. Yang. 2020. "FinBERT—A Deep Learning Approach to Extracting Textual Information." *SSRN Electronic Journal* (July).

Huang, A. H., R. Lehavy, A. Y. Zang, and R. Zheng. 2017. "Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach." *Management Science* 64, no. 6 (June): 2833–2855.

Irani, R. M., and D. Oesch. 2016. "Analyst Coverage and Real Earnings Management: Quasi-Experimental Evidence." *The Journal of Financial and Quantitative Analysis* 51, no. 2 (May): 589–627.

Islam, E., and J. Zein. 2020. "Inventor CEOs." *Journal of Financial Economics* 135, no. 2 (February): 505–527.

Jaffe, A. B., and J. Lerner. 2004. *Innovation and Its Discontents.* Princeton University Press, July.

Jennings, J., H. Seo, and M. T. Soliman. 2020. "The market's reaction to changes in relative performance rankings." *Review of Accounting Studies* 25, no. 2 (June): 672–725.

Joos, P., J. D. Piotroski, and S. Srinivasan. 2016. "Can analysts assess fundamental risk and valuation uncertainty? An empirical analysis of scenario-based value estimates." *Journal of Financial Economics* 121, no. 3 (September): 645–663.

Jung, M. J., F. M. H. Wong, and X. F. Zhang. 2018. "Buy-Side Analysts and Earnings Conference Calls." *Journal of Accounting Research* 56, no. 3 (June): 913–952.

Kaplan, S. N., and L. Zingales. 1997. "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?" *The Quarterly Journal of Economics* 112, no. 1 (February): 169–215.

Kaplan, S. N. 2018. "Are U.S. Companies Too Short-Term Oriented? Some Thoughts." *Journal of Applied Corporate Finance* 30, no. 4 (December): 8–18.

Keating, E. K., T. Z. Lys, and R. P. Magee. 2003. "Internet downturn: Finding valuation factors in Spring 2000." *Journal of Accounting and Economics* 34, nos. 1-3 (January): 189–236.

Kogan, L., D. Papanikolaou, L. Schmidt, and B. Seegmiller. 2020. "Technological Change and Occupations over the Long Run." *SSRN Electronic Journal* (May).

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman. 2017. "Technological Innovation, Resource Allocation, and Growth." *The Quarterly Journal of Economics* 132, no. 2 (May): 665–712.

Koh, P.-S., and D. M. Reeb. 2015. "Missing R&D." *Journal of Accounting and Economics* 60, no. 1 (August): 73–94.

Kothari, S. P., N. Mizik, and S. Roychowdhury. 2016. "Managing for the moment: The role of earnings management via real activities versus accruals in SEO valuation." *Accounting Review* 91, no. 2 (March): 559–586.

Lakhani, K., and M. Iansiti. 2020. *Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World.* Harvard Business School Press.

Law, K., and M. Shen. 2020. "How Does Artificial Intelligence Shape the Audit Industry?" *Nanyang Business School Research Paper* 20-31 (July).

Lee, C. M., S. T. Sun, R. Wang, and R. Zhang. 2019. "Technological links and predictable returns." *Journal of Financial Economics* 132, no. 3 (June): 76–96.

Leland, H. E., and D. H. Pyle. 1977. "Informational Asymmetries, Financial Structure, and Financial Intermediation." *The Journal of Finance* 32, no. 2 (February): 371–387.

Lev, B., and T. Sougiannis. 1996. "The capitalization, amortization, and value-relevance of R&D." *Journal of Accounting and Economics* 21 (1): 107–138.

Lev, B., and P. Zarowin. 1999. "The Boundaries of Financial Reporting and How to Extend Them." *Journal of Accounting Research* 37 (2): 353.

Li, F. 2008. "Annual report readability, current earnings, and earnings persistence." *Journal of Accounting and Economics* 45 (2-3): 221–247.

———. 2010. "The Information Content of Forward-Looking Statements in Corporate Filings-A Naïve Bayesian Machine Learning Approach." *Journal of Accounting Research* 48, no. 5 (December): 1049–1102.

Li, F., R. Lundholm, and M. Minnis. 2013. "A Measure of Competition Based on 10-K Filings." *Journal of Accounting Research* 51, no. 2 (May): 399–436.

Li, K., F. Mai, R. Shen, and X. Yan. 2021. "Measuring Corporate Culture Using Machine Learning." *The Review of Financial Studies* 34, no. 7 (June): 3265–3315.

Livdan, D., and A. Nezlobin. 2017. "Investment, Tobin's Q, and Vintage Capital: Theory and Evidence." *SSRN Electronic Journal* (November).

Ma, P., J. E. Shin, and C. C. Wang. 2021. "rTSR: Properties, Determinants, and Consequences of Benchmark Choice." *Harvard Business School Working Paper* 19 (112).

Martinez, S., J. S. Lowe, J. Roy, F. Gonzalez, and A. Chapa. 2019. "UBS Report on US Bank Technology Spending." *UBS Global Research.*

Matsumoto, D., M. Pronk, and E. Roelofsen. 2011. "What Makes Conference Calls Useful? The Information Content of Managers' Presentations and Analysts' Discussion Sessions." *The Accounting Review* 86, no. 4 (July): 1383–1414.

Matta, M., R. Vijayaraghavan, and H. Cavusoglu. 2019. "Board's IT Expertise and Firm Innovation." *SSRN Electronic Journal* (October).

McClintock, M., E. Reilly, and K. Short. 2016. "Target Corp. Diving Deep into the Digital Abyss." *Barclays Equity Research,* 1–19.

McNichols, M., M. V. Rajan, and S. Reichelstein. 2014. "Conservatism correction for the market-to-book ratio and Tobin's q." *Review of Accounting Studies* 19, no. 4 (February): 1393–1435.

Merkley, K., R. Michaely, and J. Pacelli. 2017. "Does the Scope of the Sell-Side Analyst Industry Matter? An Examination of Bias, Accuracy, and Information Content of Analyst Reports." *The Journal of Finance* 72, no. 3 (June): 1285–1334.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings* (January).

Myers, S. C., and N. S. Majluf. 1984. "Corporate financing and investment decisions when firms have information that investors do not have." *Journal of Financial Economics* 13, no. 2 (June): 187–221.

Nissim, D., and S. Penman. 2001. "Ratio analysis and equity valuation: From research to practice." *Review of Accounting Studies* 6:109–154.

O'Brien, P. C., and Y. Tian. 2006. "Financial Analysts' Role in the 1996-2000 Internet Bubble." *SSRN Electronic Journal* (December).

Oster, E. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business and Statistics* 37 (2): 187–204.

Palepu, K. G., and G. D. Elcock. 2001. "The Role of Capital Market Intermediaries in the Dot-Com Crash of 2000." *Harvard Business School Case 101-110.*

Pennington, J., R. Socher, and C. D. Manning. 2014. "GloVe: Global Vectors for Word Representation." *Empirical Methods in Natural Language Processing (EMNLP),* 1532–1543.

Peters, R. H., and L. A. Taylor. 2017. "Intangible capital and the investment-q relation." *Journal of Financial Economics* 123, no. 2 (February): 251–272.

Philippon, T. 2016. "The FinTech Opportunity." *NBER Working Paper* 22476 (August).

Piotroski, J. D., and D. T. Roulstone. 2004. *The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices,* 4, October.

Rajgopal, S., M. Venkatachalam, and S. Kotha. 2002. "Managerial Actions, Stock Returns, and Earnings: The Case of Business-to-Business Internet Firms." *Journal of Accounting Research* 40, no. 2 (May): 529–556.

Rauh, J. D. 2006. "Investment and financing constraints: Evidence from the funding of corporate pension plans." *Journal of Finance* 61, no. 1 (February): 33–71.

Resutek, R. 2021. "The relation between R&D, earnings growth, operating leverage, and stock returns." *Working Paper.*

Rock, D. 2019. "Engineering Value: The Returns to Technological Talent and Investments in Artificial Intelligence." *SSRN Electronic Journal* (August).

Roe, M. J. 2018. "Stock market short-termism's impact." *University of Pennsylvania Law Review* 167, no. 1 (December): 71–121.

Seamans, R., and M. Raj. 2018. "AI, Labor, Productivity and the Need for Firm-Level Data." *NBER Working Paper,* 1–13.

Shumway, T. 1997. "The Delisting Bias in CRSP Data." *The Journal of Finance* 52, no. 1 (March): 327–340.

Shumway, T., and V. A. Warther. 1999. "The delisting bias in CRSP's Nasdaq data and its implications for the size effect." *Journal of Finance* 54, no. 6 (December): 2361–2379.

Soliman, M. T. 2008. "The Use of DuPont Analysis by Market Participants." *The Accounting Review* 83, no. 3 (February): 823–853.

———. 2005. "Using Industry-Adjusted DuPont Analysis to Predict Future Profitability." *SSRN Electronic Journal* (November).

Spence, M. 1973. "Job Market Signaling." *The Quarterly Journal of Economics* 87, no. 3 (February): 355–374.

Stein, J. C. 1989. "Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior." *The Quarterly Journal of Economics* 104, no. 4 (November): 655.

Tambe, P. 2014. "Big Data Investment, Skills, and Firm Value." *Management Science* 60, no. 6 (June): 1452–1469.

Tambe, P., L. Hitt, D. Rock, and E. Brynjolfsson. 2020. "Digital Capital and Superstar Firms." *NBER Working Paper 28285* (Cambridge, MA) (December).

Tambe, P., L. M. Hitt, D. Rock, and E. Brynjolfsson. 2019. "IT, AI and the Growth of Intangible Capital." *SSRN Electronic Journal* (July).

To, T. Y., M. Navone, and E. Wu. 2018. "Analyst coverage and the quality of corporate investment decisions." *Journal of Corporate Finance* 51 (August): 164–181.

Tobin, J. 1969. "A General Equilibrium Approach To Monetary Theory." *Journal of Money, Credit and Banking* 1, no. 1 (May): 15–29.

Tobin, J., and W. C. Brainard. 1977. "Asset Markets and the Cost of Capital." *Cowles Foundation Discussion Papers.*

Trueman, B. 1994. "Analyst Forecasts and Herding Behavior." *The Review of Financial Studies* 7, no. 1 (January): 97–124.

Trueman, B., F. M. H. Wong, and X. J. Zhang. 2001. "Back to basics: Forecasting the revenues of internet firms." *Review of Accounting Studies* 6 (2-3): 305–329.

Trueman, B., M. H. F. Wong, and X. J. Zhang. 2000. "The Eyeballs Have It : Searching for the Value in Internet Stocks." *Journal of Accounting Research* 38 (Supplement): 137–162.

Webb, M. 2020. "The Impact of Artificial Intelligence on the Labor Market." *Working Paper.*

Welch, I. 2000. "Herding among security analysts." *Journal of Financial Economics* 58, no. 3 (December): 369–396.

Whited, T. M., and G. Wu. 2006. "Financial Constraints Risk." *Review of Financial Studies* 19, no. 2 (June): 531–559.

Womack, K. L. 1996. "Do Brokerage Analysts' Recommendations Have Investment Value?" *The Journal of Finance* 51, no. 1 (March): 137–167.

Wu, L., L. Hitt, and B. Lou. 2020. "Data analytics, innovation, and firm productivity." *Management Science* 66, no. 5 (May): 2017–2039.

You, H., and X. j. Zhang. 2009. "Financial reporting complexity and investor underreaction to 10-k information." *Review of Accounting Studies* 14, no. 4 (September): 559–586.

Zhang, M. 2021. "Determinants and Consequences of Human Capital Management Disclosure." *SSRN Electronic Journal* (October).