Decrypting the Digital Economy: The Digital Alpha and Its Origins*

Mingyi Hua†

February 10, 2022

[Click for the latest version]

Abstract

I study the stock market consequences of digitalization. I propose a novel dynamic measure of digitalization that holistically captures a firm's exposure to computers, data analytics, and programming. I find that digital firms, compared to non-digital firms, have annual realized excess returns which are 6.5% higher over the past two decades. This digital alpha does not appear to be explained by well-known stock return predictors nor firm characteristics such as age, size, profitability, or R&D intensity. Instead, the digital alpha is concentrated in firms which focus on users, a historically neglected party in the production-consumption chain. User-centricity complements digitalization within firms. The digital alpha rises to 9.0% for user-focused firms. I conclude that this figure likely compensates for risk, as user-focused digital firms have greater systematic risk.

Keywords: Digitalization, users, digital economy, GDPR, O*NET, excess returns

JEL Classification: G10, G14, O30.

*I would like to thank Pat Akey, Jean-Noël Barrot, Jean-Edouard Colliard, François Derrien, Denis Gromb, Johan Hombert, Augustin Landier, Hugues Langlois, Xitong Li, Stefano Lovo, Song Ma, Ioanid Rosu, Daniel Schmidt, Christophe Spaenjers, and especially my advisor, Thierry Foucault, as well as participants at the Nova Finance PhD Final Countdown and HEC Brownbag seminar, and anonymous conference reviewers for valuable feedback and comments. All errors are my own.

†HEC Paris, Finance Department, 1 rue de la Liberation, Jouy en Josas 78350, France. Email: mingyi.hua@hec.edu.

1 Introduction

A dramatic shift in the composition of the largest firms in our economy has occurred over the past century. In 1917, many of the largest firms in the United States dealt with commodities such as steel, oil, and rubber. By 1967, while oil companies retained their prominence, manufacturing firms such as Kodak and General Motors, as well as retail firms such as Sears, crept into the top ten (Kauflin, 2019). The five largest companies in 2017, however, were all technology companies which did not exist in 1967 (Raul, 2014).

The digital revolution has engendered a structural change in the economy. People are shifting their consumption to flow through digital channels. Nowadays, people may communicate via Facetime, read books on their Kindle, and order dinner through Uber Eats. Digitalization is thus not only manifesting through new digital services, but also through the transformation of traditional physical goods markets. For instance, only 0.5% of total retail sales in 1999 were completed through e-commerce, but by 2019, this share had increased to 10.7% (US Census Bureau, 2021). At the center of this structural transformation is the changing relationship between the firm and its customers. Before, as exemplified by the manufacturing era, the focus was on the product. Consumers passively awaited the release of these products. In the digital era, the focus shifts from a physical product to a service and the associated user experience. This is the trend of servitization.[1] "User" is also a more accurate descriptor of the final audience in the digital era: people *use* and interact with these digital offerings rather than merely *consume* them.

In this paper, I ask, does digitalization have asset pricing implications, and if so, what could be driving the evolution of these digital firms in the digital economy? In particular, how does digitalization relate to the increasing user-centricity in business? While the digital revolution may have profound and far-reaching consequences on society, it is unclear whether it should be reflected in financial markets. Perhaps market participants already have the

[1]Servitization, first defined by Vandermerwe and Rada (1988), has become a vibrant and ever-expanding field of research in management. For example, see Lay (2014) or Mastrogiacomo et al. (2019) for surveys of this literature.

tools necessary to evaluate this new generation of digital firms. However, this is still an unanswered question. A large part of the difficulty is rooted in the challenge of appropriately measuring the digital economy and especially the extent of digitalization in firms and industries. Oftentimes researchers can only measure one aspect well – such as ICT or robot adoption within firms – or settle for binary industry classifications.

I propose a holistic measure of digitalization. My starting point is occupations because a firm's ability to offer digital goods and services depends on its workers. The digital revolution should be reflected through the increasing importance of various digital-related tasks, as surveyed by the U.S. Department of Labor's O*NET program. One facet of job tasks – work activities – provides a compelling illustration. Among dozens of work activities such as "caring for others" or "handling or moving objects", the activity "interacting with computers" has dramatically grown in importance concurrently with the digital revolution. In the earliest 2002 O*NET vintage, the importance of this activity to the average occupation was ranked at the 20th percentile. Yet in the O*NET survey in 2020, the importance of "interacting with computers" exceeded 94% of all other work activities.

My digitalization measure is based on three foundational concepts of the digital revolution: infrastructure, big data, and programming. First, digital firms must have the computer infrastructure necessary to support digitalization. These computers should be fully utilized and integral to workers' responsibilities. Would a firm be a digital firm if its workers rarely interacted with computers? Second, these workers should be handling and analyzing data. How can a firm participate in the digital revolution if it is not attempting to take advantage of the influx of big data generated by new technologies and the resulting associated interactions? Third, digital firms should be participating in the process of shaping digitalization as the latter is still very much an evolving phenomenon. Thus my measure aims to capture the intersection of technological advancements brought by digitization and the power of user data in generating firm value. Digitalization is not simply productivity improvements due to the addition of computers into the production function, nor the digitization of data records if the latter are not harnessed to improve firm performance. Indeed, digital firms must not only possess the raw big data capable of generating value and the technological infrastructure necessary to make this transformation, but they also need the human expertise necessary to exploit these assets.

I build my measure of digitalization in three steps. I begin by computing a digitalization score for almost 800 occupations based on O*NET survey data. I include attributes of a job based on three aspects: (i) interaction with computers, (ii) prevalence of data handling and analysis, and (iii) contribution to further digitalization. Then I aggregate these scores to the industry level by weighting each occupation in an industry by its economic share. There are approximately 300 industries per year. Finally, for each firm, I exploit the distribution of sales generated across the firm's various business segments to transform the industry-level digitalization measure to a firm-level one. This procedure yields a time-varying digitalization score for an unbalanced panel of 5,771 firms from 1999 to 2018. I validate this measure using passage of the salient European General Data Protection Regulation (GDPR). This regulation increases the costs for firms which handle user data and thus impacts two of the three pillars behind my measure of digitalization: computers and data. Performing an event study, I confirm that digital firms indeed experience relatively lower returns around the GDPR announcements: cumulative abnormal returns are 0.5% lower for firms in the top quintile of digitalization, compared to the bottom quintile.

Using this measure, I find that digital companies post better stock market performance in our increasingly digital world. A strategy which is long (short) digital (non-digital) firms earns 6.5% per annum beyond common risk factors over the sample period July 2000 to June 2019. This figure is comparable in magnitude to other intangible alphas documented in the literature. I investigate several non-digital explanations for these excess returns by performing double sorts. While digital firms are younger, smaller, less profitable, and more R&D intensive, these characteristics cannot individually explain digital firms' outperformance. I delve further into the predictive ability of digitalization for stock returns using cross-sectional Fama-MacBeth regressions. Despite the inclusion of well-known predictors, digitalization retains its predictive power.

After establishing the existence of a digital alpha, I posit that this alpha is linked to the importance of users to digital firms. The motivation is the ongoing economy-wide structural change. Dating back to the industry revolution, the most valuable stage in the production-consumption chain, as exemplified by the largest firms in the economy, has been continually moving from the raw materials stage towards the end stage of users. The digital revolution has facilitated the interaction between digital firms and their users. Users can now actively contribute towards the creation of the valuable digital companies in the economy. In fact, this active contribution by users is likely the root of digital firms' outperformance.

There are two ways that digital firms can benefit from the digital revolution. The first way is purely technical: digitalization brings productivity improvements. For example, firms can hire programmers to write code to streamline the production processes, automate previously manual tasks, or efficiently categorize and label customer feedback. The second way firms can take advantage of the digital revolution is by exploiting the new digital tools to create additional value. This way has the potential to be far more valuable. Beyond finding out what customers think about a firm, the firm could go one step further and identify gaps between customers' expectations and the delivered results, thus giving rise to entirely new revenue streams. Also, since users are more likely to interact directly with digital firms, digital firms can build a richer profile of its customer base. In both cases, though, the firm must consciously consider its users. In other words, the firm would contribute to the general structural change towards user-centricity.

I only consider user-centricity within the context of digitalization. I do not make claims about general customer-centricity for digital and non-digital firms. The reason is because I seek to understand the digital alpha. Indeed, customer-centricity will manifest as user-centricity for digital firms because by the very nature of digital firms, many – if not all – of their consumers will arrive through digital channels as digital users, instead of as customers through physical stores. For simplicity, I will refer to digital users as users henceforth. One striking consequence of users and their interactions on digital channels is that it creates network effects. Greater network effects have positive effects on future growth when the market is large (Rietveld and Ploog, 2021), which is the case as digital firms and markets become more popular due to the digital revolution. Thus I hypothesize that the increasing importance of users to digital firms drives the digital alpha and this alpha should be larger for user-focused digital firms.

I present evidence consistent with users being particularly important for digital firms in three steps. First, digital firms do consider users to be important: within mandatory disclosures, digital firms are more likely to emphasize users to their investors. Second, aggregate new user demand helps explain digital firms' revenue growth but not non-digital firms' revenue growth. Finally, as predicted by the magnifying effect of network effects, the annual digital alpha is much larger for user-focused firms at 9.0%, compared to a non-significant 2.5% for non-user-focused firms. Interestingly, this divergence by user focus occurs only after 2009. This timing coincidences with the explosion of the smartphone and apps markets, which epitomizes the trend towards user-centricity. Furthermore, a dissection of the digital alpha reveals that user-centricity complements and magnifies digitalization: among firms which are not user-centric, digitalization does not predict higher stock returns. In short, digital firms' stock market outperformance can be traced to the unexpected yet growing importance of users in business.

Is the digital alpha – especially in user-focused firms – due to a risk premium or mispricing? One possible motivation for a digitalization risk premium is the gradual disappearance of the value premium, as documented in Smith and Timmermann (2021). To the extent that the value factor fails to adequately capture firms' intangible assets – of which digital firms have more – there may exist a digital factor which corrects the Fama-French value factor towards the "true" value factor. Digitalization could also be a mispricing. Digitalization relies on technology that is relatively new in human history, which may cause errors in analysts' forecasts or cause investors to overlook some aspects of digital firms' performance.

I test a risk premium explanation for the user-focused digital alpha as follows. If digitalization is a risk premium, then it should be priced in the cross-section of stock returns, so ex ante expected returns should be unbiased estimates of ex post realized returns. This implies that analysts' forecasts of digital firms' earnings should be unbiased on average and investors should not be consistently surprised around quarterly earnings announcements (Barrot et al. (2019), Bretscher (2020)). A mispricing story predicts the opposite. In this case, an intangible factor creates superior accounting performance. Yet this intangible factor is overlooked until the announcements of these better-than-expected realizations, at which point these intangible-intensive firms accumulate excess returns (Edmans, 2011). To explain the 9.0% user-focused digital alpha, I focus on user-centric firms and find that digitalization does not predict different analyst forecast errors for these firms. Nor are user-focused digital firms more likely to accumulate excess returns around their earnings announcements, except for the largest digital firms. Moreover, I discover that the stock returns of user-focused digital firms covary more with one another than with non-digital firms, and the covariance is also higher among digital firms than among non-digital firms. Therefore, the 9.0% user-focused digital alpha is likely a risk premium.

This paper contributes to the discussion on how to measure the digital economy. IMF (2018) discusses the challenges of classifying digital sectors and Muro et al. (2017) propose a method to identify digitalization in the workforce. In response to a scarcity of data regarding the adoption of advanced technologies, in 2018, the U.S. Census Bureau introduced new survey questions in their Annual Business Survey to gauge the extent of firm adoption of artificial intelligence (AI), cloud computing, robotics, and the digitization of information (Zolas et al., 2021). I build upon Muro et al.'s approach of using occupations which interact with computers. I expand the encompassment of *digitalization* from computers and mechanical *digitization* to its applications and future progression: I add activities which describe how digitalized tools would aid firms seeking to harness the value of their data as it becomes increasingly important, as well as activities which actively promote further digitalization. Furthermore, I transform this occupation-level categorization into a firm-level metric.

There are a number of studies in the economics and information systems literature which explore consequences of digital innovations such as AI, robots, and big data analytics. In the recent literature pertaining to AI and robots in firms, the focus is usually on outcomes related to labor. For example, Acemoglu et al. (2020a) study how the prevalence of AI affects employment and leads to associated organizational changes in firms, while Dixon et al. (2021) and Acemoglu et al. (2020b) examine similar outcomes arising from the use of robots. Other labor-related outcomes include the returns to technological talent in AI (Rock, 2019) and changes in the knowledge production function due to new combinations of data and labor (Abis and Veldkamp, 2020). There are also papers which study data analytics and find a positive impact on productivity (e.g., Brynjolfsson and McElheran (2016) and Müller et al. (2018)) and innovation (Wu et al., 2020). Tambe et al. (2020) propose a method to measure digital capital in firms and document its evolution over time. In this paper, I adopt a holistic view of digital firms – encompassing technological infrastructure, big data, and programming as the labor input – and approach these firms from a financial markets perspective as I explore how these firms are priced by financial markets. Morever, this literature often studies people as inputs into the production function, whereas I focus on people as the end users and target audience of firms.

I also contribute to the emerging literature on data, digital economics, and financial markets. Begenau et al. (2018) examine big data in finance and find that as technology improves over time, data processing on large firms becomes more fruitful, which lowers their cost of capital and allows these firms to grow even larger. This size-based asymmetry is also found by Farboodi et al. (2020), who document that price informativeness improves for large growth stocks but not for smaller ones as data becomes more abundant. Instead of focusing on firm size as a differentiating characteristic, I quantify digitalization for firms and, based on this novel measure, find that it affects firms' cost of capital.

Finally, this paper contributes to the literature on intangibles and asset pricing. Edmans (2011) discovers that firms with high employee satisfaction subsequently outperform in the stock market as they produce superior accounting performance, so employee satisfaction is mispriced. Excess returns may also arise due to a risk premium, which has been found to be the case for pollution (Hsu et al., 2020), import competition (Barrot et al., 2019), and offshoring (Bretscher, 2020). Related to technology, there is a debate regarding whether and how R&D affects future excess returns (e.g., Chambers et al. (2002), Zhang (2002), Donelson and Resutek (2012), Gu (2016)). Kamssu et al. (2003) delve directly into how Internet dependency affects stock returns. However, faced with the increasing prevalence and importance of digital advances such as digitalization, big data, AI, and robotics, there is a gap in knowledge regarding how these attributes within firms are priced in financial markets and whether investors and analysts have correct estimates about these new types of firms. My paper aims to provide some answers regarding how digitalization in firms is evaluated by market participants. I find that the digital alpha is concentrated in user-focused firms and that for these firms, digitalization may be compensating for risk, such as greater systematic risk among more digital firms.

The remainder of this paper is structured as follows. Section 2 describes the construction of the digitalization measure and presents summary statistics and a validation test. Section 3 analyzes digitalization from an asset pricing perspective and demonstrates the existence of a digital alpha. I uncover the importance of users to digital firms in Section 4. As the digital alpha appears to be concentrated in user-focused firms, I test whether the user-focused digital alpha could be a risk premium or mispricing in Section 5. Finally, Section 6 concludes.

2 Methodology and data

2.1 Constructing the digitalization measure

As digitalization is reflected in work activities, for example, through an increased use of computers, digitalization in firms is based on digitalization in occupations. I proceed in three steps: (i) I begin by constructing an occupation-level measure of digitalization based on O*NET attributes and scores, (ii) since an industry is composed of different occupations, I calculate an industry's digitalization score as the weighted average of the scores of all of the occupations which contribute to the wage bill of that industry in a given year, and (iii) I transform this time-varying industry-level measure into a firm-level one through the proportion of sales derived from each industry segment in a firm.

2.1.1 Occupation level

First, I calculate occupation-specific digitalization scores based on the U.S. Department of Labor's O*NET program. This program collects information about the knowledge, skills, work characteristics, tools and technology, and education required for various jobs. Digitalization in the workforce and the generation and use of data is related to several occupation characteristics. Given that digital storage and manipulation requires computers, I include a work activity ("Interacting with Computers") and a knowledge variable ("Computers and Electronics") as in Muro et al. (2017). Beyond these two aspects, in order to be affected by digitalization, there must also be work characteristics which involve recording and using data. Thus I also include two additional work activities ("Analyzing Data or Information" and "Documenting/Recording Information") and another knowledge variable ("Clerical"). Finally, it is also important to capture how this occupation may contribute to further digitalization as a whole, thus I include a skill variable ("Programming"). These characteristics are summarized in Table 1.

[Insert Table 1 here]

O*NET surveys workers, occupation experts, and occupation analysts in each identified occupation along many dimensions, including the "importance" (I) and "level" (L) of each characteristic in an occupation. The importance of a characteristic refers to its importance and frequency of use, while the level of a characteristic denotes the level of expertise required for the occupation. O*NET standardizes this information so that I and L each vary from 0 to 100. For example, the work activity "Interacting with Computers" is equally important to travel agents and network systems analysts; the importance is 80 for both occupations. Yet the level of required expertise differs: travel agents require a level of 40, as assessed by O*NET, whereas the level for network systems analysts is 74. Following Blinder et al. (2009) and Bretscher (2020), I apply Cobb-Douglas weights of two-thirds and one-third to the importance and level, respectively, of each occupation characteristic and build an occupation-specific digitalization score as follows:

$$d_o = \frac{1}{C} \sum_{c=1}^{C} I_{o,c}^{\frac{2}{3}} \times L_{o,c}^{\frac{1}{3}} \tag{1}$$

where $I_{o,c}$ and $L_{o,c}$ are the importance and level, respectively, of occupation characteristic c in occupation o. As the Cobb-Douglas product of each characteristic is averaged over the set of characteristics summarized in Table 1, d_o varies from 0 to 100, with a higher number indicating a higher level of digitalization. I am able to calculate d_o for 795 occupations. Insofar as the impact of technological change on a given occupation is similar across the different industries to which this occupation contributes, the fact that d_o is static should not be a problem as I will transform it into a time-varying measure using changes in the relative economic importance of occupations over time.

2.1.2 Industry level

In the second step, I aggregate these occupation-specific scores to the industry level by weighting each occupation by its economic importance. The U.S. Bureau of Labor Statistics' Occupational Employment Statistics (OES) program tracks employment and wages up to the 3-digit SIC or 4-digit NAICS level. Using the National Crosswalk Service Center to bridge the O*NET and OES data, an industry i's digitalization score in year t is:

$$D_{i,t} = \sum_{o} d_o \times \frac{emp_{o,i,t} \times wage_{o,i,t}}{\sum_o emp_{o,i,t} \times wage_{o,i,t}} \tag{2}$$

where $emp_{o,i,t}$ and $wage_{o,i,t}$ are the employment and average annual wage, respectively, in occupation o for industry i in year t. The O*NET and OES samples cover 1999 to 2018, with an average of 300 industries per year.

2.1.3 Firm level

Finally, I build a firm-level measure of digitalization by weighting each industry in which a firm operates by the share of sales from that industry, following Kim and Kung (2017). A firm f's digitalization score in year t is:

$$D_{f,t} = \sum_{i} D_{i,t} \times \frac{|sales_{i,f,t}|}{\sum_i |sales_{i,f,t}|} \tag{3}$$

where $sales_{i,f,t}$ is the sales obtained in industry i for firm f in year t. I use the absolute value of sales because even negative sales signals a firm participation's in a particular industry i, hence that firm would be exposed to digitalization changes occurring in industry i.[2] The data for the industries in which a firm operates and the resulting sales are obtained from Compustat Business Segments. For single-industry firms or firms missing segments information, this digitalization score, $D_{f,t}$, is equal to the digitalization score of the firm's industry, $D_{i,t}$.

[2] Still, negative sales are rare, occurring in less than 0.02% of the sample.

2.2 Other data

I obtain monthly stock returns data from the Center for Research in Security Prices (CRSP). The sample includes firms with ordinary shares (CRSP share codes of 10 or 11) traded on the Nasdaq, NYSE, or NYSE American (formerly AMEX). Following Beaver et al. (2007) and Hou et al. (2020), I adjust monthly returns for delistings as follows. In the month of the delisting, I compound daily returns with the delisting return reported in CRSP. If the delisting return is missing, then I use the average delisting return across all delistings in the previous 60 months which have the same delisting type and are listed in the same stock exchange.

Using the CRSP/Compustat Merged database, I add on annual accounting information from the Compustat Fundamentals database. Firms with missing sales, total assets, or industry information, or have fewer than two years of data are dropped, and all accounting variables are winsorized at 1% and 99%. Firms are sorted into quintiles based on their digitalization status in the previous fiscal year. I begin collecting OES data from 1999 for two reasons. First, although the OES data begins in 1998, from 1998 to 1995, each industry is only surveyed once every three years (OES, 2013). There is no data for 1996. Second, there is a major occupation classification system change beginning in 1999. To avoid spurious findings arising from this classification system change, I opt to drop data from 1997 and 1998. In the later analysis involving users, I supplement the above dataset with 10-K data from Loughran and McDonald (2011). This dataset of parsed text files ends in 2018. Thus my sample covers July 2000 to June 2019, with 525,314 stock-month observations belonging to 5,771 distinct stocks.

2.3 Summary statistics

Table 2 lists the ten most and least digitalized occupations. The most digitalized occupations either involve direct work with computers (e.g., programmers and database administrators), or are data-intensive and supported by digitalization (e.g., information security analysts). On the other hand, the least digitalized occupations have little contact with data or computers. Instead, these occupations involve contact with non-computer mechanical machines (e.g., garbage collectors) or other humans (e.g., attendants).

[Insert Table 2 here]

This digitalization gap is mirrored in the industries built from occupations. Table 3 lists the top five and bottom five industries by digitalization in 1999 and 2017, respectively. Only industries with at least five firms are displayed; industries with fewer than five firms are not displayed but retained for all analyses. In both snapshots, industries which handle a large amount of data or use computers form the top five, whereas industries which provide close-contact physical services or use non-computer machinery form the bottom five. However, there is one important difference between the two snapshots. In 1999, the top five industries contain computer programming and also industries which handle a large amount of data records, such as engineering specialties. But by 2017, this field is dominated by industries related to computer services, software, and data hosting, which typify the digital revolution.

[Insert Table 3 here]

Digitalization is somewhat persistent at the firm level. Table 4 reports the transition probabilities for firms in a given quintile of digitalization. Persistence is strongest at either end of the spectrum: these firms have a 93% to 95% probability of remaining in their quintile after one year, compared to a 86% to 88% probability for firms in the middle quintiles. More movement occurs after five years. Nearly a quarter of firms in the highest and lowest quintiles change quintiles. This figure rises to upwards of 40% for firms in the middle quintiles. Thus digitalization is dynamic.

[Insert Table 4 here]

Figure 1 plots the distribution of firms' digitalization scores, pooled across the entire sample period. To illustrate how firms are scored according to my measure, I highlight the digitalization scores of some of the largest companies in 2018. Restaurants and manufacturing firms such as Starbucks and Nike are in the bottom quintile of digitalization. In the middle quintiles are pharmaceutical companies like Merck & Co and also companies which manufacture electronic components like Texas Instruments. Software industries, represented by Adobe, are in the top quintile.

[Insert Figure 1 here]

However, these trends do not necessarily mean that a firm's industry determines its digitalization status. Supplemented by business segments information, a firm's amalgamated digitalization score differs from that of its primary industry in over one third of the sample of 44,153 firm-year observations. This difference is significant in 16% of the entire sample: for these observations, the firm-level digitalization quintile is at least two quintiles away from its industry-level digitalization quintile. As an example, consider the case of Ralph Lauren Corporation. In 2016, its primary 4-digit NAICS industry was 4481 (Clothing Stores), from which it derived 53% of its revenue. However, this firm also collected 45% and 2% of its revenue from two other industries: 3152 (Cut and Sew Apparel Manufacturing) and 5331 (Lessors of Nonfinancial Intangible Assets), respectively. If only Ralph Lauren's primary industry from Compustat was recorded, then its digitalization score would be 42.8. This firm would – perhaps mistakenly – belong to the middle quintile of digitalization. Yet with the additional information from all of its business segments, especially the low digitalization score of the Cut and Sew Apparel Manufacturing industry, Ralph Lauren's digitalization score becomes 37.5. This score corresponds to the bottom quintile. Thus, my digitalization measure is informative of firm-level activities.

Table 5 compares the summary statistics of firms in the different quintiles of digitalization. Several trends emerge. Digital firms are smaller: their market capitalization is lower and the average firm in the highest quintile employs one third the number of people employed by the average firm in the lowest quintile. Digital firms are also younger. Consistent with digitalization being associated with newer technology which may be harder to capture when valuing firms, digital firms have lower book-to-market ratios and tangibility. Digital firms also invest more and are more R&D intensive, but not through increased leverage – instead, they have lower and negative return on assets (ROA) on average. Furthermore, digitalization is associated with being more financially constrained according to the Whited-Wu index (Whited and Wu, 2006) for financial constraints.

[Insert Table 5 here]

2.4 Validation using GDPR

Since computers and data form two of the three main pillars of my digitalization measure, a natural way to validate this measure is to examine firms by digitalization level around a significant computer- and data-related event. Hence I perform an event study of stock returns around the passage of the General Data Protection Regulation (GDPR). The GDPR is intended to protect European Union (EU) users' rights to their own data by requiring firms to receive explicit consent before storing user data, give prompt notification of data breaches, and allow users the ability to request a copy of the data collected on them – and in some cases, to delete this information. Violators can be fined up to 20 million Euros or 4% of their annual worldwide turnover (Albrecht, 2019). The GDPR effectively makes it more costly for firms to handle users' data. Importantly, due to the borderless nature of the Internet, this EU legislation was salient in the United States. While California passed a similar data privacy act, the California Consumer Privacy Act (CCPA), it attracted much less attention than the GDPR: the peak Google search interest for the CCPA was less than a quarter of that of the GDPR, even in California itself. Thus though the GDPR governs the use of EU residents' data, it serves as a negative shock to all companies which deal with users and their data.

The GDPR is a suitable event study for digital firms because it serves as a shock to measure how much firms value the data that they have. As securing data is costly, the easiest way to comply with the GDPR is to delete all collected user data. However, many firms' business models are clearly dependent on user data. Thus data has value to these firms. The facility of EU users to request deletion of their data compromises the ability of these firms to extract value from user data, so this regulation should affect digital firms negatively. Moreover, this data regulation also pertains to the computer pillar of my digitalization measure because the GDPR concerns the handling and storage of data, and thus presumes that any affected firms possesses sufficient computer infrastructure.

I use three event dates for the passage of the GDPR:

1. March 12, 2014: the European Parliament adopts the GDPR.

2. December 15, 2015: the European Parliament, Council, and Commission reach an agreement on the GDPR.

3. April 8, 2016: the European Council adopts the GDPR.

My variables of interest are based on stock returns as the market reaction to unexpected events has been shown to reflect firms' exposure to the underlying risk, such as during the Panama Papers leak (O'Donovan et al., 2019), U.S.-China trade war announcements (Huang et al., 2019), and former President Trump's surprise election win (Wagner et al., 2018). I examine whether the cumulative abnormal returns (CARs) of digital firms differ from those of non-digital firms in a narrow window around the events. Daily abnormal returns are calculated relative to the estimates of a market model estimated from 252 to 21 days prior to each event, with a minimum of 100 observations required for the estimation window.

The first validation test traces the difference in CARs based on a matched sample. Firms in the top digitalization quintile are matched to the closest firm in size (the logarithm of market capitalization) in the bottom quintile. Figure 2 plots how this difference in CARs evolves from three days prior to six days following the average GDPR event, along with 95% confidence intervals. The most digital firms have CARs which are up to 0.9% lower immediately following the GDPR announcements. In addition, this negative difference persists for several days afterwards. Thus the GDPR exerts a negative impact on digital firms.

[Insert Table 6 and Figure 2 here]

As an alternative validation check, I study digital firms' CARs around the GDPR events with control variables, namely, lagged size, profitability, leverage, and book-to-market. In Table 6, cumulative abnormal returns are calculated geometrically within a symmetric three- or five-day window around each event date. Table 6 reaffirms that digital firms indeed have worse stock performance around the GDPR events. A one-standard-deviation increase in the digitalization score of a firm causes average cumulative abnormal returns around the three GDPR events to be 0.16% to 0.19% lower. In other words, firms in the highest quintile of digitalization experience three-day cumulative abnormal returns which are 0.5% lower than firms in the lowest quintile of digitalization, even after controlling for firm characteristics. These results are reassuring as, by construction, digital firms are likely to handle lots of data which will become subject to the GDPR, so the relative negative stock market impact upon announcement reflects the increased cost or decreased value of data to these digital firms.

3 Digital alpha

3.1 Digitalization's excess returns

I begin exploring the asset pricing implications of digitalization by studying whether digitalization is already captured by common risk factors. I analyze the excess abnormal returns over the six-factor model (Fama and French, 2018) of portfolios formed by levels of digitalization. After sorting stocks into quintile portfolios based on their digitalization score in the previous year, I calculate the monthly returns for each portfolio using two weighting schemes. In one, stocks' returns are weighted by their market capitalization. I also use equal-weighting to check the impact of small firms on results. The excess abnormal returns of each quintile portfolio is the alpha of the six-factor model with standard errors adjusted for heteroscedasticity and autocorrelation using the Newey-West procedure with 12 lags. The six factors are market, size, value, profitability, investment, and momentum, and are obtained from Kenneth French's website. To test the difference between quintiles, I also explore the excess abnormal returns of a portfolio that is long (short) firms in the highest (lowest) digitalization quintile.

Digitalization produces abnormal excess returns. A long-short portfolio with equally weighted firms produces a significant annualized alpha of 10.3%. Importantly, the alpha resulting from the value-weighted long-short portfolio is still significant – at 6.5% annually. So small firms are not driving these results. These alpha values are also in line with previous studies in the literature on various intangible factors.[3]

[Insert Table 7 here]

Figure 3 plots the value-weighted alpha from July 2000 to June 2019. The series is volatile but steadily increasing over time. In addition, the trend is mostly unrelated to economic recessions. This steady increase suggests that digitalization is not a fad, but rather is part of a larger ongoing trend.

[Insert Figure 3 here]

3.2 Double sorting on other explanations

Table 5's summary statistics demonstrate that digital firms are smaller, younger, less profitable, and more R&D intensive. Do these characteristics explain digital firms' outperformance over the 19-year sample period? I use double sorts to rule out this explanation.

I form 5×3 independently double-sorted portfolios based on digitalization and size, age,

[3]As summarized in Edmans (2011), studies have found that alphas of up to 4-6% are possible in this literature. Long-short alphas should be halved for comparison. Recent work document a similar range in alpha, for example, long-short alphas of 4.4% for pollution (Hsu et al., 2020), 7.0% for import competition (Barrot et al., 2019), and 8.4% for offshoring (Bretscher, 2020) have been discovered.

profitability, or R&D intensity, and repeat the analysis of Section 3.1. Size is based on market capitalization while firm age is calculated from the first year in which a company appears in Compustat. Profitability is proxied by ROA, which is operating income after depreciation scaled by total assets. R&D intensity is defined as a firm's R&D expenses scaled by market value following Gu (2016). These four variables are sorted into three groups as follows. For size, profitability, and R&D intensity, I use terciles based on the value in the previous year. Thresholds for the current age are set at 10 and 20 years.

Table 8 presents the value-weighted annualized excess returns from double-sorted portfolios over the six-factor model. Just as digital firms are typically smaller, Table 8 shows that the long-short strategy yields higher excess returns in the lower size terciles. Firms in the lowest size tercile have an annual value-weighted long-short alpha of 11.8% compared to 6.2% for firms in the largest size tercile. Yet all alphas are significant regardless of the size tercile. Thus size cannot explain digital firms' outperformance. A similar story occurs when focusing on firm age, therefore the digital alpha is not subsumed by an age-related alpha. When sorting by ROA, the digital alpha in the most unprofitable tercile is larger than that in the most profitable tercile and this difference is weakly significant. Nonetheless, the long-short alpha is significant in all terciles and hence the profitability (or lack thereof) of digital firms is not the reason behind their excess returns. Finally, the digital alpha exists only in the bottom two terciles of R&D intensity. However, there is no significant difference across the three terciles. The results for the equally-weighted double sorted portfolios are similar, as seen in Table A1. Together, these results indicate that while digitalization's excess returns may be in part due to smaller, younger, less profitable, and more R&D intensive firms, none of these attributes encapsulate the peculiarities of digital firms nor can explain digital firms' long-term excess returns.

[Insert Table 8 here]

3.3 Fama-MacBeth regressions

Not only does the six-factor model fail to describe the stock returns of digital firms, but also, in this section, I show that digitalization's predictive ability on stock returns survives other well-known predictors. I run cross-sectional Fama-MacBeth regressions of monthly individual stock returns (annualized by multiplying by 12) on the continuous digitalization measure, $D_{f,t}$, with several controls: stocks' historical market beta, market capitalization, book-to-market, return on assets, investment rate, leverage, and financial constraints. To ensure that the independent variables are known in year t, the digitalization measure is from year $t-1$, the historical beta is based on the previous 60 months, market capitalization is from the previous month, and all other control variables are from the end of fiscal year $t-2$. Digitalization is a continuous measure which can vary from 0 to 100, with a higher score indicating a greater level of digitalization for a given industry. All control variables are winsorized at 1% and 99% to reduce the impact of outliers.

Table 9 shows that even after controlling for known predictors of stock returns, digitalization still predicts higher stock returns. For instance, a one standard deviation increase in a firm's digitalization score predicts a significant 1.9% increase in their annualized stock return. The difference between the average digitalization score in the top and bottom quintiles is approximately 2.8 standard deviations. So this figure corresponds to a return spread of 5.3%, or about half of the equal-weighted six-factor alpha of 10.3% in Table 7.[4] As the explanatory variables are standardized, their relative predictive power can be compared. Although the predictive ability of firm size and financial constraints exceed that of digitalization, the latter explains more of firms' stock returns than book-to-market, ROA, investment rate, and leverage.

[Insert Table 9 here]

Due to the special characteristics of digital firms, I further explore how profitability and

[4] As Fama-MacBeth regressions weight each observation equally, I compare this return spread to the equal-weighted digital alpha of 10.3% rather than the value-weighted digital alpha of 6.5%.

size affect the relationship between digitalization and stock returns. I find that digitalization is most predictive of stock returns in large firms, that is, firms with above-median market capitalization in the previous month. For these firms, a one-standard-deviation increase in digitalization is associated with a 2.7% increase in their annualized stock return. Digitalization does not predict stock returns in the subsample of small firms. This result may link to Begenau et al. (2018)'s finding that big data benefits large firms more than small firms because big firms have more historical information, which can be quickly processed as technology improves. It is possible that, with less information available, the difference between small digital firms and small non-digital firms is less apparent to financial market participants. Digitalization remains predictive of stock returns across both profitability subsamples. These results are after controlling for well-known predictors and adjusting the standard errors using the Newey-West procedure with 12 lags to allow for heteroscedasticity and autocorrelation. Thus, in general, digitalization has unique predictive power beyond known predictors.

4 Digitalization and users

The technological innovations of digitalization – such as the engineering marvel of smartphones – may be the most eye-catching, but it is crucial to remember the purpose behind these innovations. Ultimately, these new technologies are to serve people. Digital firms offer products and services that fulfill people's desires: to save time or money, for convenience, to create social connections, and more. This is the trend of servitization. Businesses are shifting from selling one-off physical goods to adding an inextricable service component. Thus digital firms must consider the human factor in their business operations: the end user.

In this section, I uncover the underlying driving force behind digital firms: users. First, from the firm's perspective, digital firms indeed place more emphasis on users. Second, users have a real effect on digital firms: new user demand has predictive power for the revenue of digital firms only. Finally, I demonstrate that the digital alpha is concentrated in user-focused firms.

4.1 Firms' consideration of users

The notion of a "user" in computer systems is a foundational concept that is on the same level as the "system" (Moran, 1981). In digital markets, users correspond to the target audience and consumers.[5] Firms may have both digital users and offline customers. Do digital firms differ from non-digital firms with regards to their focus on users?

I answer this question by studying firms' annual 10-K filings. The SEC requires the vast majority of public U.S. firms to file an annual Form 10-K. These 10-Ks are generally more detailed than annual reports as the SEC mandates disclosure on numerous topics, such as a company's business, potential risk factors, operating performance, financial statements, and governing board (SEC, 2011). Importantly, companies are required to discuss their business description at length in their 10-Ks. I exploit this requirement by analyzing firms' discussion of users in these filings. I expect that digital firms consider users to be more important and hence are more likely to discuss users within their 10-Ks.

My sample includes all 10-K, 10-KSB, and 10-K405 filings (collectively referred to as 10-K filings) from 1999 to 2018, as obtained from Loughran and McDonald (2011). I use two measures of firms' discussion of users: the number of times "user(s)" is mentioned in a filing, and the number of words in paragraphs which mention "user(s)". Many firms do not mention users at all, but some firms repeat this term hundreds of times. For example, Twitter mentions users over 370 times in every one of its 10-K filings. Thus I normalize the two measures of user discussion by taking the natural logarithm of one plus the respective measure.

Table 10 confirms that digital firms consider users to be more important as these firms

[5]For some firms, most notably firms which offer "free" goods such as Google or various social media companies, users and consumers are not identical. However, users are still paramount to these firms because without a healthy demographic of users, these firms would be unable to attract consumers (i.e., advertisers).

choose to dedicate more of their 10-K disclosure to discussion of users. I control for the total 10-K word count as lengthy reports may be statistically more likely to mention various words, including users. I also control for firm size and time-varying industry effects as larger firms or firms in specific industries may have norms regarding terminology in 10-K filings. Despite controlling for these factors, digitalization within firms is associated with more references to users in these filings to investors, as well as lengthier related discussion. Thus from the perspective of firms, digital firms place more emphasis on users.

[Insert Table 10 here]

Technological advancements during the digital revolution have also contributed to digital firms' increasing emphasis on users. One such milestone was cloud computing. Beginning in 2006, Amazon opened its Amazon Web Services (AWS) to outside developers. Suddenly digital firms no longer needed to make large upfront hardware investments in order to produce their digital offerings as they could flexibly rent such hardware from AWS (Ewens et al., 2018). This change likely allowed firms to worry less about various technical considerations and instead redirect their energy towards their digital offerings to users.

4.2 Real effect of user demand for digital firms

In this section, I show that digital firms do not discuss users because it is trendy to do so. Rather, new user demand materially affects the revenue growth path for digital firms. I proxy new user demand by Google search interest for apps. I focus on apps because mobile phone usage and apps have exploded in importance in people's lives. Figure A1 traces how the average U.S. user increased their daily mobile usage time from 0.3 hours in 2008 to 3.6 hours in 2018. Over the same time period, computer and laptop usage remained constant at approximately 2 hours. This increased time on mobile devices is not only spent communicating or playing games: mobile retail e-commerce sales increased fivefold from 2013 to 2018, to $207 billion USD (see Figure A2). Moreover, e-commerce through apps has become increasingly dominant. In the fourth quarter of 2017, mobile apps became the most popular e-commerce channel, exceeding both desktop and mobile web channels by commanding 44% of all e-commerce transactions (Keyes, 2018).

Google search interest is especially suited for capturing new user demand. If a person searches for "shoes", it is likely that they want to purchase shoes rather than research the history of shoes or look at pictures of shoes. Likewise, searching for "app" likely reflects the desire to download apps and make purchases from these apps. I survey general search interest in apps. This method circumvents the need to place assumptions on how users may word their queries for specific companies, for example, whether people would search for "facebook app" versus "fb app". Making simultaneous search interest queries would not work since the popularity of one term – either at the same point in time, or in a different point in time – may eclipse the other (or even itself) and cause an uninformative search interest of 0 or "< 1". Google search interest varies from 0 (not enough data) to 100 (peak popularity) and represents the relative popularity of a search term when compared to all searches made in a given region at a certain point in time. Thus Google search interest for "app" represents the size of the demand from new users who intend to download apps and are likely to contribute to mobile app revenues. Figure 5 tracks the search interest in "app" over time.

[Insert Figure 5 here]

I estimate the following relationship:

$$\begin{aligned} y_{f,q} = \alpha &+ \beta_1 \times Digital_{f,t-1} \times NewInterest_q \\ &+ \beta_2 \times Digital_{f,t-1} + \beta_3 \times NewInterest_q + \gamma \times Controls + \varepsilon_{f,q} \end{aligned} \tag{4}$$

where $y_{f,q}$ indicates whether firm f's quarterly revenue has increased from quarter $q-1$ to quarter q, $Digital_{f,t-1}$ indicates whether firm f is in the top digitalization quintile or has above-median digitalization in year $t-1$, and $NewInterest_q$ is the average U.S. Google search interest for "app" over the three months ending in quarter q. I include industry-year fixed effects as firms in booming industries may experience a rise in revenue regardless of user interest in apps. Standard errors are clustered at the monthly level since the Google search interest data is available at the monthly frequency from 2004 onwards. I choose to examine whether revenue has increased or not, instead of actual revenue growth because new users only contribute to part of firms' revenue. I do not hypothesize that offline consumers and existing users contribute zero revenue growth so that total revenue growth should track new user demand in digital firms. Instead, I test whether new user demand constitutes a sufficiently significant portion of revenue such that it impacts the direction of total revenue growth in digital firms. I expect $\beta_1 > 0$: digital firms target users and their app is one of the main channels through which users can connect with digital firms. Therefore, new user demand should translate into a significant impact on total revenue.

Table 11 shows that new user demand, as proxied by search interest for apps, has a differential effect on the direction of revenue growth for digital firms. For digital firms, an increase in the aggregate new user demand for apps predicts a higher likelihood of positive revenue growth in the same quarter. This finding holds true whether we compare firms in the top quintile of digitalization against all other firms, or above-median versus below-median firms by digitalization. Since Google search interest is not seasonally adjusted, I do not adjust the revenue figures nor make year-on-year comparisons. As a robustness check, because Figure 5 shows minimal search interest prior to 2008, I repeat the above analysis for the 2008 to 2019 subsample in columns (3) and (4) of Table 11. The results are unaffected by this sample restriction. Thus user demand, and in particular, *new* user demand, impacts digital firms more than non-digital firms. In other words, users have a real effect on digital firms.

[Insert Table 11 here]

4.3 User-focused digital alpha

After establishing the importance of users to digital firms, I turn to the asset pricing implications of digital firms' focus on users. First, I define user-focused firms to be firms which mention "user" at least five times in their 10-K filing in a given year. This choice is to minimize false positives as the term "user" in 10-Ks is not always informative at lower frequencies. Ford Motor is a representative example. In its 2017 filing, the company mentions "user" twice but the term is used to refer to readers of the 10-K document.[6] Home Depot also mentions "user" twice in its 2019 filing and at first glance, it appears promising as the company wishes to create and maintain an appealing user interface.[7] However, upon a closer read, these terms are used generically within the context of providing a better customer experience and guarding against mechanical disruption risks, without any further elaboration on users.

At higher frequencies, the term "user" is more likely to reflect a firm's consideration of user experience and the importance of users to the firm. For example, Electronic Arts mentions "user" five times in its 2018 filing. The company discusses: (i) how they are impacted by government regulations on user privacy, (ii) risks which may degrade user experience, and (iii) how their relationship with some of their customers are conducted through other companies' platforms and can thus be unilaterally affected by such platforms.[8] Therefore, I use a threshold of five to balance screening out false positives against being overly aggressive in rejecting user-focused firms. In untabulated tests, I confirm the following results are robust to different thresholds, from three to eight.

Repeating the analysis from Section 3.1, I find that the digital alpha is concentrated in user-focused firms in Table 12. I define user-focused (not user-focused) firms to be firms which mention "user" at least (less than) five times in the previous year's 10-K filing. The

[6] https://www.sec.gov/Archives/edgar/data/37996/000003799618000015/f1231201710-k.htm

[7] https://www.sec.gov/Archives/edgar/data/354950/000035495019000010/hd_10kx02032019.htm

[8] https://www.sec.gov/Archives/edgar/data/712515/000071251518000024/ea3312018-q410kdoc.htm

value-weighted digital alpha is 9.0% annually for the former group but only 2.5% for the latter group and not significant. Recall that the value-weighted digital alpha across all firms is 6.5%. I suggest the following explanation for why digital firms which also focus on users produce better risk-adjusted returns. Digital firms and users are characterized by network effects, a connection which is often prized as firms seek to grow. Network effects can be a double-edged sword, though, by magnifying outcomes depending on the size of the market (Rietveld and Ploog, 2021). Given the expanding size of digital markets and the positive digital alpha, the magnifying effect of users' network effects contributes to a positive feedback loop, thereby enlarging the digital alpha when focusing on user-focused firms.

[Insert Table 12 and Figure 4 here]

This difference in the digital alpha by user focus becomes even more striking when plotted. First, I plot the long and short legs of the digital alpha separately. Panel A of Figure 4 graphs the monthly evolution of the H and L portfolios of Table 12. Regardless of whether firms are user-focused, digital firms outperform non-digital firms in terms of excess returns. However, the most dramatic result is the outperformance of user-focused digital firms. A $1 investment in this portfolio yields almost $6 after 19 years, compared to $1 to $2 for the other three portfolios. In other words, user-centricity complements digitalization; user-centricity by itself does not appear to predict stock returns.

Panel B of Figure 4 plots the digital alpha by user-focus, that is, the H-L portfolios of Table 12. As in the case of the general digital alpha, both series are unrelated to economic recessions. The two series increase and follow one another closely until around 2009, but afterwards only the user-focused digital alpha continues to rise – and steadily. The timing coincides with a noteworthy manifestation of user-centricity, namely, the beginning of the smartphone and apps era. Both the Apple App Store and Google Play Store launched in 2008 (Goldsmith, 2014). Smartphone apps embody the user-centric trend in business; there is no physical product between the firm and the user, and the sole purpose of these apps is for users to interact with or through the firm. There also existed a pervasive industry-wide underestimation of the appetite for smartphones throughout the early 2010s – even when forecasting within the same year (IDC, 2017). Hence the potential growth of users, especially when factoring in network effects, was also arguably underestimated. The key message here is that digital firms which focus on users differ noticeably from digital firms which do not focus on users – at least in terms of stock returns.

Taken together, these results on users show that users are uncommonly important for digital firms. Combined with accompanying phenomena such as network effects and the underestimation of the spread of smartphones, digital firms' stock market outperformance appears to stem from those firms which focus on users.

5 Explaining the digital alpha

The previous sections demonstrate the existence of a digital alpha that is distinct from several other intangible factors or predictors of stock returns. Moreover, the digital alpha is greatest in user-focused firms. In this section, I explore whether the digital alpha, by user-centricity, is a risk premium or a mispricing. The evidence points towards the user-focused digital alpha likely compensating for risk.

5.1 Analyst forecast errors

A risk premium story and a mispricing story yield different predictions for digital firms' analyst forecast errors. If the digital alpha compensates for risk, then digitalization should be correctly priced in the cross-section of stock returns. In particular, there should be no wedge between ex ante expected returns and ex post realized returns. Using analysts' earnings forecasts as one proxy for ex ante expected returns as in Barrot et al. (2019) and Bretscher (2020), the risk premium story predicts that digital firms should not experience significant differential earnings surprises. In contrast, in a mispricing story, market participants are oblivious to these firms' superior accounting performance. For example, Edmans (2011) discovers that higher employee satisfaction in firms is ignored by the stock market until these firms produce superior tangible outcomes such as earnings. Thus the mispricing story predicts that digital firms should produce relative positive earnings surprises.

Following Barrot et al. (2019) and Bretscher (2020), I estimate:

$$FE_{f,t} = \alpha + \beta \times D_{f,t-1} + \gamma \times Controls + \varepsilon_{f,t} \tag{5}$$

where $FE_{f,t}$ is the forecast error in earnings per share (EPS) for firm f in year t, $D_{f,t-1}$ is the digitalization score of firm f in year $t-1$, and control variables include stocks' lagged historical market beta, market capitalization, book-to-market, leverage, and investment rate, and year fixed effects. I study forecast errors at the one- and two-year horizon and the long-term growth (LTG) error. At the one- and two-year horizon, the forecast error is defined as the I/B/E/S actual annual earnings per share minus the median I/B/E/S consensus forecast of annual earnings per share. This figure is normalized by lagged stock prices to control for heteroscedasticity. The consensus forecast is measured 8 (20) months prior to the end of the forecast period at the one (two) year horizon. This threshold ensures that analysts are aware of prior earnings when they make their forecasts. The LTG forecast error is the average annualized EPS growth over the preceding five years minus the median LTG forecast for EPS from 56 months prior. This figure is already in percentage terms so it is not normalized. All forecast errors are winsorized at 1% and 99%.

Digitalization appears to be correctly priced in the cross-section of stock returns for user-focused firms. In Panel A of Table 13, consistent with the findings in the literature, I find that larger firms and less leveraged firms tend to produce more positive earnings surprises. Nonetheless, regardless of whether these factors are accounted for, there is no significant difference in the forecast errors at the one-year horizon and LTG by digitalization level for user-focused firms. Although digital firms are more likely to experience positive earnings surprises at the two-year horizon, this result is only significant at the 10% level. These null results for user-focused firms stand in contrast to Panel B, which show that for non-user-focused firms, digitalization predicts positive earnings surprises in the short run. Recall, however, that the non-user-focused digital alpha of 2.5% is not significant over the sample period.

[Insert Table 13 here]

Due to the declining explanatory power of earnings for stocks returns in the 21st century (Srivastava, 2014), I also examine the forecast errors for firms' sales in Tables A2. The results are similar for user-focused firms: digital firms' sales are not more or less likely to exceed analysts' sales forecasts at any horizon. In short, analysts generally understand the effect of digitalization on firms' earnings and sales in user-focused firms. So for these firms, digitalization appears to be a source of risk that is priced in the cross-section of stock returns.

5.2 Earnings announcement returns

To determine whether there exists a wedge between ex ante expected returns and ex post realized returns, beyond analysts, it is imperative to check investors' expectations. A risk premium story predicts that investors are equally unsurprised by user-focused digital firms' earnings, resulting in no differential earnings announcement returns, as in Barrot et al. (2019) and Bretscher (2020). Indeed, my results support this prediction.

As my digitalization measure is continuous, I compare the excess stock returns around quarterly earnings events by quintiles of digitalization and user-centricity. I obtain all quarterly announcement dates over my sample period from the CRSP/Compustat Merged database. For each stock, I calculate the daily excess return as the raw return minus the risk-free rate. These daily excess returns are cumulated over two windows: the (-5,1) window ((-10,1) window) is the 6-day (11-day) window beginning five days (10 days) prior to the quarterly earnings announcement day and ending the day after the announcement day. I aggregate all cumulative excess returns within a calendar quarter according to digitalization quintiles. This aggregation procedure is either value-weighted according to the stock market capitalization at the end of the previous calendar quarter, or equal-weighted.

When using equal-weighting, user-focused firms which are more digital do not experience higher earnings announcement returns (EAR) in Panel A of Table 14, consistent with a risk premium story. However, the difference becomes significant when using value-weighting. Large digital firms which are also user-focused tend to produce greater EAR – the difference amounts to an annualized 3.1% over the (-5,1) window around quarterly earnings. It is possible that these large firms release information other than their earnings and sales performance during their earnings events. In Panel B, within non-user-focused firms, digitalization does not predict differential EAR regardless of the weighting scheme. These results are quantitatively similar when using a (-10,1) window in Table A4.

[Insert Table 14 here]

5.3 Evidence from a digital factor

Although the lack of differential analyst forecast errors indicates that digitalization is a risk premium in user-focused firms, the evidence from quarterly earnings announcement returns is less conclusive. In this section, I explore another test and provide additional evidence that supports a risk-based explanation of the 9.0% user-focused digital alpha.

Do the stock returns of digital firms covary more with other digital firms than with non-digital firms? If so, then digital firms' higher returns may reflect unobservable state variables that produce undiversifiable risks which are not captured by existing factors. This line of reasoning is put forth by Fama and French (1993, 2004) regarding their size and value factors.

To test this hypothesis, I construct a digital minus non-digital (DMN) factor. This construction mirrors the construction of the value factor in Fama and French (1993). In each year, I form six size/digital portfolios. Size is split along the median NYSE size into two groups: *Small* and *Big*. The digital breakpoints are at the 30th and 70th percentiles for NYSE stocks. The bottom (top) 30% of firms by digitalization are considered *Digital* (*Non-digital*). Within each portfolio, returns are value-weighted. Finally, the DMN factor is calculated as:

$$DMN = \frac{1}{2}(\text{Small Digital} + \text{Big Digital}) - \frac{1}{2}(\text{Small Non-digital} + \text{Big Non-digital}) \tag{6}$$

If digital firms' stock returns covary more with each other, then in a factor model, $\beta_{DMN} > 0$ for digital firms. Of course, almost by construction, $\beta_{DMN} < 0$ for non-digital firms. However, if the covariance is higher among digital firms than among non-digital firms, then $|\beta_{DMN}|$ will be higher for digital firms.

Figure 6 plots the median β_{DMN} by digitalization quintile as estimated from the following equation:

$$\begin{aligned} r_{f,t} = \alpha_f &+ \beta_{DMN} DMN_t + \beta_{MKT} MKT_t + \beta_{SMB} SMB_t + \beta_{HML} HML_t \\ &+ \beta_{RMW} RMW_t + \beta_{CMA} CMA_t + \beta_{MOM} MOM_t + \varepsilon_{f,t} \end{aligned} \tag{7}$$

where $r_{f,t}$ is the monthly stock return for firm f over the risk-free rate in month t. This equation is estimated for each stock-year pair. As per the predictions, Figure 6 illustrates how β_{DMN} increases monotonically across the digitalization quintiles, from negative to positive. This pattern is true regardless of whether firms are focused on users. Importantly, $|\beta_{DMN}|$ is highest in the most digital quintiles. In user-focused firms, β_{DMN} ranges from -0.12 in the bottom digitalization quintile to 0.64 in the top quintile. These results are suggestive of the digital alpha – especially within user-focused firms – compensating for some sort of risk, for example, greater systematic risk in digital firms.

[Insert Figure 6 here]

6 Conclusion

This paper studies digitalization from an asset pricing perspective. Digital firms – exemplified by software and various computer services firms which focus on users – earn annual excess returns of 6.5% beyond common risk factors over 2000 to 2019. Yet digital firms are not a homogeneous group. In fact, separating firms by whether they focus on users produces a stark difference: whereas the non-user-focused digital alpha wanes after 2009, the user-focused digital alpha steadily increases over the past two decades and produces annual excess returns of 9.0%. Based on the evidence from analyst forecasts, earnings announcement returns, and a digital factor, this user-focused digital alpha likely represents a risk premium.

The importance of users is no coincidence. The digital revolution catalyzed a structural change in the relationship between a firm and its customers. But smartphones and apps are only one particular manifestation of the larger ongoing structural change. Since the industrial revolution, the most valuable stage along the production-consumption chain has gradually shifted towards the end. Recall that firms dealing with raw materials were the most valuable companies in 1917. However, the economic importance of these firms were overtaken by manufacturing firms by 1967. Some of these manufacturing firms produced intermediate goods for other firms. Other manufacturing firms delivered physical goods to customers, such as cameras or cars. Although consumers were undoubtedly present in firms' plans, this era was very much product-centric. In contrast, the current era can be characterized by user-centric firms. These digital firms are not creating a product and relying on a marketing team to garner sales. Instead, digital firms are building digital offerings that target users' desires and are reliant on positive user feedback to grow. Users, no longer merely awaiting the final product, are actively participating in the product creation process and shaping digital companies – sometimes into the most valuable companies in the world. The bulk of the value added share has fully shifted to the location of the end user along the production-consumption chain.

References

S. Abis and L. Veldkamp. The changing economics of knowledge production. *Available at SSRN 3570130*, 2020.

D. Acemoglu, D. Autor, J. Hazell, and P. Restrepo. AI and jobs: Evidence from online vacancies. Technical report, National Bureau of Economic Research, 2020a.

D. Acemoglu, C. Lelarge, and P. Restrepo. Competing with robots: Firm-level evidence from France. In *AEA Papers and Proceedings*, volume 110, pages 383–88, 2020b.

J. P. Albrecht. General data protection regulation, Nov 2019. URL `https://www.europarl.europa.eu/legislative-train/theme-area-of-justice-and-fundamental-rights/file-general-data-protection-regulation`.

J.-N. Barrot, E. Loualiche, and J. Sauvagnat. The globalization risk premium. *The Journal of Finance*, 74(5):2391–2439, 2019.

W. Beaver, M. McNichols, and R. Price. Delisting returns and their effect on accounting-based market anomalies. *Journal of Accounting and Economics*, 43(2-3):341–368, 2007.

J. Begenau, M. Farboodi, and L. Veldkamp. Big data in finance and the growth of large firms. *Journal of Monetary Economics*, 97:71–87, 2018.

A. S. Blinder et al. How many US jobs might be offshorable? *World Economics*, 10(2):41, 2009.

L. Bretscher. From local to global: Offshoring and asset prices. 2020. SSRN working paper.

E. Brynjolfsson and K. McElheran. The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5):133–39, 2016.

D. Chambers, R. Jennings, and R. B. Thompson. Excess returns to R&D-intensive firms. *Review of Accounting Studies*, 7(2):133–158, 2002.

J. Dixon, B. Hong, and L. Wu. The robot revolution: Managerial and employment consequences for firms. *Management Science*, 2021.

D. C. Donelson and R. J. Resutek. The effect of R&D on future returns and earnings forecasts. *Review of Accounting Studies*, 17(4):848–876, 2012.

A. Edmans. Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, 101(3):621–640, 2011.

M. Ewens, R. Nanda, and M. Rhodes-Kropf. Cost of experimentation and the evolution of venture capital. *Journal of Financial Economics*, 128(3):422–442, 2018.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56, 1993.

E. F. Fama and K. R. French. The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives*, 18(3):25–46, 2004.

E. F. Fama and K. R. French. Choosing factors. *Journal of Financial Economics*, 128(2): 234–252, 2018.

M. Farboodi, A. Matray, L. Veldkamp, and V. Venkateswaran. Where has all the data gone? Technical report, National Bureau of Economic Research, 2020.

B. Goldsmith. The smartphone app economy and app ecosystems. In *The Routledge companion to mobile media*, pages 195–204. Routledge, 2014.

Google Trends. Explore: app. `https://trends.google.com/trends/explore?date=all&geo=US&q=app`, 2021.

L. Gu. Product market competition, R&D investment, and stock returns. *Journal of Financial Economics*, 119(2):441–455, 2016.

K. Hou, C. Xue, and L. Zhang. Replicating anomalies. *The Review of Financial Studies*, 33 (5):2019–2133, 2020.

P.-H. Hsu, K. Li, and C.-Y. Tsou. The pollution premium. *Available at SSRN 3578215*, 2020.

Y. Huang, C. Lin, S. Liu, and H. Tang. Trade networks and firm value: Evidence from the US-China trade war. 2019.

IDC. Worldwide quarterly mobile phone tracker. `https://www.idc.com/tracker/showproductinfo.jsp?containerId=IDC_P8397`, 2017.

IMF. Measuring the digital economy. Technical report, IMF Staff Report, 2018.

A. J. Kamssu, B. J. Reithel, and J. L. Ziegelmayer. Information technology and financial performance: The impact of being an Internet-dependent firm on stock returns. *Information Systems Frontiers*, 5(3):279–288, 2003.

J. Kauflin. America's top 50 companies 1917–2017, 2019. URL `https://www.forbes.com/sites/jeffkauflin/2017/09/19/americas-top-50-companies-1917-2017`. Forbes Magazine.

D. Keyes. Mobile apps were the most popular e-commerce channel in q4 2017, 2018. URL `https://www.businessinsider.com/mobile-apps-most-popular-e-commerce-channel-q4-2017-2018-2?IR=T`. Business Insider.

H. Kim and H. Kung. The asset redeployability channel: How uncertainty affects corporate investment. *The Review of Financial Studies*, 30(1):245–280, 2017.

G. Lay. *Servitization in industry*. Springer, 2014.

T. Loughran and B. McDonald. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1):35–65, 2011.

L. Mastrogiacomo, F. Barravecchia, and F. Franceschini. A worldwide survey on manufacturing servitization. *The International Journal of Advanced Manufacturing Technology*, 103(9):3927–3942, 2019.

M. Meeker and L. Wu. Internet trends 2019, 2019.

T. P. Moran. The command language grammar: A representation for the user interface of interactive computer systems. *International journal of man-machine studies*, 15(1):3–50, 1981.

O. Müller, M. Fay, and J. Vom Brocke. The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems*, 35(2):488–509, 2018.

M. Muro, S. Liu, J. Whiton, and S. Kulkarni. Digitalization and the American workforce. 2017. Brookings India.

J. O'Donovan, H. F. Wagner, and S. Zeume. The value of offshore secrets: Evidence from the Panama papers. *The Review of Financial Studies*, 32(11):4117–4155, 2019.

OES. OES estimates, 1988–1995. `https://www.bls.gov/oes/estimates_88_95.htm`, 2013. Occupational Employment and Wage Statistics.

Raul. A century of america's top 10 companies, in one chart, 2014. URL `https://howmuch.net/articles/100-years-of-americas-top-10-companies`. HowMuch.net.

J. Rietveld and J. N. Ploog. On top of the game? the double-edged sword of incorporating social features into freemium products. *Available at SSRN 3381541*, 2021.

D. Rock. Engineering value: The returns to technological talent and investments in artificial intelligence. *Available at SSRN 3427412*, 2019.

SEC. How to read a 10-K. `https://www.sec.gov/fast-answers/answersreada10khtm.html`, Jul 2011. U.S. Securities and Exchange Commission.

S. Smith and A. Timmermann. Have risk premia vanished? 2021.

A. Srivastava. Why have measures of earnings quality changed over time? *Journal of Accounting and Economics*, 57(2-3):196–217, 2014.

Statista. Mobile retail e-commerce sales in the United States from 2013 to 2020 (in billion U.S. dollars) [graph]. `https://www.statista.com/statistics/249855/mobile-retail-commerce-revenue-in-the-united-states/`, 2021. Source data is from eMarketer.

P. Tambe, L. Hitt, D. Rock, and E. Brynjolfsson. Digital capital and superstar firms. Technical report, National Bureau of Economic Research, 2020.

US Census Bureau. E-stats tables, 2021. URL https://www.census.gov/programs-surveys/e-stats/data/tables.html.

S. Vandermerwe and J. Rada. Servitization of business: adding value by adding services. *European Management Journal*, 6(4):314–324, 1988.

A. F. Wagner, R. J. Zeckhauser, and A. Ziegler. Company stock price reactions to the 2016 election shock: Trump, taxes, and trade. *Journal of Financial Economics*, 130(2):428–451, 2018.

T. M. Whited and G. Wu. Financial constraints risk. *The Review of Financial Studies*, 19 (2):531–559, 2006.

L. Wu, L. Hitt, and B. Lou. Data analytics, innovation, and firm productivity. *Management Science*, 66(5):2017–2039, 2020.

X.-J. Zhang. Discussion of "Excess returns to R&D-intensive firms". *Review of Accounting Studies*, 7(2-3):159, 2002.

N. Zolas, Z. Kroff, E. Brynjolfsson, K. McElheran, D. N. Beede, C. Buffington, N. Goldschlag, L. Foster, and E. Dinlersoz. Advanced technologies adoption and use by US firms: Evidence from the annual business survey. Technical report, National Bureau of Economic Research, 2021.

7 Figures and Tables

Figure 1: Distribution of firms' digitalization scores over the sample period, July 2000 to June 2019. This graph rounds scores to the nearest integer for display purposes. Examples of the largest firms in 2018 and their scores are marked.

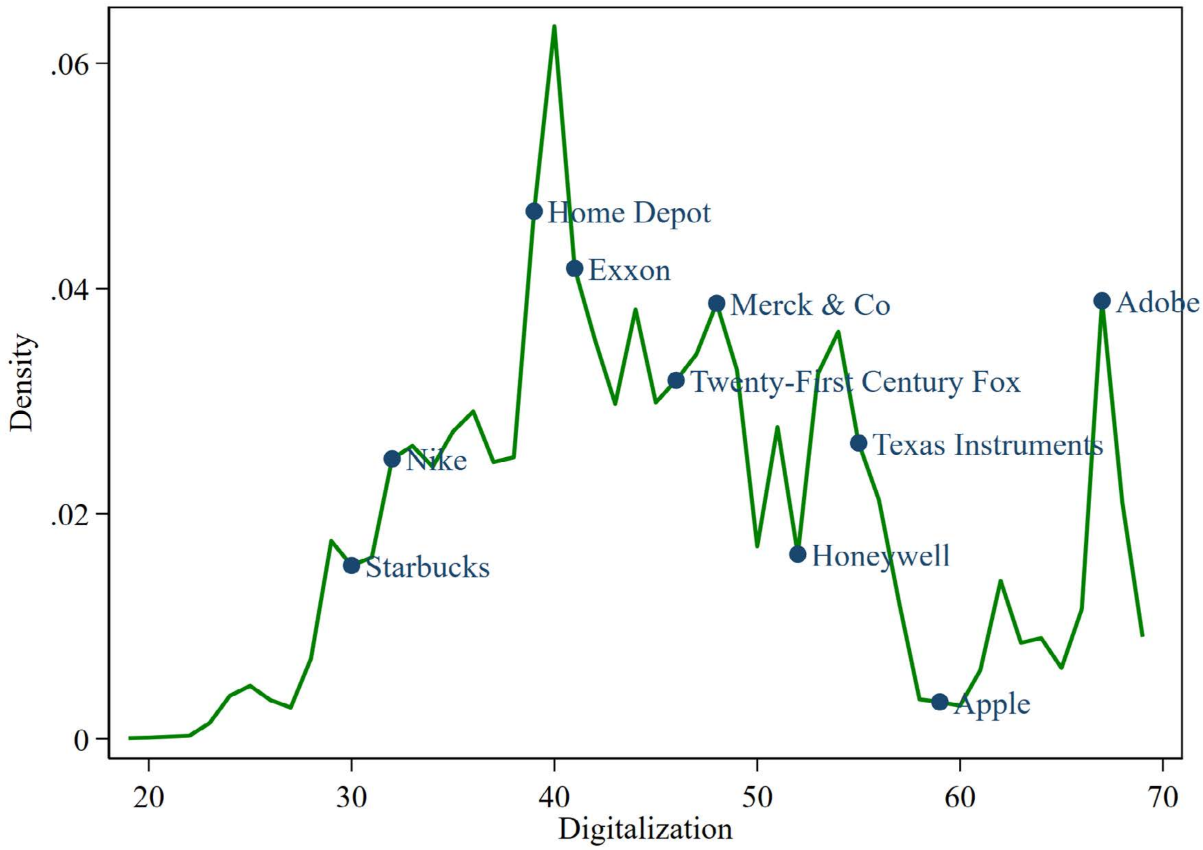


Figure 2: Evolution of the difference in cumulative abnormal returns in % around the GDPR events between firms in the top digitalization quintile and the closest matched firm in the bottom digitalization quintile based on the logarithm of market capitalization. The graph plots the average difference along with 95% confidence intervals. The three GDPR events are: (i) March 12, 2014 (the European Parliament adopts the GDPR), (ii) December 15, 2015 (the European Parliament, Council, and Commission reach an agreement on the GDPR), and (iii) April 8, 2016 (the Council adopts the GDPR). Daily abnormal returns are calculated relative to the estimates of a market model estimated from 252 to 21 days prior to each event, with a minimum of 100 observations required for the estimation window.

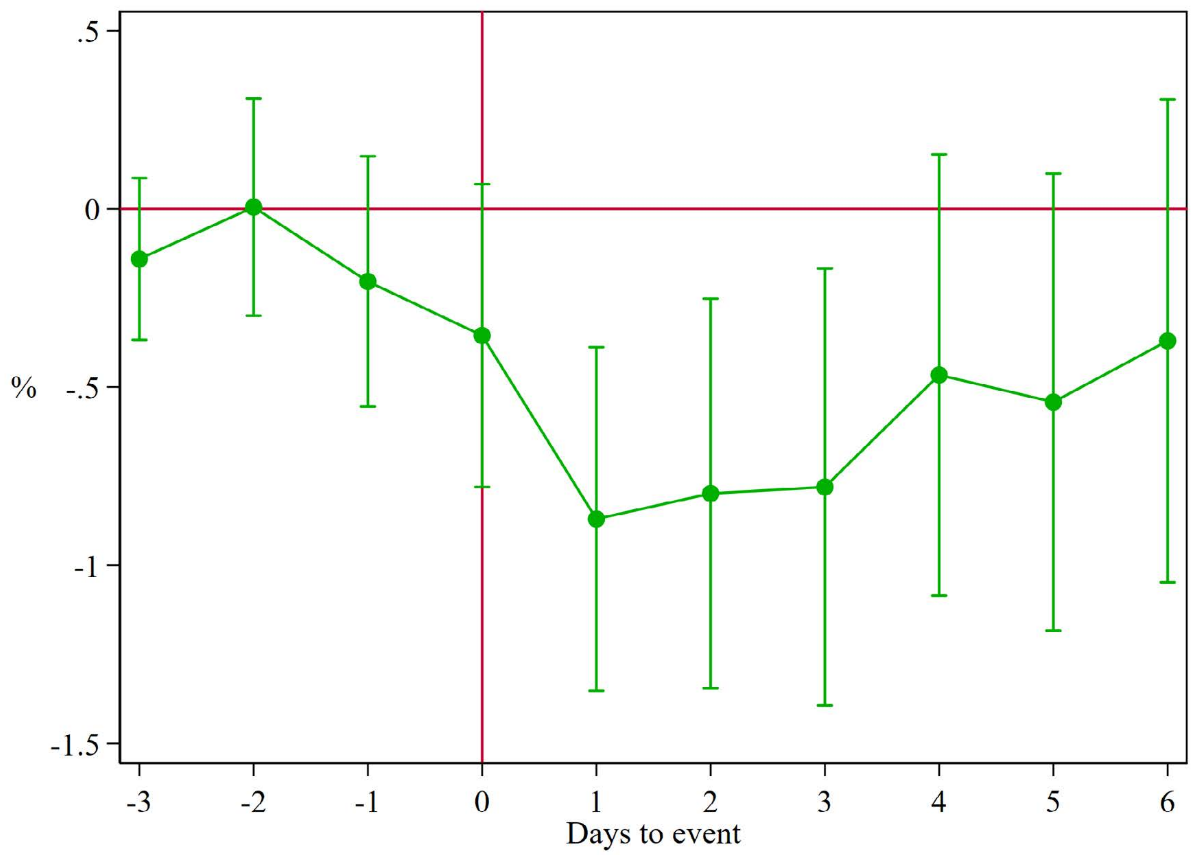


Figure 3: Cumulative excess abnormal returns of the H-L portfolio in Panel A of Table 7. This figure plots the monthly evolution over time of $1 invested in the value-weighted H-L portfolio after adjusting for the six-factor model. Shaded areas indicate recession periods according to the St. Louis Fed. The sample period is July 2000 to June 2019.

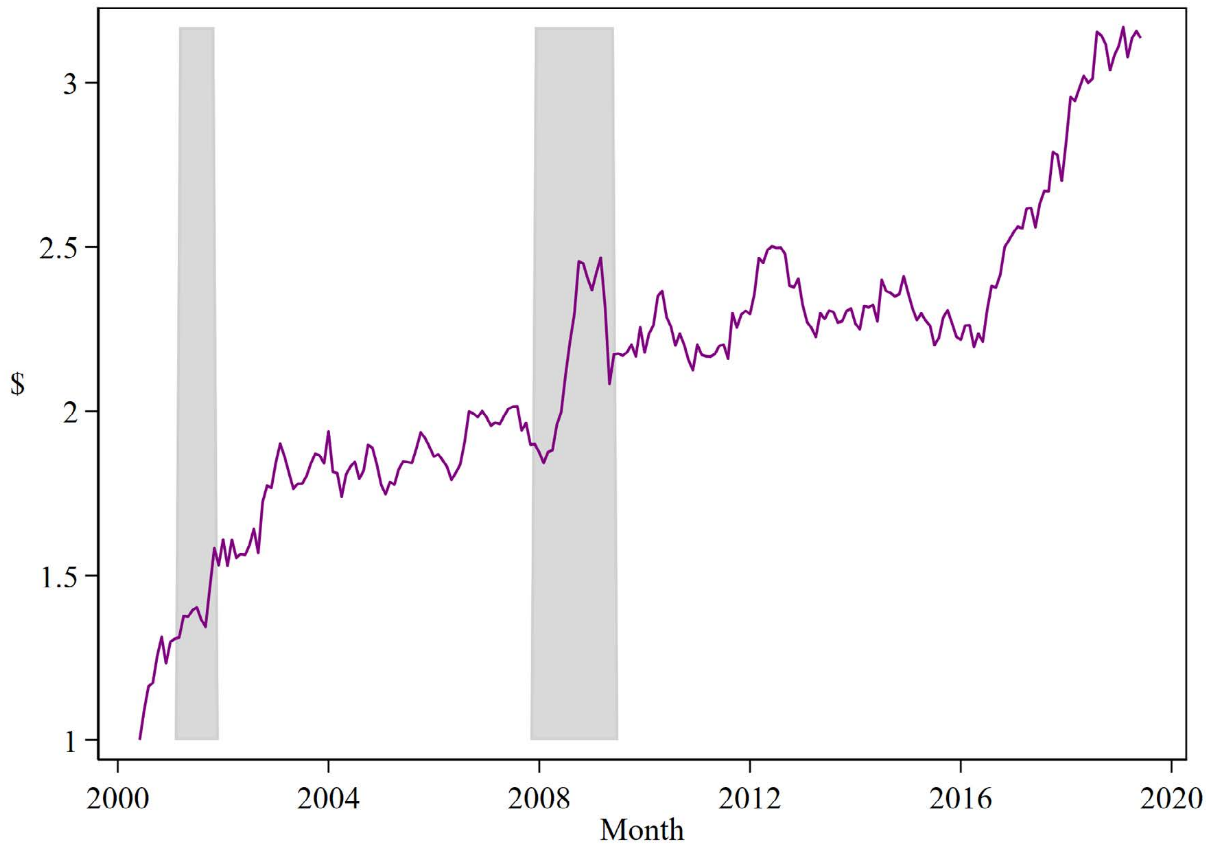
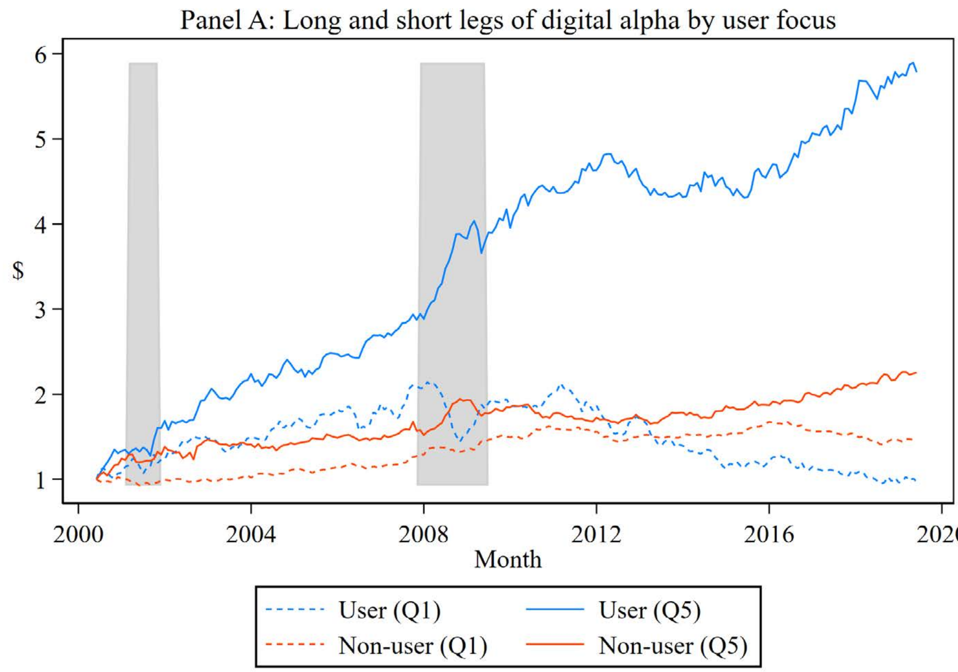


Figure 4: Cumulative excess abnormal returns of the H and L portfolios of Table 12 by whether firms are user-focused. This figure plots the monthly evolution over time of $1 invested in different value-weighted strategies after adjusting for the six-factor model. Panel A plots the long and short legs separately while Panel B plots the H-L portfolio which is long firms in the top digitalization quintile (Q5) and short firms in the bottom quintile (Q1). User-focused (non-user-focused) firms are defined to be firms which mention "user" at least (less than) five times in the preceding year's 10-K filing. Shaded areas indicate recession periods according to the St. Louis Fed. The sample period is July 2000 to June 2019.



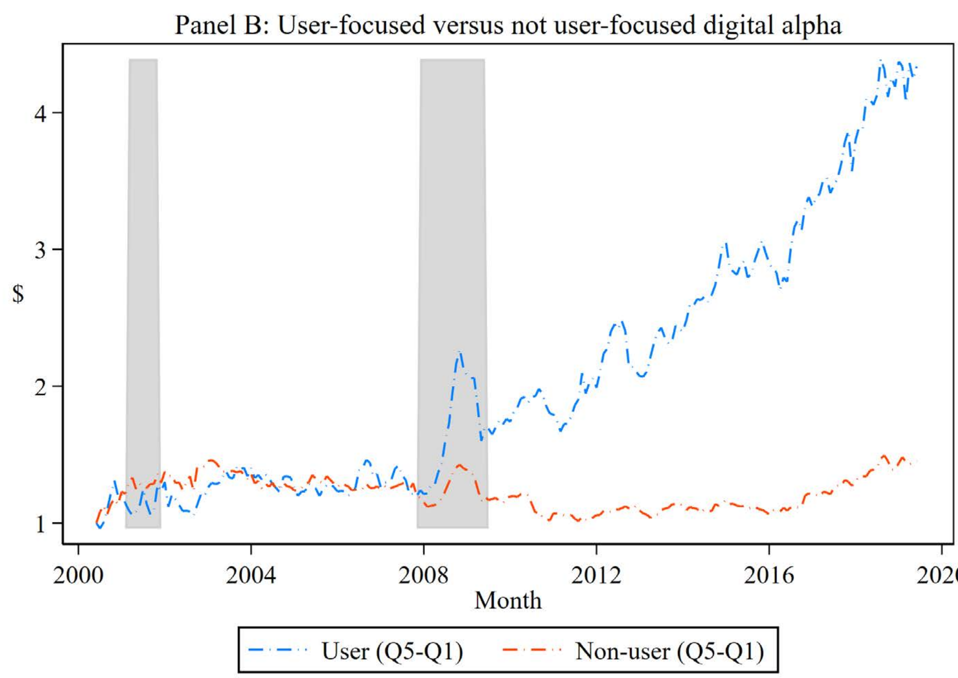


Figure 5: Google search interest for “app” in the United States from January 2004 to June 2019. Data is obtained from Google Trends (2021).

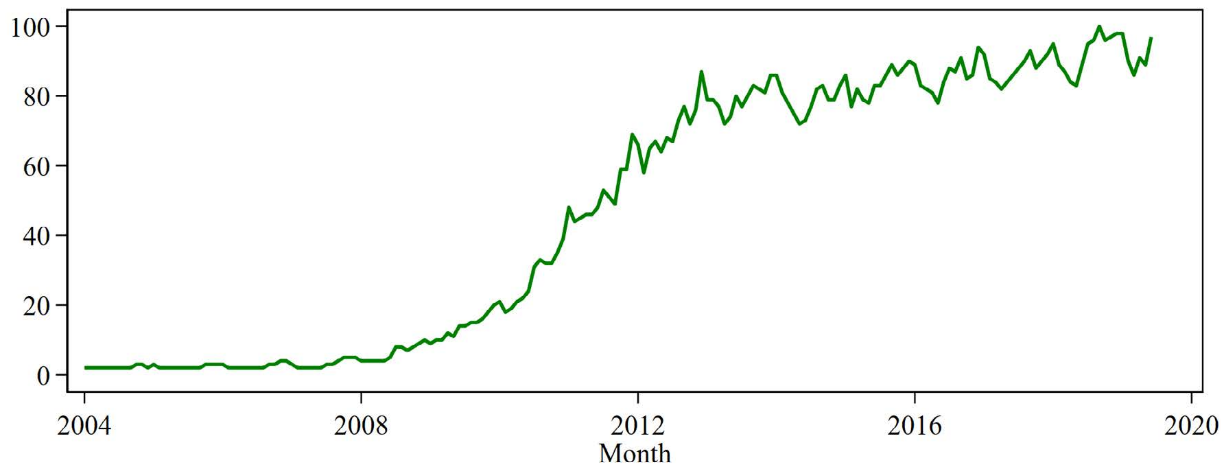


Figure 6: Betas of a DMN (digital minus non-digital) factor beyond the six-factor model by digitalization quintile and user-focus. In a given year, the following equation is estimated for firm f's monthly excess stock returns: $r_{f,t} = \alpha_f + \beta_{DMN} DMN_t + \beta_{MKT} MKT_t + \beta_{SMB} SMB_t + \beta_{HML} HML_t + \beta_{RMW} RMW_t + \beta_{CMA} CMA_t + \beta_{MOM} MOM_t + \varepsilon_{f,t}$. This graph plots the median beta within each group, pooled across the sample period, July 2000 to June 2019.

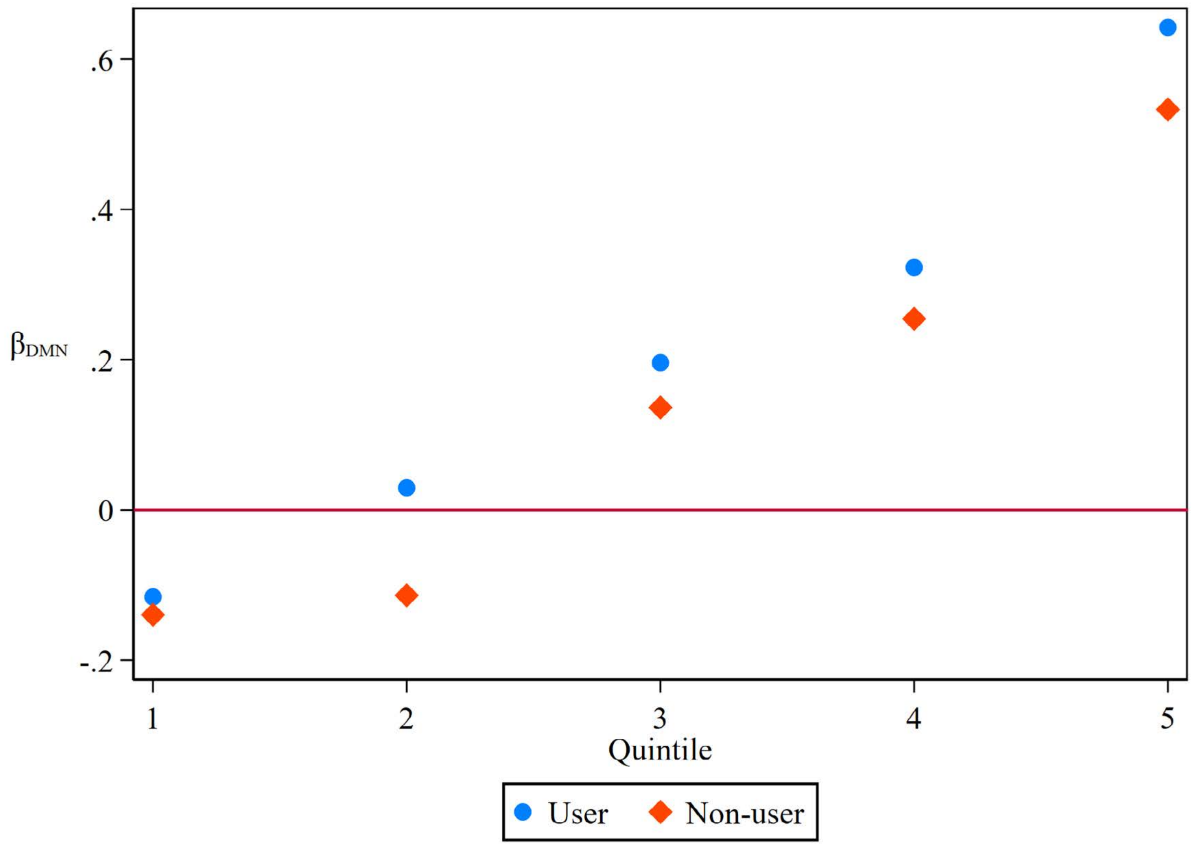


Table 1: Occupation characteristics used to calculate the digitalization level of occupations.

O*NET identifier	Occupation characteristics	Type
Computers		
4.A.3.b.1	Interacting with Computers	Work Activities
2.C.3.a	Computers and Electronics	Knowledge
Data handling and analysis		
4.A.2.a.4	Analyzing Data or Information	Work Activities
4.A.3.b.6	Documenting/Recording Information	Work Activities
2.C.1.b	Clerical	Knowledge
Contribution to digitalization		
2.B.3.e	Programming	Skills

Table 2: The top ten most and least digitalized occupations, out of a total of 795 occupations in the U.S. Department of Labor's O*NET program.

SOC	Occupation Title	d_o
15-1131	Computer Programmers	83.04
17-2061	Computer Hardware Engineers	82.34
15-1132	Computer Software Engineers, Applications	82.34
15-1133	Computer Software Engineers, Systems Software	82.34
15-1121	Computer Systems Analysts	81.05
15-1111	Computer and Information Research Scientists	77.13
15-1199	Computer Occupations, All Other	74.95
15-1141	Database Administrators	74.02
15-1134	Web Developers	73.65
15-1122	Information Security Analysts	73.34
⋮	⋮	⋮
47-2132	Insulation Workers, Mechanical	8.03
47-2131	Insulation Workers, Floor, Ceiling, and Wall	8.03
47-2151	Pipelayers	7.98
47-3011	Brickmasons, Tile and Marble Setters	7.91
47-2181	Roofers	7.58
47-3014	Painters, Paperhangers, Plasterers	7.38
47-2082	Tapers	7.08
35-9011	Dining Room Attendants and Bartender Helpers	6.54
53-7081	Refuse and Recyclable Material Collectors	5.84
51-3023	Slaughterers and Meat Packers	4.23

Table 3: The top five most and least digitalized industries in 1999 and 2017. Industries are defined at the 3-digit SIC level up to and including 2001, and at the 4-digit NAICS level beginning in 2002. *# firms* is the number of firms whose primary industry is the given industry. Only industries with at least five firms are displayed.

SIC	SIC Title	$D_{i,1999}$	# firms
737	Computer Programming, Data Processing, etc	67.33	278
357	Computer and Office Equipment	61.30	84
381	Search, Navigation, and Nautical Systems	57.58	13
366	Communications Equipment	54.48	81
871	Engineering, Architectural, and Surveying	54.28	11
⋮	⋮	⋮	⋮
251	Household Furniture	27.91	11
245	Wood Buildings and Mobile Homes	27.70	9
781	Motion Picture Production and Allied Services	25.75	6
162	Heavy Construction, Except Highway and Street	23.90	5
201	Meat Products	23.12	11

NAICS	NAICS Title	$D_{i,2017}$	# firms
5415	Computer Systems Design and Related Services	68.43	30
5112	Software Publishers	67.32	49
3341	Computer and Peripheral Equipment Manufacturing	66.07	25
5182	Data Processing, Hosting, and Related Services	65.11	46
5191	Other Information Services	62.43	83
⋮	⋮	⋮	⋮
2123	Nonmetallic Mineral Mining and Quarrying	28.50	10
5621	Waste Collection	26.51	6
2371	Utility System Construction	26.31	8
3116	Animal Slaughtering and Processing	25.69	7
2121	Coal Mining	24.36	5

Table 4: Firms' transition probabilities in % across quintiles of digitalization. The figures represent the probability that a firm remains in the same quintile or moves to a different quintile after one year or five years.

	From year t to year $t+1$				
	Q^1_{t+1}	Q^2_{t+1}	Q^3_{t+1}	Q^4_{t+1}	Q^5_{t+1}
Q^1_t	94.94	4.64	0.26	0.10	0.05
Q^2_t	4.87	87.82	6.88	0.36	0.06
Q^3_t	0.25	7.55	85.69	6.26	0.26
Q^4_t	0.08	0.21	6.25	87.04	6.42
Q^5_t	0.04	0	0.51	6.26	93.19

	From year t to year $t+5$				
	Q^1_{t+5}	Q^2_{t+5}	Q^3_{t+5}	Q^4_{t+5}	Q^5_{t+5}
Q^1_t	76.66	12.18	4.65	4.43	2.09
Q^2_t	10.47	66.69	16.14	4.58	2.13
Q^3_t	4.14	14.99	60.76	16.99	3.12
Q^4_t	3.61	4.03	11.67	62.36	18.33
Q^5_t	2.59	2.84	5.73	11.66	77.18

Table 5: Summary statistics at the firm level by digitalization quintiles. *Digitalization* is a continuous measure which can vary from 0 to 100 each year. All accounting variables are obtained from Compustat and winsorized at 1% and 99%. *Size* is the logarithm of the market value of equity (item CSHO $\times$ item PRCC_F). *Employees* is the number of employees recorded in Compustat. *Age* is the number of years since a firm's first appearance in Compustat. *Book-to-Market* is the book value of equity (item CEQ) divided by the market value of equity. *Tangibility* is property, plant, and equipment (item PPENT) divided by total assets (item AT). *R&D Intensity* is R&D expenditures (item XRD) divided by the market value of equity. *I/K* is capital expenditures (item CAPX) divided by property, plant, and equipment. *Leverage* is market leverage and equal to total debt (item DLC + item DLTT) divided by the sum of total debt and market value of equity. *ROA* is operating income after depreciation (item OIADP) divided by total assets. *WW* is the Whited-Wu index to measure financial constraints, calculated as in Whited and Wu (2006). The sample period is from July 2000 to June 2019.

	L	2	3	4	H
Digitalization	32.20	39.43	44.68	51.13	62.36
Size	6.24	6.31	6.15	5.94	5.79
Employees	12,692	16,680	8,124	6,745	5,951
Age	25.64	25.67	21.76	19.76	16.33
Book-to-Market	0.67	0.65	0.55	0.54	0.47
Tangibility	0.36	0.27	0.27	0.15	0.10
R&D Intensity	0.02	0.03	0.06	0.09	0.11
I/K	0.18	0.21	0.25	0.29	0.37
Leverage	0.28	0.25	0.23	0.16	0.10
ROA	0.07	0.06	0.01	-0.02	-0.06
WW	-0.31	-0.31	-0.27	-0.25	-0.22

Table 6: Average cumulative abnormal returns in % around three GDPR events: (i) March 12, 2014 (the European Parliament adopts the GDPR), (ii) December 15, 2015 (the European Parliament, Council, and Commission reach an agreement on the GDPR), and (iii) April 8, 2016 (the Council adopts the GDPR). In columns (1) and (2) ((3) and (4)), cumulative abnormal returns are calculated geometrically from closing on day [-1] ([-2]) to closing of day 1 (2) relative to each event date. Daily abnormal returns are calculated relative to the estimates of a market model estimated from 252 to 21 days prior to each event, with a minimum of 100 observations required for the estimation window. *Digitalization* can vary from 0 to 100 and is the digitalization score of firm f in year $t-1$. The control variables are defined in Table 5 and measured at the end of the previous year. Robust standard errors are in parentheses. Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %.

	$[-1, +1]$		$[-2, +2]$	
	(1)	(2)	(3)	(4)
Digitalization	-0.029***	-0.025***	-0.028***	-0.026**
	(0.01)	(0.01)	(0.01)	(0.01)
Size		0.108***		0.079*
		(0.03)		(0.05)
ROA		-1.329***		-2.058***
		(0.48)		(0.61)
Leverage		0.890**		1.191**
		(0.39)		(0.52)
Book-to-Market		0.259		0.153
		(0.16)		(0.21)
R^2	0.002	0.009	0.001	0.008
Observations	5,971	5,531	5,971	5,531

Table 7: Digitalization portfolios' excess returns in % over the six-factor model. Monthly returns are annualized by multiplying by 12. For a given month, stocks are value-weighted (Panel A) or equal-weighted (Panel B) into five portfolios based on their digitalization quintile in the previous year. A portfolio's monthly excess return over the risk-free rate is regressed on the market excess return (MKT), size (SMB), value (HML), profitability (RMW), investment (CMA), and momentum (MOM). The six factors are obtained from Kenneth French's website. Standard errors reported in parentheses are adjusted for heteroscedasticity and autocorrelation (Newey-West with 12 lags). Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The sample period is July 2000 to June 2019.

	Panel A: Value-weighted					
	L	2	3	4	H	H-L
α	1.904	2.828***	0.355	7.673***	8.400***	6.496**
	(1.401)	(1.040)	(1.334)	(1.639)	(2.119)	(2.664)
β^{MKT}	1.004***	1.042***	1.059***	1.035***	1.084***	0.080
	(0.031)	(0.028)	(0.037)	(0.058)	(0.051)	(0.066)
β^{SMB}	0.193***	0.314***	0.057	0.044	0.069	-0.124
	(0.062)	(0.037)	(0.072)	(0.046)	(0.085)	(0.112)
β^{HML}	0.007	0.023	0.029	-0.429***	-0.499***	-0.507***
	(0.032)	(0.043)	(0.073)	(0.054)	(0.071)	(0.088)
β^{RMW}	0.428***	0.452***	0.240***	-0.404***	-0.504***	-0.933***
	(0.053)	(0.055)	(0.075)	(0.121)	(0.144)	(0.167)
β^{CMA}	0.228***	0.217***	0.299**	0.171	-0.162	-0.390**
	(0.083)	(0.071)	(0.146)	(0.117)	(0.154)	(0.185)
β^{MOM}	-0.016	-0.049*	-0.021	-0.076	-0.162**	-0.146
	(0.045)	(0.029)	(0.044)	(0.071)	(0.076)	(0.108)

	Panel B: Equal-weighted					
	L	2	3	4	H	H-L
α	0.084	1.748	1.728	6.747***	10.402***	10.319***
	(1.475)	(1.302)	(2.442)	(2.326)	(3.179)	(3.315)
β^{MKT}	1.021***	1.015***	1.036***	0.962***	0.835***	-0.186**
	(0.062)	(0.044)	(0.060)	(0.047)	(0.092)	(0.076)
β^{SMB}	0.848***	0.979***	0.849***	0.834***	0.903***	0.055
	(0.051)	(0.049)	(0.064)	(0.062)	(0.135)	(0.127)
β^{HML}	0.244***	0.161***	0.008	-0.345***	-0.571***	-0.815***
	(0.059)	(0.044)	(0.108)	(0.077)	(0.061)	(0.094)
β^{RMW}	0.346***	0.324***	-0.032	-0.512***	-0.780***	-1.126***
	(0.113)	(0.072)	(0.075)	(0.108)	(0.136)	(0.211)
β^{CMA}	-0.019	0.033	0.089	0.187	-0.043	-0.025
	(0.119)	(0.098)	(0.125)	(0.133)	(0.188)	(0.233)
β^{MOM}	-0.221***	-0.205***	-0.211***	-0.204***	-0.365***	-0.144
	(0.066)	(0.045)	(0.046)	(0.068)	(0.140)	(0.169)

Table 8: Excess returns (α) in % over the six-factor model independently double-sorted on quintiles of digitalization and terciles of size, age, profitability (proxied by ROA), or R&D intensity. Monthly returns are annualized by multiplying by 12. Firm age is calculated from the first year in which a company appears in Compustat. The other variables are defined in Table 5. Returns are value-weighted when sorting firms every month; see Table A1 for the corresponding equal-weighted version. Standard errors reported in parentheses are adjusted for heteroscedasticity and autocorrelation (Newey-West with 12 lags). Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The sample period is July 2000 to July 2019.

	Value-weighted					
	L	2	3	4	H	H-L
			Size terciles			
T1	−0.565	−0.090	0.779	6.131*	11.195***	11.760***
	(2.763)	(3.056)	(3.787)	(3.459)	(4.240)	(4.349)
T2	0.122	2.261*	2.689	9.150***	12.768***	12.646***
	(1.273)	(1.239)	(1.692)	(2.212)	(3.650)	(4.019)
T3	2.100	3.030***	0.431	7.698***	8.294***	6.193**
	(1.447)	(1.064)	(1.345)	(1.656)	(2.142)	(2.686)
T1-T3						5.567
						(3.840)
			Age			
0-10	6.323**	7.743***	4.919*	7.079***	13.508***	7.184*
	(2.904)	(2.505)	(2.648)	(2.266)	(3.455)	(4.104)
11-20	2.350	8.138***	3.837*	8.908***	6.825***	4.476*
	(1.920)	(2.045)	(2.236)	(2.599)	(1.975)	(2.667)
21+	1.004	1.611	−1.507	7.129***	6.307***	5.303**
	(1.262)	(1.179)	(1.434)	(1.676)	(2.003)	(2.555)
T1-T3						1.881
						(3.590)
			ROA terciles			
T1	−1.900	5.926	0.856	7.212**	12.327***	14.227***
	(4.146)	(4.315)	(3.312)	(2.937)	(3.132)	(5.180)
T2	1.106	5.020***	1.150	8.046***	7.882***	6.776**
	(1.458)	(1.286)	(1.395)	(1.548)	(2.765)	(2.675)
T3	2.693	1.878	0.753	7.402***	9.012***	6.319***
	(1.757)	(1.253)	(1.580)	(2.014)	(1.604)	(2.256)
T1-T3						7.909*
						(4.626)
			R&D intensity			
T1	3.016**	3.964***	1.567	7.391***	19.330***	16.314***
	(1.487)	(1.386)	(1.429)	(2.693)	(3.716)	(3.968)
T2	−2.795	1.871	2.012	8.258***	8.275***	11.070***
	(2.563)	(1.838)	(1.642)	(2.350)	(2.345)	(3.768)
T3	0.425	−1.279	2.970	9.642***	5.159*	4.734
	(5.743)	(2.585)	(3.594)	(3.230)	(2.755)	(6.412)
T1-T3						11.580
						(7.160)

Table 9: Cross-sectional Fama-MacBeth regressions of individual monthly stock returns on digitalization and control variables. Monthly returns in % are annualized by multiplying by 12. All explanatory variables are standardized. Digitalization is a continuous measure which can vary from 0 to 100 and is the digitalization score of a firm in year $t-1$. β_{60}^{US} is the beta of a given stock's monthly return with the U.S. stock market return estimated over the preceding 60 months. *Book-to-Market* is the log of the book-to-market ratio. The remainder of the control variables are defined in Table 5. *Size* is measured at the end of the previous month. The other control variables are measured at the end of fiscal year $t-2$. The specification *Size: large* (*Size: small*) includes only stocks which have a market capitalization that is above (below) the median in the previous month. The specification *ROA: high* (*ROA: low*) includes only stocks whose ROA at the end of fiscal year $t-2$ is above (below) the median. All control variables are winsorized at 1% and 99%. Standard errors reported in parentheses are adjusted for heteroscedasticity and autocorrelation (Newey-West with 12 lags). Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The sample period is July 2000 to June 2019.

	All	Size: large	Size: small	ROA: high	ROA: low
Digitalization	1.912***	2.699***	0.890	2.342***	1.902***
	(0.354)	(0.359)	(0.615)	(0.393)	(0.590)
β_{60}^{US}	2.241***	-3.565***	7.261***	0.185	4.033***
	(0.535)	(0.491)	(0.862)	(0.623)	(0.761)
Size	-18.513***	-14.815***	-31.699***	-14.739***	-22.341***
	(0.816)	(0.944)	(1.554)	(0.982)	(1.231)
Book-to-Market	-0.500	-0.028	0.207	-0.829	0.651
	(0.419)	(0.462)	(0.664)	(0.557)	(0.632)
ROA	0.017	2.043	0.512	-0.324	-2.088
	(0.928)	(1.327)	(1.218)	(2.149)	(1.295)
I/K	0.698*	-0.095	1.415**	-0.381	1.412**
	(0.398)	(0.497)	(0.565)	(0.506)	(0.565)
Leverage	-0.647	0.845*	-1.841***	1.005*	-1.041*
	(0.407)	(0.462)	(0.607)	(0.532)	(0.589)
WW	-15.357***	-11.373***	-15.783***	-11.902***	-17.151***
	(0.800)	(0.875)	(1.267)	(0.966)	(1.187)
R^2	0.003	0.003	0.004	0.002	0.003
Observations	491,755	250,052	241,703	250,667	241,088

Table 10: Usage of the term "user" in firms' 10-K filings by digitalization level. In columns (1) and (2), the dependent variable is the natural logarithm of one plus the number of times "user(s)" is mentioned in a 10-K filing. In columns (3) and (4), the dependent variable is the natural logarithm of one plus the total number of words in paragraphs which mention "user(s)". *Digitalization* can vary from 0 to 100 and is the digitalization score of firm f in year $t-1$. *Total word count* is the natural logarithm of the total word count in firm f's 10-K filing in year t. *Size* is the natural logarithm of the market value of equity from the previous year. A constant is not reported for brevity. Data for 10-K filings is calculated based on parsed files from Loughran and McDonald (2011). Industry refers to the Fama-French 49 industries. Standard errors reported in parentheses are clustered at the year and industry level. Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The sample period is July 2000 to June 2019.

	User count		User paragraphs	
	(1)	(2)	(3)	(4)
Digitalization	0.043***	0.039***	0.022***	0.022***
	(0.009)	(0.008)	(0.005)	(0.005)
Total word count		0.403***		-0.049**
		(0.037)		(0.021)
Size		-0.017*		-0.004
		(0.009)		(0.006)
Year × Industry FE	Yes	Yes	Yes	Yes
R^2	0.383	0.415	0.312	0.314
Observations	45,993	45,983	45,993	45,983

Table 11: Logit regressions of the interaction between digitalization and Google search interest on whether firms' quarterly revenue growth is positive. The dependent variable is an indicator variable for whether firms' quarterly revenue has increased from quarter $q-1$ to quarter q. In columns (1) and (3) ((2) and (4)), *Digital* indicates whether a firm f's digitalization score in year $t-1$ is in the top quintile (above median). *New interest* is the three-month average search interest for "app" ending in quarter q. A constant is not reported for brevity. All coefficients are multiplied by 1000. Industry refers to the Fama-French 49 industries. Standard errors reported in parentheses are clustered at the month level. Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The full sample period is March 2004 to June 2019.

	2004–2019		2008–2019	
	(1)	(2)	(3)	(4)
	Top quintile	Median	Top quintile	Median
Digital × New interest	2.330***	3.341***	2.916**	5.471***
	(0.875)	(0.870)	(1.240)	(1.437)
Digital	-126.279**	-145.438**	-174.418**	-309.718***
	(51.951)	(59.765)	(84.974)	(106.867)
New interest	-13.388	-12.828	-13.267	-12.804
	(13.958)	(13.937)	(14.143)	(14.063)
Year × Industry FE	Yes	Yes	Yes	Yes
Observations	129,435	129,435	89,402	89,402

Table 12: Digitalization portfolios' value-weighted excess returns in % over the six-factor model by whether firms are user-focused. Monthly returns are annualized by multiplying by 12. This table separates Panel A of Table 7 according to whether a firm mentions "user" at least five times (Panel A) or not (Panel B) in their 10-K filing in the previous year. Standard errors reported in parentheses are adjusted for heteroscedasticity and autocorrelation (Newey-West with 12 lags). Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The sample period is July 2000 to June 2019. See Table A3 for the corresponding equal-weighted version.

	Panel A: User-focused					
	L	2	3	4	H	H-L
α	0.683	5.192***	2.113	8.122***	9.648***	8.965**
	(3.100)	(1.900)	(1.995)	(2.418)	(2.504)	(3.947)
β^{MKT}	1.310***	1.269***	1.166***	1.074***	1.082***	-0.228*
	(0.103)	(0.045)	(0.048)	(0.051)	(0.058)	(0.126)
β^{SMB}	0.441***	0.256***	-0.069	0.030	0.073	-0.368**
	(0.168)	(0.084)	(0.089)	(0.120)	(0.093)	(0.155)
β^{HML}	0.167	-0.056	0.081	-0.641***	-0.550***	-0.718***
	(0.174)	(0.070)	(0.083)	(0.122)	(0.077)	(0.188)
β^{RMW}	0.478***	0.305**	-0.142	-0.600***	-0.614***	-1.091***
	(0.156)	(0.123)	(0.170)	(0.166)	(0.163)	(0.272)
β^{CMA}	0.187	0.321**	0.011	0.057	-0.236	-0.423
	(0.190)	(0.135)	(0.166)	(0.205)	(0.177)	(0.337)
β^{MOM}	-0.116	-0.053	-0.082	-0.184**	-0.197**	-0.081
	(0.091)	(0.044)	(0.072)	(0.090)	(0.085)	(0.147)

	Panel B: Not user-focused					
	L	2	3	4	H	H-L
α	2.111	2.453**	0.090	7.343***	4.584***	2.473
	(1.411)	(1.101)	(1.484)	(1.723)	(1.407)	(2.098)
β^{MKT}	0.986***	1.008***	1.027***	1.013***	1.062***	0.076
	(0.028)	(0.029)	(0.044)	(0.065)	(0.067)	(0.085)
β^{SMB}	0.186***	0.324***	0.085	0.049	0.123	-0.063
	(0.058)	(0.037)	(0.078)	(0.053)	(0.125)	(0.168)
β^{HML}	-0.006	0.032	0.001	-0.293***	-0.397***	-0.391***
	(0.033)	(0.045)	(0.093)	(0.062)	(0.084)	(0.103)
β^{RMW}	0.425***	0.447***	0.266***	-0.291**	-0.299**	-0.724***
	(0.053)	(0.055)	(0.078)	(0.117)	(0.126)	(0.151)
β^{CMA}	0.223***	0.216***	0.309**	0.174*	0.138	-0.085
	(0.082)	(0.071)	(0.153)	(0.089)	(0.133)	(0.173)
β^{MOM}	-0.008	-0.054*	0.002	-0.020	-0.072	-0.064
	(0.044)	(0.031)	(0.040)	(0.060)	(0.070)	(0.101)

Table 13: Median analyst forecast error for earnings per share (EPS) by firms' digitalization level and whether firms mention "user" at least five times in their 10-K filing (Panel A) or not (Panel B). In columns (1) to (4), forecast error is defined as the I/B/E/S actual annual EPS minus the median I/B/E/S consensus forecast of annual EPS. The consensus forecast is calculated 8 (20) months prior to the end of the forecast period for the 1 (2) year horizon. Forecast errors are normalized by the lagged stock price at the end of the previous fiscal year to control for heteroscedasticity and then winsorized at 1% and 99%. In columns (5) and (6), forecast error is the average annualized EPS growth over the preceding five years minus the median annualized long-term growth forecast for EPS from 56 months prior. I/B/E/S data is obtained from WRDS. *Digitalization* is a continuous measure which can vary from 0 to 100 and is the digitalization score of firm f in year $t-1$. β_{60}^{US} is the beta of a given stock's monthly return with the U.S. stock market return estimated over the preceding 60 months. The remainder of the control variables are defined in Table 5. Coefficients in columns (1) to (4) are multiplied by 100. Standard errors reported in parentheses are clustered at the 3-digit SIC industry and year level. Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The sample period is July 2000 to June 2019.

Panel A: User-focused

	1 year horizon		2 year horizon		LTG	
	(1)	(2)	(3)	(4)	(5)	(6)
Digitalization	0.014	0.005	0.043*	0.048*	-0.012	-0.097
	(0.009)	(0.011)	(0.025)	(0.026)	(0.050)	(0.063)
β_{60}^{US}		-0.068		-0.852**		0.553
		(0.121)		(0.389)		(2.057)
Size		0.669***		2.082***		3.606***
		(0.091)		(0.243)		(0.291)
Book-to-Market		-0.130		0.663		-6.037***
		(0.297)		(0.791)		(2.045)
Leverage		-1.581*		-3.489**		0.342
		(0.821)		(1.415)		(5.171)
I/K		0.583		0.081		7.182*
		(0.445)		(0.794)		(3.612)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.018	0.075	0.067	0.193	0.060	0.130
Observations	5,912	5,874	5,476	5,454	3,224	3,221

Panel B: Not user-focused

	1 year horizon		2 year horizon		LTG	
	(1)	(2)	(3)	(4)	(5)	(6)
Digitalization	0.018*	0.016***	0.032	0.034**	0.031	-0.043
	(0.009)	(0.006)	(0.026)	(0.016)	(0.056)	(0.042)
β_{60}^{US}		-0.240		-1.679**		-1.774
		(0.141)		(0.652)		(1.518)
Size		0.783***		1.712***		2.707***
		(0.079)		(0.176)		(0.265)
Book-to-Market		0.404*		-0.617		-6.855***
		(0.209)		(0.595)		(1.174)
Leverage		-2.076***		-4.641***		-6.134**
		(0.588)		(1.251)		(2.270)
I/K		0.024		-1.068		7.641**
		(0.604)		(0.880)		(3.333)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.017	0.100	0.060	0.203	0.050	0.136
Observations	12,098	12,018	11,247	11,217	7,571	7,569

Table 14: Cumulative excess returns in % over a (-5,1) window around earnings announcements of firms sorted into portfolios of digitalization quintiles and by whether firms mention "user" at least five times in their 10-K filing (Panel A) or not (Panel B). Excess returns are the stock returns minus the risk-free rate, then annualized by multiplying by four. The (-5,1) window is the six-day window beginning five days prior to the quarterly earnings announcement day and ending the day after the announcement day. All cumulative excess returns within a calendar quarter are aggregated to digitalization quintiles. This aggregation procedure is either value-weighted according to the stock market capitalization at the end of the previous calendar quarter, or equal-weighted. Standard errors reported in parentheses are adjusted for heteroscedasticity and autocorrelation (Newey-West with 12 lags). Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The sample period is July 2000 to June 2019.

Panel A: User-focused											
Value-weighted						Equal-weighted					
L	2	3	4	H	H-L	L	2	3	4	H	H-L
0.86	3.91**	0.29	3.35*	3.91***	3.05**	1.53	2.52**	0.92	2.88**	2.33**	0.80
(1.22)	(1.72)	(1.18)	(1.75)	(1.24)	(1.48)	(0.99)	(1.03)	(0.96)	(1.24)	(0.91)	(0.78)

Panel B: Not user-focused											
Value-weighted						Equal-weighted					
L	2	3	4	H	H-L	L	2	3	4	H	H-L
1.18*	1.21**	0.87	2.12***	0.44	-0.74	2.40***	1.93**	1.40	2.47***	1.08	-1.32
(0.63)	(0.60)	(0.87)	(0.56)	(1.25)	(1.30)	(0.66)	(0.74)	(0.94)	(0.71)	(1.29)	(1.10)

8 Appendix

Figure A1: Daily hours spent with digital media per adult user in the United States from 2008 to 2018. Source: Bond Internet Trends (Meeker and Wu, 2019), as compiled from eMarketer data.

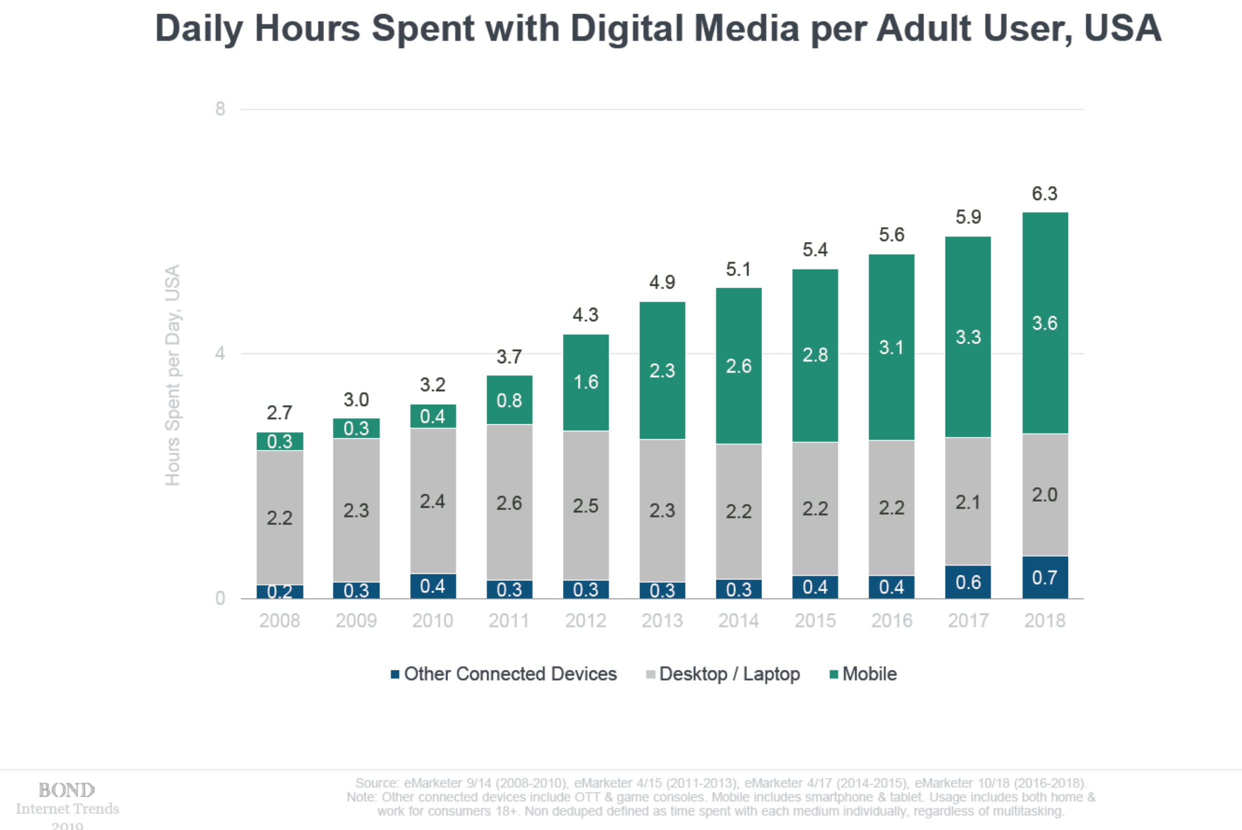


Figure A2: Mobile retail e-commerce sales in the United States from 2013 to 2020. Forecasts are denoted by *. The data includes products or services ordered using the internet via mobile devices, regardless of method of payment or fulfillment, and excludes travel and ticket sales. Source: Statista (2021), as compiled from eMarketer data.

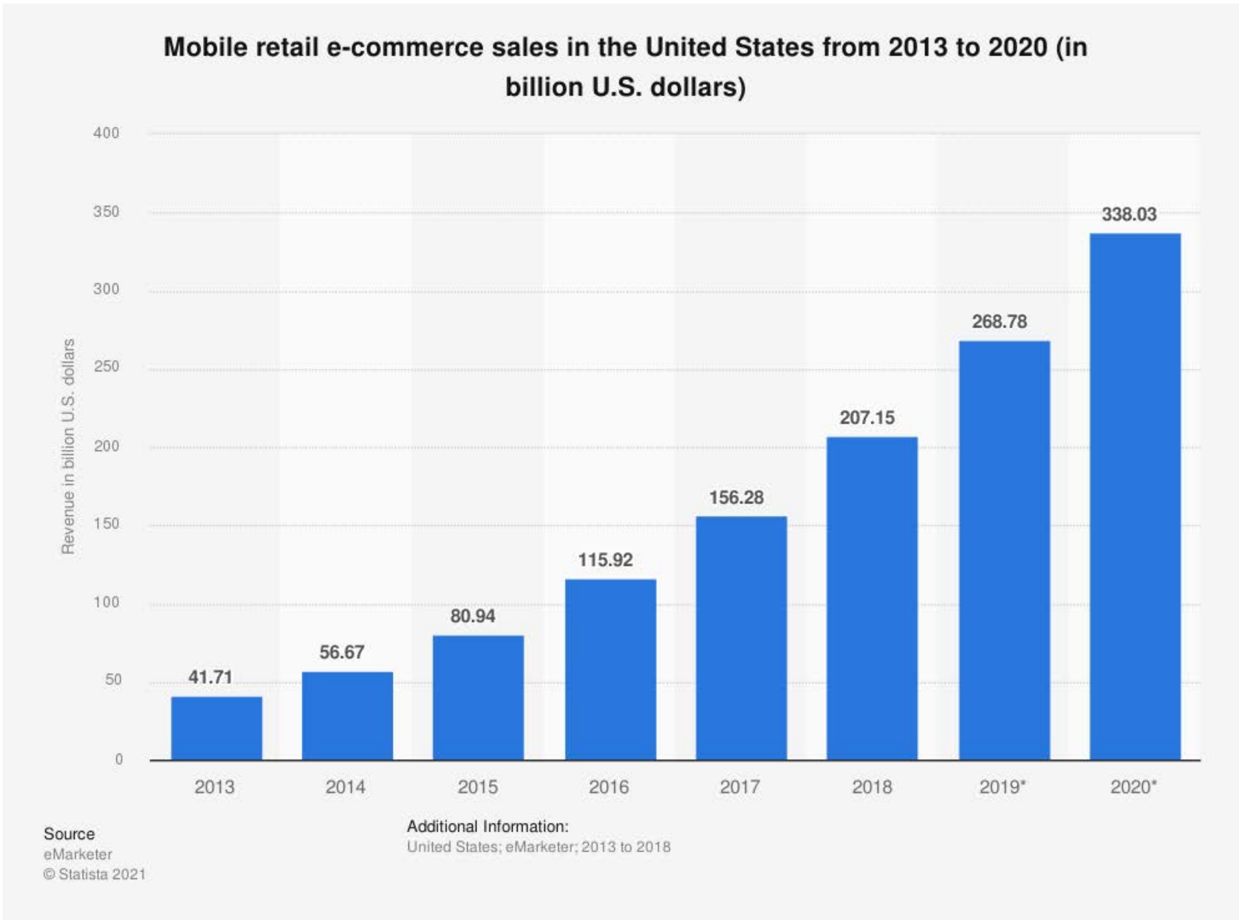


Table A1: Excess returns (α) in % over the six-factor model independently double-sorted on quintiles of digitalization and terciles of size, age, profitability (proxied by ROA), or R&D intensity. Monthly returns are annualized by multiplying by 12. Firm age is calculated from the first year in which a company appears in Compustat. The other variables are defined in Table 5. Returns are equal-weighted when sorting firms every month; see Table 8 for the corresponding value-weighted version. Standard errors reported in parentheses are adjusted for heteroscedasticity and autocorrelation (Newey-West with 12 lags). Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The sample period is July 2000 to July 2019.

	Equal-weighted					
	L	2	3	4	H	H-L
			Size terciles			
T1	−2.318	−0.081	0.632	3.140	8.282*	10.600**
	(3.620)	(3.688)	(4.853)	(4.108)	(4.725)	(4.432)
T2	−0.412	1.646	3.190	8.895***	12.177***	12.589***
	(1.394)	(1.076)	(2.024)	(2.126)	(3.681)	(4.123)
T3	2.407*	3.036***	2.729*	8.860***	10.870***	8.463***
	(1.359)	(1.062)	(1.532)	(2.069)	(2.140)	(2.616)
T1-T3						2.137
						(4.066)
			Age			
0-10	−3.204	0.704	−1.563	2.779	12.722***	15.926***
	(2.611)	(2.626)	(3.383)	(3.399)	(3.836)	(4.514)
11-20	0.668	3.416**	1.796	8.595***	6.598***	5.930**
	(2.063)	(1.722)	(2.971)	(2.532)	(2.430)	(2.746)
21+	0.463	0.486	1.473	5.969***	5.771***	5.308***
	(1.356)	(1.139)	(1.732)	(1.670)	(1.777)	(1.909)
T1-T3						10.618***
						(4.067)
			ROA terciles			
T1	−9.070***	−3.765	−3.632	2.587	7.829**	16.899***
	(3.264)	(2.972)	(4.464)	(3.386)	(3.914)	(4.510)
T2	0.963	3.007**	3.776**	8.283***	11.318***	10.355***
	(1.740)	(1.465)	(1.911)	(1.861)	(2.729)	(3.234)
T3	4.747***	3.635***	6.023***	9.913***	13.614***	8.866***
	(1.582)	(1.213)	(1.721)	(2.209)	(2.406)	(2.643)
T1-T3						8.033**
						(3.604)
			R&D intensity			
T1	1.615	2.226	3.868*	9.155***	14.582***	12.967***
	(1.683)	(1.821)	(2.179)	(2.684)	(3.890)	(4.339)
T2	−2.550	1.242	8.454***	11.287***	13.957***	16.508***
	(2.325)	(1.340)	(2.128)	(2.376)	(2.623)	(3.610)
T3	−5.227	−4.957*	−1.122	4.106	8.280**	13.507*
	(6.383)	(2.814)	(4.130)	(3.183)	(4.133)	(7.199)
T1-T3						0.540
						(5.532)

Table A2: Median analyst forecast error for sales by firms' digitalization level and whether firms mention "user" at least five times in their 10-K filing (Panel A) or not (Panel B). This table is the sales version of Table 13. In columns (1) to (4), forecast error is defined as the I/B/E/S actual annual sales minus the median I/B/E/S consensus forecast of annual sales. The consensus forecast is calculated 8 (20) months prior to the end of the forecast period for the 1 (2) year horizon. Forecast errors are normalized by the lagged stock price at the end of the previous fiscal year to control for heteroscedasticity and then winsorized at 1% and 99%. In columns (5) and (6), forecast error is the average annualized sales growth over the preceding five years minus the median annualized long-term growth forecast for sales from 56 months prior. I/B/E/S data is obtained from WRDS. *Digitalization* is a continuous measure which can vary from 0 to 100 and is the digitalization score of firm f in year $t-1$. β_{60}^{US} is the beta of a given stock's monthly return with the U.S. stock market return estimated over the preceding 60 months. The remainder of the control variables are defined in Table 5. Coefficients in columns (1) to (4) are multiplied by 100. Standard errors reported in parentheses are clustered at the 3-digit SIC industry and year level. Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The sample period is July 2000 to June 2019.

Panel A: User-focused

	1 year horizon		2 year horizon		LTG	
	(1)	(2)	(3)	(4)	(5)	(6)
Digitalization	0.023	0.023	0.117	0.095	0.079	-0.035
	(0.042)	(0.021)	(0.143)	(0.113)	(0.052)	(0.074)
β_{60}^{US}		-0.953**		-5.079**		2.673
		(0.369)		(2.063)		(1.941)
Size		1.676***		5.620***		2.666***
		(0.296)		(0.959)		(0.589)
Book-to-Market		1.229*		-0.916		-7.763
		(0.704)		(4.374)		(5.175)
Leverage		-2.601		-18.466**		-6.546
		(2.856)		(7.709)		(10.780)
I/K		1.847		-3.130		0.681
		(1.446)		(2.637)		(5.391)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.037	0.079	0.092	0.167	0.064	0.137
Observations	5,701	5,662	5,089	5,070	828	828

Panel B: Not user-focused

	1 year horizon		2 year horizon		LTG	
	(1)	(2)	(3)	(4)	(5)	(6)
Digitalization	0.029	0.019	0.155	0.107	0.042	0.004
	(0.036)	(0.025)	(0.129)	(0.082)	(0.097)	(0.096)
β_{60}^{US}		-0.502*		-6.148*		-0.757
		(0.282)		(3.296)		(3.058)
Size		1.626***		4.395***		2.383***
		(0.276)		(0.686)		(0.360)
Book-to-Market		-0.242		-9.160**		-4.879*
		(0.819)		(4.219)		(2.548)
Leverage		-3.630*		-21.293***		-4.597
		(1.815)		(5.689)		(4.303)
I/K		1.294		-4.404		2.408
		(1.941)		(4.708)		(5.728)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.026	0.061	0.083	0.163	0.055	0.130
Observations	11,167	11,093	9,718	9,689	1,485	1,485

Table A3: Digitalization portfolios' equal-weighted excess returns in % over the six-factor model by whether firms are user-focused. Monthly returns are annualized by multiplying by 12. This table separates Panel B of Table 7 according to whether a firm mentions "user" at least five times (Panel A) or not (Panel B) in their 10-K filing in the previous year. Standard errors reported in parentheses are adjusted for heteroscedasticity and autocorrelation (Newey-West with 12 lags). Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The sample period is July 2000 to June 2019. See Table 12 for the corresponding value-weighted version.

	Panel A: User-focused					
	L	2	3	4	H	H-L
α	-0.913	3.800**	4.138	9.235***	11.818***	12.730***
	(2.141)	(1.865)	(3.203)	(2.969)	(3.551)	(3.977)
β^{MKT}	1.143***	1.042***	0.955***	0.957***	0.839***	-0.304***
	(0.072)	(0.078)	(0.075)	(0.065)	(0.097)	(0.087)
β^{SMB}	1.005***	0.994***	0.877***	0.856***	0.928***	-0.077
	(0.059)	(0.084)	(0.099)	(0.091)	(0.145)	(0.159)
β^{HML}	0.262***	-0.014	-0.140*	-0.554***	-0.621***	-0.883***
	(0.098)	(0.067)	(0.083)	(0.089)	(0.070)	(0.120)
β^{RMW}	0.371**	0.189**	-0.292***	-0.854***	-0.901***	-1.272***
	(0.173)	(0.075)	(0.079)	(0.184)	(0.149)	(0.282)
β^{CMA}	-0.118	0.083	0.042	0.272	-0.156	-0.038
	(0.160)	(0.139)	(0.166)	(0.211)	(0.209)	(0.266)
β^{MOM}	-0.245***	-0.288***	-0.309***	-0.297***	-0.392**	-0.147
	(0.078)	(0.058)	(0.087)	(0.093)	(0.158)	(0.196)

	Panel B: Not user-focused					
	L	2	3	4	H	H-L
α	0.766	1.671	1.596	5.667**	8.318***	7.552***
	(1.531)	(1.393)	(2.395)	(2.261)	(2.744)	(2.756)
β^{MKT}	1.003***	1.010***	1.069***	0.959***	0.825***	-0.178**
	(0.062)	(0.043)	(0.064)	(0.043)	(0.086)	(0.071)
β^{SMB}	0.837***	0.975***	0.844***	0.837***	0.866***	0.028
	(0.058)	(0.048)	(0.065)	(0.053)	(0.123)	(0.126)
β^{HML}	0.237***	0.188***	0.043	-0.231***	-0.466***	-0.703***
	(0.060)	(0.044)	(0.119)	(0.083)	(0.064)	(0.093)
β^{RMW}	0.351***	0.337***	0.046	-0.323***	-0.463***	-0.814***
	(0.114)	(0.078)	(0.090)	(0.076)	(0.110)	(0.183)
β^{CMA}	-0.011	0.026	0.086	0.093	0.236	0.247
	(0.120)	(0.099)	(0.123)	(0.102)	(0.149)	(0.188)
β^{MOM}	-0.227***	-0.197***	-0.193***	-0.159***	-0.289***	-0.062
	(0.070)	(0.047)	(0.044)	(0.056)	(0.099)	(0.130)

Table A4: Cumulative excess returns in % over a (-10,1) window around earnings announcements of firms sorted into portfolios of digitalization quintiles and by whether firms mention "user" at least five times in their 10-K filing (Panel A) or not (Panel B). Excess returns are the stock returns minus the risk-free rate, then annualized by multiplying by four. The (-10,1) window is the 11-day window beginning 10 days prior to the quarterly earnings announcement day and ending the day after the announcement day. All cumulative excess returns within a calendar quarter are aggregated to digitalization quintiles. This aggregation procedure is either value-weighted according to the stock market capitalization at the end of the previous calendar quarter, or equal-weighted. Standard errors reported in parentheses are adjusted for heteroscedasticity and autocorrelation (Newey-West with 12 lags). Significance levels are denoted by * = 10%, ** = 5%, and *** = 1 %. The sample period is July 2000 to June 2019.

Panel A: User-focused											
Value-weighted						Equal-weighted					
L	2	3	4	H	H-L	L	2	3	4	H	H-L
2.64**	4.82***	0.54	6.00**	6.06***	3.43**	3.09**	5.33***	2.70*	4.51**	3.58**	0.49
(1.18)	(1.67)	(2.07)	(2.55)	(1.57)	(1.65)	(1.27)	(1.73)	(1.61)	(1.88)	(1.56)	(1.25)

Panel B: Not user-focused											
Value-weighted						Equal-weighted					
L	2	3	4	H	H-L	L	2	3	4	H	H-L
2.29***	2.15***	1.40	3.10***	1.22	-1.07	3.49***	3.44***	2.50**	4.11***	2.83*	-0.66
(0.80)	(0.70)	(0.99)	(0.88)	(1.26)	(1.36)	(1.05)	(1.09)	(1.21)	(1.06)	(1.64)	(1.18)