Journal of Economic Dynamics & Control 181 (2025) 105217

Contents lists available at ScienceDirect

Journal of Economic Dynamics and Control

journal homepage: www.elsevier.com/locate/jedc

The “digital” premium: Why does digitalization drive stock returns?[1]

Katharina Drechsler[a,*], Sebastian Müller[b], Heinz-Theo Wagner[c]

[a] University of Cologne, Cologne Institute for Information Systems, Germany
[b] Technical University of Munich, TUM School of Management, Campus Heilbronn, Center for Digital Transformation, Germany
[c] Neu-Ulm University of Applied Sciences (HNU), Institute for Digital Innovation (IDI), Germany

ARTICLE INFO

Keywords:
Digital orientation
Asset pricing
Market efficiency
Textual analysis
Digitalization
Digital innovation

ABSTRACT

To explore whether the rise of digital innovation across firms might point to the existence of a new risk compensation in asset pricing, we construct a text-based measure of the socio-technical phenomenon of digitalization, called digital orientation, by using the MD&A section of annual firm reports from 1996 to 2020. We find that firms with a high digital orientation (digital leaders) are systematically different along several key characteristics like valuation, sales growth, and profitability, forming a peer group of digitally leading firms across traditional industry boundaries. A digital orientation strategy, which is long (short) stocks with high (low) digital orientation, earns an equally weighted (value-weighted) monthly six-factor alpha of 0.41 % (0.28 %) per month, both statistically significant at 5 %. These results are robust to various sensitivity checks, including alternative constructions of the digital orientation measure, controls for industry membership, changes to the sample dataset, and the use of an alternative dictionary. Additionally, the results remain comparable if we construct an alternative investment strategy based on the portfolio holdings of funds with a digital innovation focus.

“There is no alternative to digital transformation. Visionary companies will carve out new strategic options for themselves – those that don’t adapt, will fail.” – Jeff Bezos

1. Introduction

In recent years, firms have been under pressure to transform digitally by reconsidering their business model, devising new digital business strategies, and using digital technologies to innovate (Vial, 2019; Wu et al., 2019). Against this backdrop, recent studies in Information Systems indicate that increasing digitalization — “a sociotechnical process of applying digitizing techniques to broader social and institutional contexts that render digital technologies infrastructural” (Tilson et al., 2010, p. 749) — blurs industrial

[1] We wish to express our thanks to Paul Zimmermann and the participants at the 28th Annual Meeting of the German Finance Association (DGF) and at the 12th Portuguese Finance Network Conference 2023 for valuable feedback. Furthermore, we would like to thank Maximilian Sölch for his excellent research assistance. An earlier version of this article was included in the first author’s doctoral dissertation (Drechsler, 2023)

* Corresponding author.
E-mail addresses: drechsler@wiso.uni-koeln.de (K. Drechsler), sebastian.mueller.hn@tum.de (S. Müller), heinz-theo.wagner@hnu.de (H.-T. Wagner).

https://doi.org/10.1016/j.jedc.2025.105217
Received 8 October 2024; Received in revised form 4 November 2025; Accepted 11 November 2025
Available online 12 November 2025

boundaries and challenges the separation of firms along traditional industrial categories. These industrial boundaries are blurring because digital innovations shift the focus from single products toward smart, connected products (physical devices connected to the internet with computing power), integrate heterogeneous areas of knowledge, and require firms to find ways to access and recombine knowledge rooted in different industries and contexts (Yoo et al., 2012; Seo, 2017). Moreover, the editable and reprogrammable nature of digital technology enables the development and deployment of digital innovation across organizations and industries (Kallinikos et al., 2013). Thus, digital innovations, defined as "the creation or adoption, and exploitation of an inherently unbounded, value-adding novelty (e.g., product, service, process, or business model) through the incorporation of digital technology" (Hund et al., 2021, p. 2), transcend established industrial boundaries (Porter and Heppelmann, 2014) and lead to a convergence of products and service offerings (Nambisan et al., 2017). Hence, firms are increasingly forced to defend their market against firms from other industries (Seo, 2017; Hopp et al., 2018). For example, Microsoft has been competing with telecommunication firms, since its acquisition of Skype (Yoo et al., 2010). Digital innovation, such as artificial intelligence, the internet of things, and cloud computing, may bring together firms in product markets across traditional industry boundaries and make them competitors. At the same time, firms engaging in and driving digital innovation share similar characteristics, such as expertise in digital technology, digital devices, and digital services.

Considering these arguments, digitalization might have unexplored consequences for capital markets. For example, because digital innovation is a cross-industry phenomenon, it seems increasingly inaccurate to classify firms solely based on their industry affiliation. The heterogeneity of digital innovation and digitalization efforts among firms within one industry rather suggests the need for an additional classification to capture the digital convergence across industry boundaries leading to a peer group of digitally leading firms (Yoo et al., 2012; Seo, 2017). Because digitalization affects the entire economy, one may further argue that the ability to digitally innovate is not only a useful firm characteristic to differentiate between firms, but may also reflect a systematic digital risk with consequences for asset pricing. Accordingly, this study identifies systematic differences in firm characteristics and financial performance depending on firms' orientation toward digital innovation. A firm's orientation toward digital innovation (briefly digital orientation) reflects a firm's past actions as well as future intentions to adopt, use and create digital innovation.

We explore whether the rise of digital innovation across firms might point to the existence of a new risk compensation in asset pricing by constructing a firm-specific measure of digital orientation (*DO*) using textual data from the MD&A section of annual firm reports. We rely on the MD&A section, since it is not audited and reflects the management's thoughts and opinions. Our methodology is similar to earlier works like Loughran and McDonald (2011), Hoberg and Philipps (2010, 2016), and Hillert et al. (2014). Specifically, to obtain a *DO* measure for each firm, we follow a bag-of-words approach to parse the 10-Ks and derive vectors of words and word counts. Our digital innovation dictionary is based on a validated word list from Information Systems literature and is theoretically grounded in the notion of the layered modular architecture of digital technology embedded in digital innovation (Yoo et al., 2010). It measures the extent to which a firm creates and uses digital innovation, distinguishing four layers of digital technology: device, network, service, and content layer (Yoo et al., 2010). As pointed out by Loughran and McDonald (2016), the dictionary approach is easy to understand and easy to replicate. Nevertheless, despite its simplicity, our measure is well able to explain important differences between firms beyond industry affiliation.

Our analyses yield two major findings. First, we demonstrate that firms with a high digital orientation, which we refer to as digital leaders, have distinct characteristics compared to firms with a low digital orientation. Digital leaders have, among other characteristics, significantly higher valuation, considerably higher sales growth, substantially lower cost ratios, significantly higher volatility, and substantially lower profitability in the short term but significantly higher expected profitability in the long term as measured by long-term earnings forecasts. Consistent with our expectations, digital leaders are also significantly younger and have a significantly higher (market) risk.

While our digital orientation (*DO*) measure is thus able to describe firm heterogeneity across important dimensions, it is also a variable that we cannot easily account for by other characteristics. Multivariate regressions to explain firms' digital orientation show low R^2 values ranging between 1 % and 4 %, even though these regressions contain explanatory variables like firm size, firm age, or Nasdaq membership, among others, which seem naturally linked to a firm's digital orientation. Even after additionally controlling for industry affiliation, the explanatory power increases only to a modest R^2 of 22 %.

Additionally, we assess the ability of the *DO* measure to explain stock inclusion in innovation-focused investment funds such as the ARK Innovation Fund. Cross-sectional Fama-MacBeth regressions show that the *DO* measure is a significant predictor of fund membership, alongside traditional firm characteristics like size, analyst coverage, and R&D spending. Our findings are robust across different groups of funds focused on digital innovation and information technology. Moreover, we show that only a moderate correlation exists between the fitted value of the Fama-MacBeth regression to explain fund membership (excluding DO as explanatory variable) and the *DO* measure. Based on these analyses, we conclude that our text-based *DO* measure is a unique variable to describe a firm's digital orientation, which is not captured well by fundamentals. Additionally, we are able to show that firms with a high *DO* (i.e., digital leaders) are systematically different from other firms of their respective traditional industry category and form their own peer group beyond traditional industry categorizations.

Second, we study the performance of a portfolio strategy based on our digital orientation measure. This strategy is long (short) in stocks of firms in the highest (lowest) digital orientation quintile, as its performance is referred to as the digital leaders' premium (or briefly digital premium) in the following. The six-factor alpha of the equally weighted portfolio following a quintile (5) minus quintile (1) strategy and controlling for exposure to momentum and the five factors of Fama and French (2015), is 0.41 % per month, with a t-statistic of 2.43. Digital leaders have a high negative exposure against the profitability factor *RMW* and the *HML* value factor of Fama and French. These relations are consistent with the fact that firms with a high digital orientation have lower operating profitability and a higher valuation. We also observe a statistically significant six-factor alpha if we form a value-weighted long-short portfolio (0.28 % per month; t-statistic: 2.41). Noting that high *DO* firms have higher market betas, we also run regressions that additionally include the betting-against-beta factor (BAB) from Frazzini and Pedersen (2014) as an explanatory variable. In this case, equally weighted and value-weighted alphas increase further to 0.50 % and 0.34 % per month, with both estimates being statistically significant at the 1 %-level.

To verify that our results are robust, we conduct several additional tests regarding changes in the calculation of the *DO* measure (e. g., accounting for the importance of each term, using alternative sections of the 10-K reports), controls for industry membership, changes in the sample data set (e.g., excluding the 1998/99 dot.com bubble and periods of high sentiment), and the use of an alternative dictionary. We find that the size of the six-factor alphas may increase or decrease slightly across the robustness tests, but the estimates are overwhelmingly statistically significant (see Section 4.3 for details). Thus, the documented pattern appears robust to reasonable changes in the methodology.

Additionally, we also compute portfolio returns based on stock membership in digital innovation-focused investment funds to compare them with the results for our *DO*-based strategy. Our results indicate that portfolios based on both measures earn statistically significant positive risk-adjusted returns, which is consistent with the idea that being a digital leader commands a risk premium.[1] Finally, we demonstrate how *DO* can be combined with profitability and stock momentum in double sorting approaches to enhance traditional factor strategies.

This research moves beyond the insights of existing literature in the following distinct ways. First, we contribute to the fast-growing field of research using textual analysis (Loughran and McDonald, 2016). Earlier research has relied on textual analyses to assess the link between firms' financial performance and accounting documents' readability (e.g., Li, 2008, Guay et al., 2016), the adoption of seven specific digital technologies (Chen and Srinivasan, 2023), changes in the language and construction of annual reports (e.g., Cohen et al., 2020), document similarity, and the sentiment prevailing in annual reports or media (e.g., Tetlock, 2007; Loughran and McDonald, 2011; Hillert et al., 2014; Hillert et al., 2024). We use a bag-of-words approach, on which sentiment analysis is also based, but use the approach to assess the extent to which firms have used and created digital innovation.

The most closely related study regarding the use of textual analysis is the contemporaneous paper by Chen and Srinivasan (2023) who apply a word list encompassing seven digital technologies. Our paper differs in several significant ways. We adopt a sociotechnical perspective on digitalization by drawing on recent literature on digital innovation, which highlights its socio-technical nature. This perspective encompasses not only the technology but also, for instance, services and contents that it entails. Our measurement of digital orientation is based on validated a word list from Information Systems literature, which is theoretically grounded in the notion of the layered architecture of digital technology embedded in digital innovation (Yoo et al., 2010).

Existing research has relied on firms' R&D expenditure or patent data to study innovation (e.g., Fang et al., 2014). Due to the distinct nature of digital innovation, such traditional approaches to measuring intangible assets cannot fully account for firms' innovation activities that rely on digital technologies. The reasons are as follows: Firms' R&D expenditures are representative of innovation activities, particularly in the manufacturing sector. However, in the service sector (e.g., financial industry), R&D expenditure is frequently not reported. Similar problems exist for patents, another frequently used measure of firms' innovation activities. While patents measure product innovation, they cannot account for other types of innovation, such as business model or process innovations, which constitute significant components of firms' innovation activities. By following a bag-of-words approach, we can derive a comprehensive measure of firms' orientation towards digital innovation. The measure, which is well grounded in extant knowledge on the socio-technical phenomenon of digital innovation in the Information Systems field, allows us to draw conclusions about firms' digital orientation toward digital innovation and its connection to firms' financial performance.

Second, we contribute to the digital innovation literature, which argues that the very nature of digital technology blurs industry boundaries (Kallinikos et al., 2013; Nambisan et al., 2017; Seo, 2017), by demonstrating empirically that digitalization occurs among a group of firms across traditional industry boundaries. With our study, we are among the first to test the theoretical argument of industry convergence and show that there is a convergence driven by digital orientation. Thus, we demonstrate that traditional industry classification does not fully explain higher returns among digital leaders. At the same time, we shed more light on the phenomenon of convergence by showing that it is not so much about industries (according to traditional categories) that are converging, but firms across industries that form a new peer group beyond traditional industry categories. We show that firms with a high digital orientation (digital leaders) exhibit distinct characteristics compared to firms with a low digital orientation, such as considerably higher sales growth, and substantially lower cost ratios.

Third, we contribute to the literature on asset pricing for intangible assets by proposing a *DO* factor. Thus far, existing asset pricing models, such as the Fama and French (1993) three-factor model, the Fama and French (2015) five-factor model, or the Fama and French (2018) six-factor model, still seem to fail to account for a firm's intangible assets, and in particular digital innovation. Yet, intangible assets account for around one-third of investment volume in the U.S. market, making them economically significant

[1] While most funds in our sample are passive ETFs, some funds like the ARK Innovation Fund have an active management philosophy. It is therefore possible that at least parts of the alpha which we observe for the strategy based on fund membership can be attributed to managers' skills. For this reason, our *DO* measure arguably represents a more accurate proxy for capturing digital orientation relative to fund holdings, especially since fund managers with an active orientation likely also consider other stock characteristics in the selection process. Moreover, compared to using portfolio holdings, the *DO* measure offers several additional advantages: It is not a binary variable but captures varying degrees of digital orientation; it enables a layer-based analysis of digital technologies; and, in principle, it can also be applied to smaller stocks and to equities from other markets.

(Corrado and Hulten, 2010). Accordingly, existing research has pointed to our incomplete understanding of the link between intangible assets and asset pricing in financial markets (Daniel and Titman, 2006). The study by Chen and Srinivasan (2023) demonstrates that nontech firms, which are early adopters of digital technologies show similarities with tech firms in terms of economic performance, and have 0.44 % higher risk-adjusted stock returns per month. Our paper differs in that we refer to asset pricing literature to demonstrate that digital orientation represents a priced risk factor that may enhance portfolio diversification when combined with other well-known factor strategies, particularly profitability and momentum. The returns associated with this factor arguably reflect the compensation investors require for bearing systematic exposure to the digital economy's evolution.

The paper is organized as follows. Section 2 outlines the theoretical foundation by reviewing the literature on digital innovation from the field of Information Systems and builds the foundation for constructing our text-based *DO* measure, described in Section 3. Additionally, Section 3 provides details of the data sources, descriptive statistics, and the analysis of our measure of digital orientation. Section 4 compares digital leaders, firms with a high digital orientation, to digital laggards, firms with a low digital orientation. We also present the results of portfolio tests based on the *DO* measure and historical *DO* factor exposure. Additionally, we conduct robustness tests. Section 5 concludes.

2. Theoretical foundation

Extant literature in Information Systems discusses the characteristics and effects of digital innovation and argues, among others, that the distinct properties of digital technology support the creation, diffusion, and use of digital innovation. These digital innovations are eventually deeply embedded in most organizational processes and market offerings, shape strategy, and unfold disruptive potential changing entire industries (Bharadwaj et al., 2013; Vial, 2019). This disruptive potential stems from digital innovations' ability to enable the blurring of industrial boundaries, as we outline in the following. Since firms are under pressure to "[…] harness outside expertise and ingenuity on an unprecedented scale" (Tiwana et al., 2010, p. 676), organizations need to facilitate increasingly distributed ways to create digital innovation (e.g., artificial intelligence, the internet of things, and cloud computing). These distributed sources of digital innovation are mainly associated with digital technologies that enable various actors to cooperate within innovation networks across organizational boundaries (Lyytinen et al., 2016), often beyond the control of the original innovator (Bogers and West, 2012), and regardless of industry boundaries. In addition, digital innovations shift the focus from single products toward connected products, which transcend established industrial boundaries (Porter and Heppelmann, 2014) and challenge the separation of firms along industrial areas.

Moreover, firms strive to embed digital technology in physical products ("smart products") (Yoo et al., 2012) and to provide more and more functions in the form of digital objects that can be easily changed and enhanced (Faulkner and Runde, 2019). Consequently, similar digital technology is used in different products, regardless of industry boundaries. In that regard, firms across industries build up similar technological competencies and rely on the same group of suppliers. In a similar vein, digital devices process any type of digital information (Tilson et al., 2010), which leads to the similarity of devices across industries. Eventually, smart products lead to the convergence of user experiences, as demonstrated by smartphones that bring together various communication, entertainment, and computation experiences (Yoo et al., 2012). For example, as already mentioned, Microsoft competes with established telecommunication firms since the acquisition of Skype (Yoo et al., 2010). Consequently, organizations from different industries compete with each other, while relying on similar digital technologies.

At the same time, the creation of digital innovation leads to a fundamental transformation in firms' IT, structure, and strategy (Tumbas et al., 2018; Vial, 2019). These transformation processes are driven by the new competitive situation that firms face, which causes them to shift their value proposition and value creation process (Vial, 2019; Wessel et al., 2021). Accordingly, firms' digitalization is characterized by socio-technical changes, ranging from major technological adaptations to shifts in firms' structure and processes. These changes affect all firms and may therefore impact the valuation of their stocks systematically. On the one hand, digitalization offers firms the potential to conquer new markets and discover new revenue streams. On the other hand, wide-ranging changes to a firm's structure and processes are also associated with risks that the invested resources cannot be reaped in the long run. Accordingly, digital innovation is not only limited to product innovation. It also encompasses radical business model innovation, on the one hand, and incremental process innovation within firms, on the other hand.

The disruptive and transformational nature of digital innovation can be traced back to distinct architectural characteristics of digital technology. Yoo et al. (2010) identify that digital technology embedded in digital innovation follows a layered modular architecture consisting of the four layers of devices, networks, services, and contents. These loosely coupled layers are product-agnostic (Yoo et al., 2010) and pave the way to be applied across diverse industries and markets. Yoo et al. (2010) discuss the four layers in detail. The device layer encompasses physical machinery, such as computer hardware, and control and maintenance capabilities, such as an operating system. For example, the physical smartphone and its operating system would be assigned to the device layer. The network layer contains physical transport mechanisms, such as cables and transmitters, and logical network protocols, such as TCP/IP. Staying with the smartphone example, 5 G connectivity and the respective communication protocols form examples of the network layer.

The content layer comprises data, such as videos, but also directory information on where to find which data, copyright, or encoding methods. Photos on a smartphone or movies on servers are examples of this layer. The service layer provides the functionality for users to deal with content, such as storing, creating, or retrieving it. The four layers are only loosely coupled, and each layer can be designed with low consideration of the other layers. This allows firms to offer products and services that concentrate on certain layers, cover several layers, or combine their offerings with other firms' offerings focusing on the same or other layers. Thus, digital innovation might be created within a certain layer or across layers. In sum, this architecture spurs an unprecedented level of generativity, where one digital innovation is used to create another one, which in turn enhances the first one, and so on (Yoo et al., 2010). This, in turn, leads to fluid product boundaries, which also enables serving different markets.

The described four layers of the modular layered architecture form the starting point for our analysis. In the next section, we create a measure of digital orientation that assesses firms' innovation activities across all four layers in order to derive an overall measure of firms' orientation toward digital innovation. We then use the derived measure of digital orientation to explore the link between firms' level of digital innovation and asset pricing (see Sections 4 and 5).

3. Data and methodology

We use firms' 10-K Management's Discussion and Analysis (MD&A) and an easy-to-use and replicable bag-of-words approach to compute firms' digital orientation (*DO*) for each firm-year. We then match this measure with data on all common stocks traded on the main stock exchanges (NYSE, AMEX, and NASDAQ) in the United States available in both CRSP and Compustat between 1996 and 2020. To benchmark our *DO* measure, we additionally collect portfolio holdings from Morningstar Direct for 26 funds with a digital innovation focus at the end of June each year. These funds include, for instance, the ARK Innovation ETF and L&G Artificial Intelligence ETF (for details of the funds, see Appendix D). We outline our approach in more detail in the following.

3.1. Digital orientation and text analysis of firms' 10-Ks

We rely on computer-aided text analysis to measure the extent to which firms create and use digital innovation by deriving a quantitative measure from firms' annual reports. We web-crawl the SEC Edgar database for all 10-K, 10K-405, and 10K-KSB filings, excluding amended documents. This results in a sample of 158,631 reports. Our sample is reduced to 91,151 reports when matching the reports with data from Compustat and CRSP (see Section 3.2.). Our data collection is restricted by the availability of the annual reports in the SEC Edgar database, since electronic filing was only mandated in 1996.

Following extant research (e.g., Loughran and McDonald, 2011), our baseline text analysis is constricted to the MD&A section since it is not audited and reflects the management's thoughts and opinions. We use regular expressions in Python to extract the relevant MD&A section and firm information, such as the central index key (CIK), the report type, fiscal year, and report date for each 10-K. To account for the different formats of the reports and to ensure that the right section of the report was covered for all firms in the sample, we iteratively revised our Python code until we were confident that the regular expressions we used could identify the MD&A section for all 10-Ks.

In the next step, we quantify the MD&A by using a dictionary approach to measure *DO*. Thus, we follow a bag-of-words approach to parse the 10-Ks and derive vectors of words and word counts. This approach has been used in the Finance discipline, for instance, to measure negative sentiment (Tetlock, 2007; Loughran and McDonald, 2011; Hillert et al., 2014; Loughran and McDonald, 2016; Hillert et al., 2024).

Since our approach relies on correctly identifying the MD&A section boundaries within 10-K filings, we acknowledge certain limitations of our parsing algorithm. In some instances, the algorithm may not precisely detect the start and end of the MD&A section, and occasionally encounters filings that contain only references to external PDF documents rather than embedded text. To assess whether these parsing errors affect our main results, we conduct robustness tests using various word count thresholds to exclude observations likely to have been incorrectly parsed, thereby demonstrating the robustness of our results (see Table 10).

Our dictionary is based on a word list developed and validated by Kindermann et al. (2021), who originally measured firms' digitalization in general across four dimensions: digital technology scope, digital capabilities, digital ecosystem coordination, and digital architecture configuration. We match each term in the word list to the layered modular architecture model by Yoo et al. (2010) and its four loosely coupled layers: device, network, service, and content layers. For this categorization of each term, we rely on the description of each layer, as outlined in Section 2. We also include synonyms and alternative spellings in our dictionary. Moreover, we add terms that are related to existing elements of the word list and that are significant in the context of digital innovation. For instance, we include the terms "internet protocol" and "local area network" on the network layer as they build essential components of the "internet". On the content layer, we complement the list with terms related to "data", such as "data analytics" and "metadata". Occupational titles or roles, such as "or "programmer" or "chief information officer" are excluded, as they cannot be clearly assigned to a distinct layer. Moreover, terms where no clear link to digital innovation exists, such as "ubiquitous" and "advanced technology" are excluded in our dictionary.

Table 11 depicts our dictionary of digital innovation with its four dimensions and the corresponding word list. The number of terms related to digital innovation in a firm's MD&A section reflects management's past actions as well as future intentions to use and develop digital innovation. Accordingly, we call the derived measure digital orientation or *DO*. Using this dictionary, we derive a measure of firms' digital orientation and explore whether convergence processes across industries occurring due to the rise of digital innovation can explain similarities across firms.

In order to derive a comparable digital orientation measure, we divide the number of terms counted by the overall lengths of the MD&A section without stop words (e.g., the, is, at, which). Thus, our measure of digital orientation (*DO* measure) is computed using the following formula:

$$DO = \frac{No.\ words\ digital\ innovation}{No.\ words\ in\ MD\&A\ section\ without\ stopwords} \tag{1}$$

Table 2 depicts the summary statistics for the *DO* measure. Across the 91,151 10-K reports analyzed in our final sample, the average value of the *DO* measure is 0.0024 (0.24 %). The device and content layer account for larger proportions of the total measure than the network and service layer. Across firm-years, we see a considerable variation, where firms with the highest *DO* value deviate significantly from the mean score across all firm-years. Further, many firm-years have a *DO* value of 0, leading to a right-skewed distribution of the measure. This is especially evident in the 25th percentile. Except for the aggregated measure of *DO* (Total), the 25th percentile only includes firms with a *DO* value of 0.

Fig. 1 illustrates the measure's development over time, both for the aggregated *DO* measure and the distinct layers. We observe a peak of the aggregated measure around 2000, which can be attributed to the dot-com bubble, excessive speculation with internet-related stocks that reached its peak in March 2000. After this peak, the value of the *DO* measure across firms decreases and then stabilizes at a high level. Starting around 2010, we can observe a slight increase in the *DO* measure that has been more pronounced in recent years. Fig. 1 also illustrates that this evolution of the aggregated *DO* measure can be attributed to different developments across the layers. The strong peak around 2000 is driven primarily by a sharp increase in the device and network layer. Thereafter, the device and network layer measures initially drop and then continue to decrease slightly over time. In contrast, the service and content layer measures are characterized by a steady increase over the observation period.

Table 3 presents the 30 largest firms in the S&P 500 from our sample, as of 2020, alongside data on their DO. We rank by market capitalization, with Apple exhibiting the highest one. For each firm, we present the percentile based on their DO, as well as their classification according to the Fama and French (1997) 48-industry system (FF48). We only included firms whose MD&A section in the annual report was at least 500 words long. Among the 30 largest firms in our sample, there are many with a high digital orientation that might be expected to be digitally leading, such as Microsoft, Amazon, Alphabet, Nvidia, Adobe, Netflix, Oracle, and Salesforce.com. In 2020, these firms were placed in the 90th percentile or higher based on their DO. Consequently, a concentration of digital leading firms is evident in the category "business services" of the FF48 classification. Conversely, the 30 largest companies include some unexpected digital leaders, such as UnitedHealth Group, AT&T, and Comcast.

Of the 30 largest firms in the S&P 500 in our sample, we only identified one digital laggard: Thermo Fischer Scientific, which focuses on measuring and control equipment. Other companies with a comparably low digital orientation include firms, such as Procter & Gamble, Johnson & Johnson, and Eli Lilly, suggesting a concentration of such firms in the pharmaceutical products industry.

Table 4 depicts the most frequent terms in firms' 10-Ks across time. Unsurprisingly, very generic terms, such as "software", "data", "internet", "computer", and "hardware", top the list. At the same time, we can also find more specific terms associated with *DO* in this list, such as "cloud", "analytics", and "broadband". Table 4 also illustrates the earlier observations with respect to the evolution of the *DO* measure across time. Terms belonging to the device layer, such as "software", "computer", and "hardware", were most frequent between 1996 and 2010; after that, their frequency decreased but remained at a high level. For the most frequent terms of the network layer, i.e., "internet", "online", and "wireless", the frequency peak primarily occurred slightly delayed between 2001 and 2010. Some terms belonging to the content and service layer, however, reached their frequency peak between 2001 and 2010 (e.g., "digital", "web", "database") and others between 2011 and 2020 (e.g., "data", "website", "e-commerce", "cloud").

Since generic terms, such as software, data, internet, computer, and hardware, occur with high frequency (see Table 4), we also calculate an alternative *DO* measure. This alternative measure accounts for the importance of a term within the entire sample by adjusting for the so-called term-frequency-inverse document frequency (tf-idf) based on Loughran and McDonald (2011). The weighted measure of each term $\mathrm{w_{ij}}$ is calculated using the following formula:

$$\mathrm{w_{ij}} = \left(1 + \log\left(tf_{i,j}\right)\right) * \log\left(\frac{N}{df_i}\right) \tag{2}$$

$tf_{i,j}$ is the raw count of the i^{th} term in the j^{th} annual report N measures the total number of annual reports and df_i represents the number of annual reports where the i^{th} term occurs at least once.

3.2. Financial data and analysis

We match each 10-K report and the derived *DO* measure to data of the CRSP and COMPUSTAT database using the CIK and report date. We perform multiple rounds of matching by first using the exact report date and then time windows around the report and filing date for the merge to Compustat/CRSP data. Since the database Edgar also included 10-Ks of firms only traded over-the-counter (OTC), firms temporarily not listed on a stock exchange, and asset-backed partnerships, which are required to file with the SEC but are not included in Compustat/CRSP (CRSP 2021), our initial sample is reduced during the matching process. We also filter out firm-years that did not belong to common stocks traded on the main stock exchanges (NYSE, AMEX, and NASDAQ) in the United States. Overall, the final sample constitutes 91,151 firm-years with 24 time periods and 11,613 unique firms. In order to prevent that outliers drive our results, we use winsorization by capping all variables derived from financial statements in the bottom and top one percentiles to the value of the variables at the 1st and 99th percentile.[2]

To study the determinants of *DO*, we run cross-sectional Fama/MacBeth regressions (Fama and MacBeth, 1973) using firm age, firm size, membership in the S&P 500 and Nasdaq, and analyst coverage as explanatory variables. The construction of these variables

[2] Variables derived from financial statements include explanatory variables, such as firm size, analyst coverage, and book-to-market. In contrast, stock returns and the digital orientation measure remain unaltered and are used in their original form.

Table 1
Digital innovation dictionary.

Device Layer	3-D printer, 3D printer, 3-D printers, 3D printers, computer, computers, compute, computing, control system, control systems, cybernetics, cyber physical system, cyber physical systems, desktop, desktops, digital device, digital devices, drone, drones, electronic, electronics, hardware, information system, information systems, information technology, information technologies, informatics, integrated solution, integrated solutions, IT infrastructure, IT infrastructures, IT solution, IT solutions, IT system, IT systems, operating system, operating systems, phone, phones, resource planning system, resource planning systems, robot, robots, sensor, sensors, smartphone, smartphones, tablet, tablets
Network Layer	bandwidth, bandwidths, bluetooth, broadband, connectivity, highspeed, high-speed, high speed, internet-based, internet, IP, internet protocol, LAN, local area network, mobile, network infrastructure, network infrastructures, network service, network services, network standard, N network standards, online, on-line, peer-to-peer protocol, P2P protocol, wifi, wi-fi, wireless
Service Layer	3-D printed, 3D printed, 3D printing, 3-D printing, additive manufacturing, algorithm, algorithmic, algorithms, AI, analytics, analytical tool, analytical tools, app, apps, app-based, API, APIs, application programming interface, application programming interfaces, artificial intelligence, autonomous, automated, automating, automation, blockchain, bot, bots, cloud, cloud-based, cloudbased, cyberspace, cyber space, cybersecurity, cyber security, digital platform, digital platforms, deep learning, ecommerce, e-commerce, fintech, homepage, homepages, home page, home pages, information security, insurtech, internet of things, internet-of-things, IoT, legaltech, machine learning, open source, open-source, robotics, robotic, SaaS, self-driving, smart, software as a service, software-as-a-service, suptech, technology platform, technology platforms, telematic, telematics, telemedicine, web, web-based, webs, website, websites
Contents Layer	big data, data, data analytics, data network, data networks, data service, data services, data transmission, data-dependent, data-driven, data-enabled, data-intensive, database, databases, digital, digitalization, digitalisation, digitally, digitization, digitisation, digitalized, digitalised, interface, GUI, graphical user interface, metadata, meta data, multichannel, multi-channel, omnichannel, omni-channel, real time, real-time, realtime, remote monitoring, social media, social technology, social technologies, streaming, user experience, UX, user interface, UI, virtual, virtualization, virtualisation, virtuality, virtualities, virtualize, virtualized

Table 2
Summary statistics of the digital orientation measure.

	Mean	Standard Deviation	Minimum	25th Percentile	Median	75th Percentile	Max
Total	0.0024	0.0042	0.0000	0.0002	0.0010	0.0028	0.0909
Device Layer	0.0008	0.0015	0.0000	0.0000	0.0002	0.0008	0.0407
Network Layer	0.0005	0.0017	0.0000	0.0000	0.0000	0.0002	0.0405
Service Layer	0.0003	0.0010	0.0000	0.0000	0.0000	0.0001	0.0313
Content Layer	0.0009	0.0028	0.0000	0.0000	0.0003	0.0009	0.0909

Table 2 shows the summary statistics of the digital orientation measure in total and separately for each layer. The digital orientation measure represents the frequency of terms in the digital innovation dictionary (see Table 1) relative to the overall number of words in the MD&A section of a firm's 10-K without stop words. The analysis is based on 91,151 10-Ks covering the timespan 1996 to 2020 included in the final sample of the analysis. Columns may not sum exactly because of rounding.

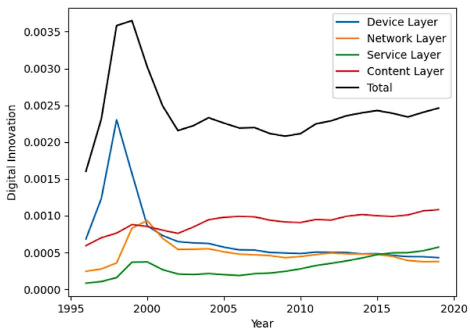

Fig. 1. Digital orientation over time – Total and layers.
This figure illustrates the measure of digital orientation in total and separately for each layer over time. The digital orientation measure represents the frequency of terms in the digital innovation dictionary (see Table 1) relative to the overall number of words in the MD&A section of a firm's 10-K without stop words. The analysis is based on 91,151 10-Ks covering the timespan 1996 to 2020 included in the final sample of the analysis.

follows earlier literature and is described in detail in Appendix A. Our baseline model uses the following regression formula:

$$\ln(1+\boldsymbol{DO}) = \boldsymbol{\alpha} + \boldsymbol{\beta}_1 \ln(\boldsymbol{age}) + \boldsymbol{\beta}_2 \ln(\boldsymbol{size}) + \boldsymbol{\beta}_3\ \boldsymbol{S\&P}\ 500 + \boldsymbol{\beta}_4\ \boldsymbol{NASDAQ} + \boldsymbol{\beta}_5 \ln(1+\ \boldsymbol{analyst}) + \epsilon_{\boldsymbol{DO}} \tag{3}$$

Table 5 depicts the results of the cross-sectional Fama-MacBeth regressions to explain firms' (natural logarithm of) *DO*. We start with univariate regressions that contain only one of the above-mentioned explanatory variables. As illustrated by specification (1), a firm's age offers explanatory power for a firm's *DO* by adding an adjusted R^2 of 2.27 %. With increasing age, the *DO* of a firm decreases. Whereas a firm's size, measured as the firm's natural logarithm of market capitalization, and the membership in the S&P 500 offer low explanatory power, Nasdaq membership is another important determinant of a firm's *DO* with a R^2 of 2.44 %, as the univariate specifications (2) to (4) illustrate. Analyst coverage offers no explanatory power for the *DO* (see specification (5)).

Table 3
Digital orientation of 30 biggest firms in the S&P 500 in 2020.

Name	Digital Orientation	Percentile based on Digital Orientation	FF48
Apple Inc	0.0037	77 %	Computers
Microsoft Corp	0.0116	95 %	Business Services
Amazon.com Inc	0.0085	93 %	Business Services
Alphabet Inc	0.0148	97 %	Business Services
Johnson & Johnson	0.0010	38 %	Pharmaceutical Products
Walmart Inc	0.0060	87 %	Retail
Visa Inc	0.0071	90 %	Business Services
Procter & Gamble Co	0.0008	30 %	Consumer Goods
Mastercard Inc	0.0032	73 %	Business Services
UnitedHealth Group Inc	0.0058	87 %	Insurance
The Home Depot Inc	0.0037	77 %	Retail
NVIDIA Corp	0.0133	96 %	Electronic Equipment
AT&T Inc	0.0203	99 %	Communication
Adobe Inc	0.0245	99 %	Business Services
Bank of America Corp	0.0035	76 %	Banking
The Walt Disney Co	0.0013	47 %	Business Services
Netflix Inc	0.0158	98 %	Entertainment
Cisco Systems Inc	0.0030	72 %	Electronic Equipment
Merck & Co Inc	0.0012	43 %	Pharmaceutical Products
PepsiCo Inc	0.0012	45 %	Candy & Soda
Comcast Corp	0.0139	97 %	Communication
Oracle Corp	0.0209	99 %	Business Services
Salesforce.com Inc	0.0135	96 %	Business Services
Abbott Laboratories	0,0020	62 %	Wholesale
Eli Lilly and Co	0,0011	40 %	Pharmaceutical Products
Thermo Fisher Scientific Inc	0,0004	19 %	Measuring and Control Equipment
Amgen Inc	0,0013	46 %	Pharmaceutical Products
T-Mobile US Inc	0,0056	87 %	Communication
Danaher Corp	0,0012	43 %	Wholesale
Nike Inc B	0,0048	83 %	Apparel

Table 3 shows the 30 largest firms in the S&P 500, alongside their *DO* data as of 2020. We also computed the percentile that each firm falls into based on its *DO*. Each firm is classified according to the Fama and French (1997) 48-industry classification system (FF48). Using our parsing algorithm, we ensured comprehensive coverage of the MD&A section by only considering firms whose annual reports consisted of a minimum of 500 words in this section. We consider firms in the 80th percentile or higher to be digital leaders, while firms in the 20th percentile or lower can be considered digital laggards.

Our baseline model is depicted in the specification (6). We find that the five independent variables can explain 4.30 % of *DO*'s variance, a relatively modest value.

As illustrated in specifications (7) and (8), the results are also robust to the inclusion of operating profitability (OP) and R&D intensity. For operating profitability, we observe a slightly negative link with digital orientation and a slight explanatory power in specification (7). R&D intensity has the expected positive and statistically significant coefficient.[3] With the inclusion of industry controls of the Fama and French (1997) 48 industry classification system (FF48) in the specification (9), we observe a significant increase in the proportion of correlation explained by the explanatory variables and controls to 22.14 %. The coefficients are similar in direction and slightly smaller in size and statistical significance, compared to the baseline regression without industry controls in specification (6).

Yet, even with the inclusion of industry controls, the regression results suggest that the selected firm characteristics can explain only a small proportion of *DO*'s variance. These findings demonstrate the benefits of inferring a firm's digital orientation directly from the MD&A section with the help of textual analysis, instead of using secondary variables, such as industry membership, that are only

[3] We do not include R&D intensity in the baseline regression, since data on R&D expenditure is only available for about half of firm-years in our sample. Especially, for firms in the service sector (e.g., banks and insurances), R&D expenditure is not reported. Estimating specification (8) based on a small R&D reporting subsample of companies is also likely to explain the sign reversal for the operating profitability coefficient.

Table 4
Thirty most frequent terms of the digital innovation dictionary in firms' 10-Ks.

Terms	Total	1996–2000	2001–2010	2011–2020	Layer
software	353,186	78,352	172,760	102,074	Device
data	322,501	34,316	138,779	149,406	Content
internet	79,136	16,737	47,738	14,661	Network
computer	76,546	35,888	29,603	11,055	Device
hardware	67,576	16,796	30,071	20,709	Device
digital	60,526	6044	27,286	27,196	Content
wireless	55,714	6212	34,244	15,258	Network
online	48,932	5066	22,337	21,529	Network
electronic	44,659	7666	22,749	14,244	Device
information technology	42,824	6836	18,357	17,631	Device
mobile	40,842	2180	15,409	23,253	Network
web	28,549	5731	17,242	5576	Service
electronics	27,999	3587	13,920	10,492	Device
information systems	23,364	8846	9629	4889	Device
broadband	20,933	1261	13,076	6596	Network
website	18,197	884	7596	9717	Service
e-commerce	16,449	1626	7051	7772	Service
computing	14,921	2149	6333	6439	Device
cloud	12,327	29	240	12,058	Service
database	11,884	2340	6879	2665	Content
computers	11,564	4378	5011	2175	Device
IP	11,450	715	6173	4562	Network
automation	11,243	1699	5383	4161	Service
websites	10,679	401	3795	6483	Service
automated	10,072	1996	4782	3294	Service
connectivity	8515	1051	3897	3567	Network
phone	8378	915	4485	2978	Device
analytics	8067	78	1517	6472	Content
smart	8065	609	2941	4515	Service
desktop	6506	1885	2783	1838	Device

Table 4 shows the 30 most frequently used terms of the digital innovation dictionary in the MD&A section of firms' 10Ks in total and separately for each layer. For each term, the absolute frequency in the sample, the frequency for the three time periods 1996–2000, 2001–2010, 2011–2020, and the respective layer in the layered architecture of digital innovation (see Section 2) are depicted. The analysis is based on 91,151 10-Ks covering the timespan 1996 to 2020 included in the final sample of the analysis.

weak proxies for *DO*.

3.3. Comparison with fund holdings

We next validate our *DO* measure by assessing whether our measure can explain a stock's inclusion in funds with a digital innovation focus like the ARK Innovation Fund (Ticker: ARKK) and a set of 26 additional exchange-traded funds (ETFs) designed to target firms leading in digital innovation or digital technology.[4] Further fund details can be found in Appendix D. Table 6 presents cross-sectional Fama-MacBeth regressions that estimate the likelihood of a firm being included in such innovation-focused investment funds. In addition to the *DO* measure, we include traditional firm characteristics that may help explain fund membership. In regression (1), the dependent variable indicates membership in the ARK Innovation Fund. To avoid forward-looking bias, a stock is only considered part of the fund from the year it was first included in the fund's holdings. Our results show that the *DO* measure is a positive and statistically significant explanatory factor for fund inclusion (t-statistic: 6.89). However, the significance of other firm-level characteristics, such as firm size, analyst coverage, and Nasdaq membership, suggests that digital orientation is only one of several factors influencing fund membership.

In regressions (2) and (3), the dependent variable indicates membership in a peer group of exchange-traded funds (ETFs) focused on digital innovation, comparable to the ARK Innovation Fund. The variable takes the value 1 if a stock is included in any of these ETFs. The results are broadly consistent with specification (1), although some variation is observed in the set of statistically significant firm

[4] The following investment funds are included as being focused on digital innovation: ARK Innovation ETF, ARK Next Generation Internet ETF, ARK Fintech Innovation ETF, ARK Autonomous Technology & Robotics ETF, Defiance Quantum ETF, First Trust Cloud Computing ETF, Invesco AI and Next Gen Software ETF, iShares Blockchain and Tech ETF, iShares Future Cloud 5G and Tech ETF, L&G Artificial Intelligence ETF, SPDR® FactSet Innovative Technology ETF, WisdomTree Cloud Computing ETF.For funds focused on information technology, we additionally consider the following investments funds: Fidelity MSCI Information Tech ETF, First Trust NASDAQ-100-Tech Sector ETF, First Trust Dow Jones Internet ETF, Invesco Dorsey Wright Technology MomtETF, Invesco QQQ Trust, Invesco S&P SmallCap Info Tech ETF, Invesco S&P 500® Equal Weight Tech ETF, iShares Expanded Tech-Software Sect ETF, iShares Expanded Tech Sector ETF, iShares U.S. Tech Independence Fcs ETF, iShares US Technology ETF, SPDR® S&P Semiconductor ETF, SPDR® S&P Software & Services ETF, Vanguard Information Technology ETF, The Technology Select Sector SPDR® ETF. Please see Appendix D for details.

Table 5
Multivariate regression to explain digital orientation.

Variables	Firm Age (1)	Firm Size (2)	S&P 500 (3)	Nasdaq (4)	Analyst (5)	Baseline (6)	Baseline + OP (7)	Baseline + OP & R&D (8)	Baseline + Industry Controls (9)
Constant	0.0046***	0.0032***	0.0029***	0.0020***	0.0026***	0.0035***	0.0033***	0.0048***	0.0008***
	(8.84)	(8.85)	(14.00)	(17.65)	(10.30)	(7.68)	(9.69)	(9.09)	(3.46)
Firm Age	−0.0007***					−0.0005***	−0.0005***	−0.0009***	−0.0002***
	(−4.24)					(−4.82)	(−5.87)	(−8.34)	(−4.78)
Firm Size		−0.0001*				−0.0001**	−0.0001***	−0.0001*	−0.0000
		(−1.80)				(2.51)	(−2.62)	(−1.71)	(−0.45)
S&P 500			−0.0005*			0.0003*	0.0002**	0.0004**	0.0001*
			(−2.16)			(1.94)	(1.97)	(2.52)	(0.98)
Nasdaq				0.0014***		0.0013***	0.0013***	0.0016***	0.0009***
				(6.05)		(13.21)	(16.22)	(11.36)	(10.38)
Analyst					0.0001***	0.0005***	0.0006***	0.0003***	0.0002***
					(2.69)	(6.83)	(7.59)	(3.94)	(6.24)
Operating Profitability (OP)							−0.0004	0.0008***	−0.0003*
							(−1.44)	(3.17)	(−1.81)
R&D Intensity (R&D)								0.0031***	
								(6.37)	
No. of Obs.	91,096	90,499	91,096	91,096	91,096	90,449	86,852	32,110	86,852
R-squared	0.0227	0.0014	0.0013	0.0244	0.0000	0.0430	0.0451	0.0747	0.2214

Table 5 shows the results Fama-MacBeth regressions to explain digital orientation, measured as ln (1+ (no. terms digital innovation/no. words MD&A section without stop words)). Regression (6) shows our baseline model, which we used to derive the residual digital orientation measure. *Firm age* is the natural log of a firm's age in years. *Firm size* is defined as the natural log of market capitalization. *S&P 500* and Nasdaq are dummy variables indicating the membership of the firm in the S&P 500 index or its listing on the Nasdaq. *Analyst* is defined as the natural log of (1+no. earnings estimates) and describes a firm's coverage by analysts. *Operating profitability (OP)* is based on Fama and French (2015) and defined by the annual revenues minus the cost of goods sold, interest expenses, selling, general, and administrative expenses divided by the book equity. *R&D intensity (R&D)* is a firm's R&D expenditure in relation to its assets. In regression (9) we further integrate industry controls using the Fama and French (1997) 48 industry classification system. The sample period is 1996–2020. Following Fama and Macbeth (1973), coefficients are calculated as time-series averages of yearly estimates. t-statistics (shown in parentheses) are based on time-series average coefficients and standard deviation, and adjusted for serial autocorrelation using Newy and West (1994) standard errors with a lag of two years, based on the formula $\lfloor [4 * (0.01 * T)^{\frac{2}{9}}] \rfloor$ with $T = 24$. The significance level is indicated as follows: * significant at the 10 % level; ** significant at the 5 % level; and *** significant at the 1 % level.

Table 6
Explanatory factors for stocks' membership in digital innovation funds.

Variables	(1) ARK Innovation Fund Membership	(2) Digital Innovation Fund Membership	(3) Digital Innovation Fund Membership (no DO)	(4) IT Fund Membership
Constant	−0.0297***	−0.0176	0.0193	0.0704***
	(−7.20)	(−1.50)	(1.33)	(7.18)
Digital Orientation	1.4972***	5.0309***		9.2609***
	(6.89)	(7.44)		(12.00)
Firm Age	−0.0081***	−0.0020	−0.0026	−0.0118***
	(−8.29)	(−1.16)	(−1.49)	(−5.16)
Firm Size	0.0077***	0.0138***	0.0138***	0.0449***
	(12.31)	(7.62)	(7.36)	(42.29)
S&P 500	0.0217***	0,0002	−0.0008	0.0075
	(7.28)	(0.10)	(−0.49)	(1.16)
Nasdaq	0.0183***	0.0235***	0.0265***	0.1143***
	(9.96)	(6.70)	(7.16)	(26.46)
Analyst	0.0040***	0.0008	0.0021	0.0171***
	(5.75)	(0.54)	(1.39)	(11.84)
Industry Dummies	Yes	Yes	Yes	Yes
No. of Obs.	16,418	48,857	49,501	48,857
R-squared	0.0664	0.1223	0.1057	0.3198

Table 6 presents the results of Fama–MacBeth regressions aimed at explaining membership in funds focused on digital innovation and information technology. Regression (1) uses membership in the ARK Innovation Fund as the dependent variable. Regressions (2) and (3) use a dependent variable that captures membership in a peer group of exchange-traded funds (ETFs) with a focus on digital innovation comparable to the ARK Innovation Fund. Building on these, regression (4) extends the analysis by considering membership in a broader set of ETFs focused on the information technology sector more generally. *Firm age* is the natural log of a firm's age in years. *Firm size* is defined as the natural log of market capitalization. *S&P 500* and Nasdaq are dummy variables indicating the membership of the firm in the S&P 500 index or its listing on the Nasdaq. *Analyst* is defined as the natural log of (1+no. earnings estimates) and describes a firm's coverage by analysts. The sample period varies based on the inception dates of the respective investment funds: it spans 2015–2020 for regression (1), and 2006–2020 for all other regressions. Following Fama and Macbeth (1973), coefficients are calculated as time-series averages of yearly estimates. t-statistics (shown in parentheses) are based on time-series average coefficients and standard deviation, and adjusted for serial autocorrelation using Newy and West (1994) standard errors with a lag of two years, based on the formula $\lfloor [4 * (0.01 * T)^{\frac{2}{9}}] \rfloor$. The significance level is indicated as follows: * significant at the 10 % level; ** significant at the 5 % level; and *** significant at the 1 % level.

characteristics. Comparing specifications (2) and (3)—where specification (3) excludes the *DO* measure—reveals only a slight difference in the R-squared values. Additionally, we find only a moderate correlation (0.2776) between the fitted values from regression (3) and the *DO* measure, indicating that digital orientation reflects unique information not fully captured by conventional firm characteristics. Overall, we therefore conclude that the *DO* measure captures a distinct dimension of firms' engagement with and development of digital technologies, separate from traditional firm fundamentals.

Finally, regression (4) uses the membership in a broader group of exchange-traded funds that focus not only on digital innovation but on information technology (IT) more generally as the dependent variable. The results are consistent with the previous specifications and further reinforce our earlier interpretation. Notably, the coefficient for the *DO* measure remains positive, statistically significant (t-statistic: 12.00), and is larger in magnitude compared to prior models. Likewise, the coefficients of other statistically significant firm characteristics also increase, suggesting stronger explanatory power in this broader context. This may be partly attributable to greater variation in the sample, given the inclusion of a larger and more diverse set of funds.

4. Results

4.1. Digital leaders vs. digital laggards

We next seek to understand how digital leaders differ from digital laggards. We use two approaches to identify digital orientations' effect on key firm characteristics. First, we use the raw data of firms' aggregate *DO* measure when sorting stocks into quintiles. Since a large proportion of firm-years has a digital orientation value of zero, the first quintile containing these firm-years is larger than the remaining four quintiles. Second, we derive a residual *DO* measure by following Hillert et al. (2014) and Hong et al. (2000). Thus, we use regression specification (6) in Table 5, where we control for the firm characteristics, firm age, firm size, S&P 500, Nasdaq membership, and analyst coverage, to obtain the residual *DO* measure. This second approach might be perceived as advantageous because it allows us to eliminate possible dependencies between digital orientation and firm characteristics by following a simple methodology. We note, however, that one should expect more than modest differences in the results, as the low explanatory power indicates a high correlation between the raw *DO* and residual *DO* measure. The variables used in the quintile analysis are described in detail in Appendix A. Table 7 illustrates the distribution of firm characteristics after sorting stocks into quintiles based on the raw *DO*

Table 7
Firm characteristics for portfolios sorted by raw and residual digital orientation measure.

Panel A: Quintile sorts based on the raw digital orientation measure											
Variables	Book-to-market	Long-term Forecast	Stock Volatility	Market Beta	ROA	ROE	Short-term Forecast	Operating Profitability	COGS/ Assets	XOPR/ Assets	Sales Growth
Q1	0.71	14.41	0.12	0.91	0.03	0.02	0.04	0.18	1.99	2.58	0.15
Q2	0.76	14.47	0.14	1.00	0.01	−0.04	0.01	0.12	2.14	2.66	0.18
Q3	0.72	15.54	0.14	1.01	0.01	−0.05	0.01	0.10	1.98	2.55	0.19
Q4	0.66	17.13	0.16	1.11	0.01	−0.06	0.00	0.09	1.71	2.35	0.20
Q5	0.56	19.93	0.17	1.23	0.00	−0.12	−0.01	0.05	1.20	1.97	0.24
Q5-Q1	−0.15***	5.53***	0.05***	0.32***	−0.03***	−0.14***	−0.05***	−0.14***	−0.79***	−0.60***	0.09***
t-stat Q5-Q1	−25.07	146.68	217.54	149.68	−20.52	−24.26	−94.20	−31.25	−50.05	−32.96	17.47
Panel B: Quintile sorts based on the residual digital orientation measure											
Variables	Book-to- market	Long-term Forecast	Stock Volatility	Market Beta	ROA	ROE	Short-term Forecast	Operating Profitability	COGS/ Assets	XOPR/ Assets	Sales Growth
Q1	0.72	17.48	0.15	0.97	0.00	−0.07	−0.01	0.06	1.84	2.41	0.23
Q2	0.72	15.27	0.14	0.99	0.02	−0.03	0.02	0.13	1.96	2.53	0.17
Q3	0.72	14.43	0.13	1.01	0.02	0.00	0.03	0.16	2.02	2.57	0.15
Q4	0.71	15.84	0.14	1.06	0.02	−0.01	0.02	0.15	1.88	2.46	0.15
Q5	0.58	20.31	0.17	1.21	0.00	−0.09	−0.01	0.06	1.25	2.00	0.23
Q5-Q1	−0.13***	2.83***	0.03***	0.24***	0.01***	−0.02***	0.00	0.00	−0.60***	−0.41***	−0.01
t-stat Q5-Q1	−18.60	17.75	28.87	27.73	3.55	−3.09	0.71	−0.10	−28.09	−17.16	−0.91

Table 7 shows time-series averages of firm variables' yearly mean for quintile portfolios based on the raw digital orientation measure (panel A) and the residual digital orientation measure (panel B). The raw digital orientation measure is defined as no. terms digital innovation/no. words MD&A section without stop words. The residual digital orientation measure is derived from Fama-MacBeth regressions with firm age, firm size, S&P500, Nasdaq and analyst coverage as explanatory variables (see Table 5) using time-series averages of yearly estimates. *Book-to-Market* is the book value of equity divided by market value of equity following (Greenwood and Hanson, 2012). *Long-term forecast* is measured as the average of analysts' forecast of the expected long-run earnings growth rate per share over the next five years following La Porta (1996). *Stock volatility* is measured using the monthly return over the previous 60 months. *Markt beta* denotes the market factor beta from the 3-factor model, measured over the last 60 months. *Return on assets* (ROA) is calculated as the ratio between income before extraordinary items and assets. *Return on equity* (ROE) is calculated following Greenwood and Hanson (2012). *Short-term forecast* is measured as the average of analysts' earnings forecast for the fiscal year annual earnings scaled by the price per share. *Cost ratios* are defined as cost of goods sold (cogs) over assets and operating expenses (xopr) over assets. *Sales growth* is measured as the log change in sales over one year. *Operating profitability* is based on Fama and French (2015) and defined by the annual revenues minus the cost of goods sold, interest expenses, selling, general, and administrative expenses divided by the book equity. The sample period is 1996–2020. The final rows in each panel display the difference between the fifth and first quantile as well as the t-statistic associated with this portfolio difference. The significance level is indicated as follows: * significant at the 10 % level; ** significant at the 5 % level; and *** significant at the 1 % level.

measure (Panel A) and the residual *DO* measure (Panel B).

The comparison of *DO* portfolios sorted by raw values shows notable and highly significant differences in firm characteristics. Digital leaders, i.e., firms in the high *DO* portfolio, are characterized by a lower book-to-market ratio. This finding is in line with Chen and Srinivasan (2023)'s study on nontech firms' adoption of digital technologies. Furthermore, digital leaders are estimated to have higher profitability in the long run, as the measure of long-term earnings forecasts shows. This finding corresponds to the notion that the value of digital innovation unfolds over time, because it depends on particular use contexts and how digital innovation can be combined with other innovation (Henfridsson et al., 2018). Developing the respective capabilities to create digital innovation needs several years and may involve the creation of networks transcending the traditional organizational structure (Vial, 2019). In a similar vein, Svahn et al. (2017) report that building new capabilities requires breaking away from established practices and creating new capabilities, which includes rethinking existing norms. Building new capabilities requires a considerable amount of time, but also fosters firms' intangible value and long-term performance (Bharadwaj et al., 1999).

At the same time, developing capabilities bears risks as the value of digital innovation is not pre-determined but depends on, first, the design choices regarding the combination of technologies and the affordances thereof, and, second, on the use context (Henfridsson et al., 2018). In addition, "(a)s physical resources - including products - become comparatively less relevant than services, as consumers contribute to influence trends related to the use of digital technologies, and as value networks become broader and more complex, firms experience higher levels of uncertainty" (Vial, 2019, p. 133). This reasoning corresponds with our finding that digital leaders are also characterized by significantly higher volatility, measured as the monthly return over the previous 60 months, and higher market beta, calculated as the market factor beta from the 3-factor model, measured over the last 60 months. The differences in market beta are quite pronounced. Firms in quintile 1 have an average market beta of 0.91, whereas firms in quintile 5 exhibit an average market beta of 1.23. Digital leaders also exhibit lower short-term operational returns, as illustrated by lower ROA, ROE, short-term forecasts, and operating profitability. For instance, for operating profitability, we observe that firms in the low *DO* portfolio perform more than three times as well, compared to firms in the high *DO* portfolio.

While developing capabilities for digital innovation and creating digital innovation impact the development of a firm's intangible value and future prospects, they are subject to uncertainty, as discussed above, and, in the short term, require considerable investment and management attention. Yet, accounting measures do not consider "time lags necessary for realizing the potential of capital investments" (Bharadwaj et al., 1999, p. 1009). Accordingly, we find lower values for digital leaders regarding short-term profitability and operating profitability.

In summary, the development of digital innovations, in particular radical business model or product innovations, is associated on the one hand with significant risks, as shown by the higher volatility of digital market leaders, and on the other hand with high short-term investments that lead to lower short-term returns. In the long term, however, the risks taken for the development of digital innovations may pay off, enabling digital market leaders to generate higher operational profitability in the long term.

Furthermore, firms placed in the high *DO* portfolio also have notably smaller cost ratios (COGS/assets) than digitally lagging firms. In that regard, studies show that firms increase their operational efficiency and cost positions by employing digital technologies to automate and improve business processes, by using cloud computing to provide on-demand services instead of owning data centers with high fixed costs, and by applying big data and analytics to speed up decision-making processes (Vial, 2019).

Moreover, sales growth is almost double in size for firms placed in the high *DO* portfolio based on the raw measure. Extant research shows that the growth of entrepreneurial firms is nonlinear and similar to digitally oriented firms that use digital technologies to shape entrepreneurial processes and outcomes (Vial, 2019). Namely, developing digital façades – the digital coupling of firms' activities with customers and partners as a first step towards building up digital capabilities – enables growth (Tumbas et al., 2015) by attracting new customers.

In untabulated results, we find that the differences between digitally leading firms and digitally lagging firms are equally evident when sorting stocks into quintiles based on the alternative measure tf-idf *DO*. The analysis shows, for instance, a significantly lower book-to-market ratio, higher long-term profitability, higher volatility, and lower short-term profitability for digital leaders. As expected, for portfolios sorted by residual values, Panel B mostly echoes these results. While some differences in firm characteristics between digital leaders and digitally lagging firms disappear (i.e., sales growth), most of the described differences in firm characteristics for firms in the extreme portfolios 1 and 5 remain after controlling for firm size, age, analyst coverage, S&P 500, and Nasdaq membership. Finally, in untabulated results, we find that an analysis based on residual values of the tf-idf *DO* measure also shows similar results.

4.2. Portfolio tests based on the level of digital orientation

The previous section highlights systematic differences between firms with high and low values for the text-based *DO* measure. In this section, we analyze whether firms with high and low *DO* levels also have different stock returns. At the end of June of every year, we group stocks into quintile portfolios based on their *DO* value. The measure of *DO* is extracted from the firms' most recent fiscal year financial statements ending in the previous calendar year. For example, our sorts for June 2009 rely on data from financial statements ending in the calendar year 2008 (in most cases December 2008).

Because a large proportion of stocks has a *DO* measure of zero, we continue to group these stocks in the first quintile. The remaining stocks are then equally spread across the other groups according to their *DO* value. Our sorts are based on the absolute *DO* value instead of the residual value from the regression Eq. (1), because other firm variables have low explanatory power to explain *DO* (see Section 3.2).

We calculate equally weighted and value-weighted quintile (5) minus quintile (1) long-short portfolio returns from July 1st until

June 30th of next year, when the portfolio is rebalanced.[5] The transaction costs associated with the *DO* strategy are supposedly minimal because the annual rebalancing frequency implies a comparatively low portfolio turnover.

Low limits-to-arbitrage are even more emphasized for the second weighting scheme based on market values. Here, we use a capped value-weighting approach that winsorizes all market capitalizations at the 80th percentile of the stock universe before calculating the stock weights. The capped value weighting is advocated by Jensen et al. (2021) because it avoids a dominating influence of a few mega caps on return calculations, like the FAANG stocks (Facebook, Apple, Amazon, Netflix, Alphabet), and hence produces more robust factors. The benefits of a capped weighting scheme seem particularly relevant in our context, as many of the dominating mega-cap stocks at the end of the sample period also tend to have a high *DO*.[6]

The long-short portfolio returns are sequentially regressed on a set of commonly used factors (e.g., Fama and French, 1993, 2015) to infer the abnormal, i.e., unexplained, return of the *DO* portfolio (also referred to as the digital premium or digital leaders' premium). In Table 7, we observe that digital orientation is meaningfully associated with several firm characteristics, such as operating profitability from Fama and French (2015) and the book-to-market equity ratio. Controlling for known equity factors like value and profitability is therefore important in order to isolate the true effect of digital orientation on stock returns. Moreover, Table 7 also shows that high *DO* firms exhibit substantially higher market betas. While this relation is not surprising, it is well established that firms with high market betas tend to earn disproportionately low stock returns. In one regression specification, we therefore also control for this effect, commonly referred to as the low-risk or betting-against-beta anomaly, by additionally including the betting-against-beta factor (BAB) from Frazzini and Pedersen (2014) as an explanatory variable. Results for equally weighted (value-weighted) portfolio schemes are shown in Table 8, Panel A (Panel B).

In Panel A of Table 8, we observe statistically insignificant raw returns of 0.2 % per month for the equally weighted long-short strategy in column (1). CAPM alphas and Fama and French three-factor alphas are also statistically insignificant, as seen in columns (2) and (3). However, as can be seen in column (4) of Panel A, the five-factor alpha of a regression that controls for the market, size, value, profitability, and investment factors is 0.32 % per month (approximately 3.8 % annualized) and statistically significant with a t-statistic of 1.78. The six-factor alpha, which additionally controls for stock momentum and is reported in column (5), is slightly larger, with 0.41 % per month and statistically significant at the 5 %-level with a t-statistic of 2.43. When we also control for the betting-against-beta factor based on Frazzini and Pedersen (2014), as shown in column (6) of Panel A, the monthly alpha increases to 0.50 and becomes highly statistically significant at the 1 %-level (t-statistic: 3.04).

The pronounced increase in the economic and statistical significance of the alpha in columns (4) to (6) is largely explained by a significantly negative correlation with the profitability factor RMW, the momentum factor UMD, and the betting-against-beta factor BAB. We also observe a negative correlation with the value factor HML. The correlations line up with findings from Section 4.1, which shows that firms with a high *DO* tend to be growth firms with lower profitability but high volatility and higher market beta. After accounting for these relations, the results in Panel A show that digital leaders earn about 6 % higher returns annually than digital laggards. This return differential may be interpreted as compensation for bearing systematic risk associated with digital transformation that is not captured by conventional asset pricing factors.

The same conclusion can be largely drawn from Panel B of Table 8, which shows corresponding regression results for the capped value-weighted *DO* strategy. For columns (1) to (3), we again observe statistically insignificant value-weighted raw returns and alphas. These regressions suffer from omitted variable bias because the negative correlations between the *DO* factor with respect to RMW, UMD, and BAB are not considered. When doing so, we observe a statistically significant Fama and French (2015) five-factor alpha of 0.24 % per month (t-statistic: 2.02) in column (4). The alpha of a six-factor regression, which additionally controls for the momentum factor, is 0.28 % per month (t-statistic: 2.41). When additionally including the betting against beta factor, the alpha increases to 0.34 % per month or around 4.1 % annualized. This spread is statistically significant at the 1 %-level with a t-statistic of 2.99.

Our finding of slightly diminished effects in value-weighted returns, as opposed to equally weighted returns, may suggest a greater emphasis on risk compensation among smaller firms. This aligns with the observations made by Grammig and Jank (2016), who find that smaller firms might be compelled to offer elevated expected returns to offset the risk associated with technological disruption from innovation, whereas larger firms possess a greater capacity to shield themselves from the impacts of such creative destruction.

In Fig. 2a and b, we further illustrate the cumulative abnormal return of the digital orientation for different factor asset pricing models across time, both for equally weighted (left) and value-weighted portfolios (right). Thus, we use the alphas depicted in Table 8 to illustrate the cumulative wealth effects across time in Fig. 2a and b. We measure the wealth effect as the cumulative abnormal return on a long-short investment of $1 over time. Consistent with the results in Table 8, the figures show that the strongest wealth effects of about 3.5 can be observed after controlling for profitability, momentum, and betting against beta factors. We observe similar effects for both the equally weighted (Fig. 2a) and the value-weighted portfolio (Fig. 2b).

The results point to a practical challenge in implementing a strategy that invests in high-*DO* firms. Without appropriate hedging against well-known asset pricing factors, such a strategy would yield only minimal excess returns relative to firms with low digital orientation. The reason is that investors in high-*DO* firms tend to obtain negative exposures to factors that have historically delivered high returns. In other words, investing in high-*DO* firms often coincides with holding stocks characterized by low profitability, negative momentum, and high systematic market risk. The regressions in Table 8, as well as Fig. 2a and b, show that sophisticated investors who

[5] Because we use financial statements for fiscal years from 1996 onwards, portfolio returns are calculated from July 1997 onwards. Factor data is obtained from Kenneth French's website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/) as of June 2020, which yields a total portfolio return sample period of 275 months from July 1997 to May 2020.

[6] For example, all of the five mentioned stocks (Facebook, Apple, Amazon, Netflix, Alphabet) belong to the highest *DO* quintile in 2020.

Table 8
Abnormal returns of the digital orientation (*DO*) portfolio.

Panel A: Equally weighted portfolios

	No factor (1)	CAPM (2)	FF 3-Factor (3)	FF 5-Factor (4)	FF 6-Factor (5)	FF 6-Factor+ BAB (6)
Alpha	0.0020	−0.0007	−0.0010	0.0032*	0.0041**	0.0050***
	(0.68)	(−0.28)	(−0.50)	(1.78)	(2.43)	(3.04)
mktrf		0.4780***	0.3619***	0.1774***	0.1078**	0.1055**
		(8.24)	(8.11)	(3.86)	(2.45)	(2.50)
smb			0.4525***	0.2019***	0.2535***	0.2728***
			(7.03)	(3.21)	(4.29)	(4.81)
hml			−0.8042***	−0.4523***	−0.5965***	−0.5128***
			(−13.30)	(−5.97)	(−8.08)	(−7.05)
rmw				−0.7868***	−0.7275***	−0.5660***
				(−9.27)	(−9.16)	(−6.84)
cma				−0.2028*	−0.1371	−0.1243
				(−1.84)	(−1.33)	(−1.26)
umd					−0.2254***	−0.1682***
					(−6.58)	(−4.83)
bab						−0.2472***
						(−4.97)
N	275	275	275	275	275	275
R-squared	0.000	0.199	0.563	0.669	0.715	0.740

Panel B: Value-weighted portfolios

	No factor (1)	CAPM (2)	FF 3-Factor (3)	FF 5-Factor (4)	FF 6-Factor (5)	FF 6-Factor+ BAB (6)
Alpha	0.0006	−0.0019	−0.0021	0.0024**	0.0028**	0.0034***
	(0.23)	(−0.80)	(−1.39)	(2.02)	(2.41)	(2.99)
mktrf		0.4297***	0.3263***	0.1283***	0.0978***	0.0962***
		(8.50)	(9.65)	(4.29)	(3.28)	(3.39)
smb			0.3758***	0.1274***	0.1501***	0.1644***
			(7.70)	(3.11)	(3.74)	(4.31)
hml			−0.8277***	−0.4413***	−0.5045***	−0.4420***
			(−18.06)	(−8.94)	(−10.07)	(−9.04)
rmw				−0.7939***	−0.7679***	−0.6472***
				(−14.37)	(−14.24)	(−11.63)
cma				−0.2847***	−0.2559***	−0.2464***
				(−3.96)	(−3.65)	(−3.71)
umd					−0.0988***	−0.0560**
					(−4.24)	(−2.39)
bab						−0.1846***
						(−5.52)
N	275	275	275	275	275	275
R-squared	0.000	0.209	0.673	0.818	0.829	0.843

Table 8 shows the results of regressions to investigate the returns of equally weighted (Panel A) and capped value-weighted (Panel B) quintile 5-minus-1 portfolios based on our digital orientation (DO) measure. Column (1) shows the raw returns per month of this *DO* factor. The remaining columns show the results of regressing the monthly long–short *DO* portfolio returns on the excess market return (MKTRF); on the Fama and French (1993) size and value factors (SMB, HML); on the Fama and French (2015) profitability and investment factors (RMW, CMA); the momentum factor from Kenneth French (UMD), and the betting against beta factor (BAB) from Frazzini and Pedersen (2014); t-statistics are reported in parentheses. *, ** and *** indicates significance at the 10 %, 5 %, and 1 % levels, respectively.

are able to neutralize these negative exposures through appropriate hedging positions can generate highly attractive alphas. Nevertheless, as we demonstrate in Section 4.4, even for investors that are not as sophisticated as hedge funds, digital orientation still proves to be a useful firm characteristic in constructing portfolios, specifically to enhance the risk-adjusted performance of established equity strategies. This advantage can be mainly explained by the diversification benefits provided by the additional inclusion of the DO factor in a portfolio context.

4.3. Robustness tests

We conduct several robustness tests to demonstrate the reliability of our results when measuring digital orientation using other parts of the firms' 10-Ks, controlling for firms' industry affiliation or Nasdaq exposure, using alternative data samples, and different dictionaries. All robustness tests are summarized in Tables 9, 10, and 11. In Table 9 we compare the results across all layers to the individual layers of digital innovation. Thus, column (1) in Panels A and B depict monthly equally weighted and capped value-weighted six-factor alphas across all layers, respectively. Columns (2) to (5) in Table 9 show the six-factor alphas for the four layers where firms may use or create digital innovation. In Tables 10 and 11, we use modified and alternative *DO* measures, additional controls, and alternative samples to demonstrate the robustness of our findings. Here, columns (1) and (2) depict monthly equally weighted and capped value-weighted six-factor alphas, which control for exposure to the excess market return (MKTRF), the Fama and

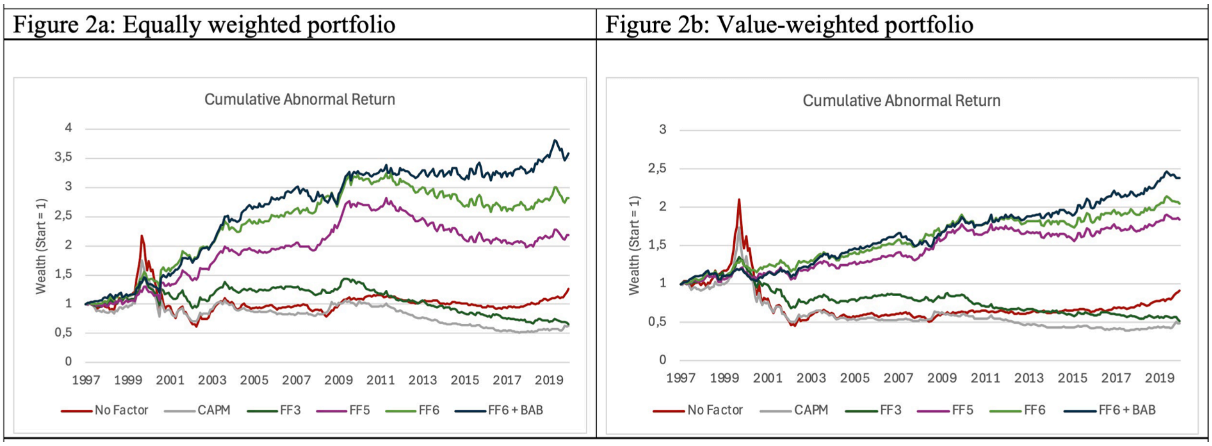

Fig. 2a and 2b. Cumulative abnormal return of the DO for different factor models.
Fig. 2a and b illustrate the cumulative abnormal return of the digital orientation for different factor models across time. Both figures use the results of regressions shown in Table 8 to investigate the profitability of equally weighted (left) and capped value-weighted (right) quintile 5-minus-1 portfolios based on the dictionary-based digital orientation measure (Do). The data series *No Factor* shows the raw returns per month of this *DO* factor. The data series *CAPM* accounts for the excess market return (MKTRF); while the Fama-French three-factor model *FF3* also integrates the Fama and French (1993) size and value factors (SMB, HML). The Fama-French five-factor model *FF5* further integrates the Fama and French (2015) profitability and investment factors (RMW, CMA). *FF6* and *FF6+ BAB* extend the Fama-French three factor model by accounting for the momentum factor and the betting against beta factor (BAB) from Frazzini and Pedersen (Frazzini and Pedersen, 2014). For the factor models, the monthly abnormal return is calculated as the constant (alpha) of the regression model plus the regression residual.

Table 9
Abnormal returns for different layer-specific digital orientation (*DO*) portfolios.

Panel A: Equally weighted portfolios					
	All layers (1)	Device layer (2)	Network layer (3)	Service layer (4)	Content layer (5)
Alpha	0.0041**	0.0050***	0.0056***	0.0059**	0.0031**
	(2.43)	(3.54)	(2.89)	(2.37)	(2.28)
N	275	275	275	275	275
R-squared	0.715	0.632	0.604	0.618	0.694
Panel B: Value-weighted portfolios					
	All layers (1)	Device layer (2)	Network layer (3)	Service layer (4)	Content layer (5)
Alpha	0.0028**	0.0035***	0.0051***	0.0041**	0.0020**
	(2.41)	(2.90)	(3.39)	(2.17)	(2.13)
N	275	275	275	275	275
R-squared	0.829	0.722	0.747	0.728	0.837

Table 9 shows the results of regressions to investigate the performance of equally weighted (Panel A) and capped value-weighted (Panel B) quintile 5-minus-1 portfolios based on digital orientation measure. Whereas for column (1) portfolios were sorted based on the aggregate DO measure, columns (2) to (5) use portfolio sorts based on layer-specific DO measures. We measure performance by regressing the monthly long–short *DO* portfolio returns on the excess market return (MKTRF); on the Fama and French (1993) size and value factors (SMB, HML); on the Fama and French (2015) profitability and investment factors (RMW, CMA); and on the momentum factor from Kenneth French (UMD); t-statistics are reported in parentheses. *, ** and *** indicates significance at the 10 %, 5 %, and 1 % levels, respectively.

French (1993) size and value factors (SMB, HML), the Fama and French (2015) profitability and investment factors (RMW, CMA), and the momentum factor from Kenneth French (UMD).[7] We outline our findings in the following.

First, we show how the different layers contribute to the stock returns for firms with high and low *DO* levels. Results for equally weighted (value-weighted) portfolio schemes are shown in Table 9, Panel A (Panel B). Overall, we find that all layer-specific DO portfolios yield positive performance. For example, our analysis in columns (2) to (4) of Panel A shows monthly six-factor alphas of 0.50 % (t-statistic: 3.54), 0.56 % (t-statistic: 2.89), and 0.59 % (t-statistic: 2.37) for device, network, and service layers, respectively. The six-factor alpha for the content layer in column (5) is slightly lower at 0.31 %, but also statistically significant (t-statistic: 2.28).

The same conclusion can be drawn from Panel B of Table 9, although the abnormal returns for the value-weighted long-short

[7] We refrain from including an additional control for the betting-against-beta factor in these robustness tests, as the alphas differ only marginally and the six-factor regressions are an established standard in the literature.

Table 10
Abnormal returns for various robustness tests.

	Equally Weighted Returns (1)	Value-Weighted Returns (2)
Panel A: Modification of the *DO* calculation		
1. tf-idf DO	0.0043**	0.0031***
	(2.50)	(2.62)
2. DO based on section 1 (Business Description)	0.0053***	0.0041**
	(2.75)	(2.54)
3. DO based on entire 10-K report	0.0063***	0.0042***
	(3.20)	(2.62)
Panel B: Control for industry membership and Nasdaq exposure		
4. FF48 industry-adjusted returns	0.0021	0.0007
	(1.50)	(0.96)
5. Nasdaq as Market Index	0.0027	0.0013
	(1.59)	(1.18)
6. DGTW-characteristic-adjusted returns	0.0027**	0.0013
	(2.42)	(1.37)
Panel C: Modification of the sample		
7. Exclusion of dot.com-bubble (98/99)	0.0032*	0.0023*
	(1.84)	(1.96)
8. Exclusion of high-sentiment periods	0.0032**	0.0032***
	(2.24)	(2.62)
9. Stocks with MV equity>10 Mio USD & P>1 USD	0.0029**	0.0026**
	(2.20)	(2.31)
Panel D: Exclusion of firms with very short or long MD&A section		
10. MD&A section length ≥ 100 words	0.0037**	0.0045***
	(2.00)	(3.05)
11. MD&A section length ≥ 100 and ≤ 50,000 words	0.0037**	0.0045***
	(2.00)	(3.04)
12. MD&A section length ≥ 500 words	0.0035*	0.0050***
	(1.80)	(2.85)

Table 10 shows the results of our robustness tests. We investigate the performance of equally weighted (Panel A) and capped value-weighted (Panel B) quintile 5-minus-1 portfolios based on the digital orientation measure. We measure performance using a six-factor model which controls for exposure to the Fama and French (1993) three-factor alphas, the Fama and French (2015) profitability and investment factors (RMW, CMA), and the momentum factor from Kenneth French (UMD). In Panel A modified measures of *DO* are used. Here, tf-idf stands for the term-frequency-inverse document frequency based on Loughran and McDonald (2011), which accounts for the importance of each term in the dictionary to measure *DO*. Panel B controls for industry membership, Nasdaq exposure and benchmark adjustments. FF48 indicates the Fama and French (1997) 48 industry classification system. The benchmark adjustments in analysis (6) rely on the set of characteristics proposed by Daniel et al. (1997), referred to as DGTW. Panel C conducts robustness tests based on modified samples. Panel D presents robustness tests that address potential parsing errors in extracting MD&A sections from 10-K filings. We impose minimum word count thresholds of 100 and 500 words to exclude filings where the MD&A section may have been incompletely parsed, and additionally test a specification with both a minimum of 100 and maximum of 50,000 words to exclude abnormally long sections that may indicate incorrect parsing. t-statistics are reported in parentheses. *, ** and *** indicates significance at the 10%, 5%, and 1% levels, respectively.

portfolio are again lower. Here, the largest six-factor alphas are observed for the network and service layers with 0.51 % (t-statistic 3.39) and 0.41 % (t-statistic 2.17), followed by the device and content layers with 0.35 % (t-statistic 2.90) and 0.20 % (t-statistic 2.13), respectively.

These slightly different results may be due to the diverging importance of digital activities within the different layers of digital technology across industries. Innovating within the device and network layers has been an important activity across industries, especially in the early years of the sample. In recent years, services have become a prominent element of the business models of many firms across all industries. However, services require continuous enhancements to the existing network to extend their reach and coverage, for example, to new customers. Similarly, devices need to adapt to the demands of new services and the opportunities presented by new technological developments. In contrast, competition at the content layer, and thus the focus on content creation, has been concentrated in a few firms and industries (e.g., media, entertainment), which may explain the weaker effect.

Second, we calculate equally weighted and value-weighted portfolio returns using modified *DO* measures (Table 10, Panel A). In our first robustness test, we measure firms' *DO* by accounting for the importance of each term in the dictionary (such as "software" or

Table 11
Alternative Digital Orientation Measures.

Panel A: Alternative dictionary (Hype Cycle)		
	Equally Weighted Returns	**Value-Weighted Returns**
1. Full sample dictionary	0.0027*	0.0023**
	(1.95)	(2.45)
2. Rolling extended dictionary	0.0033**	0.0026**
	(2.08)	(2.26)
Panel B: Portfolios formed based on membership in digital innovation-focused funds (2006-2020)		
	Equally Weighted Returns	**Value-Weighted Returns**
3. Full sample	0.0054***	0.0050***
	(2.99)	(3.25)
4. Large caps	0.0051***	0.0049***
	(3.32)	(3.11)

Table 11 shows the results of additional robustness tests. We investigate the performance of equally weighted (left) and capped value-weighted (right) quintile 5-minus-1 portfolios based on alternative DO measures. We measure performance using a six-factor model which controls for exposure to the Fama and French (1993) three-factor alphas, the Fama and French (2015) profitability and investment factors (RMW, CMA), and the momentum factor from Kenneth French (UMD). Panel A uses the alternative digital innovation dictionary derived using the Gartner Hype Cycle (see Appendix B for details). In Panel B portfolio returns are computed based on stock membership in digital innovation-focused investment funds. Long positions are taken in stocks included in such funds, while short positions are assigned to stocks not represented in these funds. The large-caps subsample in specification (4) only includes stocks in the NYSE size decile greater than 2.The sample period for Panel B is 2006-2020 since the digital innovation-focused funds were first created in 2006. t-statistics are reported in parentheses. *, ** and *** indicates significance at the 10%, 5%, and 1% levels, respectively.

"cloud") using the so-called term-frequency-inverse document frequency (tf-idf) based on Loughran and McDonald (2011). The correlation between both *DO* measures is 0.59, which suggests that sorts lead to somewhat different compositions for the stock portfolios. Additionally, for robustness tests (2) and (3), we derive *DO* measures using the 10-K's Section 1, which contains the business description, and the entire 10-K report.

The results in Panel A of Table 10 confirm that *DO* constitutes a distinct and statistically significant asset pricing premium, even when considering alternatively derived *DO* measures. The equally weighted six-factor alpha ranges between 0.43 % and 0.63 % per month for the overall *DO* measure. For value-weighted returns, the six-factor alpha is slightly lower and ranges between 0.31 % and 0.42 %. These results are comparable in economic and statistical significance to the results presented in Table 8 for the *DO* measure derived using the MD&A section. Specifically, all six-factor alphas of the portfolios are statistically significant at the 1 %- or 5 %-level. This indicates the robustness of our results when using alternatively derived *DO* measures.

Third, we control for stocks' industry affiliation, exposure to the Nasdaq index, and style. Because digitalization is a cross-industry phenomenon, we expect that the abnormal returns associated with digital orientation are not entirely explained by industry affiliation. To formally test this expectation, we calculate equally weighted and value-weighted portfolio returns based on industry-adjusted stock returns for robustness analysis (4).For the industry adjustment, we use industry returns that are calculated as the stock's return minus its industry return, based on the Fama and French (1997) 48 industry classification system, denoted as FF48. The results are displayed in Panel B of Table 10.

For equally weighted returns, the six-factor alpha of the industry-adjusted *DO* strategy is 0.21 % per month using the FF48 industry classification. The alpha is lower and no longer statistically significant. One potential explanation for this could be that our *DO* is better at capturing some industry-specific digital activities than others, particularly those related to the creation of digital content using digital technologies. Thus, the creation of new digital content as an essential part of firms' business may be highly concentrated in certain industries, such as media and publishing. Therefore, when controlling for industries, the importance of creating and using new content by relying on digital technologies may no longer play a significant role in returns.

In analysis (5), we further test the robustness of our results to firms' exposure to the Nasdaq. Specifically, we use the excess return of the Nasdaq Composite Index from Refinitiv instead of the broader CRSP value-weighted index as the market factor (mktrf). If the *DO* measure only serves as a proxy for technology and high cash flow duration, we expect the outperformance to disappear after using the Nasdaq as a benchmark. However, we find that this is only partly the case. The alpha is 0.27 % and just short of being statistically significant (t-statistic: 1.59). We conclude that firms' digital orientation is not satisfactorily described by being listed on the Nasdaq and that Nasdaq membership cannot fully explain the outperformance of digital leaders.

In robustness test (6), we follow the procedure by Daniel et al. (1997) and benchmark-adjust the stock returns before estimating the alpha of the equally weighted and value-weighted *DO* portfolio returns. The stock returns are benchmarked based on the characteristics of size, book-to-market, and prior return (denoted as *DGTW*). For equally weighted returns, we obtain an alpha of 0.27 % per month for the aggregate measure. This is statistically significant at the 5 % level.

Fourth, we use modified samples to ensure that our results are not driven by the dot.com bubble or high-sentiment periods (Table 10, Panel C). In analysis (7), we exclude the years 1998 and 1999 to control for the possibility that the dot.com bubble influences our results. Relying on the sentiment index suggested by Baker and Wurgler (2006), we conduct an additional robustness test to account for the possible influence of high sentiment periods in analysis (8). Thus, we exclude periods with a sentiment located in the top 20 % percentile of the index. Finally, we also run an analysis using a sample that excludes micro-caps. In particular, for the robustness test (9), we include only stocks with a market value equity of at least 10 Mio. USD and a stock price larger than 1 USD in our sample.

We find that our results are robust to excluding specific periods in the sample (the dot.com bubble and high sentiment periods). The equally weighted six-factor alpha is 0.32 % per month for both robustness tests. For value-weighted returns, the six-factor alphas are 0.23 % per month when excluding the dot.com bubble and 0.32 % for a sample without high-sentiment periods. These results are again comparable in economic and statistical significance to the results presented in Table 8 for the *DO* measure over the entire sample period, further indicating our results' robustness. When calculating *DO* stock returns in a sample that excludes microcaps, we find alphas that are smaller in size but remain highly statistically significant at the 5 % level. The six-factor alpha of the *DO* portfolio is now 0.29 % per month for equally weighted returns and 0.26 % per month for value-weighted returns.

Fifth, to ensure that our results are not driven by parsing errors, Panel D of Table 10 reports estimates using alternative word count restrictions. We first impose a minimum threshold of 100 words to exclude filings where the MD&A section may be incompletely captured. The results remain virtually unchanged, with an equally weighted (value-weighted) six-factor alpha of 0.37 % (0.45 %) per month, statistically significant at the 5 % (1 %) level. When we additionally impose a maximum threshold of 50,000 words to exclude abnormally long sections that likely reflect issues of incorrect parsing, the findings are similarly robust with an equally weighted (value-weighted) six-factor alpha of 0.37 % (0.45 %) per month, statistically significant at the 5 % (1 %) level. Finally, using a more conservative minimum threshold of 500 words also yields comparable results. These specifications demonstrate that our main findings are not affected by potential imperfections in the MD&A section identification algorithm.

Sixth, we use an alternative dictionary to measure firms' orientation toward digital innovation. This dictionary is based on the yearly published Hype Cycle by Gartner, which tracks the maturity of a new, promising digital innovation across time. The Hype Cycle displays a digital innovation's progress from the market entrance to maturity by considering both a technology's expectations and achievements. We compile this alternative dictionary by considering all digital innovations (and their synonyms) mentioned in the Hype Cycle in the years 1995 to 2020 (see Appendix B for details). The results are shown in Panel A of Table 11. Robustness test (1) uses a *DO* measure that is created by applying the full sample list of terms to all MD&A sections. The *DO* measure for robustness test (2) is based on a word list that contains only terms that have appeared in the Hype Cycle periodicals so far, and is extended every year with new terms. In other words, for the robustness test (2), we evaluate the digital orientation of every MD&A section only based on terms that have already appeared in prior publications.

Appendix B shows the dictionary based on the Hype Cycle and the thirty most frequent terms of the Hype Cycle dictionary in firms' 10-Ks. Compared to our baseline dictionary, the Hype Cycle dictionary is substantially longer, but terms are more specific in nature, which translates into fewer word counts within the MD&A section. In fact, >50 % of the MDA sections contain zero terms from the Hype Cycle. Consequently, instead of forming long-short portfolios based on quintiles, we calculate equally weighted and capped value-weighted returns of a portfolio, which is long in stocks with at least one word hit and short in stocks with zero word hits.

Despite conceptual differences between the Hype Cycle dictionary and our baseline dictionary, robustness tests (1) and (2) in Table 11 show that the results and conclusions are remarkably similar. The equally weighted six-factor alpha is 0.27 % per month (t-statistic 1.95) for the full sample word list, and 0.33 % (t-statistic 2.08) per month for the rolling extended dictionary. Value-weighted six-factor alphas are also significant at the 5 %-level, with 0.23 % per month and a t-statistic of 2.45 for robustness test (1) and 0.26 % per month and a t-statistic of 2.26 for robustness test (2). Moreover, we see that the differences between the first approach (full sample word list) and the second approach (continuously expanding word list) are only marginal.

Seventh, we compute both equally weighted and value-weighted portfolio returns based on stock membership in digital innovation-focused investment funds. Long positions are taken in stocks included in such funds, while short positions are assigned to stocks not represented in these funds.[8] Table 11, Panel B, reports the estimated alphas obtained by regressing the resulting portfolio returns on standard risk factors. The results indicate that the strategy yields a significantly positive alpha for both the full sample (robustness test (3)) and the large-cap subsample (robustness test (4)).

The robustness tests presented in Table 11 support the conclusion that digital innovation is associated with higher returns. While both the fund-based DO portfolio and the DO portfolio based on text data from 10-K filings yield positive performance, the baseline *DO* measure offers several distinct advantages. First, it is not a binary indicator, but rather allows for more nuanced differentiation in firms' levels of digital orientation. Second, its differentiation between different layers of digital technologies' architecture allows for a more in-depth analysis of firms' use and development of digital innovation, yielding richer insights. Finally, the *DO* measure is broadly applicable and can be used in less developed markets where comparable ETFs may not exist, thereby enabling the assessment of digital orientation even in the absence of sophisticated investment products.

In Figures C1a to C1d in Appendix C, we further illustrate the robustness of our results by demonstrating the cumulative abnormal return of the digital orientation for different factor asset pricing models across time using alternative *DO* measures. We show that the cumulative wealth effects across time are similarly large when using an alternative dictionary based on the Hype Cycle or digital innovation-focused fund holdings.

Finally, we conduct additional robustness tests to identify explanations of why digital leaders have above-average returns. In the following, we provide a summary of our results; detailed results are available upon request. We do not find evidence for the so-called discount rate channel hypothesis, which proposes that high *DO* stocks profited to a larger extent from a low-interest rate environment

[8] On the long side, we only consider stocks that are currently or have previously been listed as fund holdings at each point in time. This ensures that the strategy does not suffer from look-ahead bias.

and loose monetary policy in recent decades (e.g., Chakraborty et al., 2020). Next, to test if the digital premium is due to mispricing, we relate abnormal three-day cumulative stock returns around quarterly earnings announcements to firms' digital orientation. We also compute the analyst forecast bias (actual quarterly earnings per share (EPS) minus the median analyst forecast EPS, scaled by stock price) to test if analysts consistently underestimate the earnings of stocks with a high digital orientation. Our tests do not provide evidence supporting the mispricing hypothesis.

A third potential explanation for the above-average returns of firms with high *DO* value is that higher returns constitute a systematic risk compensation. Such a risk compensation may exist because new information about the potential of new digital technologies and future cash flows is first revealed in leading firms, which thereby act as early indicators for the growth prospects of other firms. Thus, digital leaders may have a risk premium because they are the first to provide a resolution of uncertainty that comes from the digital transformation of the overall economy. To test if a *DO* factor should be considered as a priced-in systematic risk, we run "characteristics-vs.-covariances" tests. Our findings show that exposure to the *DO* factor predicts stock returns, independent of a stock's *DO* characteristic. This evidence is consistent with the *DO* factor being a new systematic risk and that exposure to this risk is priced in the cross-section of stock returns.

This result fits with the hypothesis that digitalization is a process that affects the entire economy and is not limited to individual sectors. The resulting benefits and risks from investments in digital technology are therefore not idiosyncratic in nature, but shared across the board.

The existence of a risk premium for digital leaders is also broadly consistent with the premium for leading industries as argued by Croce et al. (2023). In their study, the authors theoretically argue and empirically show that leading industries, i.e., those that are first to resolve uncertainty about the overall economic growth, carry a risk premium. In the spirit of this model, digital leaders are the first to provide a resolution of uncertainty, which comes from digital innovation, and investors who are exposed to this uncertainty require compensation. Extending Croce et al. (2023), we show that such a leading premium exists not only for traditional industries, but also for digital leaders across all industries.

4.4. *Exploitability of strategy*

So far, our results show that digital orientation is associated with higher stock returns after controlling for well-known asset pricing factors. These effects are substantial, amounting to 6 % (4.1 %) per year for equally weighted (capped value-weighted) long-short portfolios, reflecting compensation for exposure to digitalization risk that is distinct from traditional risk factors. However, implementing a digital orientation strategy may pose also two challenges for investors. First, digital orientation is negatively related to other established return predictors, particularly profitability and momentum. Hence, investors seeking not only high alphas but also high raw returns need to appropriately hedge against such negative factor exposures. Second, to implement the baseline *DO* strategy in real

Table 12
Enhanced and weakened momentum and profitability strategies using digital orientation.

Panel A: Dictionary-based digital orientation				
	Weakened strategy (1)	Enhanced strategy (2)	Difference (2)-(1)	Difference (2)-(1) alphas for FF-6 Factor plus BAB
1. Momentum	-0.0037	0.0030	0.0067*	0.0037
	(-0.89)	(0.63)	(1.92)	(1.22)
2. Operating Profitability	-0.0027	0.0035	0.0062*	0.0043
	(-0.90)	(1.17)	(1.74)	(1.30)
Panel B: Alternative dictionary (Gartner Hype Cycle)				
	Weakened strategy (1)	Enhanced strategy (2)	Difference (2)-(1)	Difference (2)-(1) alphas for FF-6 Factor plus BAB
3. Momentum	-0.0028	0.0034	0.0062***	0.0050**
	(-0.68)	(0.76)	(2.82)	(2.47)
4. Operating Profitability	-0.0008	0.0023	0.0032	0.0016
	(-0.33)	(1.01)	(1.38)	(0.79)
Panel C: Digital innovation-focused fund holdings				
	Weakened strategy (1)	Enhanced strategy (2)	Difference (2)-(1)	Difference (2)-(1) alphas for FF-6 Factor plus BAB
5. Momentum	-0.0069	0.0051	0.0120***	0.0099***
	(-1.60)	(1.13)	(3.53)	(3.19)
6. Operating Profitability	-0.0057**	0.0046*	0.0103***	0.0091***
	(-2.03)	(1.93)	(3.56)	(3.59)

Table 12 shows the (abnormal) returns of a quintile 5-minus-1 portfolios based on a double sorting strategy, where stocks are sorted based on their digital orientation and their momentum or operating profitability. Panel A uses our baseline dictionary-based digital orientation measure and Panel B the alternative digital innovation dictionary using the Gartner Hype Cycle on a rolling basis for sorting firms based on their digital orientation. In contrast, Panel C sorts firms based on digital innovation-focused fund holdings and their momentum or profitability (Panel C). Column (1) depicts the raw returns of weakened momentum or profitability strategy, which is long in stocks with high profitability or momentum but low digital orientation, and short in stocks with low profitability or momentum but high digital orientation. Column (2) shows the raw returns for an enhanced momentum or profitability strategy, which is long in stocks with high digital orientation and high momentum, and short in stocks with low digital orientation and low momentum strategy. Column (3) and (4) depict the performance difference between the enhanced and weakened factor strategies. Whereas column (3) shows raw returns, column (4) depicts the alphas of a regression that accounts for the asset pricing factors of the Fama-French six-factor model plus the betting-against-beta factor. t-statistics are reported in parentheses. *, ** and *** indicates significance at the 10%, 5%, and 1% levels, respectively.

time, investors would need to continuously update the *DO* dictionary, whereas we have relied on the same version over time.

Our previous robustness tests to measure digital orientation based on the rolling Gartner Hype Cycle and stock membership in digital innovation-focused investment funds demonstrate that a *DO* measure can also be successfully implemented in real time. Therefore, it appears unlikely that look-ahead bias can explain the positive relation between risk-adjusted returns and digital orientation. Nevertheless, the question remains how *DO* can be usefully integrated in an investment strategy.

We aim to provide an answer by demonstrating how existing profitability and momentum strategies can be enhanced when accounting for a firm's digital orientation. To rule out any potential influence of look-ahead bias, we also focus on approaches that are free from this concern, i.e., the rolling Gartner Hype Cycle and fund membership as alternative measures of digital orientation.

Our proposed enhanced profitability and momentum strategies are based on an independent double sorting approach. The enhanced profitability strategy is long in stocks with high digital orientation and high profitability, and short in stocks with low digital orientation and low profitability. Similarly, the enhanced momentum strategy is long in stocks with high digital orientation and high momentum, and short in stocks with low digital orientation and low momentum. We compare the results for these enhanced factors to the returns obtained for weakened factors. For example, the weakened profitability strategy is long in stocks with high profitability but low digital orientation, and short in stocks with low profitability but high digital orientation. We compare the performance for the enhanced and weakened factor strategies in Table 12, columns (1) and (2).

Column (2) of Table 12 shows the long-short returns of the enhanced strategies. It illustrates that we can earn positive returns based on both strategies for portfolios long in stocks with a high digital orientation and momentum (specifications (1) and (3)) or profitability (specifications (2) and (4)) for measures of digital orientation based on the baseline *DO* dictionary (Panel A) and the alternative dictionary of the Gartner Hype Cycle (Panel B). We can contrast these returns with the performance of the weakened strategies, as shown in Column (1). Here, we observe negative, non-significant returns for portfolios formed based on the dictionary-based approach and the alternative dictionary measure.

Similarly, when considering the digital innovation-based fund holdings to form portfolios, as illustrated in Panel C, we observe positive returns for the enhanced factor strategies (Column (2)) and negative returns for the weakened strategies (Column (1)).

Columns (3) and (4) depict the performance difference between the enhanced and weakened equity strategies. While Column (3) focuses on the difference in monthly raw returns, Column (4) shows the difference for a model that accounts for the asset pricing factors of the Fama-French five-factor model plus momentum and betting-against-beta factors. We observe sizeable, positive differences in returns based on all approaches to measure digital leaders and digital laggards in Column (3). The spreads are statistically significant for the portfolio based on the baseline *DO* dictionary (Panel A) for both momentum (specification (1)); t-statistic: 1.92) and profitability doubling sorting strategy (specification (2); t-statistic: 1.74) as well as the returns for the momentum doubling sorting strategy for the alternative Gartner Hype Cycle dictionary (t-statistic 2.82), as illustrated in Panel B, specification (3). In line with these results, the returns for a portfolio strategy based on the differences between enhanced and weakened factor strategies and using the innovation-based fund holdings to form portfolios are also statistically significant both for the momentum and profitability double sorting strategy (t-statistics 3.53 and 3.56), as shown in Panel C, specifications (5) and (6). The return spreads amount to >1 % per month in Panel C, pointing to economically meaningful differences between the weakened and enhanced strategies.

We also observe positive alphas for the double sorting profitability and momentum strategy when accounting for the asset pricing factors of the Fama-French six-factor model plus the betting-against-beta factor, as shown in Column (4). The alphas are statistically significant for the double sorting momentum strategy using the alternative Gartner Hype Cycle dictionary (see Panel B, specification (3); t-statistic: 2.47) and the double sorting momentum and profitability strategy using digital innovation-focused fund holdings (see Panel C; t-statistics: 3.19 and 3.59).

Overall, we can observe that the double sorts, which include digital orientation in the portfolio selection process, can improve the performance of traditional equity strategies like momentum and profitability considerably. Enhanced factor strategies outperform weakened strategies by up to 1 % per month. This observation is also illustrated in Appendix E, where we graphically show the cumulative abnormal returns of the double sorting momentum and profitability strategy.

5. Conclusion

Although digital innovations have led to the disruption of business models in firms and challenged the boundaries of existing industries, we know little about the impact of digital innovation in explaining asset returns. Relying on a bag-of-words approach, we analyze annual reports of US firms in the period 1996–2020 to derive a text-based, objective, and easy-to-implement measure of firms' orientation toward digital innovation (called digital orientation). We show that firms' digital orientation has significant implications for the capital market and investors' valuation of stocks. Our findings illustrate that our measure of digital innovation can explain heterogeneity in stock returns across firms that cannot be accounted for by other firm characteristics.

Based on this insight, we create a new digital orientation strategy, which has significant explanatory power for expected stock returns. Against the backdrop that digitalization and the impact of digital innovation are unlikely to disappear in the next decade(s), our findings have significant implications for studying capital markets in the digital age. Moreover, they may offer forward-looking investors the potential to profit from a digital investment style.

Data availability statement

The Digital Orientation (DO) factor data generated in this study are publicly available through Mendeley Data at: https://doi.org/10.17632/wggxrnvf8s.1. The dataset includes equally weighted and capped value-weighted monthly quintile (5) minus quintile (1) long-short returns based on the DO measure from July 1997 to May 2020.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jedc.2025.105217.

References

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. J. Finance 61 (4), 1645–1680.
Bharadwaj, A.S., Bharadwaj, S.G., Konsynski, B.R., 1999. Information technology effects on firm performance as measured by Tobin's q. Manage. Sci. 45 (6), 1008–1024.
Bharadwaj, A.S., El Sawy, O.A., Pavlou, P.A., Venkatraman, N., 2013. Digital business strategy: toward a next generation of insights. MIS Q. 37 (2), 471–482.
Bogers, M., West, J., 2012. Managing distributed innovation: strategic utilization of open and user innovation. Creat. Innov. Manag. 21 (1), 61–75.
Chakraborty, I., Goldstein, I., MacKinlay, A., 2020. Monetary stimulus and bank lending. J. Financ. Econ. 136 (1), 189–218.
Chen, W., Srinivasan, S., 2023. Going digital: implications for firm value and performance. Rev. Account. Stud.
Cohen, L., Malloy, C., Nguyen, Q., 2020. Lazy prices. J. Finance 75 (3), 1371–1415.
Corrado, C.A., Hulten, C.R., 2010. How do you measure a "technological revolution"? Am. Econ. Rev. 100 (2), 99–104.
Croce, M.M., Marchuk, T., Schlag, C., 2023. The leading premium. Rev. Financ. Stud.
Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. J. Finance 52 (1), 1–33.
Daniel, K., Titman, S., 2006. Market reactions to tangible and intangible information. J. Finance 61 (4), 1605–1643.
Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. J. Finance 52 (3), 1035–1058.
Drechsler, K., 2023. Promoting digital innovation and transformation – Innovation champions in the digital age. Otto-Friedrich-University, Bamberg, Germany.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33 (1), 3–56.
Fama, E.F., French, K.R., 1997. Industry costs of equity. J. Financ. Econ. 43 (2), 153–193.
Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116, 1–22.
Fama, E.F., French, K.R., 2018. Choosing factors. J. Financ. Econ. 128 (2), 234–252.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. J. Polit. Econ. 81 (3), 607–636.
Fang, V.W., Tian, X., Tice, S., 2014. Does stock liquidity enhance or impede firm innovation? J. Finance 69 (5), 2085–2125.
Faulkner, P., Runde, J., 2019. Theorizing the digital object. MIS Q. 43 (4), 1278–1302.
Frazzini, A., Pedersen, L.H., 2014. Betting against beta. J. Financ. Econ. 111 (1), 1–25.
Grammig, J., Jank, S., 2016. Creative destruction and asset prices. J. Financ. Quant. Anal. 51 (6), 1739–1768.
Greenwood, R., Hanson, S.G., 2012. Share issuance and factor timing. J. Finance 67 (2), 761–798.
Guay, W., Samuels, D., Taylor, D., 2016. Guiding through the fog: financial statement complexity and voluntary disclosure. J. Account. Econ. 62 (2–3), 234–269.
Henfridsson, O., Nandhakumara, J., Scarbrough, H., Panourgias, N., 2018. Recombination in the open-ended value landscape of digital innovation. Inf. Organ. 28 (2), 89–100.
Hillert, A., Jacobs, H., Müller, S., 2014. Media makes momentum. Rev. Financ. Stud. 27 (12), 3467–3501.
Hillert, A., Niessen-Ruenzi, A., Ruenzi, S., 2024. Mutual fund shareholder letters: flows, performance, and managerial behavior. Manage. Sci.
Hoberg, G., Phillips, G., 2010. Product market synergies and competition in mergers and acquisitions: a text-based analysis. Rev. Financ. Stud. 23 (10), 3773–3811.
Hoberg, G., Phillips, G., 2016. Text-based network industries and Endogenous Product differentiation. J. Polit. Econ. 124 (5), 1423–1465.
Hong, H., Lim, T., Stein, J.C., 2000. Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies. J. Finance 55 (1), 265–295.
Hopp, C., Antons, D., Kaminski, J., Oliver Salge, T., 2018. Disruptive innovation: conceptual foundations, empirical evidence, and research opportunities in the digital age. J. Prod. Innov. Manag. 35 (3), 446–457.
Hund, A., Wagner, H.-T., Beimborn, D., Weitzel, T., 2021. Digital innovation: review and novel perspective. J. Strateg. Inf. Syst. 30 (4).
Jensen, T.I., Kelly, B.T., Pedersen, L.H., 2021. Is there a replication crisis in finance? J. Finance 78 (5), 2465–2518.
Kallinikos, J., Aaltonen, A., Marton, A., 2013. The ambivalent ontology of digital artifacts. MIS Q. 37 (2).
Kindermann, B., Beutel, S., Garcia de Lomana, G., Strese, S., Bendig, D., Brettel, M., 2021. Digital orientation: conceptualization and operationalization of a new Strategic orientation. Eur. Manag. J. 39 (5), 645–657.
La Porta, R., 1996. Expectations and the cross-section of stock returns. J. Finance 51 (5), 1715–1742.
Li, F., 2008. Annual report readability, current earnings, and earnings persistence. J. Account. Econ. 45 (2), 221–247.
Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. J. Finance 66 (1), 35–65.
Loughran, T., McDonald, B., 2016. Textual analysis in accounting and Finance: a survey. J. Account. Res. 54 (4), 1187–1230.
Lyytinen, K., Yoo, Y., Boland, R.J., 2016. Digital product innovation within four classes of innovation networks. Inf. Syst. J. 26 (1), 47–75.
Nambisan, S., Lyytinen, K., Majchrzak, A., Song, M., 2017. Digital Innovation Management: reinventing Innovation Management research in a Digital world. MIS Q. 41 (1), 223–238.
Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. Rev. Econ. Stud. 61 (4), 631–653.
Porter, M.E., Heppelmann, J.E., 2014. How smart, connected products are transforming competition. Harv. Bus. Rev. 92 (11), 64–88.
Seo, D., 2017. Digital business convergence and emerging contested fields: a conceptual framework. J. Assoc. Inf. Syst. 18 (10), 687–702.
Svahn, F., Mathiassen, L., Lindgren, R., 2017. Embracing digital innovation in incumbent firms: how Volvo cars managed competing concerns. MIS Q. 41 (1), 239–253.
Tetlock, 2007. Giving content to investor sentiment: the role of Media in the stock market. J. Finance 62 (3), 1139–1168.
Tilson, D., Lyytinen, K., Sørensen, C., 2010. Digital infrastructures: the missing IS research agenda. Inf. Syst. Res. 21 (4), 748–759.
Tiwana, A., Konsynski, B., Bush, A.A., 2010. Platform evolution: coevolution of Platform architecture, governance, and environmental dynamics. Inf. Syst. Res. 21 (4), 675–687.
Tumbas, S., Berente, N., vom Brocke, J., 2018. Digital innovation and Institutional entrepreneurship: chief Digital officer perspectives of their emerging role. J. Inf. Technol. 33 (3), 188–202.
Tumbas, S., Berente, N., Seidel, S., Vom Brocke, J., 2015. The 'Digital façade' of rapidly growing entrepreneurial organizations. In: Proceedings of the 36th International Conference on Information Systems, pp. 1–19.
Vial, G., 2019. Understanding digital transformation: a review and a research agenda. J. Strateg. Inf. Syst. 28 (2), 118–144.
Wessel, L., Baiyere, A., Ologeanu-Taddei, R., Cha, J., Jensen, T.B., 2021. Unpacking the difference between Digital transformation and IT-enabled organizational transformation. J. Assoc. Inf. Syst. 22 (1), 102–129.
Wu, L., Hitt, L., Lou, B., 2019. Data analytics, innovation, and firm productivity. Manage Sci. 66 (5), 2017–2039.
Yoo, Y., Henfridsson, O., Lyytinen, K., 2010. The new organizing logic of digital innovation: an agenda for Information Systems Research. Inf. Syst. Res. 21 (4), 724–735.
Yoo, Y., Boland, R.J.J., Lyytinen, K., Majchrzak, A., 2012. Organizing for innovation in the digitized world. Organ. Sci. 23 (5), 1398–1408.