

Data Analytics Strategy and Internal Information Quality

Katie Wei Lem

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2023

Reading Committee:

Weili Ge, Chair

Elizabeth Blankespoor

Dawn Matsumoto

Sarah McVay

Program Authorized to Offer Degree:
Business Administration – Accounting

©Copyright 2023

Katie Wei Lem

University of Washington

Abstract

Data Analytics Strategy and Internal Information Quality

Katie Wei Lem

Chair of the Supervisory Committee:

Weili Ge

Business Administration – Accounting

I examine whether a strategic focus on data analytics improves firms' internal information quality, the decision-usefulness of information. Using textual analysis of firm disclosures to identify a data analytics strategy, I first document that firm, leadership, and operating environment characteristics are all important determinants of the decision to adopt a data analytics strategy. To infer the impact of a data analytics strategy on internal information quality, I examine its implications for various operating and financial reporting outcomes. I find that a data analytics strategy is primarily associated with better operating efficiency. More specifically, when firms employ a data analytics strategy, they invest and utilize existing resources more efficiently. I also find evidence that a data analytics strategy is associated with improvements in management forecasts. Overall, the results are consistent with a data analytics strategy improving firms' internal information quality.

1. Introduction

Over the past decade, companies have collected vast amounts of data, drastically increasing the importance of data analytics. Data analytics refers to technology designed for the transformation of structured and unstructured data, using quantitative analysis, predictive models, and visualization techniques, into useful information for decision-makers (Schneider, Dai, Janvrin, Ajayi and Raschke 2015). Prior research suggests that data analytics is likely to have an important impact on the firm's internal information environment (Davenport 2006; Schneider et al. 2015; Warren, Moffit, and Brynes 2015) and change the role of financial analysts within the firm. However, there has been limited large sample research in this area (Rikhardsson and Yigitbasioglu 2018), in part due to the unobservability of data analytics at the firm level. In this study, I use textual analysis of firm disclosures to examine whether a data analytics strategy improves internal information quality.

I focus on a data analytics strategy, which I define as a direction set forth by firm leadership to employ the use of data analytics technology in creating firm value, for two important reasons. First, most firms perform some level of data analysis (e.g., using Excel), but I am specifically interested in firms that use the type of technology designed to overcome the challenges of big data, that of high volume, velocity, and variety (McAfee and Brynjolfsson 2012). This technology enables users to better prepare and integrate data of different structures, visualize data insights, and perform statistical analyses to inform decision-making.¹ Second, effectively leveraging data analytics requires an integrated approach, which goes beyond simply adopting the technology. Prior research suggests that firms with technological initiatives driven by leadership are more likely to take this integrated approach and realize the benefits (Venkatesh, Morris, G. Davis, and

¹ Examples of data analytics technology, such as Tableau, Alteryx, and Power BI are included in Appendix B.

F. Davis 2003; Neufeld, Dong, and Higgins 2007). Thus, I focus on firms for which the use of data analytics is supported by its executives.

I examine whether a data analytics strategy improves firms' internal information quality, i.e., the decision-usefulness of information for managers. Many factors can generate differences in internal information quality, and understanding these differences is important to both internal and external firm stakeholders. However, internal information quality is difficult to observe. Prior literature suggests a strong relation between a firm's internal information quality and its decision processes, such as evaluating investment opportunities and forming financial estimates (e.g., Keller and Staelin 1987; Goodman, Neamtiu, Shroff, and White 2014; Gallemore and Labro 2015). Following this line of literature, I use outcomes of operating and financial reporting decisions to infer the quality of the information that managers have.

I propose that three key features of data analytics strategy can help reduce the marginal cost of processing data into useful information. First, data analytics technology can help users prepare and incorporate various types of data more easily (e.g., online browsing data for forecasting demand). As unprecedented amounts of unstructured data are collected at faster speeds, analyzing data requires more cleaning and standardization. Second, data analytics can help users visualize patterns and anomalies in ways that are easy for decision-makers to understand (e.g., interactive graphs highlighting business risks and exposure). Third, data analytics can simplify statistical modeling and its adaptation to new information and assumptions (e.g., comparing alternative scenarios to optimize the use of existing assets). With the support of firm leadership, initiatives to leverage these capabilities could increase the feasible volume and quality of data analysis, resulting in more decision-useful information and thus, better operating and reporting decisions.

On the other hand, it is possible that a data analytics strategy does not measurably improve or even hurts internal information quality. A successful strategy should align with corporate structure, resources, and culture (Higgins 2005), and anecdotal evidence suggests that some firms are “diving in” when it comes to data analytics without understanding the necessary steps to realize its benefits. For example, firms must have the existing infrastructure, such as centralized systems, to collect and organize the inputs to data analytics. Firms also need to have employees with the skills to utilize the technology. Lastly, firms must overcome the cultural challenges associated with changing existing processes. That is, leveraging data analytics requires employees at every level of the decision-making chain to have the necessary incentives and the right mindset to learn from the data. Devoting resources to these challenges can take away from other value-enhancing activities. Furthermore, overreliance on data analytics and information overload could both be detrimental to decision-making. Thus, a data analytics strategy might not improve the quality of information ultimately used by decision-makers.

I use textual analysis of earnings conference call transcripts to identify firms that employ a data analytics strategy. I search conference call transcripts for the keywords “data analytics” and “big data”, the input to data analytics, and consider firms to adopt a data analytics strategy once they have discussed these topics.² This approach requires assumptions that disclosures are indicative of 1) a firm’s actual use of data analytics and 2) leadership’s strategic focus on data analytics. As such, I start by reading the transcripts for a random sample of 100 conference calls to provide descriptive insight on how firms talk about data analytics. Although firms’ disclosures on data analytics vary in frequency and specificity, most disclosures indeed suggest the use of data

² I choose these terms from a larger list of data analytics keywords based on their relevance and their prevalence in disclosures. When I use the larger list of data analytics terms, which includes features of the technology as well as the names of popular data analytics providers, the results are inferentially similar. Refer to section 4 for details on the alternative dictionary of terms. Examples of firm disclosures are provided in Appendix D.

analytics for decision making. To gain assurance that data analytics disclosure is a signal of true technological involvement rather than “cheap talk”, I next turn to job postings data. I find that hiring a high percentage of employees with data analytics skills is a significant determinant of my conference call measure, providing evidence that firms discussing data analytics in their disclosures concurrently undertake meaningful actions consistent with employing a data analytics strategy.

I begin my analysis by modeling the decision to adopt a data analytics strategy as a function of the firm’s organizational, leadership, and operating environment characteristics following prior literature on information technology (IT; e.g., Kobelsky, Richardson, Smith, and Zmud 2008; Oliveira and Martins 2011). I find that while larger firms are more likely to adopt a data analytics strategy, growing firms are also more likely to do so. Having a larger board of directors and directors with technology experience are also strong predictors of adopting a data analytics strategy. Lastly, firms are more likely to adopt a data analytics strategy when they operate in certain industries where the technology is expected to have a major impact (e.g., retail), and consistent with its increasing popularity, more firms are adopting a data analytics strategy over time.

In tests of my main hypotheses, I examine whether a data analytics strategy is positively associated with measures of operating and financial reporting outcomes to infer internal information quality. I specifically consider areas that are subject to a high degree of judgment and estimation as data analytics is well suited to improve these outcomes. The results suggest that a data analytics strategy improves investment expenditure efficiency, particularly related to acquisitions. A data analytics strategy also increases efficiency with respect to the use of existing resources. That is, a data analytics strategy is associated with improvements in both fixed and intangible asset turnover.

I also document improvements to certain aspects of financial reporting quality. Namely, when firms employ a data analytics strategy, they issue more accurate earnings guidance. On average, however, a data analytics strategy does not appear to impact the propensity to forecast, forecast timeliness, or routine accounting estimates, potentially because limited resources (e.g., skilled employees) are first being used for operating decisions.³ Although use of the technology directly in financial reporting processes appears to be limited, I find some evidence that firms adopting data analytics are less likely to have fraud-related restatements. This result suggests that improvements in internal transparency may serve as a deterrent for intentional misreporting. Overall, the results are consistent with a data analytics strategy improving internal information quality, as evidenced by both operating outcomes and, to a lesser extent, financial reporting outcomes.

One challenge with identifying a relation between a data analytics strategy and internal information quality is that implementing and disclosing a data analytics strategy are choices of the firm, and firms that make these choices could be inherently different from firms that do not. Thus, I employ entropy balancing in my main analyses to correct for selection on observable characteristics (Hainmueller 2012). I triangulate my results using firm fixed effects, the Heckman two-stage procedure (Heckman 1979), double clustering, and OLS. An important caveat is that it is possible that the firms that see results from data analytics are more likely to disclose their use of the technology. Thus, the operating efficiency and forecast improvements that I document could represent the upper bound of potential benefits.

³ Consistent with this explanation, untabulated analyses of job postings data confirm that only approximately 2% of jobs requiring data analytics skills are in accounting roles, whereas 18% are in finance roles. For context, 1% of all jobs in the LinkUp.com data relate to accounting roles, and 4% relate to finance roles.

I conduct several additional analyses to test whether the inferences are sensitive to other research design choices. I use a broader list of keywords to identify data analytics adoptions, including more sophisticated analytics terms (e.g., artificial intelligence). I search for data analytics disclosures through alternative channels (e.g., analyst and shareholder meetings). I also include a falsification test using randomly generated placebo adoption dates. The results from these analyses support the main conclusion that adopting a data analytics strategy is associated with improvements in internal information quality.

This study fills important gaps in the literature on data analytics in accounting. Existing studies contemplate the impact of data analytics on firms' information environments (e.g., Schneider et al. 2015), but archival evidence has been limited thus far. While contemporaneous research finds that digital technologies are associated with higher market valuations (Chen and Srinivasan 2022), I explore the reasons why these market valuations are justified, providing evidence consistent with data analytics improving internal information quality.⁴ My study also complements Labro, Lang, and Omartian (2022), who use Census data to study the impact of predictive analytics on the manufacturing-employment relation. I use publicly available information to show how data analytics impacts broader operating and financial reporting outcomes for a diverse sample of firms. Lastly, I provide novel descriptive evidence on disclosures about data analytics, and the results of my study serve to validate the disclosure of technology topics as a proxy for true technological involvement.

My study also contributes to the broader IT literature. First, the determinants analysis provides evidence that adopting a data analytics strategy is a function of a firm's organizational,

⁴ This study shows that the market outcomes that Chen and Srinivasan (2022) document (e.g., higher valuations, stock price co-movements with technology firms, and higher ERCs) are not just the result of perceived similarities between digital firms and technology firms. For firms that employ data analytics, these market outcomes are supported by real operating and financial reporting improvements.

leadership, and operating environment characteristics, as information systems theory predicts (e.g., Oliveira and Martins 2011). Second, prior research has examined the impact of other technologies, such as Enterprise Systems (ES), on various firm outcomes (e.g., Pincus, Tian, Wellmeyer, and Xu 2017, Dorantes, Li, Peters, and Richardson 2013). However, while the primary goals of ES relate to data storage and integrity, those of data analytics relate to data processing and insights. As such, the mechanisms through which firm outcomes are impacted by these technologies are different. This study illustrates how data analytics and ES serve important but distinct roles in the firm.⁵ Lastly, I provide preliminary evidence that currently evolving technologies such as machine learning are likely to yield similar improvements in information quality.

2. Background and Hypothesis Development

Internal Information Quality and Operating and Financial Reporting Decisions

Internal information quality refers to the decision-usefulness of information for managers (Gallemore and Labro 2015). Decision-useful information is accessible, accurate, and relevant. There are many internal and external factors that can generate differences in the quality of information that managers have, such as organizational complexity, the knowledge and effort of employees within the firm, and the availability of relevant peer financial statements (Badertscher, Shroff, and White 2013; Jennings, Seo, and Tanlu 2019; Call, Campbell, Dhaliwal, and Moon 2017; Shroff 2017).

Internal information quality can impact a wide range of decisions (e.g., Keller and Staelin 1987), including operating and financial reporting decisions. Prior literature shows managers who

⁵ Compared to the implementation of data analytics, the implementation of ES is a rudimentary step on the path to digital maturity (vom Brocke, Fay, Petry, and Teinzer 2017). ES were widely deployed in the late 1990s to the early 2000s, and today, most firms consider these a fundamental technology (vom Brocke, Maab, Buxmaan, Maedche, Leimeister, and Pecht 2018). As such, results during my sample period should not be confounded by the effects of having strong ES.

acquire and process better information (i.e., those with higher internal information quality) make better capital investments (e.g., Goodman et al. 2014; Shroff 2017). Based on similar reasoning that managements' earnings forecasts are a direct reflection of the internal information environment, studies also use earnings forecast properties to answer questions about internal information quality (Dorantes et al. 2013; Gallemore and Labro 2015). Following this line of literature, I examine how a data analytics strategy impacts outcomes of both operating and financial reporting decisions to infer the quality of information within the firm.

Research on Data Analytics and Other Information Technologies

Research in information systems and accounting suggests data analytics has many important applications from understanding consumer interactions to detecting fraud (e.g., Chau and Xu 2012; Abbasi, Albrecht, Vance, and Hansen 2012). Prior research also describes how data analytics will impact the internal information environment, positing that improvements to the quality and relevance of information will help decision-makers understand different perspectives (Schneider et al. 2015; Warren et al. 2015). Labro et al. (2022) find that manufacturing firms that use predictive analytics (based on Census survey data) indeed delegate decision-making differently. Consistent with the notion that better information leads to improved outcomes, initial archival studies find evidence that data analytics is associated with greater sales for specific cross-sections of firms (Müller, Fay, and vom Brocke 2018; Wu, Hitt, and Luo 2019).⁶ I build on this research by exploring the determinants of a data analytics strategy as well as the specific operating and reporting benefits more broadly.

A closely related working paper by Chen and Srinivasan (2022) documents a positive relation between the use of digital technologies (including data analytics, cloud computing, and

⁶ Specifically, Muller et al. (2018) examine technology firms, and Wu et al. (2019) focus on innovation improvements for a survey sample.

artificial intelligence) and capital market outcomes. More specifically, they find that digital firms have higher market valuations, stock price co-movements with technology firms, and earnings response coefficients (ERCs). At the same time, they find mixed evidence on whether these firms improve their overall financial performance. As such, the market outcomes they observe could reflect investor perceptions of technological sophistication or real improvements to firm value. My study complements Chen and Srinivasan (2022) by exploring the latter explanation. More specifically, digital firms employ a data analytics strategy that leads to better internal information quality. Resulting improvements in operating efficiency and financial reporting quality, which prior research has not examined, could both justify the higher market valuations and ERCs that Chen and Srinivasan (2022) document.

Empirical research in accounting also documents the effects of earlier information technologies such as Enterprise Systems (ES), including Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) systems, on various operating and reporting outcomes (e.g., Dorantes et al., 2013). This stream of research finds that ES implementations, which were widespread during the 1990s, improve operating performance (Hitt, Wu, and Zhou 2002; Hendricks, Singhal, and Stratman 2007). The use of ES is also associated with a higher likelihood of issuing management forecasts, more accurate forecasts, and more timely filings (Brazel and Dang 2008; Dorantes et al. 2013; Pincus et al. 2017). Lastly, studies find mixed evidence on how ES impact earnings quality.⁷

Although ES and data analytics are both information technologies, they impact information quality differently. ES (and related internal controls) are primarily focused on gathering, storing,

⁷ While Brazel and Dang (2008) find that ERP systems lead to larger discretionary accruals, Pincus et al. (2017) find that ES implementations lead to fewer financial misstatements and lower audit fees.

and protecting data (Grabski, Leech, and Schmidt 2011).⁸ In contrast, data analytics is concerned with transforming data into useful information by allowing users to better integrate and analyze unstructured data as well as illustrate patterns and anomalies (E.g., Davenport 2006; Cao, Chychyla, and Stewart 2015). See Figure 1. As such, it is unclear whether the findings on ES generalize to data analytics and more advanced technologies. While the change from manual data collection and storage (e.g., Excel) to ES is likely to improve data integrity, the focus on data analytics is likely to result in better insights and decision-making.

Hypothesis Development

I propose that four important features of a data analytics strategy can improve the quality of information used in firm decision-making by reducing the marginal cost of processing data into useful information. I begin by describing these features in the context of operating decisions, which contribute to firms' core objectives.⁹ First, data analytics can make it easier to prepare and incorporate unstructured data from different sources into existing analyses. For example, firms can analyze social media activity in deciding on a potential acquisition target or new building location. Second, data analytics can help users present information in a variety of ways that make it easier for decision-makers to understand. For example, visualizations of equipment productivity data can aid managers in identifying inefficiencies. Third, data analytics can simplify data modeling and its adaptation to the arrival of new information. For example, as managers collect information on the progress of different research and development (R&D) projects, they can easily create and adjust statistical analyses to determine which projects are the most viable and which to discontinue.

⁸ A related stream of research has looked at firms with internal control weaknesses (ICW) and IT control weaknesses (ITCW). Firms with ICWs issue less accurate management guidance (Feng, Li, and McVay 2009, Li, Peters, Richardson, and Watson 2012) and have lower earnings quality (Doyle, Ge, and McVay 2007; Ashbaugh-Skaife, Collins, Kinney, and LaFond 2008). Although ITCW can impact data integrity, only 216 firm-years in my sample have these underlying issues (e.g. systems access). Inferences are robust to dropping these firm-years.

⁹ There are numerous ways in which data analytics could be applied within the firm. Appendix C contains additional examples as they relate to the specific outcomes that I examine in the paper.

The fourth important element of a data analytics strategy, which is not a feature of the technology itself, is that the use of the technology is endorsed by firm leadership. Technology initiatives that are implemented by a firm's executives are more likely to gain cooperation at all levels of the organization and ultimately achieve their intended outcome (Venkatesh et al. 2003; Neufeld et al. 2007; Lim, Stratopoulos, and Wirjanto 2012). With executive support, data analytics is likely to help firms process more data to its fullest extent, improving internal information quality.

My first hypothesis is stated in the alternative form:

HYPOTHESIS 1. Employing a data analytics strategy improves internal information quality, leading to better operating decisions.

Although anecdotal evidence suggests that improving operating decisions is the primary purpose of using data analytics, the capabilities described above could also generate financial reporting benefits. These benefits could arise through spillovers in information quality from operating decisions. For example, predictive analytics intended to better forecast product demand could help managers model future earnings, and incorporating additional information gathered on customer creditworthiness could allow managers to improve estimates of the allowance for doubtful accounts. Financial reporting benefits could also arise if firms have enough resources to deploy data analytics directly in financial reporting processes, as recent articles suggest (e.g., Morganti, Schloemer, and Panth 2019; Maurer 2020). My second hypothesis is stated in the alternative form:

HYPOTHESIS 2. Employing a data analytics strategy improves internal information quality, leading to better financial reporting decisions.

While there are many potential benefits of a data analytics strategy, there are also several challenges that could prevent firms from realizing these benefits (Brynjolfsson and McElheran 2016) and even ways in which data analytics could negatively impact decision-making. To realize

the rewards of a data analytics strategy, firms need to go beyond just purchasing the software. They must have the appropriate infrastructure, including effective enterprise systems, to gather and transfer data as well as employees with the necessary skills to use the technology. Furthermore, they must establish a culture that is open to data-driven decision-making. That is, employees at every level of the decision-making process need to have the desire to use the technology and the mindset to learn from it, even if it requires change or reveals inefficiencies. Addressing these infrastructure, employee, and cultural challenges can divert time and money from other value-enhancing activities, potentially offsetting any upside from using the technology. Moreover, firms using sophisticated technology face the risks of information overload and overreliance. Data analytics makes it easier to incorporate more types of information, but too much information can be difficult for employees to process (Edmunds and Morris 2000). Lastly, there is still a need for human judgment, particularly with factors that cannot be easily quantified. Overreliance on data without considering these factors could be detrimental to decision-making. Thus, it remains an empirical question whether data analytics positively impacts internal information quality.

3. Research Design

Measuring a Data Analytics Strategy

To measure a data analytics strategy, I use textual analysis to identify firms that discuss data analytics topics on earnings conference calls and consider firms to adopt a data analytics strategy once they do so. I focus on conference calls as opposed to 10-K/10-Q filings as managers generally have more discretion over the content of conference calls (Davis, Ge, Matsumoto, and Zhang 2015). Consistent with this notion, based on a random sample of 100 10-K/10-Q filings containing data analytics terms, I observe that the majority of these disclosures pertain to the firms' product or service rather than the firms' use of the technology in decision making (e.g., Oracle

Corporation’s 2018 10-K contains the phrase, “*We offer big data solutions...*”). I begin by compiling a broad list of keywords that includes data analytics synonyms, features of the technology, and names of popular software providers. See Table 1 panel C. Based on the prevalence of the most direct terms capturing my construct of interest, “data analytics”, and the input to data analytics, “big data”, I focus on these two terms to identify a data analytics strategy in the main analyses. I explore the sensitivity of my results to using the complete dictionary of keywords in section 4. Recent papers use similar textual analysis methods to proxy for technological involvement (e.g. Chen and Srinivasan 2022; Yen and Wang 2021).

Using disclosure to identify a strategic focus on data analytics allows for a larger sample, but it also requires certain assumptions. The first assumption is that firms that do not discuss data analytics topics on conference calls do not employ a data analytics strategy. Although firms can certainly use data analytics without disclosure, focusing on firms that disclose their use serves to capture data analytics strategies that are supported and emphasized by the firm’s executives. Furthermore, to the extent that the benefits of data analytics are not conditional on disclosure, failure to identify certain data analytics strategies due to a lack of disclosure should bias against finding results.

The second assumption is that firms that discuss data analytics actually adopt the technology. That is, disclosures of data analytics use are not “cheap talk”. To provide some assurance that my measure captures a true emphasis on technology, I obtain job postings data from LinkUp.com to see whether firms hire employees with data analytics skills during the periods in which their disclosures suggest they employ a data analytics strategy.¹⁰ For each firm-year with

¹⁰ I do not use the job postings data in my main analyses because the data does not provide consistent coverage until 2014. By using the conference call measure, I increase my sample size, which is particularly important as the discussion of data analytics on conference calls begins to increase in 2010. See Figure 2.

job postings data, I calculate the percentage of jobs filled that list terms and skills related to data analytics in the description, *PctDAJobs*, to capture whether firms are hiring consistently with a data analytics strategy.¹¹ To account for the likely non-linear relation between hiring and disclosure, I create an indicator variable, *HighPctDAJobs*, equal to one if *PctDAJobs* is above the pooled median and zero otherwise. As seen in Appendix E, both *PctDAJobs* and *HighPctDAJobs* are significant determinants of my conference call measure, *DataAnalytics*. These results provide some assurance that the disclosure of data analytics on conference calls is associated with firms undertaking real actions consistent with a data analytics strategy. In other words, these disclosures (on average) reflect true technological involvement.

To gain comfort that conference call disclosures capture my construct of interest, I read the transcripts for a random sample of 100 disclosures containing the keywords “data analytics” or “big data”. Appendix D summarizes this hand-collected sample and provides examples of different types of disclosures. Based on this sample, 71% of data analytics disclosures capture instances in which discussions of the technology suggest it is being used to improve decision making within the firm. Among this 71%, 59% of disclosures suggest that the firm uses data analytics to improve general (i.e., non-product related) decisions. These disclosures differ in specificity and are made by firms operating various business models.¹² See examples A1 and A2 in Appendix D. The other 13% of disclosures in this category suggest that the firm uses data analytics to improve a product or service. For example, Humana’s CEO discusses how they use member analytics to improve their telehealth recommendations (Appendix D example B). I classify another 5% of disclosures

¹¹ Refer to Appendix E for details.

¹² While some disclosures in the hand-collected sample provide details of the firm’s data analytics goals (e.g., sales and marketing initiatives, Appendix D Example A1) or specific applications (e.g., in 2014, Magellan Health Inc. discusses using data analytics to “refine clinical programs and the model of care”), the majority of disclosures are high-level or general in nature (e.g., “applying digital technologies like... big data analytics to improve decision quality”, Appendix D example A2).

as “vague” (e.g., instances in which a manager characterizes an analyst’s question as a “big data project”). It is unclear whether these vague disclosures suggest the use of data analytics for decision making.

Notably, a portion of disclosures containing the terms “data analytics” and “big data” are made by firms that 1) provide analytics as a product or service (12%) or 2) provide a product or service related to analytics, e.g., servers or semiconductors (11%). For example, Integrated Device Technology’s CEO discusses how demand for one of their products is driven by trends in data analytics (Appendix D example D). As these firms arguably have lower costs to adopt data analytics for internal decision making, the likelihood that they do so is high. However, for these firms, disclosures concerning data analytics are unlikely to indicate the timing of when they adopt the technology for internal decision making (i.e., what I intend to capture). Thus, including these disclosures likely adds noise to my measure. To reduce this noise, I rely on an earlier observation that the majority of data analytics disclosures in the 10-K/10-Q filings come from this same group – firms providing analytics or analytics-related products or services. Therefore, I exclude periods in which managers simultaneously mention “data analytics” or “big data” in the 10-K/10-Q and the conference call from my sample.¹³

Determinants of Data Analytics Strategy

I begin my analyses by modeling the decision to adopt a data analytics strategy based on prior literature on IT adoption using equation (1), where *AdoptDataAnalytics* is an indicator equal

¹³ In my random sample of 100 conference call disclosures, the exclusion of simultaneous 10-K/10-Q disclosures eliminates 40% of these noisy observations (i.e., from providers of analytics or analytics-related products/services; groups C and D in Appendix D). I do not attempt to eliminate these disclosures by industry as within my hand-collected sample, these groups comprise 10 different industries, even at the two-digit SIC level. Inferences are similar if I do not exclude firm-years with overlapping conference call and 10-K/10-Q disclosures. Specifically, adoption of a data analytics strategy is associated with improvements with investment efficiency (R&D rather than acquisitions), resource utilization (both fixed and intangible), and forecast accuracy. However, in this specification, a data analytics strategy is no longer associated with a reduced likelihood of restatement.

to one in the first year a firm mentions “data analytics” or “big data”. This literature suggests that the successful implementation of innovative IT is driven by a firm’s leaders, organizational structure, and external environment (Kobelsky et al. 2008; Oliveira and Martins 2011). To capture the effects of firm leadership, I include an indicator variable equal to one if any member of the board of directors has prior work experience in the technology industry, *BoDTechExpert*.¹⁴ I also include board size, *BoDSize*, and independence, *BoDInd*, as broader measures of firm leadership and governance. Next, I include fundamental firm characteristics, *FirmChars*, that are likely to influence the ability and willingness to adopt a data analytics strategy (i.e., size, age, growth, cash, and profitability). I also include variables for organizational change, *OrgChange*, and firm complexity, *Complexity*. While it is likely to be harder for firms undergoing significant change (e.g., business expansion) and inherently complex firms to successfully implement new technology (Oliveira and Martins 2011), it is also possible that these firms could have a greater need for it. Following Feng et al. (2009), *OrgChange* is the principal component of asset growth, sales growth, leverage, and merger activity, and *Complexity* is the principal component of geographic segments, operating segments, the existence of foreign transactions and restructuring activity.

$$\begin{aligned}
P(\text{AdoptDataAnalytics}_{i,t}=1) = & \beta_1 + \beta_2 \text{BoDTechExpert}_{i,t-1} + \beta_3 \text{BoDSize}_{i,t-1} + \beta_4 \text{BoDInd}_{i,t-1} \\
& + \beta_5 \text{OrgChange}_{i,t-1} + \beta_6 \text{Complexity}_{i,t-1} + \beta_7 \text{HighDAIndustry}_{t-1} + \beta_8 \text{DATProviderGeo}_{i,t-1} \\
& + \beta_9 \text{TimeTrend}_t + \sum \beta \text{FirmChars}_{i,t-1} + \varepsilon_{i,t}
\end{aligned} \tag{1}$$

To capture the effects of the operating environment, I include variables based on industry and location. Based on a review of technology articles, I identify a list of eight industries where data analytics is expected to have a high impact. *HighDAIndustry* is an indicator variable equal to one if the firm operates in the Software, Hardware, Communications, Medicine, Retail,

¹⁴ Prior literature uses similar measures based on executives’ backgrounds to capture technical competency (e.g., Boritz, Efendi, and Lim 2017).

Professional Services, Real Estate, or Insurance industry. Firms in these industries are more likely to employ a data analytics strategy for various reasons (e.g., the costs to do so are relatively low, their primary operations rely heavily on data, etc.). For geographic effects, I include an indicator for whether a firm is located in the same area as data analytics software provider, *DATProviderGeo*, as knowledge and technology spillovers can be facilitated by physical proximity (e.g., Tambe and Hitt 2014; Dougal, Parsons, and Titman 2015). Lastly, I include a time trend variable, *TimeTrend*, to capture the increasing popularity of data analytics over time.

Data Analytics Strategy and Internal Information Quality

I test whether a data analytics strategy improves internal information quality, resulting in better operating and financial reporting decisions, using equation (2).¹⁵ *DataAnalytics* is an indicator equal to one if the firm has mentioned “data analytics” or “big data” in the current period or any prior period.¹⁶

$$\begin{aligned}
 IIQ_{i,t} = & \beta_1 + \beta_2 DataAnalytics_{i,t} + \sum \beta DADeterminants_{i,t-1} + \sum \beta IIQSpecificControls_{i,t} \\
 & + \sum \beta IndustryFE + \sum \beta YearFE + \varepsilon_{i,t}
 \end{aligned} \tag{2}$$

One challenge with identifying the relation between a data analytics strategy and internal information quality archivally is that a data analytics strategy is a choice of the firm. This choice is potentially driven by firm characteristics that could simultaneously impact internal information quality. Therefore, I employ entropy balancing to correct for selection bias following Hainmueller (2012). Specifically, I balance the treatment and control samples on the first three moments of the

¹⁵ Balance sheet controls are measured as of the beginning of the year, and income statement controls are measured over the current year to capture the underlying economics of the current period.

¹⁶ Based on reading firm disclosures, it appears that most firms discuss data analytics in a way that implies they have already deployed the technology. Thus, I consider firms to employ a data analytics strategy beginning in the first period they discuss data analytics on a conference call. Since several of my dependent variables, which are measures of firm performance, are also autocorrelated, I cluster standard errors in equation (2) by firm. I discuss alternative clustering specifications in section 4.

determinants of data analytics adoption (i.e., the variables included in equation (1)). I triangulate my results using alternative specifications in section 4.

Hypothesis 1 predicts that improvements in internal information quality will manifest in better operating outcomes, and Hypothesis 2 predicts that improvements in internal information quality will manifest in better financial reporting outcomes. Below, I discuss the specific operating and reporting measures that I examine.

Operating Efficiency Consequences

The first set of *IIQ* proxies that I examine are measures of operating efficiency to capture areas in which data analytics has the largest potential impact. I begin by examining whether a data analytics strategy improves a firm's investment decisions. Allocating resources to the right investments is an important task for profit maximizing firms and one that is well-suited to benefit from data analytics. For example, firms can use predictive modeling to compare the expected payoffs of different R&D projects or acquisition targets, improving the quality of information related to potential investments. These improvements in information quality can facilitate better investment decisions. Furthermore, prior research finds that moral hazard and adverse selection problems can also prevent managers from investing efficiently (e.g., Biddle, Hilary, and Verdi 2009). To the extent that improvements in information quality facilitate monitoring and reduce information asymmetry, they could also mitigate inefficient investment stemming from these agency problems.¹⁷

I measure investment efficiency using *UEInvestment*, which captures the unexpected portion of investments in capital expenditures, acquisitions, and R&D. Prior literature models the expected (i.e. efficient) level of investment based on investment opportunity, proxied by *TobinsQ*,

¹⁷ More specifically, an improvement in monitoring could help prevent over-investment caused by moral hazard, and a reduction in information asymmetry could help prevent under-investment caused by financing constraints.

cash flows from operating activities, asset growth, and prior investments (e.g., Goodman et al. 2014; Shroff 2017). I adapt this model as follows. First, I scale the level of investment by total assets at the beginning of the year rather than average total assets to avoid any mechanical relation between the current period investment and average total assets. Second, I use sales growth, *SGrowth*, instead of asset growth (consistent with Biddle et al. 2009), again to avoid any mechanical relation between the current period investment and asset growth. The unexpected investment, *UEInvestment*, is the absolute value of the residual from equation (3), estimated at the industry-year level.¹⁸ An underlying assumption of this measure is that the efficient level of investment is determined by the collective actions of industry peers (i.e., on average, firms in the industry invest at the optimal rate based on their available resources, investment opportunity set, and current rate of growth).¹⁹ In untabulated regressions of future return on assets and buy-and-hold abnormal returns, I find that *UEInvestment* is negatively associated with future performance. This is consistent with prior literature (e.g., Goodman et al., 2014) and the interpretation that larger values of *UEInvestment* are indeed inefficient.

$$Investment_{i,t} = \beta_1 + \beta_2 TobinsQ_{i,t-1} + \beta_3 CFO_{i,t} + \beta_4 SGrowth_{i,[t-1,t]} + \beta_5 Investment_{i,t-1} + \varepsilon_{i,t} \quad (3)$$

I examine the firm's total investments of capital expenditures, acquisitions, and R&D together (*UEInvestment*) as well as the individual components separately (*UECapex*,

¹⁸ Another common approach to measuring investment efficiency looks at firms' sensitivity to their investment opportunities, proxied by Tobin's-Q (Hayashi 1982). The model I employ incorporates the fundamentals of investment-Q sensitivity and considers additional factors that prior literature has shown to impact investment decisions, such as growth and the cost of financing (e.g., Fazzari, Hubbard, and Petersen 1988). These proxies of investment efficiency each have different strengths and limitations. Therefore, I carefully consider each aspect of the models used by prior literature. Moreover, I attempt to validate the interpretation of my measure using ex-post performance. Lastly, I triangulate the inferences on overall operating efficiency using alternative measures that do not rely on such models, such as asset turnover.

¹⁹ One concern is that the extent to which firms in any industry invest efficiently depends on the prevalence of data analytics in that industry. While I expect that industry peers are a better benchmark for efficient investment in industries where a high percentage of firms employ data analytics, I do not expect the underlying *drivers* of investment (i.e. the variables in the model) to differ based on the prevalence of data analytics. In my main analyses, I use industry fixed effects in regressions where *UEInvestment* is the dependent variable to mitigate this concern.

UEAcquisitions, and *UER&D*).²⁰ If a data analytics strategy helps decision-makers invest more efficiently, I expect the coefficient on *DataAnalytics* (β_2 in equation (2)) to be negative.

Unexpected investment captures one aspect of operating efficiency, with respect to a firm's expenditures, but it does not capture how well a firm leverages its *existing* resources. Therefore, I next examine various measures of resource utilization, which should also benefit from data analytics. For example, firms could improve fixed asset turnover by visualizing data on equipment productivity; firms could also improve human capital utilization by incorporating different types of data (e.g., from employee surveys) into hiring and staffing decisions. Since data analytics has the potential to impact efficiency in multiple areas, I begin broadly with total asset turnover, *ATO*, and I also examine more specific measures of efficiency related to different resources. Namely, I assess fixed asset turnover, *FixedAssetTO*, and intangible asset turnover, *IntangibleAssetTO*, as data analytics could help firms redeploy underutilized fixed or intangible assets. I also consider revenue over average headcount, *RevenueEmployee*, following Banker, Chang, and Natarajan (2005). For *ATO*, *FixedAssetTO*, *IntangibleAssetTO*, and *RevenueEmployee*, I expect the coefficient on *DataAnalytics* (β_2 in equation (2)) to be positive.

Financial Reporting Consequences

In the next section, I explore the effects of a data analytics strategy on internal information quality by examining financial reporting outcomes. As predictive ability is one of the primary strengths of data analytics, I begin by examining properties of management earnings forecasts. Specifically, I test whether a data analytics strategy increases managers' confidence in forward-

²⁰ When I examine R&D alone, I modify equation (3) to use cash from operations with the current period R&D expense added back. To the extent that the current period R&D expense is predominantly cash, this mitigates the negative mechanical correlation with the current period cash from operations.

looking estimates, making them more likely to issue an earnings forecast, *Forecast*.²¹ If so, I expect the coefficient on *DataAnalytics* (β_2 in equation (2)) to be positive. Among firm-years with earnings forecasts, I also test whether forecasts are more accurate or timely, consistent with improvements to existing forecasting processes. *AvgError* is the absolute value of the difference between actual and forecasted EPS, scaled by beginning of period price following prior literature (e.g., Ajinkya, Bhojraj, and Sengupta 2005; Goodman et al. 2014).²² *Horizon* is the number of days between the forecast date and the fiscal period end.²³ If data analytics helps managers examine forecasting data more effectively and efficiently, I expect the coefficient on *DataAnalytics* (β_2 in equation (2)) to be negative for *AvgError* and positive for *Horizon*.

Lastly, I explore whether a data analytics strategy improves information quality in the financial statements. Managers make a variety of estimates that appear in the financial statements in the form of accruals, which could also derive benefits from the use of data analytics. For example, predictive analytics on utilities or employee retirement contributions could improve the accuracy of accrued expenses. Thus, I examine the impact of data analytics on accruals quality, *AccQuality*, which captures how well economic estimates map into cash flow realizations, measured following Dechow and Dichev (2002) and McNichols (2002). Since *AccQuality* is a measure based on the standard deviation of residuals, lower values of *AccQuality* are better. I also examine the likelihood of non-reliance restatements as a measure of earnings quality. A data analytics strategy could decrease the likelihood of restatement through improved estimates or by facilitating monitoring within the firm. For example, the visualization of account patterns could

²¹ For binary outcomes (i.e. *Forecast*, *Restate_NF*, and *Restate_F*) I report a linear probability model to avoid the incidental parameters problem associated with including fixed effects in logit or probit models (Cameron and Trivedi 2005). This approach is consistent with prior research (e.g. deHaan, Madsen, and Piotroski 2017).

²² If managers issue a range forecast, I utilize the upper bound of the range as managers' expected earnings following Ciconte, Kirk, and Tucker (2014).

²³ I exclude forecasts issued after the fiscal year end date, as incentives may differ for pre-announcement strategies.

help managers detect and investigate anomalies or unexpected trends (e.g., a transitory spike in accounts receivable). Thus, I examine both non-fraud-related restatements, *Restate_NF*, and fraud-related restatements, *Restate_F*. If data analytics helps managers improve earnings quality, I expect the coefficient on *DataAnalytics* (β_2 in equation (2)) to be negative for *AccQuality*, *Restate_NF*, and *Restate_F*.

Control Variables

Across all internal information quality regressions, I include the determinants of a data analytics strategy as control variables. These determinants capture a broad set of important firm characteristics that could impact various measures of operating efficiency and financial reporting quality. In addition, I include a distress indicator (*Loss*), as differences in operating environments for loss firms are likely to impact both operating efficiency and reporting quality. I also include the number of analysts following the firm (*Ln_Analysts*), representing the external information environment, as well as additional measures of governance including institutional ownership (*IO*) and an indicator for whether the firm has a Big 4 auditor (*Big4*). Prior literature shows that these factors can impact operating performance, management forecast properties, and earnings quality (e.g., Ajinkya et al. 2005, Chen, Cheng, and Wang 2015).

Next, I include control variables specific to each measure of internal information quality. For investment efficiency, I add determinants of investment found by prior literature such as the firm's asset composition, tangibility, and various measures of the firm's ability to invest (i.e., financial leverage, dividends, cash flows to sales, and slack) (Biddle and Hilary 2006; Biddle et al. 2009). For resource utilization, I add profit margin to account for different operating strategies naturally yielding different levels of asset turnover. For all management forecast properties, I include measures to capture uncertainty such as analyst forecast dispersion and the surprise implied

by the management forecast, following Feng et al. (2009). I control for whether the firm issued equity during the year, as firms have incentives to reduce information asymmetry to lower the cost of capital (e.g., Healy and Palepu 1993), and high litigation risk, which may discourage disclosure (Healy and Palepu 2001). I also include the presence of internal control weaknesses, which increases forecast errors (Feng et al. 2009). Lastly, I include forecast horizon in examining forecast error, as forecasts issued earlier in the year are likely to be less accurate. For earnings quality, I add the length of the operating cycle and the magnitude of accruals, following Dechow and Dichev (2002), as these are significant determinants of accruals quality. I also include the presence of internal control weaknesses, as these are likely to reduce the quality of accounting estimates (Doyle et al. 2007). Lastly, for all highly autocorrelated measures (*ATO*, *FixedAssetTO*, *IntangibleAssetTO*, *RevenueEmployee*, *Forecast*, and *Horizon*), I include the lagged value of the dependent variable (DV_{t-1}). All variables are defined in Appendix A.

4. Data and Results

Sample Selection

My sample begins with 132,760 firm-years with basic financial information (i.e., total assets) on Compustat for the period 2004-2018. I require firm-years to have all the necessary data to calculate the measures in my determinants model. 57,252 firm-years have non-missing values for zip code, SIC code, and all of the variables necessary to compute *OrgChange*, *Complexity*, *Size*, *Age*, *BTM*, *Cash*, and *RoA*. After merging the Compustat sample with BoardEx for *BoDSize*, *BoDInd*, and *TechBoDExp*, I have 33,774 firm-years. Lastly, I merge the sample with conference call transcripts from StreetEvents for data on my variable of interest, *DataAnalytics*. My main sample consists of 13,619 firm-years. Compared to the Compustat universe, my sample consists

of larger, more established firms. The firm-years in my sample are also characterized by fewer losses (untabulated).²⁴

Descriptive Statistics

Table 2 panel A presents descriptive statistics for the variables included in the determinants model. Consistent with the growing popularity of data analytics over time, firm-years with a data analytics strategy group occur on average later in the sample period (*TimeTrend*). Data analytics strategy firm-years are characterized by larger (*BoDSize*) and more independent boards (*BoDInd*). These years also exhibit more complexity (*Complexity*), and growth (lower *BTM*), potentially capturing differences in the need for sophisticated technology. Firms employing a data analytics strategy tend to be larger (*TotalAssets*), with median total assets of \$2.4B, and there are more firms employing a data analytics strategy in the high-impact industries (*HighDAIndustry*), consistent with the cost-benefit ratio being lower in these industries. Lastly, firm-years with a data analytics strategy are characterized by closer geographic proximity to a software provider (*DATProviderGeo*) and a higher likelihood of having a technology expert on the board of directors (*BoDTechExpert*).

Table 2 panel A also presents descriptive statistics for my dependent variables partitioned by *DataAnalytics*. The univariate tests of means provide mixed evidence on operating efficiency. For example, while firm-years with a data analytics strategy exhibit more efficient capital expenditures (*UECapex*), they also exhibit lower intangibles turnover (*IntangibleAssetTO*) and revenues generated per employee (*RevenueEmployee*). Turning to the financial reporting quality variables, firm-years with a data analytics strategy have more earnings forecasts (*Forecast*), and

²⁴ The median total assets of firm-years in Compustat (my sample) is \$226 million (\$1.4 billion). The median age of firms in Compustat (my sample) is 8 (22). Approximately 46% (26%) of firms in Compustat (my sample) report negative net income. Lastly, the median book-to-market ratio for firm-years in Compustat (my sample) is .52 (.54).

the forecasts issued during these years are characterized by smaller errors (*AvgFError*) and longer horizons (*Horizon*). Data analytics strategy firm-years also appear to better accruals quality (lower *AccQuality*) and fewer restatements (*Restate_NF*). Overall, the univariate statistics provide some support of my hypotheses that a data analytics strategy improves internal information quality.

Results of Determinants Model

Table 3 reports the results of the determinants analysis. As predicted, firm-specific characteristics are significant determinants of data analytics adoption. Consistent with the findings of Chen and Srinivasan (2022), I find that larger firms (*Size*) are more likely to adopt a data analytics strategy. Firms that are complex (*Complexity*), growing (lower *BTM*), and have more resources (*RoA*) are more likely to adopt a data analytics strategy, capturing both the need and ability to leverage the technology. Firms are also more likely to adopt a data analytics strategy when they are not undergoing significant organizational change (*OrgChange*). As theory predicts, firm leadership and governance also have a significant impact on a firm's data analytics strategy. Specifically, firms with larger boards (*BoDSize*) and firms with a technology expert on the board (*BoDTechExpert*) are more likely to adopt a data analytics strategy. Lastly, firms are more likely to adopt a data analytics strategy when they operate in certain high-impact industries (*HighDAIndustry*).

To better illustrate the economic significance of the determinants, Table 3 Column (2) presents the marginal effect of an interquartile change in each variable on the predicted probability, with all other variables held at their mean (scaled by the unconditional probability of adoption in any given period for interpretation). This analysis reveals that outside of industry membership and firm size, firm leadership is one of the most important determinants. For example, having a board member with experience in the technology industry more than doubles the likelihood that a firm

adopts a data analytics strategy that year. The area under the Receiver Operating Characteristic (ROC) curve is .7511, indicating a “fair” ability to identify the decision to adopt a data analytics strategy (Hosmer and Lemeshow 2000). Overall, firm, leadership, and operating environment characteristics are all important determinants of whether a firm adopts a data analytics strategy, and more firms are doing so over time.

Data Analytics Strategy and Operating Efficiency

Next, I examine whether a data analytics strategy improves internal information quality using measures of operating efficiency (Hypothesis 1). The results of my investment efficiency analysis are presented in Table 4. Column (1) reports the regression of unexpected investment on *DataAnalytics* using the aggregate investment measure, the sum of capital expenditures, acquisitions, and R&D. Columns (2) - (4) report the impact of data analytics on each type of investment separately. The coefficient on *DataAnalytics* in Column (1) is negative and significant ($\beta = -0.021$; t-stat = -3.76), suggesting that when firms employ a data analytics strategy, they exhibit smaller deviations from the efficient investment level. This decrease represents 25.4% of the unconditional mean of unexpected investment (.082). Columns (2) - (4) show that this result is primarily driven by more efficient acquisitions (rather than capital expenditures or R&D).^{25,26} Of the three investments, acquisitions likely warrant the most thorough due diligence, providing greater opportunity for improvement through data analytics (see e.g., Vandeventer, Vigiano, and Guschwan 2016). Turning to the control variables, firms undergoing organizational change have larger deviations from expected levels of investment (determinants of data analytics are not

²⁵ Results are not sensitive to adjusting capital expenditures for proceeds from the sale of property, plant, and equipment (as in e.g., Richardson 2006).

²⁶ In untabulated analyses, I separately examine whether data analytics helps mitigate over-investment, under-investment, or both. The results appear to be driven by data analytics reducing over-investment (e.g., among firm-years with positive values of *UEInvestment*, the coefficient on *DataAnalytics* is -0.037 (t-stat = -3.38).

tabulated for brevity). Firms with more volatile past investments (*SDInvestment*) also have larger deviations from the efficient level of investment.

I next explore how a data analytics strategy impacts operating efficiency with respect to the firm's existing resources. These results are presented in Table 5. The impact of a data analytics strategy on total asset turnover is not statistically significant, potentially because total assets include a wide range of asset types. Thus, in columns (2) – (4), I examine a subgroup of turnover measures, capturing the revenue generating ability of specific productive resources. Firms implementing a data analytics strategy see improvements in both fixed asset turnover ($Ln_FixedAssetTO$ $\beta = 0.021$, $t\text{-stat} = 2.22$) and intangible asset turnover ($Ln_IntangibleAssetTO$ $\beta = 0.051$, $t\text{-stat} = 2.93$). These results suggest that data analytics helps shed light on how to improve the utilization of various types of existing resources, from patents to equipment. In terms of economic magnitude, improvements in operating efficiency are significant, 2.1% ($Ln_FixedAssetTO$) to 5.2% ($Ln_IntangibleAssetTO$; e.g. $\exp(\beta) - 1 = \exp(0.051) - 1 = 0.052$). One concern is that the types of firms with fewer recorded assets may be more likely to adopt a data analytics strategy, and these firms mechanically have higher asset turnover ratios. As I control for size, measured as the natural log of total assets, directly in my regression (determinants of data analytics are not tabulated for brevity) and utilize entropy balancing, this is unlikely to explain the results.²⁷ Moreover, to the extent that this variation is captured by industry, using industry fixed effects should mitigate this concern.

²⁷ In untabulated analyses, I include control variables capturing firm business strategy (i.e., Defenders versus Prospectors) following the method in Higgins, Omer, and Phillips (2015) to further mitigate the concern that firm type drives both the decision to adopt data analytics and improvements in performance. When I include these control variables in Tables 4-6, the effects of a data analytics strategy on management forecast error (Table 6) is no longer significant, but all other inferences remain the same. When these controls are included, firms adopting a data analytics strategy also exhibit more efficient R&D investments (table 4) and total asset turnover (table 5) ($p\text{-value} < .10$).

Overall, the results in Tables 4 and 5 are consistent with a data analytics strategy having a positive impact on internal information quality, as evidenced by improvements in operating efficiency. Firms employing a data analytics strategy invest more efficiently, particularly when it comes to investments that require extensive analysis such as acquisitions. These firms also better use their resources to generate revenues, and this is apparent for both fixed and intangible assets.

Data Analytics Strategy and Financial Reporting

In Table 6, I investigate improvements in internal information quality using outcomes of financial reporting decisions (Hypothesis 2). I begin with the results for management forecasts in columns (1) – (3). For firms that issue at least one earnings forecast during the sample, a data analytics strategy does not appear to have an impact on the likelihood of issuing an earnings forecast in any given period. As documented in prior literature, the decision to issue a forecast is highly autocorrelated as indicated by the coefficient on DV_{t-1} . On the other hand, within periods that have an earnings forecast, a data analytics strategy does appear to improve average forecast error ($AvgError$ $\beta = -0.003$, $t\text{-stat} = -2.97$), consistent with data analytics helping managers form better forward-looking estimates. In general, control variables load in the expected directions. For example, the negative coefficient on *ForecastDispersion* in column (1) is consistent with the provision of a forecast decreasing in uncertainty, and the positive coefficients on *ForecastDispersion* and *ForecastSurprise* in column (2) are consistent with forecast errors increasing in uncertainty.

Columns (4) through (6) of Table 6 report the results for earnings quality measures of internal information quality. The coefficients on data analytics for accruals quality, *AccQuality*, and non-fraud related restatements, *Restate_NF*, are statistically insignificant. However, there is some evidence that firms that adopt a data analytics strategy are less likely to subsequently restate

their financial statements due to fraud ($Restate_F$ $\beta = -0.002$, $t\text{-stat} = -1.82$). This result suggests that the implementation of data analytics may serve as a deterrent to intentional misreporting by increasing visibility within the firm. Overall, the evidence in Tables 4-6 indicates that firms primarily use data analytics to improve operating efficiency. In Table 6, there is some evidence that firms also use data analytics to improve reporting quality, but taken together with anecdotal evidence, there appears to be room for broader use of data analytics within financial reporting.

Emphasis of a Data Analytics Strategy

In this section, I explore whether the consequences of a data analytics strategy vary based on the extent to which firms emphasize the technology. I expect that firms placing a greater emphasis on data analytics invest more resources in reaping its benefits and thus, are more likely to overcome implementation challenges and better leverage the technology.

To approximate the emphasis on a data analytics strategy, I use the cumulative percentage of conference calls on which a firm has discussed data analytics since adoption. In my main analysis, if the firm discusses data analytics on any conference call during the year, I consider them to adopt a data analytics strategy going forward. This approach does not take into account variation in the frequency of firms' data analytics disclosures. For example, while 79% of data analytics firms discuss the technology in one quarter of the adoption year, 21% discuss the technology in more than one quarter, and 3% even do so in all four quarters. Similarly, while 58% of data analytics firms discuss the technology in fewer than three years, 32% discuss the technology in four to six years, and 3% even do so in ten or more years during the sample period (Table 1 panels A-B). These statistics suggest that some executives place a greater emphasis on the technology than others. To exploit this variation, I create separate treatment variables for low (high) emphasis firms, *DataAnalyticsLow* (*DataAnalyticsHigh*), using the median cumulative percentage of

conference calls on which the firm has discussed data analytics since adoption. I estimate equation (4) (in which I replace *DataAnalytics* in equation (2) with *DataAnalyticsLow* and *DataAnalyticsHigh*) to examine whether the effects of data analytics vary based on emphasis. If firms capture incremental benefits from placing a greater emphasis on data analytics, I expect 1) the coefficient on *DataAnalyticsHigh* (β_3) to be significant across a greater number of outcomes than the coefficient on *DataAnalyticsLow* (β_2), 2) the coefficient on *DataAnalyticsHigh* (β_3) to be larger in magnitude than the coefficient on *DataAnalyticsLow* (β_2) for any given outcome, or both.

$$IIQ_{i,t} = \beta_1 + \beta_2 DataAnalyticsLow_{i,t} + \beta_3 DataAnalyticsHigh_{i,t} + \Sigma \beta DADeterminants_{i,t-1} + \Sigma \beta IIQSpecificControls_{i,t} + \Sigma \beta IndustryFE + \Sigma \beta YearFE + \varepsilon_{i,t} \quad (4)$$

Table 7 reports the results of the emphasis analysis. For each measure of internal information quality in Tables 4-6, I tabulate the coefficients on *DataAnalyticsLow* and *DataAnalyticsHigh* in columns (1) and (2), respectively. In panel A, while the results for total investment efficiency (*UnexpectedTotalInvest*) are present in both low and high emphasis firms, only high emphasis firms also see independent improvements in acquisition efficiency, (*UnexpectedAcquisitions* $\beta = -0.011$, t-stat = 3.95). An F-test indicates that, for *UnexpectedAcquisitions*, the coefficient on *DataAnalyticsHigh* is significantly larger in magnitude than the coefficient on *DataAnalyticsLow*.²⁸ Similarly, the results for intangible asset turnover (*Ln_IntangibleAssetTO*) are present in both low and high emphasis firms, but only high emphasis firms also see improvements in fixed asset turnover, (*Ln_FixedAssetTO* $\beta = .024$, t-stat = 2.30).

In panel B, both high and low emphasis firms exhibit more accurate earnings forecasts (*AvgError*). However, only firms that place a high emphasis on data analytics exhibit a lower likelihood of restatement (*Restate_F* $\beta = -0.002$, t-stat = 1.87). Overall, the results in Table 7 are

²⁸ Although it appears low emphasis firms see a marginally significant improvement in R&D efficiency (*Unexpected R&D*), the difference in the coefficients on *DataAnalyticsLow* and *DataAnalyticsHigh* is statistically insignificant.

consistent with certain benefits of a data analytics strategy being greater in both scope and magnitude for firms that place a higher emphasis on the technology.

Additional Analyses

Data Analytics Keywords

In the main analyses, I refine my initial list of data analytics keywords to “data analytics” and “big data” and consider firms to employ a data analytics strategy when they discuss either of these terms on a conference call. In contrast, recent papers that use similar methods to proxy for technological involvement include up to 20 keywords (e.g., Yen and Wang 2021). To examine whether my results are sensitive to this design choice, I re-define my measure of a data analytics strategy using the complete list of keywords initially compiled. In addition to “data analytics” and “big data”, this list includes other synonyms and underlying characteristics (e.g. “visualization tools”, “unstructured data”) as well as specific software names (e.g., Tableau, Alteryx). I also add terms related to more recent developments in analytics. Specifically, I include “artificial intelligence”, “robotic process automation”, “machine learning”, “natural language processing” and “predictive algorithm”. These newer types of analytics are broadly characterized by the ability to mimic the functions of intelligent beings. They also have potential to improve internal information quality and present similar challenges (e.g., requiring a balance between employee buy-in and overreliance), albeit potentially to a greater degree. The full list of these terms is presented in Table 1 Panel C.

In untabulated analyses, when I consider a firm to employ a data analytics strategy when managers mention any of the keywords included in Table 1 panel C, the inferences from my main analyses are similar but weaker.²⁹ As technology and the related nomenclature continually evolves,

²⁹ Specifically, firm, leadership, and external characteristics are still significant determinants of a data analytics strategy. A data analytics strategy, defined using the larger dictionary of terms, is associated with improvements (p-

the inclusion of other keywords likely shifts my measure across time. For example, terms that gained popularity earlier such as “data driven” and “business intelligence tools” could indicate the use of more naïve types of analytics (e.g., Excel), whereas terms such as “artificial intelligence” likely indicate the earliest adoptions of more sophisticated analytics. Future research could examine the longer-term effects of these recent and more autonomous analytics tools such as machine learning. Given that capturing the timing of adoption is important to my research design, my primary measure only retains the keywords “data analytics” and “big data”.

Additional Disclosure Channels

Although the earnings conference call has long been a prominent voluntary disclosure mechanism rich in useful information (e.g., Frankel, Johnson, and Skinner 1999; Matsumoto, Pronk, and Roelofsen 2011), firms could also communicate their use of data analytics through other interactions with stakeholders. To the extent that there are systematic differences between firms that choose one disclosure channel over another, only looking at conference call disclosures could bias inferences about the benefits of data analytics. To mitigate this concern, I expand my search of firms’ disclosures to include transcripts from other corporate events captured in the StreetEvents database such as analyst and shareholder meetings.³⁰ Appendix F presents descriptive statistics on the prevalence of data analytics disclosures through these additional channels. There is wide variation in the frequency with which managers discuss data analytics during these events. For example, managers discuss data analytics in 2.4% (23%) of guidance conference calls (analyst meetings). In untabulated analyses, I consider firms to adopt a data analytics strategy once they

value < .10) in both operating efficiency (investment efficiency) and reporting quality (management forecast accuracy).

³⁰ In addition to the Earnings Conference Call/Presentation, I include the following events from the StreetEvents database: Guidance Conference Call/Presentation, Analyst Conference Call, Analyst Meeting, Shareholder Meeting, Conference Presentation, Corporate Conference Call/Presentation, M&A Conference Call/Presentation, Sales Conference Call/Presentation, and Other Corporate Conference Events.

discuss “data analytics” or “big data” in any of these disclosure settings, resulting in an additional 170 treatment firms. Using this alternative measure, although a data analytics strategy no longer appears to impact the likelihood of restatement, it is still associated with improvements in both operating efficiency (investment and resource efficiency) and financial reporting quality (forecast accuracy). I do not utilize these additional events to construct my primary data analytics measure as restricting the sample to firms with these events could further limit the generalizability of my results.

Falsification and Robustness Tests

In the main analyses, I carefully control for observable characteristics that could affect my specific measures of internal information quality, and I rely on entropy balancing to mitigate the effects of selection bias. To the extent that my determinants model captures sufficient variation in the decision to adopt a data analytics strategy, weighting the sample on these variables should address the selection bias concern. However, entropy balancing is not a perfect solution. It is possible that unobservable firm characteristics (e.g., firm culture) impact both a data analytics strategy and internal information quality. Any matching method also relies on correctly identifying the determinants of the treatment. Thus, I triangulate my results using alternative specifications.

I begin by removing the entropy balancing to confirm that my results are not induced by the matching procedure. When I use OLS, a data analytics strategy is still associated with improvements in both operating efficiency (investment efficiency and asset turnover) and financial reporting quality (restatements). Next, I run my analyses using firm fixed effects to control for time-invariant unobservable characteristics. When I include firm fixed effects, the investment and resource efficiency inferences again remain the same. However, data analytics does not have a significant impact on forecast error or the likelihood of restatement. I also implement the Heckman

two-stage procedure as an alternative method for correcting selection on unobservable characteristics.³¹ Using the Heckman approach, the primary improvements to investment and resource efficiency remain the same. Across these alternative specifications, the results are consistent with robust improvements to operating outcomes and weaker improvements to financial reporting outcomes. These results also illustrate that a data analytics strategy is associated with improvements in internal information quality not only across firms but also within firms.

I also explore whether my results are robust to different clustering specifications. In the main analyses, I correct for cross-sectional dependence using year fixed effects. However, to the extent that the time effect is not fixed, specifications with time fixed effects can still yield dependent standard errors (Petersen 2009). Therefore, I test whether my results are robust to clustering standard errors by industry-year and firm. I use industry-year groupings to 1) account for the spread of the technology across different industries occurring at different times and 2) to ensure enough clusters (Petersen, 2009). Using double clustering, the operating efficiency and financial reporting inferences remain the same.

Lastly, I conduct a falsification test using randomized data analytics adoption dates. Specifically, for treatment firms, I randomly generate a placebo adoption year during my sample period. I code a new *DataAnalytics* treatment variable equal to one starting in the placebo adoption year and re-estimate equation (2). Using this specification, a data analytics strategy is not associated with improvements in investment efficiency, resource utilization, forecast accuracy or restatement likelihood at conventional significance levels. Overall, these robustness tests support

³¹ In the Heckman analysis, I use *BoDTechExpert*, capturing a board member's experience in the technology industry, as the instrumental variable. As shown in the determinants analysis, this variable is a strong predictor of adopting a data analytics strategy. Furthermore, there is no reason that a board member's technology experience should affect information quality other than through the firm's technology strategy. That is, it should satisfy the exclusion restriction requirement. I employ the Heckman approach as a robustness test as it operates under strict functional form assumptions (Lennox, Francis, and Wang 2012).

the inferences from Tables 4 – 6, consistent with a data analytics strategy improving internal information quality.

Conclusion

In this study, I examine the determinants and consequences of employing a data analytics strategy. Using textual analysis of firm disclosures to identify a data analytics strategy, I find that large firms as well as growing firms are more likely to employ a data analytics strategy. The technology experience of leadership and the operating environment are also important determinants. The results of the main analyses suggest that a data analytics strategy positively impacts operating efficiency and to a lesser extent, financial reporting quality, consistent with improvements in internal information quality. Although I use entropy balancing to mitigate selection effects, I cannot rule out the possibility that firms that successfully reap the benefits of data analytics are more likely to disclose their use of the technology. However, my results are robust to a battery of alternative specifications. My findings contribute to existing Accounting and Information Systems research, documenting the benefits of data analytics for a broad sample of firms. These findings also demonstrate that on average, disclosure of technology topics is associated with true technological involvement. This is important as textual analysis of disclosure content becomes more common in empirical research. Lastly, the objectives of a data analytics strategy and the difficulties associated with its implementation are likely extend to newer forms of analytics, such as artificial intelligence. I leave the long-term effects of these more sophisticated analytics to future research.

References

- Abbasi, A., C. Albrecht, and A. Vance, and J. Hansen. 2012. Metafraud: a meta-learning framework for detecting financial fraud. *MIS Quarterly*: 1293-1327.
- Ajinkya, B., S. Bhojraj, and P. Sengupta. 2005. The association between outside directors, institutional investors and the properties of management earnings forecasts. *Journal of Accounting Research* 43(3): 343-376.
- Ashbaugh-Skaife, H., D. W. Collins, W. R. Kinney Jr, and R. LaFond. 2008. The effect of SOX internal control deficiencies and their remediation on accrual quality. *The Accounting Review* 83(1): 217-250.
- Badertscher, B., N. Shroff, and H. D. White. 2013. Externalities of public firm presence: Evidence from private firms' investment decisions. *Journal of Financial Economics* 109 (3): 682-706.
- Banker, R. D., H. Chang, and R. Natarajan. 2005. Productivity change, technical progress, and relative efficiency change in the public accounting industry. *Management Science* 51 (2): 291-304.
- Biddle, G. C., G. Hilary, and R. S. Verdi. 2009. How does financial reporting quality relate to investment efficiency?. *Journal of Accounting and Economics* 48 (2-3): 112-131.
- Boritz, J. E., J. Efendi, and J. H. Lim. 2017. The Impact of Senior Management Competencies on the Voluntary Adoption of an Innovative Technology. *Journal of Information Systems* 32 (2): 25-46.
- Brazel, J. F., and L. Dang. 2008. The effect of ERP system implementations on the management of earnings and earnings release dates. *Journal of Information Systems* 22 (2): 1-21.
- Brynjolfsson, E., and K. McElheran. 2016. The rapid adoption of data-driven decision-making. *American Economic Review* 106 (5): 133-39.
- Call, A. C., J. L. Campbell, D. S. Dhaliwal, and J.R. Moon Jr. 2017. Employee quality and financial reporting outcomes. *Journal of Accounting and Economics* 64 (1): 123-149.
- Cameron, A.C., and P. K. Trivedi. 2005. *Microeconometrics: methods and applications*. Cambridge university press.
- Cao, M., R. Chychyla, and T. Stewart. 2015. Big Data analytics in financial statement audits. *Accounting Horizons* 29 (2): 423-429.
- Chau, M., and J. Xu. 2012. Business intelligence in blogs: Understanding consumer interactions and communities. *MIS quarterly*: 1189-1216.
- Chen, W., and S. Srinivasan. 2022. Going Digital: Implications for Firm Value and Performance. Working paper, Hong Kong University of Science and Technology and Harvard Business School.
- Chen, X., Q. Cheng, and X. Wang. 2015. Does increased board independence reduce earnings management? Evidence from recent regulatory reforms. *Review of Accounting Studies* 20 (2): 899-933.
- Ciconte, W., M. Kirk, and J. W. Tucker. 2014. Does the midpoint of range earnings forecasts represent managers' expectations?. *Review of Accounting Studies* 19 (2): 628-660.

- Davenport, T. H. 2006. Competing on analytics. *Harvard Business Review* 84 (1): 98.
- Davis, A. K., W. Ge, D. A. Matsumoto, and J. L. Zhang. 2015. The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies* 20 (2): 639-673.
- Dechow, P. M., and I. D. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review* 77 (s-1): 35-59.
- Dechow, P. M., R. G. Sloan, and A. P. Sweeney. Detecting earnings management. *The Accounting review* (1995): 193-225.
- Dehaan, E., J. Madsen, and J. D. Piotroski. 2017. Do Weather-Induced Moods Affect the Processing of Earnings News?. *Journal of Accounting Research* 55 (3): 509-550.
- Dorantes, C. A., C. Li, G. F. Peters, and V. J. Richardson. 2013. The effect of enterprise systems implementation on the firm information environment. *Contemporary Accounting Research* 30 (4): 1427-1461.
- Dougal, C., C. A. Parsons, and S. Titman. 2015. Urban vibrancy and corporate growth. *The Journal of Finance* 70 (1): 163-210.
- Doyle, J. T., W. Ge, and S. McVay. 2007. Accruals quality and internal control over financial reporting. *The Accounting Review* 82 (5): 1141-1170.
- Edmunds, A., and A. Morris. 2000. The problem of information overload in business organizations: a review of the literature. *International Journal of Information Management* 20 (1): 17-28.
- Fazzari, S., R. G. Hubbard, and B. Petersen. 1988. Investment, financing decisions, and tax policy. *The American Economic Review* 78 (2): 200-205.
- Frankel, R., Johnson, M. and Skinner, D.J., 1999. An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research*, 37(1): 133-150.
- Feng, M., C. Li, and S. McVay. 2009. Internal control and management guidance. *Journal of Accounting and Economics* 48 (2-3): 190-209.
- Gallemore, J., and E. Labro. 2015. The importance of the internal information environment for tax avoidance. *Journal of Accounting and Economics* 60 (1): 149-167.
- Goodman, T. H., M. Neamtiu, N. Shroff, and H. D. White. 2014. Management forecast quality and capital investment decisions. *The Accounting Review* 89 (1): 331-365.
- Grabski, S. V., S. A. Leech, and P. J. Schmidt. 2011. A review of ERP research: A future agenda for accounting information systems. *Journal of Information Systems* 25 (1): 37-78.
- Hainmueller, J. 2012. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* (2012): 25-46.
- Healy, P. M., and K. G. Palepu. 1993. The effect of firms' financial disclosure strategies on stock prices. *Accounting Horizons* 7 (1): 1.
- Healy, P. M., and K. G. Palepu. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics* 31 (1-3): 405-440.

- Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society* (1979): 153-161.
- Hendricks, K. B., V. R. Singhal, and J. K. Stratman. 2007. The impact of enterprise systems on corporate performance: A study of ERP, SCM, and CRM system implementations. *Journal of Operations Management* 25 (1): 65-82.
- Higgins, D., Omer, T. C., & Phillips, J. D. 2015. The influence of a firm's business strategy on its tax aggressiveness. *Contemporary Accounting Research*, 32(2): 674-702.
- Higgins, M. C. 2005. Career imprints: Creating leaders across an industry (Vol. 16). John Wiley and Sons.
- Hitt, L. M., D. J. Wu, and X. Zhou. 2002. Investment in enterprise resource planning: Business impact and productivity measures. *Journal of Management Information Systems* 19 (1): 71-98.
- Hayashi, F. 1982. Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica: Journal of the Econometric Society* (1982): 213-224.
- Hosmer, D. W., and S. Lemeshow. 2000. Applied Logistic Regression, 2nd edition. John Wiley and Sons. *New York*.
- Hu, D., J. L. Zhao, Z. Hua, and M. C. Wong. Network-based modeling and analysis of systemic risk in banking systems. *MIS Quarterly* (2012): 1269-1291.
- Jaffe, A. B., M. Trajtenberg, and R. Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics* 108 (3): 577-598.
- Jennings, J. N., H. Seo, and L. Tanlu. 2019. The effect of organizational complexity on earnings forecasting behavior. Working Paper, Washington University in St. Louis, Purdue University, and Washington and Lee University.
- Keller, K. L., and R. Staelin. 1987. Effects of quality and quantity of information on decision effectiveness. *Journal of Consumer Research* 14 (2): 200-213.
- Kobelsky, K. W., V. J. Richardson, R. E. Smith, and R. Zmud. W. 2008. Determinants and consequences of firm information technology budgets. *The Accounting Review* 83 (4): 957-995.
- Kothari, S. P., A. J. Leone, and C. E. Wasley. 2005. Performance matched discretionary accrual measures. *Journal of Accounting and Economics* 39 (1): 163-197.
- Labro, E., M. H. Lang, and J. D. Omartian. 2022. Predictive analytics and centralization of authority. *Journal of Accounting and Economics*, forthcoming.
- Lennox, C. S., J. R. Francis, and Z. Wang. 2012. Selection models in accounting research. *The Accounting Review* 87 (2): 589-616.
- Lim, J. H., T. C. Stratopoulos, and T. S. Wirjanto. 2012. Role of IT executives in the firm's ability to achieve competitive advantage through IT capability. *International Journal of Accounting Information Systems* 13 (1): 21-40.
- Li, C., G. F. Peters, V. J. Richardson, M. W. Watson. 2012. The consequences of information technology control weaknesses on management information systems: the case of

- Sarbanes-Oxley internal control reports. *Management Information Systems Quarterly* 36 (1): 179-203.
- Maurer, M. 2020. Finance Teams Adapt to Closing the Books Remotely Amid Coronavirus. *The Wall Street Journal*, March 31, https://www.wsj.com/articles/finance-teams-adapt-to-closing-the-books-remotely-amid-coronavirus-11585647017?mod=hp_minor_pos7, accessed April 5, 2020.
- Matsumoto, D., Pronk, M. and Roelofsen, E., 2011. What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *The Accounting Review*, 86(4):1383-1414.
- McAfee, A., E. Brynjolfsson, T.H. Davenport, D. J. Patil, and D. Barton. 2012. Big data: the management revolution. *Harvard Business Review* 90 (10): 60-68.
- McNichols, M. F. 2002. Discussion of the quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review* 77 (s-1): 61-69.
- Morganti, T., J. Schloemer, and A. Panth. 2019. Digital Tech Poised to Revolutionize Financial Reporting. *The Wall Street Journal*, September 5, <https://deloitte.wsj.com/cio/2019/09/05/digital-technology-poised-to-revolutionize-reporting/>, accessed May 1, 2020.
- Müller, O., M. Fay, and J. vom Brocke. 2018. The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems* 35 (2): 488-509.
- Neufeld, D. J., L. Dong, and C. Higgins. 2007. Charismatic leadership and user acceptance of information technology. *European Journal of Information Systems* 16 (4): 494-510.
- Oliveira, T., and M. F. Martins. 2011. Literature review of information technology adoption models at firm level. *Electronic Journal of Information Systems Evaluation* 14 (1): 110.
- Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies* 22 (1): 435-480.
- Pincus, M., F. Tian, P. Wellmeyer, and S. Xu. X. 2017. Do Clients' Enterprise Systems Affect Audit Quality and Efficiency?. *Contemporary Accounting Research* 34 (4): 1975-2021.
- Puhani, P. 2000. The Heckman correction for sample selection and its critique. *Journal of Economic Surveys* 14 (1): 53-68.
- Richardson, S. 2006. Over-investment of free cash flow. *Review of Accounting Studies* 11 (2-3): 159-189.
- Rikhardsson, P., and O. Yigitbasioglu. 2018. Business intelligence and analytics in management accounting research: Status and future focus. *International Journal of Accounting Information Systems* 29: 37-58.
- Schneider, G. P., J. Dai, D. J. Janvrin, K. Ajayi, and R. L. Raschke. 2015. Infer, predict, and assure: Accounting opportunities in data analytics. *Accounting Horizons* 29 (3): 719-742.
- Schniederjans, M. J., J. L. Hamaker, and A. M. Schniederjans. 2010. *Information Technology Investment: Decision-Making Methodology* Second Edition. World Scientific Publishing Company.

- Shroff, N. 2017. Corporate investment and changes in GAAP. *Review of Accounting Studies* 22 (1): 1-63.
- Tambe, P., and L. M. Hitt. 2014. Job hopping, information technology spillovers, and productivity growth. *Management Science* 60 (2): 338-355.
- Vandeventer, R., A. Vigiano, and M. Guschwan. 2016. Data Analytics for M&A Due Diligence. *LinkedIn*, November 15, <https://www.linkedin.com/pulse/data-analytics-ma-due-diligence-reuben-vandeventer/>, accessed June 3, 2020.
- Venkatesh, V., M. G. Morris, G. B. Davis, and F. D. Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS Quarterly* (2003): 425-478.
- vom Brocke J., Fay M., Schmiedel T., Petry M., Krause F., Teinzer T. 2017. A journey of digital innovation and transformation: the case of Hilti. In: Oswald G., Kleinemeier M. (eds) *Shaping the digital enterprise*. Springer, Cham, pp 237–251
- vom Brocke, J., Maab, W., Buxmann, P., Maedche, A., Leimeister, J.M., and Pecht, G. 2018. Future work and enterprise systems. *Business & Information Systems Engineering*, 60: 357-366.
- Warren Jr, J. D., K. C. Moffitt, and P. Byrnes. 2015. How Big Data will change accounting. *Accounting Horizons* 29 (2): 397-407.
- Wu, L., L. Hitt, and B. Lou. 2020. Data analytics, innovation, and firm productivity. *Management Science* 66, no. 5 (2020): 2017-2039.
- Yen, J. C., and T. Wang. 2021. Stock Price Relevance of Voluntary Disclosures about Blockchain Technology and Cryptocurrencies. *International Journal of Accounting Information Systems* 40 (2021): 100499.

APPENDIX A
Variable definitions

Variable	Definition
<i>DataAnalytics</i>	Indicator equal to one if the firm has mentioned “data analytics” or “big data” on a conference call in the current year or a prior year
<i>AdoptDataAnalytics</i>	Indicator equal to one in the first year the firm mentions “data analytics” or “big data” on a conference call
<i>Determinants</i>	
<i>BoDTechExpert</i>	Indicator equal to one if any board member (including the CEO and CFO) has previously served on the board for a firm in "Information Technology Hardware" or "Software & Computer Services" industries (BoardEx) as of the beginning of the year
<i>BoDSize</i>	The number of board members as of the beginning of the year
<i>BoDInd</i>	The percentage of independent board members as of the beginning of the year
<i>Size</i>	Log of total assets (Compustat at) as of the beginning of the year
<i>Age</i>	The number of years since the first available year on Compustat (Compustat fyear - year1) as of the beginning of the year
<i>BTM</i>	Book to market ratio (Compustat (prcc_f * csho) / teq) as of the beginning of the year
<i>Cash</i>	Cash scaled by total assets (Compustat ch / at) as of the beginning of the year
<i>RoA</i>	Return on assets (Compustat ebit / average at) for the prior year
<i>OrgChange</i>	Following Feng et al. (2009), the principal component of asset growth _[t-2,t-1] , sales growth _[t-2,t-1] , and leverage _{t-1}
<i>Complexity</i>	Following Feng et al. (2009), the principal component of total segments _{t-1} , foreign transactions _{t-1} , and restructuring charges _{t-1}
<i>HighDAIndustry</i>	Indicator equal to one if the firm operates in an industry with high expected data analytics benefits (Software, Hardware, Communications, Medicine, Retail, Professional Services, Real Estate, or Insurance; SIC codes: 283, 357, 366 - 368, 481, 483, 484, 489, 501 - 509, 511 - 521, 527, 531, 533, 539, 540 - 542, 550, 553, 560, 562, 565, 566, 570, 571, 573, 581, 590, 591, 594, 596, 599, 621 - 633, 635, 636, 650, 651, 653, 655, 731 - 737, 800, 801, 805 - 809, 871 - 874) as of the beginning of the year
<i>DATProviderGeo</i>	Indicator equal to one if the firm is headquartered in the following zip codes as of the beginning of the year (where DAT providers are located): 98052, 98103, 19073, 94065, 10504, 94306, 92612, 19406
<i>TimeTrend</i>	Compustat fyear - 2004 (the first year in the sample period)

<i>Operating Outcomes</i>	
<i>UEInvestment</i>	Unexpected investment; the absolute value of the residual from the regression of investment on Tobin's $Q_{t-1} + \text{CFO}/\text{at}_{t-1} + \Delta\text{sales}_{[t-1,t]} + \text{investment}_{t-1}$ by industry-year; Tobin's $Q = (\text{prcc}_f * \text{csho} + \text{dlc} + \text{dltt})/\text{at}$; CFO is Compustat oancf; investment = Compustat capx + xrd + acq scaled by at_{t-1}
<i>UECapex</i>	Unexpected capital expenditures; see UEInvestment, where investment = Compustat capx
<i>UEAcquisitions</i>	Unexpected acquisition expenditures; see UEInvestment, where investment = Compustat acq; missing values of acq are set to zero
<i>UER&D</i>	Unexpected R&D investment; see UEInvestment, where investment = Compustat xrd and CFO = cash from operations (Compustat oancf) plus R&D to avoid a mechanical correlation between cash from operations and R&D
<i>AssetTO</i>	Asset turnover ratio (Compustat revt / average at)
<i>Ln_FixedAssetTO</i>	Log of fixed asset turnover (Compustat revt / average ppent)
<i>Ln_IntangibleAssetTO</i>	Log of intangible asset turnover (Compustat revt / average intan)
<i>Ln_RevenueEmployee</i>	Log of revenues per employee (Compustat revt / average emp)
<i>Reporting Outcomes</i>	
<i>Forecast</i>	Indicator equal to one if the firm issued an annual (IBES fpi = 1) earnings forecast during the year
<i>AvgError</i>	$ \text{Actual EPS} - \text{Forecasted EPS} $, scaled by beginning of period price; the average for all annual earnings forecasts issued during the year
<i>Horizon</i>	The number of days between the forecast date and the fiscal year end for the first annual earnings forecast issued during the year
<i>AccQuality</i>	Accruals quality modified from Dechow and Dichev (2002) following McNichols (2002); the standard deviation (over the past 16 quarters) of the residual from the regression: $\Delta\text{WC} (\text{recch} + \text{invch} + \text{apalch} + \text{txach} + \text{aoloch}) = \text{CFO}_{t-1} + \text{CFO}_t + \text{CFO}_{t+1} + \Delta\text{sale} + \text{ppent}$ run by industry-year; all variables scaled by average total assets
<i>Restate_NF</i>	Indicator equal to one if the financial statements are subsequently restated; unrelated to fraud
<i>Restate_F</i>	Indicator equal to one if the financial statements are subsequently restated; related to fraud
<i>Control Variables (Excluding those already defined in determinants and outcomes)</i>	
<i>AbsAccruals</i>	The magnitude of accruals; the absolute value of Compustat $\text{ib} - \text{oancf}$, scaled by beginning of year total assets
<i>Beta</i>	The coefficient from the regression of CRSP daily ret on vwretd over the prior year
<i>Big4</i>	Indicator equal to one if the firm is audited by one of the Big 4 auditors as of the beginning of the year
<i>CFOSale</i>	Operating cash flows to sales (Compustat oancf / sale)

<i>Dividend</i>	Indicator for whether the firm issued dividends during the year; equal to one if Compustat dv or dvc is greater than zero
<i>DV_{t-1}</i>	The dependent variable for a given regression as of the prior year
<i>EquityIssue</i>	Indicator equal to one if the firm issued new common shares during the year
<i>ForecastDisp</i>	The standard deviation of analysts' annual earnings forecasts issued 60 days prior to the management forecast date; this window is extended up to 180 days if there are less than two analyst forecasts issued within 60 days of the management forecast
<i>ForecastSurp</i>	The surprise implied by the management forecast; the absolute value of (forecasted EPS - most recent consensus EPS) scaled by beginning of period price
<i>HighLitigation</i>	Following Francis et al. (1994) and Feng and Koch (2010), Indicator equal to one if the firm operates in one of the following high litigation risk industries: SIC 2833 - 2836, 8731 - 8734, 3570 - 3577, 7370 - 7374, 3600 - 3674, 5200 - 5961
<i>ICWeakness</i>	The number of internal control material weakness documented for the year; count_weak from the audit analytics file auditsox404
<i>IO</i>	The percentage of the firm's shares held by institutional investors (from the s34 database) as of the beginning of the year
<i>Kstructure</i>	Financial leverage (Compustat dltt / (dltt + csho*prcc_f)) following Biddle et al. (2009) as of the beginning of the year
<i>Leverage</i>	Debt to assets ratio (Compustat lt / at) as of the beginning of the year
<i>Ln_Analysts</i>	Log of the number of analysts following the firm (IBES numest) as of the beginning of the year
<i>Ln_OpCycle</i>	Log of the length of the operating cycle (Compustat (360 / (sale / average rect)) + (360 / (cogs / average invt)))
<i>Ln_Slack</i>	Log of the cash to fixed assets ratio (Compustat ch / ppent) following Biddle et al. (2009) as of the beginning of the year
<i>Loss</i>	Indicator equal to one if Compustat ni is less than zero
<i>ProfitMargin</i>	Net profit margin (Compustat ebit / revt)
<i>SDInvestment/Revenue /OCF/Earnings</i>	The standard deviation of investments (defined above), Compustat revt, oancf, or ib scaled by assets over the past 16 quarters ending at the beginning of the year
<i>Sgrowth</i>	Sales growth (Compustat revt _t - revt _{t-1}) / revt _{t-1}
<i>Tangibility</i>	Following Biddle et al. (2009), the tangibility of assets (Compustat ppent / at) as of the beginning of the year

APPENDIX B

Examples of data analytics technology features

Tableau Example 1 – Data pre-processing allows users to standardize unstructured fields (e.g., names or transaction identifiers) for seamless integration into analyses.

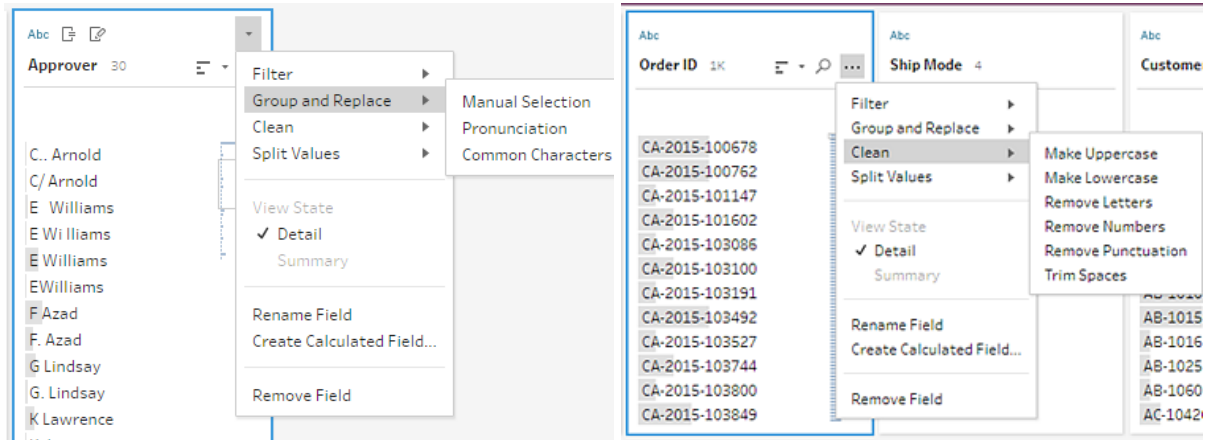
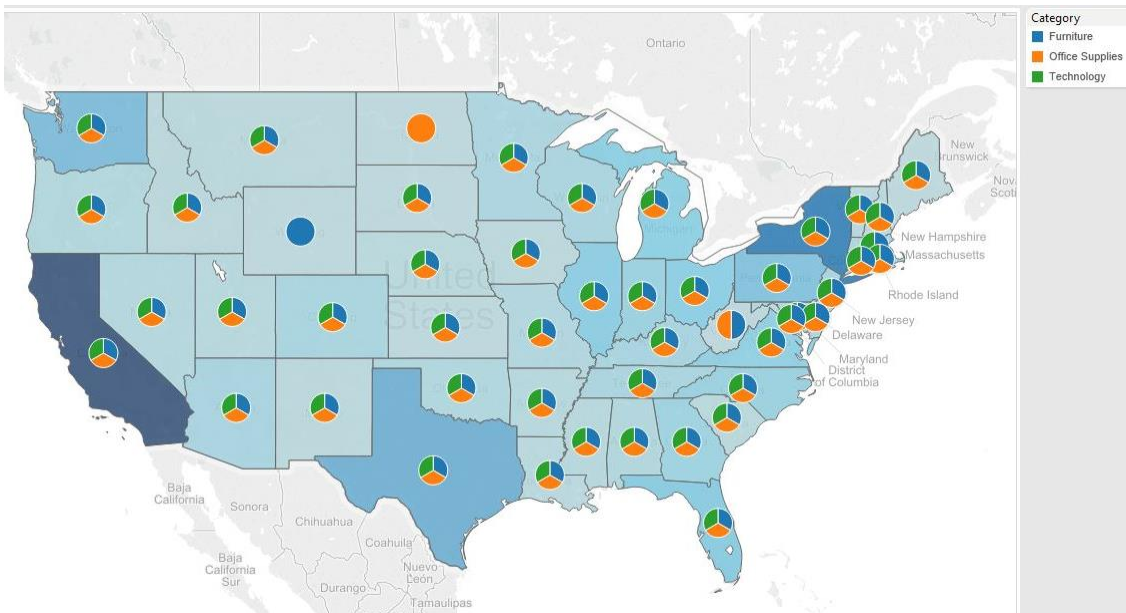


Tableau Example 2 – Advanced visualization (of e.g., sales mix by location) allows users to layer insights to identify patterns and anomalies.



APPENDIX B (continued)

Microsoft Power BI Example 1 – Scenario analysis allows users to adjust model assumptions and inputs to see the potential impact of different decisions.

The image displays a Microsoft Power BI interface. On the left, a bar chart titled "SalesAmount and Sales after Discount by Month" shows monthly sales data. The Y-axis represents sales amount in millions (0M to 6M). The X-axis lists the months from January to December. Two data series are shown: "SalesAmount" (teal bars) and "Sales after Discount" (black bars). A purple arrow points from the chart to a "Discount percentage" slicer below it, which is set to 0.30. To the right of the slicer is a "Discount percentage Value" bar chart showing a single teal bar at approximately 0.30. On the right side of the image is the "What-if parameter" configuration panel. It includes fields for Name, Data type, Minimum, Maximum, and Increment, along with a checkbox to "Add slicer to this page".

What-if parameter

Name	Cost Increase Rate
Data type	Fixed decimal number
Minimum	0
Maximum	1
Increment	0,1
Default	
<input checked="" type="checkbox"/> Add slicer to this page	

OK Cancel

APPENDIX C

Examples of diverse applications of data analytics

IIQ proxies	Examples of how data analytics could lead to improvements
<i>Investment efficiency</i>	
Capital expenditure efficiency	Incorporating different types of data (e.g., local safety, traffic) into a potential building purchase
R&D efficiency	More robust modeling of expected payoffs of different R&D projects
Acquisition efficiency	Incorporating different types of data (e.g., social media chatter) into a potential acquisition decision
<i>Asset turnover</i>	
Fixed asset turnover	Visualizing data on equipment productivity and maintenance; more robust modeling of alternative scenarios to optimize fixed asset use
Intangible asset turnover	Incorporating different types of data (e.g., transactions, demographics) into analyses of customer lists; more robust fair value estimates and thus, goodwill measurement
Revenue per employee	Incorporating different types of data (e.g., from employee surveys) into hiring and staffing decisions
<i>Management forecasts</i>	
Forecast propensity	More robust modeling of earnings, providing greater confidence in forecasting
Forecast error	Incorporating different types of data (e.g., online browsing statistics) into product demand forecasts; visualizing data on business risks and exposure
Forecast horizon	Streamlining the forecasting / data aggregation process
<i>Earnings quality</i>	
Accruals quality	More robust modeling of accrued expenses (e.g., utilities, employee retirement contributions) or allowances
Restatements	Visualizing account patterns to detect anomalies or unexpected trends (e.g., a transitory spike in accounts receivable)

Notes: Appendix C provides examples of how data analytics could impact the specific decision outcomes I examine in the paper. I choose outcomes that are subject to a high degree of managerial discretion, as these have the greatest potential to be impacted by data analytics. The examples given could impact multiple outcomes listed and additional outcomes not listed. E.g., visualizing data on equipment productivity and maintenance could also impact depreciation and maintenance expenses, analyzing online browsing statistics could also reduce marketing expenses, etc. This table is not intended to be an exhaustive list of data analytics applications.

APPENDIX D

How data analytics is discussed on conference calls

How data analytics is discussed	Types of firms	N
<i>Data analytics is used to improve internal decision making</i>		
A. Data analytics is used to improve general / non-product-related decisions	Various	59
B. Data analytics is used to improve the firm's product / service	Consulting firms (e.g., healthcare, staffing, advertising), search engines, insurance companies, etc.	13
<i>Data analytics is or is related to the firm's product / service</i>		
C. Data analytics is the firm's product / service	Software firms, firms that perform analytics as a service, etc.	12
D. Data analytics is related to the firm's product / service	Firms that manufacture computer servers, semiconductors, etc.	11
E. Vague / Unclear	Various	5
Total random hand-collection sample		100

Examples:

A1. Data analytics is used to improve general / non-product-related decisions

“I think where we get a little bit more sophisticated is in the area of big data analytics and leveraging it for sales and marketing initiatives. And the example that I used in the prepared remarks, we're going in that direction that we are really applying more sophisticated analytics to the information that we have on the millions of installed instruments. And the analytics allows us to better recognize the potential and based on the potential also use the right go-to-market approach.”

- Oliver Filliol, CEO, Mettler-Toledo International – 2016

A2. Data analytics is used to improve general / non-product-related decisions

“Secondly, in terms of IT investments, we are working on a data strategy for the company. So that means we are dealing with a question on how does our potential data and systems infrastructure and future should look like in order to be able to fully leverage new digital technologies; for example, collecting data, applying digital technologies like advanced analytics and big data analytics in order to improve decision quality and to enhance -- to reduce the manual workload, which is still quite significant.”

- Marcus Kuhnert, CFO, Merck KGaA – 2018

APPENDIX D (continued)

B. Data analytics is used to improve the firm's product / service

*“We can improve the quality of care by using **data analytics to predict when a member is at risk or an acute event** and using that data to arm the physician and in-home clinicians with comprehensive view of the member.”*

- Bruce Broussard, CEO, Humana Inc., on their in-home clinical and telehealth capabilities – 2018

C. Data analytics is the firm's product / service

*“Our CreditVision product serves to remind and motivate us that big ideas, combined with focused execution and persistence can move an industry forward in the uses and benefits of **data analytics** for both business customers and consumers.”*

- Jim Peck, CEO, TransUnion, on their credit analytics platform – 2016

D. Data analytics is related to the firm's product / service

*“With respect to memory interface...**I think where we are at and the industry is at right now is, if you take a look at the broad demand just driven by cloud data analytics** for higher and higher performance memory, that continues to have very robust demand.”*

- Greg Waters, CEO, Integrated Device Technology – 2014

E. Vague / Unclear

*“So I don't think a **big data** type project like that would be particularly relevant on our portfolio.”*

- Stan Galanski, CEO, The Navigators Group, in response to an analyst question – 2015

APPENDIX E

Data analytics hiring and data analytics disclosure

VARIABLES	(1) <i>DataAnalytics_t</i>	(2) <i>DataAnalytics_t</i>
<i>HighPctDAJobs_t</i>	0.1805** (2.2489)	
<i>PctDAJobs_t</i>		0.6341** (1.9661)
Determinants as controls	Included	Included
Clustered Standard Errors	Firm	Firm
Area under ROC Curve	0.6984	0.6966
Observations	3,493	3,493

Notes: Appendix E presents the results of the validation of the conference call measure, *DataAnalytics*, using firm hiring data (discussed in section 3). The dependent variable, *DataAnalytics*, is equal to one when the firm has discussed “data analytics” or “big data” on a conference call in the current year or in a prior year. *PctDAJobs*, is the percentage of job postings the firm filled during the year that contain data analytics keywords. To compile these keywords, I start by gathering ten articles: five on the market leaders in data analytics and five on the key skills required for analytics roles. I refine this list based on reviewing a sample of job postings for data analytics and data scientists. I include each vendor or skill appearing more than once across these articles and postings in an initial job postings scrape. I further refine the list by excluding keywords either not mentioned at all or picking up a large number of false positives in the initial scrape. The final dictionary includes the general data analytics terms in Table 1 panel C and the following skills: “data modeling”, “looker”, “sql”, “python”, “sas”, “r”, “spss”, “javascript”, “c#”, “php”, “objective c”, “ssas”, “olap”, “c++”. *HighPctDAJobs*, is an indicator if *PctDAJobs* is greater than the median for the year and zero otherwise. T-statistics are reported in parentheses with standard errors clustered by firm. All other variables are defined in Appendix A, and continuous variables are winsorized at 1% and 99% to reduce the impact of outliers.

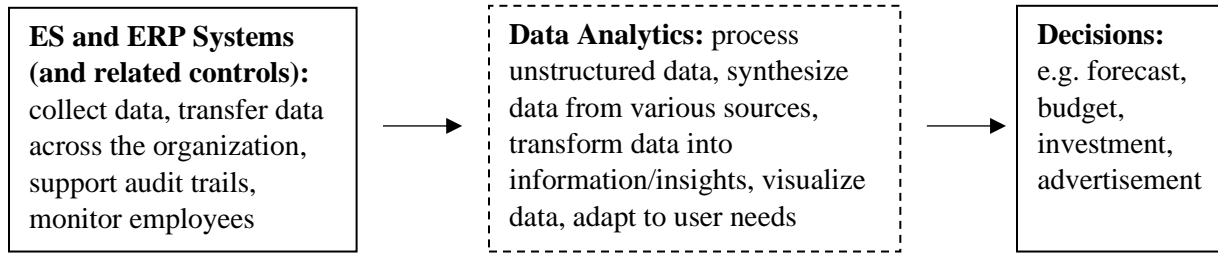
APPENDIX F

Alternative data analytics disclosure channels

Disclosure Channel / Event Type	Firm-years with event		Firm-years with DA disclosure	
	N	%	N	%
Analyst Meeting, Analyst Conference Call	1,459	10.4%	335	23.0%
Shareholder Meeting	693	5.0%	25	3.6%
Corporate Conference Call/Presentation	1,349	9.6%	46	3.4%
Guidance Conference Call/Presentation	212	1.5%	5	2.4%
M&A Conference Call/Presentation	961	6.9%	39	4.1%
Sales Conference Call/Presentation	40	0.3%	-	0.0%
Conference Presentation, Other Corporate Conference Event	6,088	43.5%	697	11.4%

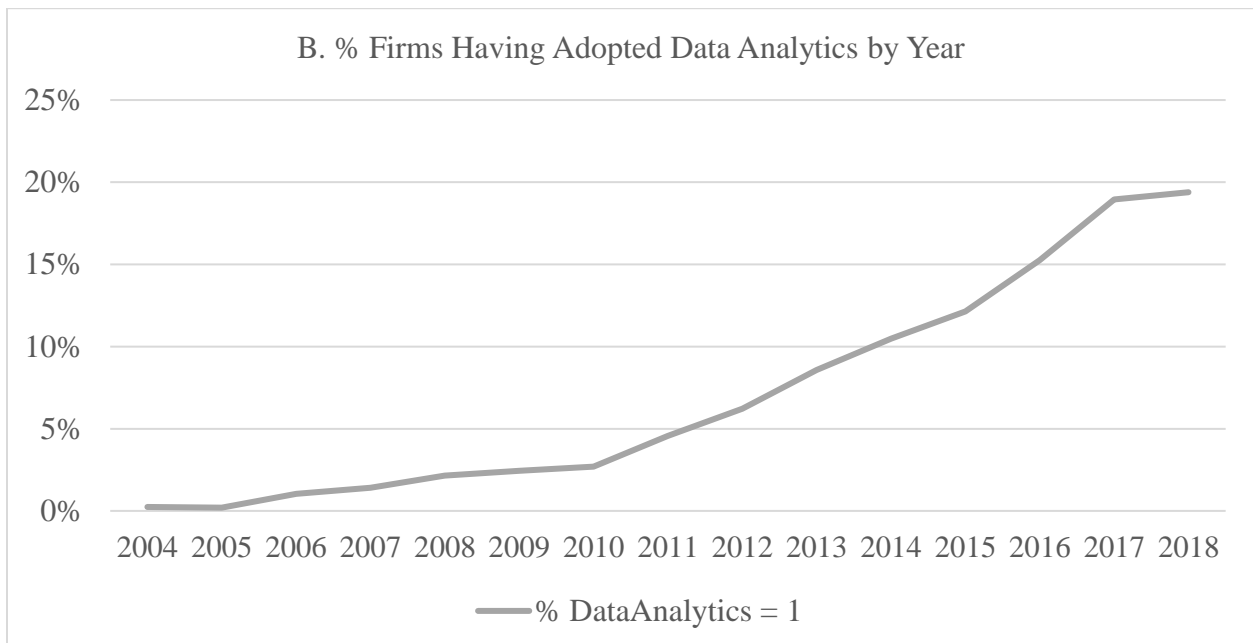
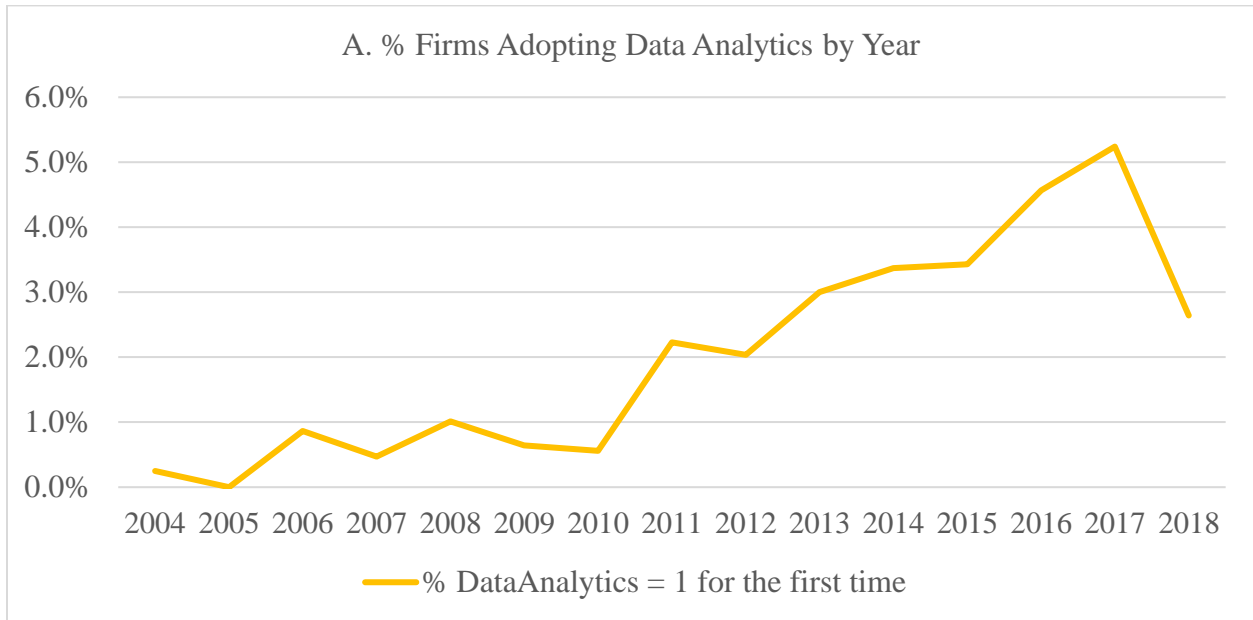
Notes: Appendix F presents the frequency of firms' disclosures of "data analytics" or "big data" through other disclosure channels with transcripts in the StreetEvents database. I exclude a small number of events labeled "Company Visit", "Conference", "Other Brokerage Event", "Syndicated Roadshow", and (blank). Including earnings conference calls, used for my primary measure, the events included in this appendix account for 99% of the events in StreetEvents for the sample period 2004-2018.

FIGURE 1
Enterprise systems and data analytics in the decision-making process



Notes: Figure 1 illustrates the relationship between Enterprise Resource Planning (ERP) systems, Enterprise Systems (ES), and Data Analytics technology as part of decision-making processes within the firm.

FIGURE 2
Data analytics adoption over time



Notes: Figure 2A plots the percentage of firms adopting data analytics (i.e., discussing data analytics topics on a conference call for the first time) in each year. Figure 2B plots the percentage of firms that have adopted data analytics (i.e., discussed data analytics topics on a conference call at least once before) over time.

TABLE 1
Conference call disclosure descriptive statistics

Panel A: Number of quarters in which data analytics is mentioned (adoption year)		
Quarters mentioned	Adopters	% Adopters
1 quarter	266	79%
2 quarters	47	14%
3 quarters	14	4%
4 quarters	11	3%
Total	338	100%

Panel B: Number of years in which data analytics is mentioned		
Years mentioned	Adopters	% Adopters
1 – 3 years	196	58%
4 – 6 years	107	32%
7 – 9 years	23	7%
10 or more years	12	3%
Total	338	100%

Panel C: Alternative keyword list; % of treatment firm-years (defined based on all words below) in which each term is mentioned			
Text string	% Firm-years	Text string	% Firm-years
Data analytics	32.8%	Data visualization	1.1%
Big data	29.7%	Natural language processing*	0.9%
Artificial intelligence*	11.6%	Tableau	0.8%
Data mining	10.3%	Hadoop	0.8%
Machine learning*	9.4%	Cloudera	0.3%
Data driven	9.1%	Robotic process automation*	0.3%
Predictive analytics	7.7%	BI tool	0.2%
Quantitative analy-	3.4%	SAP analytics cloud	0.0%
Advanced analytics	3.3%	Oracle analytics cloud	0.0%
Power BI	2.7%	IBM cognos	0.0%
Visualization t-	2.3%	Alteryx	0.0%
Predictive algorithm*	1.4%	Qlik	0.0%
Unstructured data	1.3%		

Notes: Table 1 panels A and B illustrate the variation in frequency of data analytics disclosures. Panel A (panel B) shows the distribution of the number of quarters (years) in which firms mention data analytics in the adoption year (over the sample period). Table 1 panel C presents the alternative word list discussed in additional analyses. This list includes the initial word list compiled as well as more sophisticated analytics terms (denoted with an asterisk). The frequency with which each term appears among firm-years in which at least one of the terms is mentioned

(the alternative treatment variable discussed in section 4) is presented. Text strings ending in “-” are intended to capture multiple possible phrases (e.g., “Quantitative Analy-” captures both “Quantitative Analysis” as well as “Quantitative Analytics”).

TABLE 2

Panel A: Descriptive statistics

Variable	<i>DataAnalytics = 0</i>				<i>DataAnalytics = 1</i>				Tests of Means	
	N	Mean	Med	Std Dev	N	Mean	Med	Std Dev	Diff	T-Stat
<i>Determinants</i>										
<i>BoDTechExpert</i>	12,398	0.58	1.00	0.49	1,221	0.79	1.00	0.41	0.21	17.12
<i>BoDSize</i>	12,398	12	10	7	1,221	14	11	8	2	8.9
<i>BoDInd</i>	12,398	0.71	0.73	0.16	1,221	0.74	0.77	0.14	0.04	8.71
<i>TotalAssets</i>	12,398	6,651	1,183	17,578	1,221	15,197	2,441	34,445	8,546	8.56
<i>Age</i>	12,398	26	21	17	1,221	30	25	17	4	6.81
<i>BTM</i>	12,398	0.54	0.45	0.46	1,221	0.42	0.33	0.41	-0.13	-10.17
<i>Cash</i>	12,398	0.16	0.08	0.26	1,221	0.15	0.09	0.20	-0.01	-0.85
<i>RoA</i>	12,398	0.04	0.07	0.18	1,221	0.06	0.08	0.16	0.01	3.03
<i>OrgChange</i>	12,398	-0.03	-0.03	0.01	1,221	-0.03	-0.03	0.01	-0.00	-3.1
<i>Complexity</i>	12,398	0.46	0.16	1.27	1,221	0.84	1.06	1.28	0.39	10.05
<i>HighDAIndustry</i>	12,398	0.39	0.00	0.49	1,221	0.58	1.00	0.49	0.19	13.06
<i>DATProviderGeo</i>	12,398	0.16	0.00	0.37	1,221	0.21	0.00	0.40	0.05	3.89
<i>TimeTrend</i>	12,398	8.06	8.00	4.08	1,221	11.27	12.00	2.62	3.21	38.48
<i>Internal Information Quality Measures - Operating Efficiency</i>										
<i>UEInvestment</i>	6,554	0.08	0.05	0.10	697	0.08	0.06	0.09	-0.00	-0.9
<i>UECapex</i>	11,865	0.02	0.01	0.03	1,152	0.01	0.01	0.02	-0.01	-9.12
<i>UEAcquisitions</i>	11,889	0.04	0.02	0.07	1,152	0.04	0.02	0.07	0.00	1.23
<i>UER&D</i>	6,556	0.04	0.02	0.06	697	0.04	0.02	0.05	-0.00	-0.73
<i>AssetTO</i>	12,397	0.92	0.78	0.73	1,221	0.91	0.70	0.72	-0.01	-0.62
<i>FixedAssetTO</i>	11,549	15	6	33	1,183	16	8	27	1	1.19
<i>IntangibleAssetTO</i>	10,429	36	6	120	1,099	25	3	93	-11	-3.42
<i>RevenueEmployee</i>	12,130	734	328	1,424	1,214	599	305	1,074	-136	-4.06

TABLE 2**Panel A:** Descriptive statistics (continued)

Variable	<i>DataAnalytics</i> = 0				<i>DataAnalytics</i> = 1				Tests of Means	
	N	Mean	Med	Std Dev	N	Mean	Med	Std Dev	Diff	T-Stat
<i>Internal Information Quality Measures – Financial Reporting Quality</i>										
<i>Forecast</i>	12,398	0.36	0.00	0.48	1,221	0.43	0.00	0.49	0.07	4.42
<i>AvgError</i>	4,474	0.02	0.00	0.17	520	0.01	0.00	0.02	-0.01	-5.25
<i>Horizon</i>	4,515	303	322	61	524	311	327	57	8	3.03
<i>AccQuality</i>	2,939	0.07	0.05	0.07	217	0.07	0.05	0.05	-0.01	-1.75
<i>Restate_NF</i>	12,398	0.11	0.00	0.31	1,221	0.07	0.00	0.26	-0.03	-4.34
<i>Restate_F</i>	12,398	0.00	0.00	0.04	1,221	0.00	0.00	0.04	-0.00	-0.11

Notes: Table 2 panel A presents the descriptive statistics for the determinants and consequences examined in Tables 3 – 6. The sample is partitioned by treatment (*DataAnalytics* = 1) and control (*DataAnalytics* = 0) groups, and T-Statistics are presented for tests of the differences in sample means. *DataAnalytics* is equal to one if the firm has mentioned “data analytics” or “big data” on a conference call in the current year or in a prior year. Refer to Appendix A for all variable definitions.

TABLE 2

Panel B: Pearson correlation matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) <i>DataAnalytics_t</i>		0.02	-0.05	0.05	0.01	-0.01	0.07	-0.10	-0.02	0.04	-0.12	0.08	-0.01	-0.03	0.00
(2) <i>Unexpected Total Invest_t</i>	-0.01		0.22	0.43	0.45	0.03	0.05	-0.19	-0.18	0.00	-0.01	-0.03	0.18	0.01	-0.01
(3) <i>Unexpected Capex_t</i>	-0.06	0.17		0.17	0.17	0.12	-0.28	0.04	-0.15	-0.06	0.04	-0.02	0.12	0.02	-0.01
(4) <i>Unexpected Acquisitions_t</i>	0.01	0.64	0.04		0.17	0.26	0.12	-0.18	-0.19	0.10	-0.04	-0.02	0.03	-0.01	0.00
(5) <i>Unexpected R&D_t</i>	-0.01	0.47	0.12	0.03		-0.08	0.00	-0.14	-0.17	-0.05	-0.09	0.03	0.23	0.00	-0.01
(6) <i>AssetTO_t</i>	-0.01	-0.04	0.01	0.05	-0.14		0.41	0.38	-0.18	0.11	0.12	-0.01	0.02	-0.01	0.02
(7) <i>Ln_FixedAssetTO_t</i>	0.06	0.00	-0.31	0.07	-0.10	0.45		0.02	0.10	0.03	0.10	-0.04	0.12	-0.02	0.00
(8) <i>Ln_IntangibleAssetTO_t</i>	-0.09	-0.16	0.10	-0.18	-0.08	0.39	0.06		0.16	-0.17	0.16	-0.05	0.07	-0.04	-0.01
(9) <i>Ln_RevenueEmployee_t</i>	-0.02	-0.15	-0.02	-0.07	-0.23	-0.03	0.15	0.19		-0.05	-0.06	0.02	-0.10	-0.01	0.00
(10) <i>Forecast_t</i>	0.04	-0.03	-0.10	0.06	-0.12	0.07	0.03	-0.17	-0.04		*	*	-0.19	0.01	-0.02
(11) <i>AvgError_t</i>	-0.03	0.01	-0.01	0.00	0.00	-0.01	0.01	0.01	-0.03	*		-0.09	0.09	0.04	-0.01
(12) <i>Horizon_t</i>	0.04	0.00	-0.03	0.01	0.02	0.02	-0.01	-0.08	-0.02	*	-0.03		-0.18	-0.01	0.01
(13) <i>AccQuality_t</i>	-0.02	0.20	0.07	-0.02	0.31	-0.01	0.07	0.06	-0.11	-0.16	0.00	-0.07		-0.03	-0.01
(14) <i>Restate_NF_t</i>	-0.03	0.00	0.02	0.01	-0.02	-0.01	-0.01	-0.04	-0.01	0.01	0.03	0.01	-0.03		-0.01
(15) <i>Restate_F_t</i>	0.00	-0.02	-0.01	-0.01	-0.01	0.02	0.00	-0.01	0.00	-0.02	0.00	0.01	0.04	-0.01	

Notes: Table 2 panel B presents the Pearson (Spearman) correlations on the bottom left (top right) for the variable of interest, *DataAnalytics*, and the operating and financial reporting measures of internal information quality. Bolded (bolded and italicized) correlations are significant with p-value ≤ 0.05 (0.01). Note that forecast error and horizon can only be calculated when the firm has issued a forecast.

TABLE 3
 Probit model – determinants of a data analytics strategy

VARIABLES	<i>Dependent Variable = AdoptDataAnalytics_t</i>	
	(1) Coefficient (T-Stat)	(2) Marginal Effects
<i>BoDTechExpert_{t-1}</i>	0.311*** (4.292)	2.25
<i>BoDSize_{t-1}</i>	0.014** (2.556)	1.85
<i>BoDInd_{t-1}</i>	0.228 (1.136)	0.39
<i>Size_{t-1}</i>	0.052** (2.248)	2.88
<i>Age_{t-1}</i>	0.000 (0.094)	-0.90
<i>BTM_{t-1}</i>	-0.263*** (-3.164)	-1.20
<i>Cash_{t-1}</i>	0.105 (0.777)	0.43
<i>RoA_{t-1}</i>	0.469* (1.895)	0.64
<i>OrgChange_{t-1}</i>	-8.133*** (-2.781)	-0.86
<i>Complexity_{t-1}</i>	0.062** (2.273)	1.66
<i>HighDAIndustry_{t-1}</i>	0.462*** (5.545)	6.48
<i>DATProviderGeo_{t-1}</i>	0.010 (0.114)	0.00
<i>TimeTrend_t</i>	0.138*** (14.013)	9.34
Clustered Std Errors		Firm
Area Under the ROC Curve		0.7511
Observations		10,633

Notes: Table 3 presents the determinants of adopting a data analytics strategy (equation (1)). The sample excludes years after the adoption year. *AdoptDataAnalytics* is equal to one the first year a firm discusses “data analytics” or “big data” on a conference call. T-statistics are reported in parentheses. All variables are defined in Appendix A. Continuous variables are winsorized at 1% and 99% to reduce the impact of outliers. Column (2) presents the effect of an interquartile change in each variable on the predicted probability, with all other variables at their mean (scaled by the unconditional likelihood of adoption in any given year for interpretation).

TABLE 4
OLS regressions – data analytics strategy and investment efficiency

VARIABLES	(1)	(2)	(3)	(4)
	$UEInvestment_t$	$UECapex_t$	$UEAcquisitions_t$	$UER\&D_t$
<i>DataAnalytics_t</i>	-0.021*** (-3.758)	-0.000 (-0.280)	-0.006** (-2.268)	-0.005 (-1.469)
<i>Loss_t</i>	0.011 (1.217)	-0.002* (-1.718)	0.000 (0.106)	0.001 (0.302)
<i>Ln_Analysts_{t-1}</i>	0.002 (0.320)	0.001 (0.729)	0.003 (1.186)	-0.001 (-0.464)
<i>IO_{t-1}</i>	-0.000 (-0.065)	0.001 (0.954)	0.003 (0.807)	-0.007* (-1.663)
<i>Big4_{t-1}</i>	0.002 (0.321)	-0.001 (-1.011)	0.001 (0.146)	0.005 (1.040)
<i>Tangibility_{t-1}</i>	0.007 (0.192)	0.021*** (3.535)	-0.028** (-1.965)	0.026 (1.449)
<i>Kstructure_{t-1}</i>	-0.012 (-1.423)	-0.000 (-0.238)	0.001 (0.218)	-0.009* (-1.812)
<i>CFOSale_t</i>	0.002** (2.528)	-0.000* (-1.847)	0.000** (2.040)	-0.001 (-1.490)
<i>Ln_Slack_{t-1}</i>	0.005* (1.731)	0.001** (1.975)	0.001 (0.483)	0.002 (0.916)
<i>Dividend_t</i>	0.004 (0.545)	-0.002* (-1.806)	-0.001 (-0.411)	-0.001 (-0.342)
<i>SDInvestment_[t-4,t-1]</i>	0.239*** (3.506)	0.819*** (7.251)	0.142*** (3.649)	0.427*** (2.638)
Determinants as Controls	Included	Included	Included	Included
Entropy Balancing	Yes	Yes	Yes	Yes
Fixed Effects	Ind & Year	Ind & Year	Ind & Year	Ind & Year
Clustered Std Errors	Firm	Firm	Firm	Firm
Observations	3,440	9,096	9,135	3,464
Adjusted R ²	0.121	0.279	0.0745	0.390

Notes: Table 4 presents the effects of employing a data analytics strategy on absolute deviations from expected levels of investment (equation (2)). Column (1) presents the results for the sum of capital expenditures, acquisitions, and R&D investments, and columns (2) – (4), presents each type of expenditure individually. *DataAnalytics*, is equal to one when the firm has discussed “data analytics” or “big data” on a conference call in the current year or in a prior year. T-statistics are reported in parentheses. All variables are defined in Appendix A, and continuous variables are winsorized at 1% and 99% to reduce the impact of outliers.

TABLE 5
 OLS regressions – data analytics strategy and resource efficiency

	(1)	(2)	(3)	(4)
VARIABLES	<i>AssetTO_t</i>	<i>Ln_Fixed AssetTO_t</i>	<i>Ln_Intangible AssetTO_t</i>	<i>Ln_Revenue Employee_t</i>
<i>DataAnalytics_t</i>	0.007 (1.334)	0.021** (2.217)	0.051*** (2.931)	-0.001 (-0.080)
<i>Loss_t</i>	-0.030*** (-3.622)	-0.088*** (-5.223)	-0.040 (-1.540)	-0.089*** (-7.970)
<i>Ln_Analysts_{t-1}</i>	-0.009* (-1.792)	-0.007 (-0.717)	-0.029* (-1.835)	0.012* (1.653)
<i>IO_{t-1}</i>	-0.010 (-1.523)	-0.011 (-0.834)	-0.027 (-1.033)	0.003 (0.301)
<i>Big4_{t-1}</i>	0.012 (1.487)	0.020 (1.018)	0.005 (0.180)	0.010 (0.982)
<i>SDRevenue_[t-4,t-1]</i>	0.009*** (2.613)	-0.001 (-0.170)	0.018* (1.755)	0.002 (0.332)
<i>ProfitMargin_t</i>	0.026*** (3.402)	0.049 (1.365)	0.030 (1.634)	0.067*** (2.740)
<i>DV_{t-1}</i>	0.949*** (160.049)	0.964*** (140.149)	0.951*** (131.733)	0.956*** (83.074)
Determinants as Controls	Included	Included	Included	Included
Entropy Balancing	Yes	Yes	Yes	Yes
Fixed Effects	Ind & Year	Ind & Year	Ind & Year	Ind & Year
Clustered Std Errors	Firm	Firm	Firm	Firm
Observations	9,691	9,676	9,545	9,671
Adjusted R ²	0.967	0.964	0.934	0.953

Notes: Table 5 presents the effects of employing a data analytics strategy on resource efficiency (equation (2)). Column (1) presents the results for overall asset turnover, and columns (2) – (4) present turnover as it relates to fixed assets, intangible assets, and human capital. *DataAnalytics*, is equal to one when the firm has discussed “data analytics” or “big data” on a conference call in the current year or in a prior year. T-statistics are reported in parentheses. All variables are defined in Appendix A, and continuous variables are winsorized at 1% and 99% to reduce the impact of outliers.

TABLE 6
 OLS regressions – data analytics strategy and financial reporting quality

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Management Forecasts</i>			<i>Earnings Quality</i>		
	<i>Forecast_t</i>	Avg <i>Error_t</i>	<i>Horizon_t</i>	<i>Acc Quality_t</i>	<i>Restate _NF_t</i>	<i>Restate _F_t</i>
<i>DataAnalytics_t</i>	-0.017 (-1.504)	-0.003*** (-2.973)	-0.002 (-0.001)	-0.004 (-0.703)	-0.012 (-0.475)	-0.002* (-1.818)
<i>Loss_t</i>	-0.022 (-1.043)	0.010*** (3.744)	-7.060 (-1.499)	0.002 (0.346)		
<i>Ln_Analysts_{t-1}</i>	0.026** (2.119)	-0.003** (-2.019)	3.927 (1.432)	0.005 (1.200)	-0.017 (-0.746)	0.001 (0.508)
<i>IO_{t-1}</i>	-0.002 (-0.108)	-0.003 (-1.270)	1.150 (0.356)	-0.002 (-0.216)	0.014 (0.493)	0.001 (0.297)
<i>Big4_{t-1}</i>	0.035 (1.638)	0.004 (0.790)	7.267 (0.900)	-0.001 (-0.169)	0.030 (1.167)	0.001 (1.630)
<i>SDEarnings_[t-4,t-1]</i>	-0.019*** (-2.815)	-0.000 (-0.671)	-2.271 (-1.465)			
<i>Beta_t</i>	0.005 (0.348)	0.006** (2.338)	1.276 (0.353)			
<i>ForecastDisp_t</i>	-0.517** (-2.468)	1.082*** (8.240)	-146.369 (-1.043)			
<i>ICWeakness_t</i>	-0.026** (-2.318)	0.004 (1.599)	-5.970 (-1.640)	-0.003* (-1.689)	0.047*** (3.052)	0.007* (1.814)
<i>EquityIssue_t</i>	-0.018 (-1.433)	-0.001 (-1.045)	-3.061 (-1.529)			
<i>HighLitigation_t</i>	0.041* (1.674)	0.001 (0.453)	6.805* (1.797)			
<i>DV_{t-1}</i>	0.732*** (35.450)		0.207*** (2.595)			
<i>Horizon_t</i>		0.000*** (4.312)				
<i>ForecastSurp_t</i>		0.413*** (290.146)				
<i>SDOCF_[t-4,t-1]</i>				0.016*** (3.601)	-0.029* (-1.916)	0.002** (2.018)
<i>Ln_OpCycle_t</i>				0.000 (0.048)	-0.029** (-2.144)	0.000 (0.115)

TABLE 6 (continued)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Management Forecasts</i>			<i>Earnings Quality</i>		
	<i>Forecast_t</i>	<i>Avg Error_t</i>	<i>Horizon_t</i>	<i>Acc Quality_t</i>	<i>Restate _NF_t</i>	<i>Restate _F_t</i>
<i>AbsAccruals_t</i>				0.093*** (2.628)	0.180* (1.824)	-0.012* (-1.826)
Determinants as Controls	Included	Included	Included	Included	Included	Included
Entropy Balancing	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Ind & Year	Ind & Year	Ind & Year	Ind & Year	Ind & Year	Ind & Year
Clustered Std Errors	Firm	Firm	Firm	Firm	Firm	Firm
Observations	8,464	4,652	4,260	2,203	7,034	7,034
Adjusted R ²	0.604	0.927	0.238	0.302	0.100	0.0157

Notes: Table 6 presents the effects of employing a data analytics strategy on financial reporting estimates (equation (2)). Columns (1) – (3) present the results for whether firms are more likely to issue earnings guidance and whether earnings guidance is more accurate or timely. Columns (4) – (6) present the results for whether firms improve accruals quality and reduce the likelihood of restatement. *DataAnalytics_t* is equal to one when the firm has discussed “data analytics” or “big data” on a conference call in the current year or in a prior year. T-statistics are reported in parentheses. All variables are defined in Appendix A, and continuous variables are winsorized at 1% and 99% to reduce the impact of outliers.

TABLE 7

Cross-sectional variation by emphasis on data analytics (high vs. low)

DEPENDENT VARIABLE	(1)	(2)	N	Adj - R ²	P-Value from F-Test Col (1) = Col (2)
	Coefficient on <i>DataAnalytics</i> <i>Low_t</i>	Coefficient on <i>DataAnalytics</i> <i>High_t</i>			
Panel A: Operating efficiency outcomes					
<i>UEInvestment_t</i>	-0.021** (-2.522)	-0.021*** (-3.662)	3,440	0.121	0.949
<i>UECapex_t</i>	0.001 (0.697)	-0.001 (-1.227)	9,096	0.280	0.127
<i>UEAcquisitions_t</i>	0.003 (0.482)	-0.011*** (-3.947)	9,135	0.078	0.015
<i>UER&D_t</i>	-0.009* (-1.701)	-0.004 (-1.053)	3,464	0.390	0.286
<i>AssetTO_t</i>	0.007 (0.880)	0.007 (1.291)	9,691	0.967	0.999
<i>Ln_FixedAssetTO_t</i>	0.015 (1.087)	0.024** (2.300)	9,676	0.962	0.531
<i>Ln_IntangibleAssetTO_t</i>	0.060** (2.292)	0.047*** (2.648)	9,545	0.930	0.598
<i>Ln_RevenueEmployee_t</i>	0.002 (0.173)	-0.002 (-0.280)	9,671	0.956	0.729
Panel B: Financial reporting quality outcomes					
<i>Forecast_t</i>	-0.025 (-1.531)	-0.013 (-1.040)	8,464	0.657	0.451
<i>AvgFerror_t</i>	-0.004** (-2.358)	-0.004*** (-2.840)	4,652	0.908	0.755
<i>Horizon_t</i>	0.123 (0.036)	-0.004 (-0.002)	4,620	0.239	0.967
<i>AccQuality_t</i>	-0.006 (-0.685)	-0.003 (-0.483)	2,203	0.300	0.800
<i>Restate_NF_t</i>	0.013 (0.346)	-0.024 (-0.896)	7,034	0.101	0.299
<i>Restate_F_t</i>	-0.002 (-1.484)	-0.002* (-1.873)	7,034	0.016	0.469

Notes: Table 7 presents the results of cross-sectional analyses based on the extent to which firms emphasize the use of data analytics (equation (4)). The extent to which firms emphasize the use of data analytics is measured using the cumulative percentage of conference calls on which they mention data analytics since adoption. *DataAnalyticsLow* (*DataAnalyticsHigh*) is an indicator variable equal to one if the percentage is below (above) the median. T-statistics are reported in parentheses. All variables are defined in Appendix A, and continuous variables are winsorized at 1% and 99% to reduce the impact of outliers. Panel A presents the results for the investment and resource efficiency variables, and panel B presents the results for the financial reporting variables.