9-2023

Big data analytics and management forecasting behavior

Beng Wee GOH
Singapore Management University, bwgoh@smu.edu.sg

Na LI

Tharindra RANASINGHE

ACCOUNTING HORIZONS
Vol. XX, No. XX
MONTH YEAR
pp. 1–18

American Accounting Association
DOI: 10.2308/HORIZONS-2020-145

Big Data Analytics and Management Forecasting Behavior

Beng Wee Goh
Singapore Management University

Na Li
York University

Tharindra Ranasinghe
American University

SYNOPSIS: This paper investigates whether the use of Big Data analytics by firms has a spillover effect on management forecasting behavior. Insights provided by Big Data could potentially improve firms' ability to forecast earnings (supply channel) and investor demand for earnings information is likely higher for firms engaging in data analytics (demand channel). Using a text-based measure of firms' commitments to and usage of Big Data analytics, we find that Big Data analytics usage is positively associated with the propensity to issue management earnings forecasts. Consistent with the "supply channel" explanation, we find that Big Data analytics usage is positively associated with management forecast accuracy as well. Also, supporting the "demand channel" explanation, we find that Big Data analytics usage is associated with greater analyst following. Our findings of improved disclosure following commitments to Big Data analytics highlight a potentially unintended benefit of the Big Data revolution.

Keywords: Big Data; data analytics; management forecasts; voluntary disclosure.

I. INTRODUCTION

Big Data and data analytics have become business buzzwords in recent years. The term "Big Data analytics" generally refers to the complex process of analyzing large and varied datasets to uncover information including hidden patterns, unknown correlations, market trends, and customer preferences that can help organizations make informed business decisions. Enabled by the simultaneous emergence of huge troves of data and technological tools to analyze, visualize, and make sense of insights such data present, business organizations are increasingly finding it important to deploy Big Data analytics in their business operations (Chen, Chiang, and Storey 2012; Lycett 2013; McKinsey Analytics 2018). For example, the 2018 McKinsey Global Survey on data and analytics find that an increasing share of companies is using data and analytics to generate growth, and nearly 50 percent of respondents say that analytics and Big Data have fundamentally changed business practices in their sales and marketing functions (McKinsey Analytics 2018).[1]

We thank Vernon J. Richardson (editor) and two anonymous reviewers for their helpful comments and suggestions. Tharindra Ranasinghe acknowledges financial support from the Kogod Faculty Excellence Fellowship.

Beng Wee Goh, Singapore Management University, School of Accountancy, Singapore; Na Li, York University, School of Administrative Studies, Faculty of Liberal Arts & Professional Studies, Toronto, Ontario, Canada; Tharindra Ranasinghe, American University, Kogod School of Business, Department of Accounting and Taxation, Washington, DC, USA.

Editor's note: Accepted by Vernon J. Richardson, under the Senior Editorship of Ray J. Pfeiffer.

Submitted: September 2020
Accepted: June 2023
Early Access: August 2023

[1] The online survey was conducted from March 14–24, 2017, and garnered responses from 530 C-level executives and senior managers representing the full range of regions, industries, and company sizes. To adjust for differences in response rates, the data are weighted by the contribution of each respondent's nation to global GDP.

The gathering of business intelligence and the use of information technology resources to analyze such intelligence are not new phenomena (e.g., see Sharma, Mithas, and Kankanhalli 2014). Yet, the current Big Data/data analytics revolution differs from prior business intelligence activities because it is facilitated by the generation of extremely large quantities of data (ranging from terabytes to exabytes in size) from web-based as well as mobile and sensor-based platforms and the development of storage, management, analysis, and visualization techniques which enable businesses to utilize these large datasets to generate important and novel insights and support decision making (Chen et al. 2012).[2] According to a survey of enterprise analytics and business intelligence professionals across five major economies conducted by MicroStrategy Inc., major benefits of Big Data analytics include improving efficiency and productivity, faster and more effective decision making, better financial performance, greater competitive advantage, improved customer experiences, customer acquisition and retention, and the identification and creation of new revenue streams (MicroStrategy Inc. 2018).

The objective of this paper is to investigate whether, beyond these stated goals, firms' use of data analytics also impacts their voluntary disclosure behavior. Specifically, we investigate the provision of management earnings forecasts (earnings guidance) by firms that utilize data analytics to improve their decisions. Even though enhancing voluntary disclosures is unlikely to be a primary objective of firms making commitments toward data analytics, insights provided by Big Data could potentially improve firms' ability to forecast earnings more accurately. Because firms are more likely to provide management earnings forecasts when earnings are more predictable (Waymire 1985), if Big Data analytics improve firms' ability to better predict earnings, then we could expect the likelihood of providing management earnings guidance to be higher for firms that use Big Data analytics. We label this the "supply channel." Moreover, it is also possible that investors are keen to understand how commitments to Big Data analytics impact firms' financial outcomes and therefore more likely to demand earnings guidance from firms that undertake Big Data analytics. This "demand channel" too predicts a positive association between the use of data analytics and the issuance of management earnings forecasts.

However, countervailing factors make it unclear whether Big Data analytics would lead to a greater likelihood of providing management earnings guidance. For instance, it is not obvious that insights obtained, and actions undertaken as a result of Big Data analytics at different functional levels get readily communicated to the financial reporting function in order to improve earnings forecasts in a timely manner. It is also argued that data analytics induce various functions to make speedy decisions based on statistical correlations with little to no understanding of the underlying causal relationships, thereby generating not less, but more outcome uncertainty (Clemen and Reilly 2013; Sharma et al. 2014; Strauß 2015). If so, because firms are less likely to provide earnings guidance when faced with higher uncertainty, engagement in Big Data analytics could in fact lead to a lower propensity to provide management earnings forecasts. Moreover, strategic actions undertaken following the insights provided by Big Data might generate proprietary cost concerns prompting the management to be cautious about revealing bottom-line outcomes too soon (Verrecchia 1983; Wang 2007). Hence, the association between engaging in Big Data analytics and management earnings forecasting behavior is an empirical question.

Based on the notion that credible commitments toward data analytics are likely reflected in discussions provided in 10-Ks, we capture firms' Big Data analytics activities based on the occurrences of related keywords in 10-K filings. Our sample consists of a broad group of firms and spans over the 2010–2018 period. In our main tests, we find a positive association between the measures of Big Data analytics usage and the propensity to issue management earnings forecasts.

Having established a positive association between commitments toward data analytics and propensity to issue earnings guidance, we next examine whether this finding could be attributed to either the supply channel and/or the demand channel discussed above. If, as argued in the "supply channel" explanation, Big Data analytics improve firms' ability to forecast earnings, then it is reasonable to posit that engaging in these activities is likely to be associated with greater forecast accuracy as well. Our findings reveal this indeed to be the case. Next, we investigate the "demand channel." If investors demand more information from firms that engage in Big Data analytics, it is likely that in addition to firm managers, financial intermediaries too would respond to such demand. If so, we would expect Big Data analytics to be associated with not only more management forecasts but greater analyst coverage as well. Our findings support this argument. Overall, our results suggest that the association between Big Data analytics and management forecasts can be attributed to both the lower cost of supplying information as well as greater demand for the same.

In supplementary analyses, we also examine the role of a number of firm-level factors such as size, complexity, and governance on the relationship between Big Data analytics and management forecasting propensity.[3]

[2] In this paper, we use the terms "Big Data," "Big Data analytics," and "data analytics" interchangeably.
[3] These are reported in Appendix A, "Methods and Supplemental Information."

We make several contributions to both the academic literature and the practice. As noted previously, even though firms embark on Big Data analytics with several primary objectives in mind, improving voluntary disclosure is unlikely to be one of them. Our findings suggest that independent of the main objectives, Big Data analytics are also associated with the likely unintended outcome of improving disclosure behavior. This is an important finding because higher disclosure quality is associated with several desirable firm-level outcomes such as greater stock liquidity, lower cost of capital, greater access to capital markets, and better investment decisions (e.g., Amihud and Mendelson 1986; Merton 1987; Frankel, McNichols, and Wilson 1995; Coller and Yohn 1997; Healy, Hutton, and Palepu 1999; Goodman, Neamtiu, Shroff, and White 2014). Our findings suggest corporate managers to be cognizant of less obvious, yet important effects of Big Data when making investment decisions regarding data analytics.

We also contribute to the literature on management forecast antecedents. Although prior literature on factors that influence management's forecast decision examine external influences such as legal, regulatory, analyst, and investor environments as well as several firm characteristics (e.g., see Hirst, Koonce, and Venkataraman 2008), we extend this literature by suggesting that the ongoing Big Data revolution also improves corporate voluntary disclosure behavior. Relatedly, we also contribute to and extend the literature on how IT systems and IT governance improve firms' disclosure and information environment (Li, Peters, Richardson, and Watson 2012; Dorantes, Li, Peters, and Richardson 2013; Haislip and Richardson 2018).

The remainder of this paper is organized as follows: Section II develops our primary hypothesis. Section III briefly presents the data and research design. Results are presented in Section IV. Section V concludes.

II. HYPOTHESIS DEVELOPMENT

As evident from the MicroStrategy Inc. (2018) survey discussed in the previous section, firms engage in Big Data analytics with a number of clearly articulated business objectives in mind (MicroStrategy Inc. 2018). Yet, to the best of our knowledge, no company has cited improving earnings forecastability or voluntary disclosures as a reason for using Big Data analytics. However, we argue that Big Data analytics could potentially improve a firm's propensity to issue management forecasts due to two nonmutually exclusive reasons.

First, even though firms are unlikely to engage in Big Data analytics with improving the ability to forecast earnings as a primary objective, it is quite conceivable that such benefits may accrue as a spillover effect. If data analytics endeavors are well designed with an organization-wide audience in mind and if insights obtained are effectively communicated across functional boundaries, Big Data analytics could assist in the forecasting of the overall financial performance of the firm. For example, companies are increasingly using predictive analytics to anticipate the likely effects of events such as marketing campaigns, regional changes in market conditions, price discounts, and other factors to project sales (Pelland 2017).[4] It is quite conceivable to envisage similar trends in projecting costs as well. Hence, the use of predictive analytics could help improve the accuracy of firms' earnings forecasts. Prior literature indicates fear of making inaccurate predictions to be a major deterrent to management earnings forecasts (Waymire 1985; Graham, Harvey, and Rajgopal 2005). If Big Data analytics improve managers' ability to predict earnings, it should attenuate this fear. Therefore, Big Data could increase the supply of management forecasts by reducing disclosure costs caused by potential forecast inaccuracy. We refer to this as the "supply channel" argument.

Second, disclosure theory predicts that when there is more uncertainty about firm value and information asymmetry among investors, managers have greater incentives to make more voluntary disclosures (Verrecchia 2001; Healy and Palepu 2001). Given the prevailing high level of interest in how firms are leveraging the opportunities provided by Big Data, investors are likely keen on understanding how data analytics activities are impacting financial outcomes. For example, high implementation costs could negatively impact short-term financial performance, but benefits accrued could improve performance due to both higher revenue and/or lower costs. Hence, timely projections of financial performance would help investors in understanding how firms' Big Data endeavors are impacting their bottom lines and assist investors in firm valuations. Therefore, it is likely that firms' engagement in Big Data analytics increases demand for management earnings forecasts. To the extent firms respond to this increased information demand, we would expect Big Data analytics to be positively associated with the issuance of management earnings forecasts. We refer to this as the "demand channel" argument.

However, there are also several reasons to suggest that Big Data analytics may not readily induce a higher propensity to issue management forecasts. First, because Big Data initiatives are largely made with no explicit financial reporting focus, it is unclear whether insights gained in other functional areas are speedily transmitted to the financial

[4] Predictive analytics applies statistical analysis, predictive modeling, data modeling, real-time scoring, and machine learning to discover trends in structured and unstructured data (generated within an organization and from outside sources) to forecast and rank likely events and their outcomes.

reporting function to make a discernible impact on earnings forecasting ability. If so, the spillover effects of Big Data analytics on the management earnings guidance domain could be limited. Second, several authors have suggested that increased datafication of the firm could in fact increase overall firm-level uncertainty because many tactical/functional decisions are made with exceedingly greater speed with insights obtained via pure statistical correlations without a sufficient understanding of the underlying causal links and thereby potentially exacerbating outcome uncertainty (Clemen and Reilly 2013; Sharma et al. 2014; Strauß 2015). Consequently, instead of abating, Big Data analytics could potentially magnify the likelihood of making inaccurate forecasts, thus making managers more reluctant to issue management forecasts. Third, engaging in Big Data analytics could generate significant proprietary cost concerns. For example, as reported by Rosenbush (2012) and Marr (2016), corporations such as Royal Dutch Shell keep the amounts and outcomes of their Big Data investments in oil and gas exploration activities as closely guarded secrets. Even though earnings projections are unlikely to contain significant amounts of proprietary data as a standalone piece of information, competitors might be able to combine this information with business intelligence gathered elsewhere to assess the success of a firm's Big Data initiatives and plan counter strategies accordingly. These increased proprietary costs could dissuade firms engaging in Big Data analytics from issuing management forecasts (Verrecchia 1983; Wang 2007).

Hence, whether Big Data analytics is associated with a higher incidence of issuing management earnings forecasts is a question that can only be resolved via empirical investigations. Therefore, we frame our hypothesis in the null form as follows:

Hypothesis: There is no association between Big Data analytics usage and the propensity to issue management earnings forecasts.

III. DATA AND RESEARCH DESIGN

Data

Firms typically do not disclose quantitative information on their Big Data analytics activities. However, they are comparatively more forthcoming in providing largely qualitative discussions regarding Big Data analytics in annual reports. Accordingly, we capture firms' usage of Big Data analytics by following the growing body of literature that uses textual analysis of firms' annual reports to elicit useful information that are otherwise difficult to obtain and quantify (e.g., Li 2010, 2011; Hoberg and Phillips 2010; Brown and Tucker 2011; Hoberg and Maksimovic 2015; Bushman, Hendricks, and Williams 2016; Hoberg and Lewis 2017).[5] We expect firms that significantly employ Big Data analytics to provide some discussions about them in their 10-Ks and the amount of this disclosure to correlate with the extent a company invests in and deploys Big Data.

Following the methodology of Li, Lundholm, and Minnis (2013) and Bushman et al. (2016), we count the number of occurrences of keywords in 10-K fillings that are related to Big Data analytics. These keywords include "data analytics/analysis," "cloud technology," "business intelligence," "business analytics/analysis," and "Big Data." Samples of 10-K discussions are provided in Appendix B. After extracting these keywords from WRDS SEC Analytics Suite,[6] we create two measures of the extent to which firms deploy Big Data analytics. The first measure *Analytics_Indicator* is an indicator variable that equals 1 if the firm disclosed any keywords related to Big Data analytics in its 10-K for a given year, and 0 otherwise. In the second measure (*Analytics_RankWords*), for observations with discussions of Big Data analytics in the annual reports we quartile rank the ratio of keywords on Big Data analytics to the total number of words in the 10-K. Observations with no discussion of Big Data analytics continue to be coded as 0.[7] For both treatment and control observations, we exclude firms in computer equipment/services/software industries (SIC codes: 3570–3572, 3575–3577, 7370–7374, 7377) from our analyses because instead of consumers, these firms are more likely to be vendors of Big Data products.[8] Our sample period is from 2010 to 2018.

It is worth noting that our proxies are unlikely to capture Big Data analytics usage with complete precision. For example, we cannot completely rule out the possibility that some firms may discuss Big Data in their 10-Ks rather loosely and with no real actions to back them up simply because it is fashionable to do so. Conversely, there could also

[5] For example, Hoberg and Phillips (2010) construct a measure of product similarity based on textual analysis of product descriptions in 10-K filings and then define industries as sets of sufficiently similar firms. Li et al. (2013) measure management's perceptions of the intensity of the competition they face using textual analysis of the firm's 10-K filing. Specifically, they count the number of occurrences of "competition, competitor, competitive, compete, competing." Building upon Li et al. (2013), Bushman et al. (2016) use textual analysis of banks' 10-K filings to construct a comprehensive, time-varying, bank-specific measure of a bank's competitive environment.

[6] The WRDS SEC Analytics Suite is a "one-step research platform" designed to facilitate data extraction from over 3.5 million SEC filings made since 1994. This platform enables researchers to easily mine and effectively parse any textual or tabulated information from the necessary SEC filings.

[7] In other words, *Analytics_RankWords* is a discrete variable that ranges from 0 to 4.

[8] However, none of our results are sensitive to the exclusion or inclusion of these industries.

be firms that do employ data analytics but do not discuss it anywhere in the 10-K either due to proprietary cost concerns or because their usage of Big Data is less efficient.

We provide more details of data and variable construction in Appendix A, "Methods and Supplemental Information."

Table 1, Panel A presents sample firm distribution by year. As seen in column (1), the total number of sample firms with Big Data activities steadily increases over the years. Column (2) shows the number of firms that do not disclose Big Data activities. Table 1, Panel B presents the sample composition based on two-digit SEC SIC industry classifications.

TABLE 1

Sample Selection and Descriptive Statistics

Panel A: Sample Firms Distribution by Year

Fiscal	# of Discloser	# of Nondiscloser	# of Total
2010	97	3,431	3,528
2011	111	3,277	3,388
2012	141	3,207	3,348
2013	166	3,151	3,317
2014	208	3,075	3,283
2015	247	3,056	3,303
2016	274	3,121	3,395
2017	321	3,057	3,378
2018	296	3,070	3,366
Total	1,861	28,445	30,306

Panel B: Sample Firm Distribution by Industry

Two-Digit SIC	Industry	# of Discloser	# of Nondiscloser	Total
27	Printing and Publishing	37	175	212
28	Chemical and Allied Products	103	3,688	3,791
35	Industrial Machinery and Equipment	77	1,456	1,533
38	Instruments and Related Products	103	1,713	1,816
42	Trucking and Warehousing	15	171	186
48	Communications	83	967	1,050
49	Electric, Gas, and Sanitary Services	12	1,190	1,202
56	Apparel and Accessory Stores	59	257	316
57	Furniture and Homefurnishings Stores	10	97	107
58	Eating and Drinking Places	61	330	391
59	Miscellaneous Retail	56	453	509
60	Depository Institutions	3	25	28
61	Nondepository Institutions	11	189	200
62	Security and Commodity Brokers	32	312	344
63	Insurance Carriers	39	264	303
64	Insurance Agents, Brokers, and Service	18	79	97
73	Business Services	394	3,164	3,558
79	Amusement and Recreation Services	14	263	277
80	Health Services	42	459	501
82	Educational Services	24	176	200
87	Engineering and Management Services	146	328	474
	Others	522	12,689	13,211
	Total	1,861	28,445	30,306

Panel A of this table presents the sample firm distribution over the period from 2010 to 2018. Panel B presents the sample firm distribution by industry.

Research Design

We employ the following probit regression to test our hypothesis (i and t stand for firm and year subscripts, respectively):

$$MF_{i,t+1} = \beta_0 + \beta_1 AnalyticsProxy_{i,t} + [CONTROLS]_{i,t} + [FIXED\ EFFECTS] + \varepsilon_{i,t} \quad (1)$$

The dependent variable $MF_{i,t+1}$ is an indicator variable that equals 1 if the firm issues at least one management earnings forecast following the discussion of Big Data analytics activities in the 10-K, and 0 otherwise. $AnalyticsProxy_{i,t}$ is either the indicator variable *Analytics_Indicator* or the categorical variable *Analytics_RankWords* as defined earlier. If firms that employ Big Data analytics in their business activities are more (less) likely to issue management earnings forecasts, we would expect a significantly positive (negative) coefficient on *AnalyticsProxy* (β_1).

Following prior studies, we include a number of variables to control for other determinants of propensity to issue management forecasts as well as industry and year fixed effects. Complete details of these are presented in Appendix A, "Methods and Supplemental Information."

Table 2 presents the descriptive statistics of variables used in our regressions. The table reveals that about 6 percent of the sample firms make discussions regarding Big Data analytics on 10-Ks during the sample period (*Analytics_Indicator*). Untabulated data shows that conditional on discussing data analytics, keywords related to Big Data analytics appear about four times on average. 20 percent of the sample firms have issued at least one management earnings forecast during the sample period (*MF*).

IV. RESULTS

Main Result (Hypothesis)

Table 3 presents our main results. Column (1) presents the results with the indicator variable *Analytics_Indicator* as the variable of interest. The variable of interest in column (2) is the categorical variable *Analytics_RankWords*. In

TABLE 2

Descriptive Statistics

Variables	n	Mean	Std. Dev.	Min.	0.25	Mdn	0.75	Max
$Analytics\_Indicator_{i,t-1}$	30,306	0.06	0.23	0	0	0	0	1
$Analytics\_RankWords_{i,t-1}$	30,306	0.13	0.62	0	0	0	0	4
$MF_{i,t}$	30,306	0.20	0.40	0	0	0	0	1
$MFE_{i,t}$	5,137	0.01	0.02	0	0	0	0	0.48
$AF_{i,t}$	22,776	2.15	0.78	1	2	2	3	3.69
$SF_{i,t}$	30,870	0.26	0.44	0	0	0	1	1
$Size_{i,t-1}$	30,306	6.65	2.25	−1.20	5.06	6.64	8.19	12.46
$MTB_{i,t-1}$	30,306	2.92	9.43	−116.89	1.12	1.97	3.59	95.20
$Loss_{i,t-1}$	30,306	0.35	0.48	0	0	0	1	1
$ROA_{i,t-1}$	30,306	−0.06	0.50	−29.51	−0.04	0.03	0.07	0.59
$Lev_{i,t-1}$	30,306	0.20	0.22	0	0	0.15	0.32	1.87
$RetVol_{i,t-1}$	30,306	0.13	0.08	0.01	0.07	0.10	0.16	0.65
$Litirisk_{i,t-1}$	30,306	0.52	0.50	0	0	1	1	1
$DebtIssue_{i,t}$	30,306	0.02	0.14	0	0	0	0	1
$ShrTurnover_{i,t-1}$	30,306	2.04	2.21	0.02	0.66	1.42	2.56	17.02
$BusSeg_{i,t-1}$	30,306	1.63	0.75	0	1.39	1.39	2.30	3.76
$Dedicate\%_{i,t-1}$	10,992	0.02	0.04	0	0	0	0.03	0.23

This table reports descriptive statistics for variables used in the analyses. All continuous variables are Winsorized at the 1 percent and 99 percent levels.
See Appendix C for variable definitions.

TABLE 3

Big Data Analytics and the Propensity to Issue Management Earnings Forecasts

Variables	(1) MF	(2) MF
Analytics_Indicator	**0.1755*****	
	(4.58)	
Analytics_RankWords		**0.0586*****
		(3.94)
Size	0.1621***	0.1623***
	(28.59)	(28.63)
MTB	0.0042***	0.0042***
	(3.54)	(3.53)
Loss	−0.2885***	−0.2885***
	(−8.95)	(−8.95)
ROA	0.5509***	0.5508***
	(5.72)	(5.72)
Lev	0.2477***	0.2467***
	(4.65)	(4.63)
RetVol	−4.3141***	−4.3192***
	(−17.95)	(−17.97)
ShrTurnover	0.0685***	0.0686***
	(14.26)	(14.28)
LitiRisk	−0.0325	−0.0331
	(−1.48)	(−1.51)
DebtIssue	−0.1708**	−0.1698**
	(−2.21)	(−2.19)
Intercept	−2.0894***	−2.0829***
	(−7.69)	(−7.71)
Industry fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
n	30,306	30,306
R^2	0.236	0.236

*, **, *** Indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.
This table reports the results from probit regressions of management's propensity to issue earnings forecasts on firms' usage of Big Data analytics. Column (1) presents the results using the indicator variable that equals 1 if the firm disclosed any of the keywords related to Big Data analytics in its 10-K for a given year, and 0 otherwise. Column (2) presents the results using the quartile ranking of the ratio of keywords on Big Data analytics to the total number of words in the 10-K. The dependent variable is an indicator variable that equals 1 if the firm issues at least one management earnings forecast following the discussion of Big Data analytics activities in the 10-K, and 0 otherwise. The dependent variable is measured in year t+1, whereas all the independent variables are measured in year t. All continuous variables are Winsorized at the 1 percent and 99 percent levels. All z-statistics are computed based on two-tailed tests.
See Appendix C for variable definitions.

column (1), the coefficient on *Analytics_Indicator* is positive and significant (coefficient size = 0.1755; z-statistic = 4.58). In column (2), the coefficient on *Analytics_RankWords* is positive and significant as well (coefficient size = 0.0586; z-statistic = 3.94). These results support the assertion that the propensity to issue management earnings forecasts is higher for firms that engage in Big Data analytics. In terms of economic significance, the magnitude of the coefficient on *Analytics_Indicator* indicates that at the margin, firms that use Big Data analytics are 4 percent more likely to issue management earnings forecasts compared to firms that do not.[9]

[9] The marginal effect of *Analytics_Indicator* is calculated as the difference in the probability of issuing management forecasts when *Analytics_Indicator* changes from 0 (12 percent) to 1 (16 percent), with the other explanatory variables taking the value of sample means.

Attributing Hypothesis Results to Supply and Demand Channels

Supply Channel

If, as articulated in the "supply channel" argument, Big Data analytics improve firms' overall information environment and makes it easier for them to forecast earnings, then it is reasonable to expect data analytics to be positively associated with not only the propensity to issue management forecasts but management forecast accuracy as well. Accordingly, we examine the supply channel explanation by investigating the relationship between Big Data analytics and management forecast accuracy.

Specifically, we re-run regression model (1) as an OLS regression after replacing the dependent variable with management forecast error (*MFE*).[10] The results of this test are presented in Table 4. Consistent with our expectation, Table 4, column (1) shows that the coefficient on *Analytics_Indicator* is negative and significant (coefficient size = −0.0016; t-statistic = −2.41) and Table 4, column (2) shows that the coefficient on *Analytics_RankWords* is negative and significant as well (coefficient size = −0.0005; t-statistic = −2.21). In other words, in line with the "supply channel" explanation, we find that Big Data analytics is associated with more accurate management earnings forecasts.

Demand Channel

Investors' information demands are fulfilled not only by firms' management but also by other information intermediaries in the marketplace. Therefore, if, as we argue, investors demand more information from firms that engage in Big Data analytics, it is likely that in addition to the management, other financial intermediaries such as financial

TABLE 4

Tests of the Supply Channel—Big Data Analytics and Management Forecast Accuracy

Variables	(1) MFE	(2) MFE
Analytics_Indicator	**−0.0016****	
	(−2.41)	
Analytics_RankWords		**−0.0005****
		(−2.21)
MF	0.3572***	0.3573***
	(4.81)	(4.81)
Size	−0.0005**	−0.0005**
	(−2.00)	(−2.02)
MTB	−0.0000	−0.0000
	(−0.38)	(−0.35)
Loss	0.0022	0.0022
	(1.10)	(1.09)
ROA	−0.0146	−0.0147
	(−1.34)	(−1.35)
Lev	0.0031	0.0031
	(0.89)	(0.89)
RetVol	0.0241	0.0240
	(1.53)	(1.53)
ShrTurnover	0.0008*	0.0008*
	(1.81)	(1.81)
LitiRisk	0.0007	0.0007
	(1.01)	(1.03)

(continued on next page)

[10] We measure management forecast error as the absolute value of the firm's last management earnings forecast for the year minus actual earnings per share for the year, scaled by the stock price at the beginning of the year.

TABLE 4 (continued)

Variables	(1) MFE	(2) MFE
DebtIssue	−0.0024	−0.0024
	(−1.51)	(−1.50)
Intercept	−0.0028	−0.0028
	(−0.80)	(−0.79)
Industry fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
n	5,137	5,137
R^2	0.257	0.257

*, **, *** Indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.
This table reports the results from pooled regressions of management forecast accuracy on firms' usage of Big Data analytics. Column (1) presents the results using the indicator variable that equals 1 if the firm disclosed any of the keywords related to Big Data analytics in its 10-K for a given year, and 0 otherwise. Column (2) presents the results using the quartile ranking of the ratio of keywords on Big Data analytics to the total number of words in the 10-K. The dependent variable is management earnings forecast error (MFE), measured as the absolute value of the firm's last management earnings forecast for the year minus actual earnings per share for the year, scaled by the stock price at the beginning of the year. The dependent variable is measured in year t+1, whereas all the independent variables are measured in year t. All continuous variables are Winsorized at the 1 percent and 99 percent levels. All z-statistics are computed based on two-tailed tests.
See Appendix C for variable definitions.

analysts too would respond to such demand. Accordingly, if the "demand channel" plays a role in the positive association between Big Data analytics and management forecasting propensity, it is reasonable to expect Big Data analytics to be associated with a higher analyst following as well.

We test this prediction by replacing the dependent variable in model (1) by analyst following (*AF*) and running this model as an OLS regression.[11] These results are presented in Table 5. In Table 5, column (1), we observe a positive coefficient on *Analytics_Indicator* (coefficient size = 0.1224; t-statistic = 8.69). Similarly, the coefficient on *Analytics_RankWords* reported in Table 5, column (2) is positive as well (coefficient size = 0.0363; z-statistic = 6.72). These results strongly indicate that analyst following is higher for firms that employ Big Data analytics. They are consistent with the idea of investor demand for information being higher for firms that engage in Big Data analytics (demand channel).

TABLE 5

Tests of the Demand Channel—Big Data Analytics and Analyst Following

Variables	(1) AF	(2) AF
Analytics_Indicator	**0.1224*****	
	(8.69)	
Analytics_RankWords		**0.0363*****
		(6.72)
Size	0.2243***	0.2244***
	(79.52)	(79.50)
MTB	0.0049***	0.0049***
	(10.60)	(10.60)
Loss	0.0195*	0.0196*
	(1.90)	(1.91)

(continued on next page)

[11] We measure analyst following as the natural logarithm of 1 plus the number of analysts following the firm during the year.

TABLE 5 (continued)

Variables	(1) AF	(2) AF
ROA	−0.0634***	−0.0634***
	(−3.70)	(−3.70)
Lev	0.0121	0.0119
	(0.58)	(0.57)
RetVol	−1.7167***	−1.7203***
	(−22.83)	(−22.87)
ShrTurnover	0.1207***	0.1208***
	(44.99)	(45.03)
LitiRisk	0.0934***	0.0927***
	(11.03)	(10.93)
DebtIssue	0.1244***	0.1240***
	(4.53)	(4.51)
Intercept	0.1224***	0.1532*
	(8.69)	(1.75)
Industry fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
n	22,776	22,776
R^2	0.525	0.524

*, **, *** Indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.
This table reports the results from pooled regressions of analyst following on firms' usage of Big Data analytics. Column (1) presents the results using the indicator variable that equals 1 if the firm disclosed any of the keywords related to Big Data analytics in its 10-K for a given year, and 0 otherwise. Column (2) presents the results using the quartile ranking of the ratio of keywords on Big Data analytics to the total number of words in the 10-K. Analyst following (AF) is calculated as the natural logarithm of 1 plus the number of analysts following the firm during the year. The dependent variable is measured in year *t*+1, whereas all the independent variables are measured in year *t*. All continuous variables are Winsorized at the 1 percent and 99 percent levels. All z-statistics are computed based on two-tailed tests.
See Appendix C for variable definitions.

Overall, the results reported in Tables 4 and 5 suggest that our main finding of a positive association between Big Data analytics and the propensity to issue management forecasts can be attributed to both lower cost of providing earnings forecasts due to higher accuracy (supply channel) and the investors' increased demand for earnings information (demand channel).

Additional Analyses

We conduct a number of additional analyses to gain further insights on factors that moderate the relationship between Big Data analytics and management forecasting and find that this relationship is stronger for larger, more complex, and well-governed firms when compared with those that are smaller, less complex, and less well-governed. Moreover, we find the adoption of Big Data analytics to be also associated with a higher propensity to issue management sales forecasts. Further details of these tests, as well as robustness tests conducted to alleviate endogeneity concerns, are discussed in Appendix A, "Methods and Supplemental Information."

V. CONCLUSION

As the Big Data revolution continues to advance along with the technology to obtain, store, and analyze massive troves of data, companies stand to reap numerous benefits ranging from improved operational efficiency, faster decision-making, improved customer experiences, and identification and creation of new revenue streams. Our paper highlights that Big Data activities also entail additional, less obvious benefits such as improved and more accurate communications with capital markets that are less likely to come into consideration at the time of making decisions about Big Data investments. In this light, this paper encourages business leaders to take a broader perspective in decisions regarding Big Data and its potential benefits to the firm.

In conclusion, we also note two caveats. First, our paper is an association study and our proxies for Big Data based on firms' own disclosures are relatively coarse. If made available, finer and quantifiable data on Big Data investments should facilitate further refinements and stronger establishment of causality. Second, although we focus on a potential benefit of Big Data, it should also be noted that these investments consume resources and carry significant costs. Hence whether and under what organizational contexts the benefits outweigh costs as well as firms' propensities to underinvest and overinvest in Big Data become important questions for future research.

REFERENCES

Amihud, Y., and H. Mendelson. 1986. Liquidity and stock returns. *Financial Analysts Journal* 42 (3): 43–48. https://doi.org/10.2469/faj.v42.n3.43

Bamber, L., and Y. Cheon. 1998. Discretionary management earnings forecast disclosures: Antecedents and outcomes associated with forecast venue and forecast specificity choices. *Journal of Accounting Research* 36 (2): 167–190. https://doi.org/10.2307/2491473

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther. 2010. The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics* 50 (2–3): 296–343. https://doi.org/10.1016/j.jacceco.2010.10.003

Brown, S. V., and J. W. Tucker. 2011. Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research* 49 (2): 309–346. https://doi.org/10.1111/j.1475-679X.2010.00396.x

Bushee, B. J. 1998. The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review* 73 (3): 305–333.

Bushee, B. J. 2001. Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research* 18 (2): 207–246. https://doi.org/10.1506/J4GU-BHWH-8HME-LE0X

Bushman, R. M., B. E. Hendricks, and C. D. Williams. 2016. Bank competition: Measurement, decision-making, and risk-taking. *Journal of Accounting Research* 54 (3): 777–826. https://doi.org/10.1111/1475-679X.12117

Bushman, R. M., Q. Chen, E. Engel, and A. Smith. 2004. Financial accounting information, organizational complexity and corporate governance systems. *Journal of Accounting and Economics* 37 (2): 167–201. https://doi.org/10.1016/j.jacceco.2003.09.005

Chen, H., R. H. Chiang, and V. C. Storey. 2012. Business intelligence and analytics: From Big Data to big impact. *MIS Quarterly* 36 (4): 1165–1188. https://doi.org/10.2307/41703503

Chen, S., D. Matsumoto, and S. Rajgopal. 2011. Is silence golden? An empirical analysis of firms that stop giving quarterly earnings guidance. *Journal of Accounting and Economics* 51 (1–2): 134–150. https://doi.org/10.1016/j.jacceco.2010.10.006

Clemen, R. T., and T. Reilly. 2013. *Making Hard Decisions with Decision Tools*. Mason, OH: South-Western, Cengage Learning.

Coller, M., and T. L. Yohn. 1997. Management forecasts and information asymmetry: An examination of bid-ask spreads. *Journal of Accounting Research* 35 (2): 181–191. https://doi.org/10.2307/2491359

Cotter, J., I. Tuna, and P. Wysocki. 2006. Expectations management and beatable targets: How do analysts react to explicit earnings guidance? *Contemporary Accounting Research* 23 (3): 593–624. https://doi.org/10.1506/FJ4D-04UN-68T7-R8CA

De Haes, S., and W. Van Grembergen. 2004. IT governance and its mechanisms. *Information Systems Control Journal* 1: 27–33. http://www.gti4u.es/curso/material/complementario/de_haes_y_van_grembergen_2004.pdf

Dorantes, C., C. Li, G. Peters, and V. J. Richardson. 2013. The effect of enterprise systems implementation on the firm information environment. *Contemporary Accounting Research* 30 (4): 1427–1461. https://doi.org/10.1111/1911-3846.12001

Dye, R. 1985. Disclosure of nonproprietary information. *Journal of Accounting Research* 23 (1): 123–145. https://doi.org/10.2307/2490910

Feng, M., and A. S. Koch. 2010. Once bitten, twice shy: The relation between outcomes of earnings guidance and management guidance strategy. *The Accounting Review* 85 (6): 1951–1984. https://doi.org/10.2308/accr.2010.85.6.1951

Frankel, R., M. McNichols, and G. P. Wilson. 1995. Discretionary disclosure and external financing. *The Accounting Review* 70 (1): 135–150.

Gong, G., L. Li, and H. Xie. 2009. The association between management earnings forecast errors and accruals. *The Accounting Review* 84 (2): 497–530. https://doi.org/10.2308/accr.2009.84.2.497

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. https://doi.org/10.2308/accr-50575

Graham, J. R., C. R. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40 (1–3): 3–73. https://doi.org/10.1016/j.jacceco.2005.01.002

Habib, M., D. B. Johnsen, and N. Naik. 1997. Spinoffs and information. *Journal of Financial Intermediation* 6 (2): 153–176. https://doi.org/10.1006/jfin.1997.0212

Haislip, J. Z., and V. J. Richardson. 2018. The effect of CEO IT expertise on the information environment: Evidence from earnings forecasts and announcements. *Journal of Information Systems* 32 (2): 71–94. https://doi.org/10.2308/isys-51796

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. https://doi.org/10.1016/S0165-4101(01)00018-0

Healy, P. M., A. P. Hutton, and K. G. Palepu. 1999. Stock performance and intermediation changes surrounding sustained increases in disclosure. *Contemporary Accounting Research* 16 (3): 485–520. https://doi.org/10.1111/j.1911-3846.1999.tb00592.x

Hirst, D. E., L. Koonce, and S. Venkataraman. 2008. Management earnings forecasts: A review and framework. *Accounting Horizons* 22 (3): 315–338. https://doi.org/10.2308/acch.2008.22.3.315

Hoberg, G., and C. Lewis. 2017. Do fraudulent firms produce abnormal disclosure? *Journal of Corporate Finance* 43: 58–85. https://doi.org/10.1016/j.jcorpfin.2016.12.007

Hoberg, G., and V. Maksimovic. 2015. Redefining financial constraints: A text-based analysis. *The Review of Financial Studies* 28 (5): 1312–1352. https://doi.org/10.1093/rfs/hhu089

Hoberg, G., and G. Phillips. 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies* 23 (10): 3773–3811. https://doi.org/10.1093/rfs/hhq053

Jung, W., and Y. Kwon. 1988. Disclosure when the market is unsure of information endowment of managers. *Journal of Accounting Research* 26 (1): 146–153. https://doi.org/10.2307/2491118

Kasznik, R., and B. Lev. 1995. To warn or not to warn: Management disclosures in the face of an earnings surprise. *The Accounting Review* 70: 113–134.

Lang, M., and R. Lundholm. 2000. Voluntary disclosure and equity offerings: Reducing information asymmetry or hyping the stock? *Contemporary Accounting Research* 17 (4): 623–662. https://doi.org/10.1506/9N45-F0JX-AXVW-LBWJ

Lee, S., S. Matsunaga, and C. Park. 2012. Management forecast accuracy and CEO turnover. *The Accounting Review* 87 (6): 2095–2122. https://doi.org/10.2308/accr-50220

Lennox, C., and C. Park. 2006. The informativeness of earnings and management's issuance of earnings forecasts. *Journal of Accounting and Economics* 42 (3): 439–458. https://doi.org/10.1016/j.jacceco.2006.05.001

Li, C., G. F. Peters, V. J. Richardson, and M. W. Watson. 2012. The consequences of information technology control weaknesses on management information systems: The case of Sarbanes-Oxley internal control reports. *MIS Quarterly* 36 (1): 179–204. https://doi.org/10.2307/41410413

Li, F. 2010. The information content of forward-looking statements in corporate filings—A naïve bayesian machine learning approach. *Journal of Accounting Research* 48 (5): 1049–1102. https://doi.org/10.1111/j.1475-679X.2010.00382.x

Li, F. 2011. Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature* 29: 143–165. https://www.researchgate.net/publication/228299274_Textual_Analysis_of_Corporate_Disclosures_A_Survey_of_the_Literature

Li, F., R. Lundholm, and M. Minnis. 2013. A measure of competition based on 10-K filings. *Journal of Accounting Research* 51 (2): 399–436. https://doi.org/10.1111/j.1475-679X.2012.00472.x

Lycett, M. 2013. 'Datafication': Making sense of (Big) Data in a complex world. *European Journal of Information Systems* 22 (4): 381–386. https://doi.org/10.1057/ejis.2013.10

Marr, B. 2016. *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. Hoboken, NJ: John Wiley & Sons.

McKinsey Analytics. 2018. Analytics comes of age. McKinsey & Company. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Analytics%20comes%20of%20age/Analytics-comes-of-age.ashx

Merton, R. C. 1987. A simple model of capital market equilibrium with incomplete information. *The Journal of Finance* 42 (3): 483–510. https://doi.org/10.1111/j.1540-6261.1987.tb04565.x

MicroStrategy Inc. 2018. Global State of Enterprise Analytics Report. Tysons Corner, VA: MicroStrategy Inc.

Pelland, D. 2017. Big Data needs advanced analytics to forecast the financial future. *FEI Daily*. https://daily.financialexecutives.org/optimizing-business-planning-advanced-analytics/

Raodeo, V. 2012. IT strategy and governance: Frameworks and best practice. *International Journal of Research in Economics & Social Sciences* 2 (3): 49–59. https://euroasiapub.org/category/ijress/past-ijress/volume-2-issue-3-march-2012-past-ijress/page/3/?https%3A%2F%2Feuroasiapub_org%2Fcategory%2Fijress%2Fpast-ijress%2Fvolume-2-issue-3-march-2012-past-ijress%2F

Rosenbush, S. 2012. Shell is in a 'technology race.' *The Wall Street Journal* (November 23). https://www.wsj.com/articles/BL-CIOB-1222

Sharma, R., S. Mithas, and A. Kankanhalli. 2014. Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems* 23 (4): 433–441. https://doi.org/10.1057/ejis.2014.17

Strauß, S. 2015. Datafication and the seductive power of uncertainty—A critical exploration of Big Data enthusiasm. *Information* 6 (4): 836–847. https://doi.org/10.3390/info6040836

Verrecchia, R. E. 1983. Discretionary disclosure. *Journal of Accounting and Economics* 5: 179–194. https://doi.org/10.1016/0165-4101(83)90011-3

Verrecchia, R. E. 2001. Essays on disclosure. *Journal of Accounting and Economics* 32 (1–3): 97–180. https://doi.org/10.1016/S0165-4101(01)00025-8

Wang, I. Y. 2007. Private earnings guidance and its implications for disclosure regulation. *The Accounting Review* 82 (5): 1299–1332. https://doi.org/10.2308/accr.2007.82.5.1299

Waymire, G. 1985. Earnings volatility and voluntary management forecast disclosure. *Journal of Accounting Research* 23 (1): 268–295. https://doi.org/10.2307/2490919

Yang, H. 2012. Capital market consequences of managers' voluntary disclosure styles. *Journal of Accounting and Economics* 53 (1–2): 167–184. https://doi.org/10.1016/j.jacceco.2011.08.003

APPENDIX A

Methods and Supplemental Information

Methods

Further Details on Constructing Data Analytics Variables

As noted in the "Data" section of the paper, we capture the usage of data analytics based on the occurrences of related keywords in 10-K fillings. These keywords include "data analytics/analysis," "cloud technology," "business intelligence," "business analytics/analysis," and "Big Data." Some representative examples of Big Data-related disclosures with each of the keywords are presented in Appendix B. As can be seen from these examples, firms refer to Big Data analytics in their 10-K in a variety of contexts. These include the use of Big Data analytics in business decision-making, identifying the use of data analytics by competitors as a risk factor, hiring executives with Big Data expertise, etc. Our empirical measures of Big Data analytics engagement assume firms that discuss more about Big Data in any context to be more likely to employ these tools when compared with firms that remain silent on this issue.

Data Sources and Sample Construction

With respect to other variables used in this study, we collect management earnings forecast data from I/B/E/S, financial accounting data from Compustat, stock return data from CRSP, and corporate governance data from the ISS Directors database. Our sample period is from 2010 to 2018. We start our sample in 2010 because not many firms mention Big Data in their 10-Ks prior to that.

Our textual search conducted via SEC Analytics Suite yields 4,958 firm-year observations with Big Data analytics discussions in their 10-Ks. We then eliminate 1,458 firm-year observations belonging to computer equipment/services/software industries, 363 firm-year observations with duplicate reporting records, and 1,276 firm-year observations with missing Compustat data to construct control variables. This process yields a total of 1,861 firm-year observations (643 unique firms) with Big Data activities discussions in 10-Ks. Note that our control group consists of firms from the same sample period and industries as the treatment group with the notable exception that they do not mention data analytics related words in their 10-Ks. The control group consists of 28,445 firm-year observations.

Research Design

The full specification of the probit regression employed to test our hypothesis is as follows (i and t stand for firm and year subscripts, respectively):

$$\begin{aligned} MF_{i,t+1} = {} & \beta_0 + \beta_1 AnalyticsProxy_{i,t} + \beta_2 Size_{i,t} + \beta_3 MTB_{i,t} + \beta_4 Loss_{i,t} + \beta_5 ROA_{i,t} + \beta_6 Lev_{i,t} + \beta_7 RetVol_{i,t} \\ & + \beta_8 ShrTurnover_{i,t} + \beta_9 LitiRisk_{i,t} + \beta_{10} DebtIssue_{i,t} + Industry\ FE + Year\ FE + e_{i,t} \end{aligned} \tag{A1}$$

As discussed in the "Research Design" section of the paper, the dependent variable $MF_{i,t+1}$ is an indicator variable that equals 1 if the firm issues at least one management earnings forecast following the discussion of Big Data analytics activities in the 10-K, and 0 otherwise. The variable of interest, $AnalyticsProxy_{i,t}$ is either the indicator variable *Analytics_Indicator* or the categorical variable *Analytics_RankWords*.

We follow prior studies (e.g., Kasznik and Lev 1995; Bamber and Cheon 1998; Lennox and Park 2006; Cotter, Tuna, and Wysocki 2006; Gong, Li, and Xie 2009; Feng and Koch 2010; Chen, Matsumoto, and Rajgopal 2011; Lee, Matsunaga, and Park 2012) and include an array of control variables to control for other determinants of management forecasting propensity. Specifically, we include firm size (*Size*) to control for the overall information environment and market-to-book ratio (*MTB*) to control for proprietary costs related to growth opportunities. We also control for firm performance using an indicator variable that denotes whether the firm is loss-making (*Loss*) and the firm's return on assets (*ROA*). We control for financial leverage (*Lev*) and debt financing needs (*DebtIssue*) because these characteristics could be associated with management forecast issuance (e.g., Lang and Lundholm 2000; Gong et al. 2009). We also control for information uncertainty using the standard deviation of the monthly stock returns during the year (*RetVol*). Further, we control for liquidity using monthly share turnover (*ShrTurnover*) and litigation risk (*LitiRisk*) as measured by Yang (2012). All control variables are measured contemporaneous to the variable of interest (i.e., $AnalyticsProxy_{i,t}$) so that they capture the nature of the information environment and demand for information at the time of the 10-K Big Data discussions. Complete definitions of all variables used in this study are provided in Appendix C. We also include industry fixed effects and year and fixed effects to control for industry characteristics and time trends respectively.

Additional Analyses

In this section, we report several additional analyses conducted to gain further insights into the relationship between Big Data analytics and management earnings forecasting as well as robustness tests conducted to alleviate endogeneity concerns.

The Role of Firm Size and Complexity

The incremental usefulness of Big Data analytics to improve information clarity and earnings forecastability is likely greatest for organizations that are larger and more complex. For example, organizational complexity has been shown to increase both internal and external information asymmetry (Dye 1985; Jung and Kwon 1988; Habib, Johnsen, and Naik 1997). Organizational complexity increases the costs of collecting internal information and providing external communication with shareholders. Big Data analytics' capabilities in managing large troves of data and obtaining insights are likely to be more useful for such firms. In contrast, for small firms or firms with less complex operations, the incremental benefits accrued by data analytics in this respect are likely to be lower. Accordingly, we examine whether firm size and complexity affect the positive association between Big Data analytics and the propensity to issue management forecasts. We conduct these analyses by re-running regression model (1) after partitioning the sample based on size and complexity. We measure size in terms of the log of total assets and complexity in terms of the number of business segments with high numbers representing firms with greater business complexity.

Untabulated results indicate that the positive association between Big Data analytics and the propensity to issue management earnings forecasts is stronger for larger firms when compared with their smaller counterparts irrespective of whether the engagement in Big Data analytics is measured via *Analytics_Indicator* or *Analytics_RankWords*. Similarly, we find the above relationship to be stronger for firms with greater business complexity when compared with those with lower business complexity. These findings are consistent with the notion that informational benefits accrued by Big Data analytics are likely to be incrementally greater for bigger and more complex firms.

The Role of Firm Governance

Information systems research identifies IT governance as an integral component of corporate governance (De Haes and Van Grembergen 2004; Raodeo 2012). Hence firms with better governance are expected to maximize the value of IT investments by ensuring that they align with overall corporate strategy and their benefits are distributed across functional boundaries. Extending this line of argument, one could conjecture that the benefits of Big Data analytics are more likely to spill over to other functions for better-governed firms. Prior literature suggests that firms with more dedicated institutional investors are better governed because these investors have both capability and incentives to closely monitor the management and minimize value-destroying managerial behavior (Bushee 1998, 2001). Accordingly, we examine whether the propensity of Big Data analytics to induce management earnings forecasts is greater for firms with stronger corporate governance by employing regression model (1) after partitioning the sample based on the percentage of dedicated institutional investors. Underscoring the conjectured role of governance, our results (unreported) reveal that the association between Big Data analytics and management forecasting propensity is positive and significant for firms with strong corporate governance, but not for those with weak governance.

Big Data Analytics and Sales Forecasts

The primary disclosure we investigate in our paper is management earnings forecasts. Although this choice is guided by the prior literature that highlights earnings forecasts as the major quantified voluntary financial forecast provided by managers (e.g., see Healy and Palepu 2001; Hirst et al. 2008; Beyer, Cohen, Lys, and Walther 2010), as an additional test, we also investigate whether Big Data analytics is associated with a higher propensity to issue sales forecast. We conduct this analysis by replacing the dependent variable in model (1) with an indicator variable that equals 1 if the firm issues at least one sales forecast in a given year, and 0 otherwise (*SF*). Our findings (unreported) reveal that Big Data analytics is strongly associated with the propensity to issue sales forecasts as well. In other words, it appears that greater voluntary disclosure associated with Big Data analytics is not confined to earnings forecasts only.

Robustness Tests Addressing Endogeneity Concerns

Endogeneity is a common problem in most empirical studies in accounting and financial economics. Our paper is no exception. Although fully eliminating this concern is exceedingly challenging, we conduct some robustness tests to minimize this concern.

First, we re-run our analyses with a matched control sample generated via propensity score matching (PSM). Specifically, for each year *t*, we identify firms that initiate Big Data analytics usage. We create a dummy variable *NewUser* and set it to one if a firm starts to use Big Data analytics in year *t* and 0 otherwise. We then use PSM to identify control firms that are similar but have not initiated Big Data usage during the sample period. These control firms should have the closest propensity score estimated based on the determinant model. We estimate the decision to initiate the Big Data analytics using the probit model that includes the following firm controls: firm size (*Size*), the market-to-book ratio of equity (*MTB*), firm profitability (*Std_Earn* and *ROA*), the volatility of stock returns (*RetVol*), the complexity of firm operations (*GeoDisp*), and investment intensity in fixed assets (*PPE_AT*). Using a caliper value of 0.001, this matching process yields a sample of 4,980 firm-year observations for our main test for management forecast issuance, representing 388 new users and 388 nonusers. We then re-estimate our earlier tests within this PSM sample.

Our results (untabulated) indicate that the coefficient of interest, (*Analytics_Indicator*) is positive and significant, reconfirming the previous finding that Big Data analytics usage is positively associated with the propensity to issue management earnings forecast. Moreover, we continue to find that Big Data analytics is associated with greater management forecast accuracy (supply channel) and greater analyst following (demand channel) as well. In other words, our results using the propensity score matching approach are quite consistent with those using the full sample.

Second, we note that the sample employed in the main analyses includes firms that employ Big Data analytics as well as those that do not. Employing both the pre-analytics years of firms that subsequently employ data analytics and firms that never employ data analytics as controls render, in essence, a quasi-difference-in-differences attribute to our research design. Nonetheless, to allay any concerns that this approach might lead to our coefficients picking up some inherent and time-invariant differences between firms that employ analytics and those that have never done so, we also re-run our analysis after removing firms that never employed analytics from the sample.[12] In untabulated results, we continue to find an increase in forecast propensity following the adoption of data analytics. This mitigates the concern that our finding is driven by inherent differences between data analytics and nondata analytics firms. With respect to the "channel tests," we also continue to find the adoption of Big Data analytics to be followed by an increase in analyst following (demand channel). However, we no longer find significant results with respect to changes in forecast error (supply channel).

APPENDIX B

Selected Examples of 10-K Discussions of Data Analytics

The Hartford Financial Services Group, Inc., 10-K Filed on February 24, 2017

Keywords: "data analytics," "Big Data"

Part I—Item 1. Business

Middle market

Middle market business is considered "high touch" and involves individual underwriting and pricing decisions. The pricing of Middle market accounts is prone to significant volatility over time due to changes in individual account characteristics and exposure, as well as legislative and macro-economic forces. National and regional carriers participate in the middle market insurance sector, resulting in a competitive environment where pricing and policy terms are critical to securing new business and retaining existing accounts. Within this competitive environment, The Hartford is working to deepen its product and underwriting capabilities, and leverage its sales and underwriting talent with tools it has introduced in recent years. Through advanced training and **data analytics**, the Company's field underwriters are working to improve risk selection and pricing decisions. In product development and related areas, such as claims and risk engineering, the Company is extending its capabilities in industry verticals, such as energy, construction, auto parts manufacturing, food processing and hospitality. Through a partnership with AXA Corporate Solutions, the Company offers business insurance coverages to exporters and other U.S. companies with a physical presence overseas. The Company

[12] We thank the referee for highlighting this issue.

has also added new middle market underwriters in the Midwest and Western U.S. to deepen relationships with its distribution partners.

Part I—Item 1 A. Risk Factors

Competitive activity, use of ***data analytics****, or technological changes may adversely affect our market share, demand for our products, or our financial results*

The industries in which we operate are highly competitive. Our principal competitors are other property and casualty insurers, group benefits providers and providers of mutual funds and exchange-traded products. Competitors may expand their risk appetites in products and services where The Hartford currently enjoys a competitive advantage. Larger competitors with more capital and new entrants to the market could result in increased pricing pressures on a number of our products and services and may harm our ability to maintain or increase our profitability. For example, larger competitors, including those formed through consolidation, may have lower operating costs and an ability to absorb greater risk while maintaining their financial strength ratings, thereby allowing them to price their products more competitively. In addition, a number of insurers are making use of "**Big Data**" analytics to, among other things, improve pricing accuracy, be more targeted in marketing, strengthen customer relationships and provide more customized loss prevention services. If they are able to use **Big Data** more effectively than we are, it may give them a competitive advantage. Because of the highly competitive nature of the industries we compete in, there can be no assurance that we will continue to compete effectively with our industry rivals, or that competitive pressure will not have a material adverse effect on our business and results of operations.

Xerox Corp., 10-K Filed on February 16, 2016

Keyword: "Big Data"

Innovation and Research

Xerox has a rich heritage of innovation, and innovation continues to be a core strength of the Company as well as a competitive differentiator. Our aim is to create value for our customers, our shareholders, and our people by driving innovation in key areas. Our investments in innovation align with our growth opportunities in areas like business process services, color printing and customized communication. Our research efforts can be categorized under four themes:

Usable Analytics

Transform **Big Data** into useful information resulting in better business decisions:

Competitive advantage can be achieved by better utilizing available and real-time information. Today, information resides in an ever increasing universe of servers, repositories and formats. The vast majority of information is unstructured, including text, images, voice and videos. One key research area is making sense of unstructured information using natural language processing and semantic analysis. A second major research area focuses on developing proprietary methods for prescriptive analytics applied to business processes. Here, we seek to better manage very large data systems in order to extract business insights and use those insights to provide our clients with actionable recommendations. Tailoring these methods to various vertical applications leads to new customer value propositions.

Advisory Board Co., 10-K Filed on June 15, 2009

Keyword: "business intelligence"

Business Strategy

Drive Tangible Results by Incorporating Data and Analytics into Robust, Web-Based Business Intelligence Tools

Through our best practices research, we have identified the need for **business intelligence** tools that consolidate, analyze, and report member data in order to gain visibility into areas of opportunity for operational or financial improvement. To meet this member need, we have combined commercially available and proprietary technology with our insight, industry expertise, and standardized data definitions to offer robust, web-based business intelligence tools. These tools provide valuable insights by efficiently presenting regularly updated data extracted from what are often numerous and disparate systems, benchmarking their performance versus other organizations, and allowing them to drill down to transaction-level detail. These sophisticated tools hardwire our existing best practices research into the daily and weekly process flows at the member institutions, thereby allowing a broad group of executives, managers, and front-line leaders to leverage the insights and data in both their daily and strategic decisions. This integrated approach allows our members to achieve ongoing, tangible cost and performance gains and high return on investment. We frequently update our business intelligence tools through benchmarking, research activities, and member input.

Kimball International Inc., 10-K Filed on August 30, 2016

Keyword: "business analytics"

Executive Officers of the Registrant

Mr. Nicholson was appointed Vice President, Chief Administrative Officer in February 2015 with responsibility for the human resources and information technology functions. He also served as Vice President, Chief Information Officer from January 2014 until March 2015. Throughout 2013 he served as Director, **Business Analytics** and then Vice President, **Business Analytics**, with oversight of strategic application of data analysis, social media and mobile computing in support of the Company's growth of information management into more predictive analysis in order to build greater responsiveness to customer needs and improvement of operational decision making. He also served as Director of Organizational Development from November 2011 until January 2013, and Director of Employee Engagement from November 2008 until November 2011 following other roles of advancing responsibility in the areas of application development, systems analysis, process reengineering, lean/continuous improvement and enterprise resource planning ("ERP") since joining the Company in 1986.

APPENDIX C

Variable Definitions

Variables	Definition
Variables for cross-sectional tests:	
Dependent variables:	
MF	An indicator variable that equals 1 if the firm issues at least one management earnings forecasts following the discussion of Big Data analytics activities in the 10-K, and 0 otherwise.
AF	Analyst following, calculate as the natural logarithm of 1 plus the number of analysts following the firm during the year.
MFE	Management earnings forecast error, measured as the absolute value of the firm's last management earnings forecast for the year minus actual earnings per share for the year, scaled by the stock price at the beginning of the year. For range forecasts, we use the upper-bound of the range.
SF	An indicator variable that equals 1 if the firm issues at least one annual sales forecast for the year, and 0 otherwise.
Big Data analytics disclosure proxies:	
Analytics_Indicator	Indicator variable that equals 1 if the firm disclosed any of the key words related to Big Data analytics in its 10-K for a given year, and 0 otherwise.
Analytics_RankWords	Quartile ranking of the ratio of key words on Big Data analytics to the total number of words in the 10-K.
Control variables:	
Size	Firm size, measured as the natural logarithm of firm's total assets at the end of year.
MTB	Market-to-book ratio, measured as the market value divided by the book value of common equity of the firm at the end of year.
Loss	Indicator variable equal to 1 if a firm incurs loss in the year, and 0 otherwise.
ROA	Return on assets, calculated as net income divided by total assets.
Lev	Leverage ratio, calculated as long-term debt divided by total assets.
RetVol	Stock return volatility, calculated as standard deviation of the monthly returns during the fiscal year.
ShrTurnover	Share turnover ratio, calculated as monthly total trading volume divided by shares outstanding.
LitiRisk	Indicator variable equal to 1 if firm is in of the following high-litigation-risk industries: SIC codes 2833–2836 (biotechnology), 3570–3577 and 7370–7374 (computers), 3670–3674 (electronics), 5200–5961 (retailing), and 8731–8734 (R&D service), and suffers a 20 percent or greater decrease in earnings; 0 otherwise. (Yang 2012)
DebtIssue	Indicator variable equal to 1 if book value of debt increases in the following year is more than 5 percent of current year total assets, and 0 otherwise.
Size	Firm size, measured as the natural logarithm of firm's total assets at the end of year.

(continued on next page)

APPENDIX C (continued)

Variables	Definition
Variables for cross-sectional tests:	
BusSeg	Total number of segments, calculated as the natural logarithm of 1 plus the total of business and geographic segments.
Dedicate%	Percentage of shares held by dedicated institutional investors as per Prof. Brian Bushee's Institutional Investor Classification data.
Additional variables for propensity score matching tests:	
NewUser	Indicator variable equal to 1 if a firm starts to use Big Data analytics in year *t*, and 0 otherwise.
GeoDisp	Geographic dispersion (e.g., Bushman, Chen, Engel, and Smith 2004), measured as sum of the squares of (firm sales in each geographic segment/total firm sales) minus one, then multiplied by negative one.
PPE_AT	Investment intensity in fixed assets, measured as total PPE assets value divided by total assets.
Std_Earn	Standard deviation of earnings (excluding extraordinary items) from prior four years, scaled by total assets.