The Informativeness of Dark Data for Future Firm Performance

Ewelina Forker
Emory University
ewelina.forker@emory.edu

May 1, 2023

I thank my dissertation committee Karen Sedatole (chair), Isabella Grabner, David Schweidel, and Kristy Towry for valuable guidance, insight, and encouragement. I appreciate the helpful comments and suggestions from Susanna Gallani, Hayden Gunnell, Ron Harris, Melissa Martin, Robbie Moon, Sara Palmer, Jingyi Qian, Suhas Sridharan, as well as seminar participants at the University of Amsterdam, Ole Miss, Georgia State, Notre Dame, University of Wisconsin-Madison, University of Chicago, University of Georgia, University of Virginia, University of Washington, Cornell University, and the 2022 AAA/Deloitte Foundation/J. Michael Cook Doctoral Consortium. I thank the partnering firm for supplying data and extensive access to the research site and its employees. I gratefully acknowledge financial support from Emory University and the Deloitte Foundation.

The Informativeness of Dark Data for Future Firm Performance

ABSTRACT

This study investigates whether an organization's "dark data" can predict future firm performance. "Dark data" represents the vast amounts of information that organizations collect and store but fail to use for decision-making purposes. I examine if the aggregate sentiment from employee emails, a ubiquitous form of dark data, predicts future sales incremental to traditional information sources. I further study if differences in the information environment explain cross-sectional variation in the predictive ability of aggregate email sentiment. To answer these questions, I collaborate with a large, U.S.-based medical technology company seeking to improve its planning process. The firm provides financial data as well as access to over 200,000 de-identified emails from more than 200 employees directly involved in planning across sales, operations, and accounting functions. Using the firm's dark email data, I estimate aggregate employee sentiment by product-month and test its relation to future performance. My results show that, even after controlling for known predictors of future sales, product-specific sentiment is significantly associated with future sales. Further, the predictive ability of aggregate employee sentiment for future sales is greater for growth and declining products for which information uncertainty is greater, and for emails sent by rank-and-file employees for which information sharing is more costly. This study provides novel evidence of how organizations can use dark data to identify inefficiencies in forecasting and planning processes.

I. INTRODUCTION

Technological advances in data storage and processing capabilities have led to dramatic increases in the volume, variety, and velocity of information organizations collect and store. While this information is intended to spur more data-driven decision-making within the firm, according to recent market intelligence surveys, anywhere from 75% to 90% of all data stored within an organization is "dark data" and remains unused for analytics (Splunk 2019; Taulli 2019). "Dark data" is defined as the "information assets organizations collect, process and store during regular business activities, but generally fail to use for other purposes" (Gartner 2022).[1] Dark data is typically soft, qualitative information in an unstructured format that is internally generated and incidentally collected across a large number of organizational stakeholders such as employees, customers, and suppliers (Johnson 2020). Examples of dark data include email correspondence, customer call records, and online meeting logs. One major reason behind the underutilization of dark data is its unstructured format that makes it difficult for organizations to recover, process, and analyze. Empirical evidence of whether and how dark data can be used to improve internal decision-making is limited. This study seeks to provide evidence on whether an organization's dark data can be used to improve the sales and operations planning (S&OP) process. Specifically, I examine whether dark data predicts future firm performance and how characteristics of the information environment affect its predictive ability.

The S&OP process is characterized by the planning and coordination efforts between commercial, operational, and accounting functions. The purpose of this formalized process is to forecast revenues at the product-level and coordinate resources to meet these revenue forecasts

[1] The concept of "dark data" is closely related to the emerging construct of "exhaust data" from the information systems literature. O'Leary and Storey (2020) define "exhaust data" as "data that, initially, is not core data, but may be collected as a byproduct of some event (transaction, event, search, disclosure, etc.), has unknown value, and ultimately might be used for another purpose to create value."

while maintaining or improving operational efficiency.[2] Ineffective S&OP is costly for firms, resulting in lost revenue (Watson 1987), increased supply chain costs (Oliva and Watson 2011), and unfavorable capital market consequences such as less accurate earnings forecasts (Ittner and Michels 2017) and an increased likelihood of misreporting (Kroos et al. 2021). Effective execution of the S&OP process relies on efficient information sharing within and across functions as well as the accurate aggregation of information across these different sources.

Within this important yet challenging context, I investigate if dark data embedded in employee emails possesses untapped information incremental to traditional information sources. More, specifically, I examine if a measure of aggregate sentiment from employee emails predicts future sales. Further, I examine if differences in the degree of information uncertainty and the hierarchical distance of email senders from top management explain cross-sectional variation in the predictive ability of aggregate email sentiment.

To test these questions, I collaborate with a large, U.S.-based medical technology company seeking to improve its S&OP process. Over a period of 14 months during 2021-2022, I collect the textual attributes and meta data of nearly 225,000 emails (scrubbed of sender and recipient names) from the "Sent" mail folders of over 200 employees directly involved in the S&OP process. I employ the machine learning textual analysis algorithm used in Wen et al. (2020) to obtain the sentiment of individual emails on a continuous scale from 0 (negative) to 1 (positive). I aggregate this to a product-specific email sentiment measure by product-month. I also collect both forecasts and actual performance for monthly sales by product over the same period.

Using a hierarchical linear model to accommodate the nested nature of the data structure, I empirically test the relation between aggregate email sentiment and future firm performance; that

[2] In contrast to budgeting, goal-setting, and other strategic planning processes typically studied in the accounting literature, S&OP is executed periodically, usually monthly or quarterly, rather than annually.

is, product-month actual unit sales. Email sentiment is aggregated by product-month for the month leading up to the forecast decision. To distinguish the incremental information value from aggregate email sentiment from traditional information sources, I include a control for forecast values. The forecasted values for unit sales provide a useful summary of all formal (e.g., quantitative or financial) and informal (e.g., qualitative or tacit knowledge) information known at the time of the forecast. Hence, if aggregate email sentiment is significantly associated with future performance after controlling for forecast values, the sentiment represents the amount of information that was available but not incorporated through traditional means during the formal S&OP process.

I posit that soft, qualitative information is typically more difficult and costly to share within an organization, making it more likely that soft information is excluded from formal information aggregation processes (Frame et al. 2001; Liberti and Petersen 2019). Email communication between employees provides a rich set of soft information, and textual analysis provides a means to harden this valuable but costly information. Indeed, I find that aggregate email sentiment predicts future unit sales, incremental to information incorporated into the forecast. In terms of economic significance, a one standard deviation increase in aggregate sentiment is associated with a 7.5% increase in unit sales.

In cross-sectional analysis, I investigate whether the degree of information uncertainty and the cost of information sharing moderate the relation between aggregate email sentiment and future performance. I use three proxies for information uncertainty: products in the growth phase of the product life cycle, products in the decline phase of the product life cycle, and demand uncertainty measured as the unpredictable portion of demand variation. I find that the predictive ability of aggregate sentiment for unit sales is driven by growth and decline products, suggesting that the hardened soft information from employee emails is especially valuable when there is less historical, hard data available for decision-making or when uncertainty about future prospects is greater. For growth products, a one standard deviation increase in aggregate sentiment is associated with a 21.4% increase in unit sales. For products in the decline phase, a one standard deviation increase in aggregate sentiment predicts an astounding 317% increase in unit sales. The predictive ability of aggregate email sentiment for unit sales does not, however, vary with increasing demand uncertainty.

I further examine whether the cost of information sharing moderates the relation between aggregate email sentiment and future performance using the email sender's hierarchical distance from top management as my proxy for information sharing cost. I expect and find that the predictive ability of aggregate email sentiment for unit sales is stronger for emails sent from lower-level rank-and-file employees where the cost of information sharing is greater owing to their dispersion in the organization. My findings are robust to alternative measures of sentiment based on three common dictionary-based measures.

This study contributes to the stream of accounting research focused on understanding attributes of the internal information environment (Hemmer and Labro 2008; Ittner and Michels 2017; Brüggen et al. 2021; Kroos et al. 2021). Specifically, it is the first, in any discipline, to study ways that organizations can use unstructured data to harden and quantify soft information to improve internal forecasting and planning. Understanding the conditions when hardened soft information may be more valuable provides insight into where inefficiencies exist in the aggregation of information for planning purposes. I document that in cases of greater information uncertainty and when information sharing costs are higher, employees have valuable soft information about future firm performance that is not efficiently captured through traditional planning processes.

I also contribute to research using textual analysis of qualitative information to assess firms' future prospects. One stream of this research examines the predictive value of unstructured information in firm disclosures such as 10-K filings, earnings press releases, and conference calls (Li 2010; Larcker and Zakolyukina 2012; Davis et al. 2012; Huang et al. 2014; Donovan et al. 2021). Another stream investigates the informativeness of "crowd sourced" information from sites such as Twitter, GlassDoor, and Seeking Alpha (Chen et al. 2014; Bartov et al. 2018; Hales et al. 2018). Mine is the first study to examine the predictive ability of internally-generated dark data for financial and operational performance. While prior studies find that broadly disseminated, external data has information value, I demonstrate that firms have an opportunity to capitalize on the vast troves of data that are collected and stored internally but unutilized for data-driven decisions.

Finally, the results from this study offer practical implications for firms seeking to improve the S&OP process with better information aggregation through utilization of unstructured data. There are many anecdotal examples of the potential for unstructured data to help organizations improve decision-making (Marr 2019; Harbert 2021). This study seizes on an opportunity for academic research to lead practice by providing empirical evidence on the value of email dark data in predicting future firm performance.

The next section provides background and develops hypothesis. Section III describes the research design, Section IV discusses results, and Section V concludes.

II. BACKGROUND AND THEORETICAL DEVELOPMENT

Predictive Ability of Unstructured Data

An extensive accounting and finance literature has emerged using textual analysis of unstructured, external data to predict firm performance and stock market reactions. One stream focuses on the informativeness of qualitative company disclosures (Li 2010; Larcker and Zakolyukina 2012; Jegadeesh and Wu 2013; Huang et al. 2014; Donovan et al. 2021). In terms of sentiment, Li (2010) finds that the tone of forward-looking statements in the MD&A section of 10-K filings positively predicts future earnings. Jegadeesh and Wu (2013) find that both positive and negative tone in 10-K filings is associated with market reaction for up to two weeks after the filing is released. Huang et al. (2014) analyze earnings press releases and find that abnormally positive tone is associated with negative future earnings and cash flows, positive future earnings restatements, and positive stock returns immediately after the announcement followed by negative returns in subsequent quarters. Importantly, the data that are the focus of these studies, while unstructured, are not "dark" in that they are purposefully gathered and presented for the intended use of external stakeholders.

Another stream of research examining unstructured data focuses on the value-relevance of externally generated crowd-sourced information (Chen et al. 2014; Bartov et al. 2018; Hales et al. 2018; Tang 2018; Campbell et al. 2019). A subset of this research uses unstructured data from social media sources to examine the predictive ability of aggregate sentiment on various capital market outcomes. Chen et al. (2014) find that the sentiment of Seeking Alpha articles and their corresponding user comments predict stock returns and earnings surprise, providing evidence for the value-relevance of non-professional analysts. Bartov et al. (2018) document that firm-specific Twitter sentiment predicts future firm-level earnings and announcement returns, with sentiment having stronger predictive ability for tweets related to firm fundamentals and for firms in weaker information environments. Meanwhile, Tang (2018) finds that product-specific Twitter sentiment predicts firm-level sales, with greater predictive ability for firms with major consumer customers, as opposed to major business customers, and when advertising is limited.

While this growing body of literature provides compelling evidence on the predictive ability of sentiment in externally available data on firm and market outcomes, the generalizability of the findings and applicability of the proposed mechanisms to an internal forecasting and planning setting are questionable for several reasons. First, firm disclosures are strategic communications from top managers seeking to credibly manage firm impressions. As such, rather than being predictors of outcomes such as stock returns, they are intended to affect those outcomes. In contrast, routine internal communication among employees via email is not necessarily strategic and comes from all levels of an organization, including from rank-and-file employees who may not have visibility to information across functions and levels that may materially affect the firm's performance.

Furthermore, the underlying assumption in the second stream of research relies on the existence of the "wisdom of crowds" where a large group of *independent* and *diverse* individuals can often predict outcomes more accurately than experts (Galton 1907; Surowiecki 2004; Gardner and Tetlock 2015). The "wisdom of crowds" concept is less likely to generalize to internal decision-making within a single organization. Organizations are composed of a relatively homogenous, smaller (relative to social media audiences) number of self-selected employees who engage with each other on a day-to-day basis. As such, phenomena such as "herding" and "groupthink" are common in internal information aggregation processes, coming at the expense of prediction quality (Whyte 1989; Prendergast 1993). Thus, whether the unstructured, dark data of employee email correspondence will be predictive of firm performance remains an open empirical question.

Sales and Operations Planning (S&OP)

This study examines the predictive ability of dark data in the context of a firm's sales and operations planning (S&OP) process. S&OP is a collaborative and formalized process that coordinates demand, supply, and financial planning (Lapide 2004; Oliva and Watson 2011). Demand planning typically involves sales and marketing functions predicting future customer demand based on scheduled customer orders and future potential orders influenced by changing market conditions, promotional activities, and new product introductions by the organization or its competitors. In supply planning, various operations functions within the supply chain coordinate the supply of material, labor, and equipment to efficiently fulfill demand forecasts.

S&OP is an iterative process, typically repeated monthly, that uses formal coordination mechanisms to reach consensus among different functions on sales and production plans. The process culminates in the executive approval of a feasible financial plan that dictates subsequent resource allocation and operational activities to efficiently meet customer needs. It provides an ideal setting to study the value of employee email correspondence in internal decision-making because of the importance of efficient information aggregation across functions and hierarchical levels of the organization.

Efficient information aggregation means that all known information is fully incorporated into a final decision. Efficient information aggregation depends on two key elements. First, it relies on the timely and complete collection of all known information available at the time of the decision. Second, it involves the accurate combination of available information. This combination is based on an appropriate weighting of factors dictated by the volume, reliability, and relevance of information for the decision. The efficient aggregation of information in planning processes produces more accurate forecasts that enable better resource allocation decisions, resulting in higher revenues, increased inventory turnover, and lower operational costs (Vickery et al. 2003; Droge et al. 2004; Oliva and Watson 2011). While an integrated S&OP process serves as a formal mechanism to facilitate the efficient flow of information, the question of interest for this study is whether dark, unstructured sources of data can be aggregated so as to identify any inefficiencies in this formal process.

Hard Versus Soft Information

The form of information, whether hard or soft in nature, influences the extent to which the information can be efficiently aggregated and, as a result, its usefulness for internal decision-making (Liberti and Petersen 2019). While no clear dichotomous distinction defines hard versus soft information, hard information is typically characterized as quantitative, objective, and verifiable (e.g., interpreted in the same way by different parties) with the meaning of its content being independent of its context (Stein 2002). Compared to soft information, it also has lower transaction costs for collection, processing, and transmission (Frame et al. 2001). Historical sales data are a classic example of hard information used in organizational forecasting and planning. Such data are objective, numerical values that are interpreted by the two parties in the same way.

Dark, unstructured data, however, is typically soft information (Johnson 2020). Soft information is qualitative and subjective and more difficult to directly verify. Context plays a critical role in the interpretation of its content. Further, unlike hard information that primarily relies on historical data, soft information can also include forward-looking narratives (Francis et al. 1997; Sedor 2002). An example of soft information includes customer preferences. While an organization can see if a customer purchased a particular product in the past, the historical data does not reveal why the customer made the choice and what preferences are for future purchases. In this case, a sales representative can communicate with the customer to extract information on preferences for price, quality, and functionality while assessing future intentions to buy the product.

These contrasting attributes highlight the tradeoff in the utilization of hard or soft information for decision-making. While hard information is generally easier to collect, interpret, and transmit, it suffers significant information loss about the context from which it is collected (Liberti and Petersen 2019). Thus, the exclusion of soft information, often rich in contextual content, serves as a major source of inefficiencies in information aggregation processes.

Within an organization, employee email correspondence offers a rich set of soft information with the potential to provide context missing in more traditional sources of hard data (e.g., historical sales). Email, a dominant means of communication in most organizations, provides a ubiquitous information sharing channel by which soft information is shared for internal decision-making (Mazmanian et al. 2006; Fragale et al. 2012; Wen et al. 2020). This unstructured source of dark data can be hardened through textual analysis to ascertain "sentiment," thereby extracting useful information content into a succinct, summary measure that is easier to compile and analyze.

Sentiment is a content attribute extracted from textual information reflecting the net of positive and negative opinion about a specific context (Das, Martinez-Jerez, and Tufano 2005; Das and Chen 2007). While condensing text into a quantitative index only captures a fraction of the information from the original text, recent studies exploring the importance of email content features suggest email sentiment can predict organizational outcomes such as employee job turnover (Gloor et al. 2017), individual performance evaluations (Wen et al. 2020), and even future team performance (Wasiak et al. 2011). For example, Wasiak et al. (2011) investigate how email content affects project performance. The authors assert that both positive and negative sentiment play a role in information sharing and problem solving that can motivate quicker resolution of issues to improve project management.

In the context of S&OP, I argue that sentiment from employee emails provides soft information not reflected in other, more traditional data sources that make it incrementally informative regarding future performance. I make the following prediction:

H1: *Aggregate email sentiment predicts future firm performance incremental to traditional information sources.*

There are a few key sources of tension for this prediction. Recent advances in the sophistication of predictive analytic models for internal forecasting and planning paired with an increasing volume and variety of hard information available for these models may decrease the private information advantage of employees (Labro et al. 2023). Moreover, it is plausible that the information employees convey via email is biased, noisy, or already incorporated through traditional information sources.

Information Uncertainty

Uncertainty is an inherent condition in any prediction task. The degree of information uncertainty is characterized by ambiguity of future events, stemming either from poor or limited information or from variability in expected outcomes (Zhang 2006). This uncertainty influences the efficiency by which hard information throughout the organization is aggregated (Khatri and Ng 2000). More specifically, as information uncertainty increases, the availability, timeliness, reliability, and relevance of hard information typically decreases (Mintzberg 1994).

When uncertainty is high, hard information such as historical data, provides an incomplete picture of the possible future outcomes that may be observed. Soft information becomes especially useful to fill the void because it can provide timely and relevant information specific to an ever-changing context. When the Covid-19 pandemic started, many organizations could no longer rely on historical hard information as these data quickly became outdated and, hence, irrelevant for the task of predicting future outcomes. In response, organizations sought out soft information from their customers to augment incomplete information and to improve their planning decisions (Becdach et al. 2020). While the pandemic might be an extreme example, Garcia (2013) provides empirical evidence on the value of soft information in the context of recessions. He finds that the predictive ability of news sentiment on stock returns is concentrated in times of economic recession, suggesting that the value of soft, qualitative information is stronger in times of greater uncertainty.

There is an extensive literature on the heuristics and biases that negatively affect decision quality for human judgment under uncertainty, suggesting that soft information from employees during times of greater uncertainty may be of questionable value (Kahneman et al. 1982; Dawes et al. 1989). Nonetheless, recent empirical research by Choi et al. (2022) provides evidence that employees are especially adept at anticipating future changes in product demand when environmental uncertainty is higher. Their findings suggest that employees with access to soft, qualitative information relevant to the forecasting decision, are able to capitalize on weaknesses of predictive analytics models that use hard, historical information as inputs. Similarly, I expect that the usefulness of soft information for predicting future outcomes will grow as uncertainty increases when hard, historical information is less valuable. Thus, I predict that hardened soft information in the form of aggregate email sentiment will have greater predictive ability for future firm performance when uncertainty is greater. Formally stated:

H2: *The predictive ability of aggregate email sentiment for future firm performance incremental to traditional information sources is stronger with increasing information uncertainty.*

Hierarchical Distance

The efficient aggregation of information in forecasting and planning also depends on the source of information – that is, where in the organizational hierarchy the employee who possesses the information sits. A number of theoretical models predict that that the incorporation of soft information for decision-making is more difficult with increasing hierarchical distance between an agent and principal (Radner 1993; Bolton and Dewatripont 1994; Garicano 2000; Stein 2002; Li and Suen 2004).

Information sharing of soft information across the hierarchy is difficult for two primary reasons. First, soft information possessed by employees in lower levels of the organization is costly to communicate up the hierarchy (Jensen and Meckling 1992). Verbal communication is often the most effective way to convey soft information, but lower-level employees typically have fewer opportunities for in-person interactions or face-to-face meetings with decision-makers higher up in the organization. In addition, the volume and context-specificity of soft information make it difficult to aggregate and summarize for others. Thus, soft information from employees in lower levels is more likely to be diluted and/or transformed (e.g., as in the well-known "telephone game") as it flows through different levels of the hierarchy. Finally, it is often difficult to know ahead of time what specific information may be valuable to share with others.

Second, even when employees in lower levels of the organization attempt to share soft information, that information may be overlooked or even ignored by those in higher levels of the organization. This, in turn, reduces the incentives for employees to share that information in the first place (Liberti and Petersen 2019). Liberti and Mian (2009) provide empirical evidence in support of this assertion in a lending setting, finding that greater hierarchical distance between loan officers and their superiors leads to less reliance on soft information by superiors for loan approval decisions. Moreover, lower employees that are most likely to possess soft information (e.g., customer-facing employees) often do not, themselves, have the authority to act on that information (Aghion and Tirole 1997; Stein 2002), meaning such information goes unutilized in decision-making.

Together, these arguments suggest that in hierarchical organizations, soft information from lower levels of the organizations is less likely to be included in traditional information aggregation processes. Thus, I expect the predictive ability of soft information in the form of aggregate email sentiment to increase with greater hierarchical distance between a sender and top management. I make the following prediction:

> ***H3:*** *The predictive ability of aggregate email sentiment for future firm performance incremental to traditional information sources increases with the sender's hierarchical distance to top management.*

III. RESEARCH DESIGN

Research Site

To test my hypotheses, I collaborate with a large, U.S.-based medical technology company seeking to improve its planning process. The firm supplies surgical implants and specialty tooling to its customers, composed of hospitals, surgical centers, and individual surgeons. Customers purchase the surgical implants and must use the proprietary tooling provided by the firm to safely and effectively complete surgical procedures. The surgical implants are consumable products that generate revenue for the firm, but the tooling items are equipment assets that are allocated across customers as needed.

The key planning decision for the firm involves forecasting unit sales of the surgical implants. These decisions, however, are made within the context of ensuring that there are enough tooling assets available to support demand. Tooling assets are allocated to customers either via consignment for high volume customers or through a loaner program for customers with infrequent orders of the surgical implants. Commercial teams for the firm prefer to consign tooling to customers for extended periods of time to assure asset availability rather than working through the more onerous loaner process that typically allocates an asset to a customer for a limited 5-day period and requires extensive logistics coordination between customers. Importantly, operations teams do not have the capacity or resources to make tooling that will sit idle at the customer. To maximize profit, the firm seeks to increase utilization of existing tooling rather than producing new assets.

The firm created a formal S&OP process in 2020 to improve coordination between its commercial and operations teams with objectives to improve customer satisfaction, increase asset utilization, and lower costs. Figure 1 details the S&OP process which ultimately produces monthly forecasts for unit sales based on planned asset utilization to align and coordinate execution activities within the organization.

Data Collection

Firm data span multiple levels of a nested data structure consisting of a two-level product hierarchy, with part numbers nested within brands. Figure 2 provides a visual of the multi-level structure and how data are collected from these levels for unit sales and aggregate email sentiment. The firm generates the monthly forecast and actual unit sales from July 2021 through June 2022 for 5,715 unique part numbers, resulting in 40,647 part number-month observations. The forecasts for unit sales are based on a 3-month planning horizon. Thus, to match up with actual values, the final sample covers unit sales forecasts from July 2021 through March 2022.

The firm also provides access to de-identified emails from the “Sent” mail folders of the employees directly involved in the S&OP process for messages sent between July 2021 and August 2022. Email is a ubiquitous form of communication within organizations that is used to execute day-to-day business operations. The sample contains *all* emails from employees directly involved in the planning process. This proprietary access to all emails provides a comparative advantage in research design over prior literature studying the predictive ability of unstructured data for future firm performance that tend to suffer from selection concerns.

I focus on “Sent” mail messages because they capture intentional information sharing of the employees. Further, “Inbox” messages may contain noisy, unsolicited information and their inclusion would significantly increase the computational burden. Due to privacy and sensitivity concerns, I do not obtain and store full emails. Rather, I develop a script that parses and cleans email messages, tags email with brand identifiers, and then extracts textual analysis features for each email with a non-empty message body field.[3] The script is run on company servers. I obtain a researcher output file that contains de-identified email meta-data (e.g., sender and recipient user IDs, date, email size, etc.) along with the calculated sentiment attributes for each email message. The firm also provides the organizational level and function of each sender, matched on sender ID. The email data sample is comprised of 219,198 messages sent between July 2021 and March 2022 to correspond with the time period from the forecast data sample.

[3] The script was developed in Python as a turnkey solution for the firm based on a subset of email data from a senior level manager within the firm. First, the script parses Outlook .pst files to read in email contents. Next, the script executes a series of tagging and cleaning routines, establishing email IDs, replacing sender and recipient names with unique user IDs, identifying emails with brand tags, and removing email chains and signature blocks. Finally, the script runs a series of commands to extract textual analysis features before exporting the researcher output file, which is scrubbed of any email content.

To link the individual email message data with the performance data, I aggregate email sentiment at the brand-month level. Emails are tagged by brand based on an extensive keyword search of brand names in the subject line and body of the email. The keyword search criteria for each brand has been verified by the firm. The unit sales data contains aggregate sentiment for 58 brands, resulting in 496 unique brand-month observations with repeated observations at the part number level.

Model

To test the predictive ability of aggregate email sentiment for future firm performance, I estimate a model regressing actual performance in month *m* on sentiment from month *m-3*, using the forecast in month *m-3* as a comprehensive summary of all information from traditional sources relevant to forecasting demand made at the time of the forecast. Figure 3 documents the timing sequence between these key milestones in the S&OP process. Based on the nested nature of the data, I use hierarchical linear modeling, which addresses the lack of independence across observations from the same brand and overcomes limitations of conventional techniques (e.g., aggregation bias) for estimating relations between variables at different levels (Raudenbush and Byrk 2002). To test H1, I specify the two-level model for unit sales with product part number as level 1 and brand as level 2 as follows:

Level 1: $ActualSales_{P,B,m} = \alpha_{0,B} + \alpha_{1,B}\, FcstSales_{P,B,m-3} + [\text{Controls}]_{m-3} + e_{P,B,m}$

Level 2: $\alpha_{0,B} = \beta_{0,0} + \beta_{0,1}\, Sentiment_{B,m-3} + \mu_{0,B,m}$ (1)

In Equation (1), *ActualSales* denotes the actual unit sales of part number *P* within brand *B* in month *m*. *FcstSales*, forecasted values of unit sales, is a level 1 variable that controls for information from traditional sources at the time *m-3*. *Sentiment*, a level 2 variable, is the aggregate sentiment for brand *B* from individual emails sent in month *m-3*. Higher values represent a more positive sentiment and lower values represent a more negative sentiment. The model also includes additional controls, described below, to control for seasonality, broad economic conditions, and general employee sentiment. Level 2 of the model incorporates brand random effects.[4] All variable definitions are detailed in Appendix A. Actual and forecast unit sales are winsorized at the top and bottom one percent to minimize the influence of outliers. H1 asserts that the aggregate brand-month sentiment from individual emails predicts future firm performance. Thus, I expect $\beta_{0,1}>0$.

To understand how the information environment affects the predictive ability of email sentiment, I add to Equation (1) interactions between *Sentiment* and the proxies for information uncertainty (a level 1 variable), described in detail below. To test H2, I specify the following two-level model for unit sales:

$$\text{Level 1:} \quad ActualSales_{P,B,m} = \alpha_{0,B} + \alpha_{1,B}\, FcstSales_{P,B,m\text{-}3} + \alpha_{2,B}\, InfoUncertainty_{P,B,m\text{-}3} + [\text{Controls}]_{m\text{-}3} + e_{P,B,m}$$

$$\text{Level 2:} \quad \alpha_{0,B} = \beta_{0,0} + \beta_{0,1}\, Sentiment_{B,m\text{-}3} + \mu_{0,B,m}$$

$$\alpha_{2,B} = \beta_{2,0} + \beta_{2,1}\, Sentiment_{B,m\text{-}3} + \mu_{2,B,m} \qquad (2)$$

H2 predicts stronger predictive ability with increasing information uncertainty and thus, I expect $\beta_{2,1}>0$.

To test whether the predictive ability of email sentiment varies with a sender's level in the organizational hierarchy, I replace *Sentiment* in Equation (1) with an aggregate email sentiment partitioned by senders' hierarchical distance, described below. To test H3, I specify the following two-level model for unit sales:

[4] Following suggestions of Kreft and De Leeuw (1998), I first estimate a unit sales model with only the intercept term and brand random effects. The analysis indicates that between-brand differences explain 39% of the total variation in unit sales. Further, the likelihood ratio test rejects the null hypothesis that there are no brand differences (χ^2=5650.72, p=0.000), indicating that the multilevel model is preferred over a single-level model. Together, the results from these "null model" estimates provide strong justification for fitting multilevel models in this setting.

Level 1: $ActualSales_{P,B,m} = \alpha_{0,B} + \alpha_{1,B}\ FcstSales_{P,B,m\text{-}3} + [\text{Controls}]_{m\text{-}3} + e_{P,B,m}$

Level 2: $\alpha_{0,B} = \beta_{0,0} + \beta_{0,1}\ SentimentOrg\text{-}[X]_{B,m\text{-}3} + \mu_{0,B,m}$ (3)

In Equation (3), *SentimentOrg-[X]* is the aggregated sentiment by brand-month for a sender that is *X* levels removed (i.e., lower) from the highest level of the organizational hierarchy. Higher values thus indicate a sender that is further from top management. H3 predicts stronger predictive ability with increasing hierarchical distance. Thus, I expect $\beta_{0,1,[X=2]} < \beta_{0,\,1,[X=3]} < \beta_{0,\,1,[X=4]} < \beta_{0,\,1,[X=5]}$.

Sentiment Measures

The main independent variable of employee email sentiment relies on a holistic content assessment using the naïve Bayes algorithm from Wen et al. (2020) that is built into the Condor software application. This naïve Bayes algorithm has been trained on over a billion tweets, achieving over 80% accuracy on many email corpora (Wen et al. 2020). Each email is assigned a sentiment score ranging for 0 (negative) to 1 (positive). Appendix B provides email message examples and their associated sentiment score to help readers gauge the face validity of the sentiment measure. For variable *Sentiment*, I aggregate email sentiment at the brand-month level and match it to 3-month-ahead unit sales to capture the product-specific sentiment at the time of information aggregation into the forecast in the S&OP process.

In addition to machine learning-based sentiment measures, I also construct three dictionary-based measures commonly used in prior literature. Each of the measures is based on the difference in counts of positive and negative words scaled by the total words in the message. Higher values denote more positive sentiment and theoretically could range from -1 (negative) to 1 (positive). *SentBlob* is based on the positive and negative word lists in the TextBlob Python library, *SentH4* uses the word lists from the Harvard IV-4 dictionary, and *SentLM* uses the

Loughran-McDonald master dictionary for financial text (Loughran and McDonald 2011). All three methods exclude words with negations. The measures are aggregated at the brand-month level and matched to 3-month-ahead unit sales.

For future robustness, I also construct the variable *SentBERT* based on the Google BERT algorithm, an advanced supervised deep learning model using Bidirectional Encoder Representations from Transformers (Devlin et al. 2018). This machine learning model is pre-trained on English Wikipedia and is fine-tuned on the Stanford Sentiment Treebank v2 (SST2), used for predicting sentiment from longer movie reviews. The assigned sentiment values for each email range from -1 (negative) to 1 (positive). Due to the extensive computational burden to run the model, data collection is still ongoing and is expected to be completed by Fall 2023.

Information Uncertainty

To capture information uncertainty, I use three empirical proxies. The first proxy, *Growth*, is based on the company-defined measure of product life cycle. The organization categorizes products across four stages: growth, maturity, retiring, and decline.[5] Products in the growth stage have less historical data available for information aggregation, representing one aspect of greater information uncertainty based on poor or limited information. *Growth* is an indicator variable that takes the value of 1 for growth stage products and 0 otherwise. On the other end of the product life cycle, we have products in the decline stage, which represent the second proxy *Decline*. While there is extensive historical data available for these products, its relevance for predicting future sales is less clear since future trends are distinctly different than past trends for these aging products. Uncertainty about future outcomes is often compounded by the introduction of new products that may cannibalize the demand for declining products at some uncertain rate. Together,

[5] The firm has a fifth category for obsolete products but the data sample does not contain any products from the obsolete stage because these products are not forecasted by the firm.

these characteristics represent another aspect of greater information uncertainty based on variability in expected outcomes. *Decline* is an indicator variable that takes the value of 1 for decline stage products and 0 otherwise. The third proxy, *DemandUncert,* is based on the measure for demand uncertainty used in prior literature (McConnell and Perez-Quiros 2000; Banker et al. 2014). It is calculated as the standard deviation of the *unpredictable* portion of demand uncertainty, which is derived from the residuals of a first-order autoregression model of log-changes in sales.

Hierarchical Distance

The email message dataset identifies the hierarchical distance of each sender from the CEO of the firm. Thus, the direct reports of the CEO would have a hierarchical distance of -1, their direct reports would have a hierarchical distance of -2, and so forth. Higher hierarchical distance indicates the employee is at a lower level of the organization. To evaluate the cross-sectional variation in the predictive ability of email sentiment based on the hierarchical distance of the sender, I aggregate email sentiment based on brand-month-hierarchical distance and construct *SentOrg-2*, *SentOrg-3*, *SentOrg-4*, and *SentOrg-5*. In this organization, Level -2 represents director-level employees, Level -3 represents middle managers, and Levels -4 and -5 represent rank-and-file employees.

Control Variables

The research design includes forecast information as a comprehensive summary of all formal and informal information aggregated in the S&OP process to control for traditional sources of information. I avoid inclusion of additional controls that may be related to information that employees know and share but is not incorporated into the forecast as these covariates could explain away the very effect I am investigating. As such, I include a parsimonious set of controls to account for general employee sentiment unrelated to products as well as time trends.

To distinguish product-specific sentiment from general employee sentiment, I construct *GenSentiment* which considers a sender's general, non-product related sentiment. For each month, I aggregate the email sentiment from the naïve Bayes algorithm for all messages that are not identified with a brand tag to capture employees' general sentiment in a given month. The measure is matched to 3-month-ahead unit sales. I also construct the corresponding general sentiment measures for the dictionary-based methods and create *GenSentBlob*, *GenSentH4*, and *GenSentLM* in the same manner as *GenSentiment*.

I include two controls for time trends. *Q4* is an indicator variable that takes the value of 1 if a forecast is completed for a month in the fourth quarter and 0 otherwise. *Q4* controls for seasonality that is common in the industry as consumers increase utilization of medical procedures based on insurance benefits concluding at the end of the calendar year. To control for broad macroeconomic conditions in the industry, I include the monthly producer price index, *PPI*, for the surgical and medical instrument manufacturing industry at the time of the forecast decision.

IV. RESULTS

Descriptive Statistics

Table 1 provides descriptive statistics on a variety of email characteristics for messages sent between July 2021 and March 2022. In the sample, the average user sends out 282 emails per month with a mean of 34 words per message. Senders only mark 1% of messages as "High" importance and 58% of messages are tagged with a brand identifier. Nearly three-quarters of the messages are disseminated (e.g., replies or forwards) as opposed to original emails. The mean email sentiment based on the naïve Bayes algorithm is 0.58, indicating a slightly positive set of messages. The general sentiment, which only considers messages without a brand identifier, has a mean of 0.59, indicating that general messages unrelated to products are more positive than product-specific messages.

[Table 1]

Table 2 reports user characteristics by hierarchical level and function. The hierarchical levels range from -2 to -6 while the functions cover users in Accounting, Operations, and Sales. Panel A reports the total message counts. The greatest proportion of emails come from Level -5 of the hierarchy (40%) and from the Sales function (51%). Panel B documents the number of unique senders by level and function for the 220 unique users directly involved in the S&OP process. Panel C compares the mean *Sentiment* and *GenSentiment* across groups. Univariate comparisons show that product-specific messages from the Sales function are significantly more positive than those from the Accounting function or the Operations function with messages from the Operations function having the least positive product-specific sentiment. In terms of general sentiment, messages from the Accounting function have the most positive sentiment while messages from the Operations function have the least positive general sentiment.

[Table 2]

Table 3 presents the descriptive statistics of the key variables for the multivariate analysis. On average, there are 17.59 actual unit sales per part number-month while 19.52 unit sales are forecasted, indicating that forecasts are positively biased by approximately 11%. Median sales are much lower at 3 units per part number-month for both actual and forecast values, showing that the distributions are positively skewed. In the sample, 26% of the observations come from Growth products while 0.4% come from products in the Decline product life cycle.

[Table 3]

Table 4 presents the pairwise correlations between the key analysis variables. As expected, there is a significantly positive correlation between forecast and actual unit sales. There is also a significantly positive correlation between actual sales and product-specific sentiment measures. Given the nested structure of the data, I now turn to the multilevel analysis for evaluation.[6]

[Table 4]

Aggregate Email Sentiment and Future Firm Performance (H1)

The first hypothesis predicts that aggregate email sentiment will have incremental predictive ability for future firm performance over traditional information sources. Table 5 presents the estimation results of the test of this hypothesis. H1 predicts a significantly positive coefficient on *Sentiment*. Due to the positive skewness observed in the descriptive statistics, both actual and forecast sales are log-transformed in the empirical tests. In Model (1), as expected, the coefficient on *FcstSales*, the proxy for traditional information sources, is significantly positive (β=0.812, p=0.000). Turning to Model (2), the coefficient on *Sentiment* is significantly positive (β=0.570, p=0.015), even after controlling for *FcstSales*, providing support for H1. Model (3) includes additional controls and the coefficient on *Sentiment* remains significantly positive (β=0.811, p=0.019).

[Table 5]

A one standard deviation increase in aggregate sentiment predicts a 7.5% increase in unit sales (i.e., $0.06*(e^{0.811} - 1) = 0.075$), demonstrating an economically significant predictive value for aggregate email sentiment. This product-specific sentiment represents information that was communicated by employees but ultimately was not incorporated into the formal forecast,

[6] To help rule out multi-collinearity concerns, I run a single level model of the reduced form of Equation 1 with the addition of cross-level interactions and compute variance inflation factors. In untabulated results, I find that the maximum VIF among the regressors is 3.34 with an average VIF of 1.79 for the model.

representing a missed opportunity to improve forecasting and planning with information that was available and known among employees. These results provide support for H1, where aggregate email sentiment has predictive ability for future sales, incremental to traditional information sources.

The Role of Information Uncertainty (H2)

Turning to the second hypothesis, I expect the predictive ability of aggregate sentiment to be stronger under conditions of greater information uncertainty. H2 predicts a significantly positive coefficient on the interaction term between sentiment and information uncertainty. Table 6 reports the results of the tests for H2 with Panel A presenting the estimation results for Equation (2). In Model (1), the coefficient of interest is the interaction term *Sentiment * Growth*. The results show the interaction is significantly positive (β=1.742, p=0.004), in line with the prediction. In Model (2), the coefficient of interest is the interaction term *Sentiment * Decline*. The results again show the interaction is significantly positive (β=3.715, p=0.000).

In Model (3), I include both interaction terms for *Growth* and *Decline* so that the values of zero reflect the more stable products in the mature and retiring phases. The coefficients on both interaction terms continue to be significantly positive at the one percent level. For growth products, a one standard deviation increase in aggregate sentiment predicts a 21.4% increase in unit sales (i.e., $0.06*(e^{(1.758-0.238)}-1) = 0.214$). The results suggest that the soft information that employees collect for products with limited historical data has significant predictive ability for unit sales in support of H2. For products in the decline phase of their life cycle, a one standard deviation increase in aggregate sentiment predicts an astonishing 317% increase in unit sales. While products in the decline phase may have extensive historical data, there is significant uncertainty about their future prospects. Thus, the hard data is likely less relevant for predicting future sales. In contrast, the results suggest that the soft information that employees collect based on extensive experience with the product and its customers is especially informative for these types of products with high variability in expected outcomes.

Model (4) examines the predictive role of sentiment under increasing demand uncertainty. In this case, the coefficient on the interaction term *Sentiment* * *DemandUncert* is negative and insignificant. The main effect of *Sentiment* remain significantly positive (β=0.520, p=0.006), however, it does not vary with increasing demand uncertainty.

[Table 6]

The theory underlying H2 asserts that hard information is less available, timely, relevant, and reliable under conditions of increasing uncertainty, creating an opportunity for soft information to be especially predictive. Traditional information sources for planning activities rely on hard information to generate forecasts. As such, I would expect these sources of information to have less predictive ability for future performance when uncertainty is higher.

To corroborate H2 and the underlying theory, I examine if the predictive ability of these traditional information sources varies with information uncertainty. I add interaction terms between the summary measure of traditional information sources, *ln(FcstSales)*, and each of the proxies for information uncertainty. I expect the coefficients on the interaction terms to be negative to provide additional support for H2. Table 6, Panel B reports the estimation results. In Model (1), the coefficient on the interaction term *ln(FcstSales)* * *Growth* is negative but insignificant. In Model (2), the coefficient on the interaction term *ln(FcstSales)* * *Decline* is significantly negative (β=-1.565, p=0.000). In Model (3), I combine the *Growth* and *Decline* interactions in one model and find similar results. The interaction of *ln(FcstSales)* * *Growth* is negative but insignificant while the interaction of *ln(FcstSales)* * *Decline* is significantly negative (β=-1.583, p=0.000), suggesting that traditional, hard sources of information are less predictive of future performance when uncertainty is higher. Moreover, the coefficient on the interaction term *Sentiment * Decline* remains significantly positive (β=1.703, p=0.013). Together, these results suggest that the soft information from employee emails can offset, at least partially, the loss in predictive ability of traditional, hard information sources when uncertainty is greater. In Model (4), the coefficient on the interaction term *ln(FcstSales) * DemandUncert* is also significantly negative (β=-0.667, p=0.000). Overall, I find consistent support for H2 with aggregate email sentiment having greater predictive ability for products in the growth and decline phases while finding that traditional, hard information has less predictive ability when uncertainty is greater.

The Role of Employee Hierarchical Distance (H3)

For the third hypothesis, I expect that the predictive ability of aggregate email sentiment is stronger for employees lower in the organizational hierarchy.[7] Table 7 reports the estimation results of the hypothesis tests with each of the four models using a different sub-sample of sentiment aggregated on senders' hierarchical distance.[8] H3 predicts that the coefficients on sentiment increase with hierarchical distance. While this pattern is generally seen in the table, the only statistically significant coefficient is found in Model (3) for *SentOrg-4* (β=0.302, p=0.091).

I explicitly compare the coefficients between models using the z-statistic derived by Clogg, Petkova, and Haritou (1995). The results are shown at the bottom of the panel. While there is no significant difference in coefficients between Level -2 and Level -3, the coefficient on Level -2 is

[7] The organization has a sixth level of employees within the Operations function only. Because the underlying theory for H3 is not function-specific, I do not include this level in the hypothesis tests. Previous results are robust to exclusion of email messages from Level -6.

[8] Due to multi-collinearity between the partitioned sentiment variables, I refrain from including all variables in one model. The correlations are strongest between Levels -3, -4, and -5, ranging from ρ=0.26 to ρ=0.52 (p=0.000). Given that nearly three-quarters of the emails in the sample are replies or forwarded messages, it is probable that a large portion of emails are responses between different hierarchy levels. In these cases, the email sentiment from one level likely influences the response sentiment of the other level, driving correlation between the hierarchy levels.

significantly smaller than Level -4 (z=-1.882, p=0.030). Further, Level -3 is significantly smaller than Level -4 (z=-1.459 p=0.072) while there is no statistical difference between Level -4 and Level -5. These findings suggest that the predictive ability of aggregate email sentiment for unit sales is concentrated in the communications of rank-and-file employees (Level -4) rather than with directors and middle managers (Levels -2 and -3), consistent with H3.

[Table 7]

Robustness

To ensure my results are robust to alternative measures of sentiment, I also consider proxies based on three common dictionary-based measures. Table 8, Panel A presents the descriptive statistics for these alternative measures. Not surprisingly, because each measure uses a different word list to identify positive and negative words, there are differences in the distributions. The Loughran-McDonald measures (*SentLM* and *GenSentLM)*, which are based on a word list designed for financial text, show the least variation in aggregate email sentiment.

[Table 8]

Panel B reports the estimation results of the regressions of *ActualSales* on the three dictionary-based proxies. Aggregate email sentiment is a significant predictor of unit sales in the case of *SentBlob* (β=0.892, p=0.067) and *SentH4* (β=1.779, p=0.007). The coefficient for *SentLM* is positive but statistically insignificant (β=3.575, p=0.195). Together, these results show that the findings in this study are not sensitive to the measurement of aggregate email sentiment.

As previously discussed, sales forecasts are positively biased in the sample. To address concerns of spurious correlation related to the inclusion of biased explanatory variables, I re-estimate the models using signed forecast error as the dependent variable. Table 9 presents the main results. Model (1) includes the lagged forecast error from the previous three months to control for positive autocorrelation in forecast errors. As expected, all three coefficients for lagged forecast error are significantly positive. In Models (2) and (3), increasing positive sentiment is associated with lower forecast error. Because forecasts are positively biased, this negative relation suggests employee emails contain predictive information that aligns actual sales with forecasted values and offers consistent support for H1. In untabulated results, I replace the dependent variable in previous models for H2 and H3 with forecast error and conclusions remain unchanged.

[Table 9]

V. CONCLUSION

Using a novel dataset, this study provides the first empirical evidence on the value of unstructured, internal data for identifying inefficiencies in information aggregation for forecasting and planning. I document that aggregate employee sentiment from email communications, a ubiquitous form of organizational dark data, predicts future sales incremental to traditional information sources. The predictive ability of aggregate sentiment for sales is stronger for new and aging products and for emails sent from lower levels in the organization.

A common concern with archival field studies is external validity of the findings beyond the research site. While this is an inherent limitation of this study, the methodology is the best suited to answer the research question on the value of dark data, a question that is of interest to academics and practitioners alike. Moreover, the theory underlying my predictions, rooted in information economics, is generalizable to other settings. Nonetheless, additional field-based research to understand how different contextual features such as predictive analytics sophistication, incentives, and manager characteristics is needed to shed light on the generalizability of the findings in this study.

Future research opportunities include investigating how companies can effectively incorporate dark data into the forecasting and planning process. There are a variety of reasons why information aggregation within organizations might be inefficient. First, employees may intentionally withhold information with S&OP decision makers, either because they think it is unimportant, or they think it will be ignored or dismissed. Second, even if employees share potentially valuable soft information, S&OP decision makers may, in fact, ignore it without realizing its informational value. Lastly, S&OP decision makers may incorporate soft information into the planning process, but do so in an inefficient manner. While my study is unable to distinguish between these possibilities, I nonetheless provide a compelling proof of concept for other organizations that suggests research examining the reasons for inefficient information aggregation might reveal opportunities for improvements in S&OP planning processes through the (re-)design of management control systems.

Moreover, there are opportunities to explore a variety of textual attributes beyond sentiment, such as specificity, complexity, and ambiguity, to further our understanding of how dark data can be used for data-driven decision-making. Finally, future research can explore how dark data can inform other decision processes outside of forecasting and planning related to performance evaluation, employee turnover and retention, and risk management. Altogether, there are many exciting research opportunities to explore with important practical implications for firms seeking to capitalize on their vast data assets.

REFERENCES

Aghion, P., and J. Tirole. 1997. Formal and Real Authority in Organizations. *Jourgonal of Political Economy* 105 (1): 1–29.

Banker, R. D., D. Byzalov, and J. M. Plehn-Dujowich. 2014. Demand Uncertainty and Cost Behavior. *The Accounting Review* 89 (3): 839–865.

Bartov, E., L. Faurel, and P. S. Mohanram. 2018. Can Twitter Help Predict Firm-Level Earnings and Stock Returns? *The Accounting Review* 93 (3): 25–57.

Becdach, C., B. Brown, F. Halbardier, B. Henstorf, and R. Murphy. 2020. Rapidly forecasting demand and adapting commercial plans in a pandemic. *McKinsey & Company*.

Bolton, P., and M. Dewatripont. 1994. The Firm as a Communication Network. *The Quarterly Journal of Economics* 109 (4): 809–839.

Brüggen, A., I. Grabner, and K. L. Sedatole. 2021. The Folly of Forecasting: The Effects of a Disaggregated Demand Forecasting System on Forecast Error, Forecast Positive Bias, and Inventory Levels. *The Accounting Review* 96 (2): 127–152.

Campbell, J. L., M. D. DeAngelis, and J. R. Moon. 2019. Skin in the game: personal stock holdings and investors' response to stock analysis on social media. *Review of Accounting Studies* 24 (3): 731–779.

Chen, H., P. De, Y. (Jeffrey) Hu, and B.-H. Hwang. 2014. Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. *Review of Financial Studies* 27 (5): 1367–1403.

Choi, J., E. Forker, I. Grabner, and K. L. Sedatole. 2022. Human in the loop: An examination of the value of human judgment over predictive analytics in anticipating propensity to sell for product assortment planning. *Working Paper*.

Clogg, C. C., E. Petkova, and A. Haritou. 1995. Statistical Methods for Comparing Regression Coefficients Between Models. *American Journal of Sociology* 100 (5): 1261–1293.

Das, S. R., A. Martinez-Jerez, and P. Tufano. 2005. eInformation: A Clinical Study of Investor Discussion and Sentiment. *Financial Management* 34 (3): 103-137.

Das, S. R., and M. Y. Chen. 2007. Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science* 53 (9): 1375–1388.

Davis, A. K., J. M. Piger, and L. M. Sedor. 2012. Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language*: Content of Earnings Press Release Language. *Contemporary Accounting Research* 29 (3): 845–868.

Dawes, R. M., D. Faust, and P. E. Meehl. 1989. Clinical versus actuarial judgment. *Science* 31: 1668–1674.

Devlin, J., M. W. Chang, K. Lee, and K. Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Donovan, J., J. Jennings, K. Koharki, and J. Lee. 2021. Measuring credit risk using qualitative disclosure. *Review of Accounting Studies* 26 (2): 815–863.

Droge, C., J. Jayaram, and S. K. Vickery. 2004. The effects of internal versus external integration practices on time-based performance and overall firm performance. *Journal of Operations Management* 22 (6): 557–573.

Fragale, A. R., J. J. Sumanth, L. Z. Tiedens, and G. B. Northcraft. 2012. Appeasing Equals: Lateral Deference in Organizational Communication. *Administrative Science Quarterly* 57 (3): 373–406.

Frame, W. S., A. Srinivasan, and L. Woosley. 2001. The Effect of Credit Scoring on Small-Business Lending. *Journal of Money, Credit and Banking* 33 (3): 813.

Francis, J., J. Douglas Hanna, and D. R. Philbrick. 1997. Management communications with securities analysts. *Journal of Accounting and Economics* 24 (3): 363–394.

Galton, F. 1907. Vox Populi (The Wisdom of Crowds). *Nature* 75 (1949): 450–451.

Garcia, D. 2013. Sentiment During Recessions. *The Journal of Finance* 68 (3): 1267–1300.

Gardner, D., and P. E. Tetlock. 2015. *Superforecasting: The Art and Science of Prediction*. Crown Publishers/Random House.

Garicano, L. 2000. Hierarchies and the Organization of Knowledge in Production. *Journal of Political Economy* 108 (5): 874–904.
Gartner. 2022. Dark Data.
Gloor, P. A., A. Fronzetti Colladon, F. Grippa, and G. Giacomelli. 2017. Forecasting managerial turnover through e-mail based social network analysis. *Computers in Human Behavior* 71: 343–352.
Hales, J., J. R. Moon, and L. A. Swenson. 2018. A new era of voluntary disclosure? Empirical evidence on how employee postings on social media relate to future corporate disclosures. *Accounting, Organizations and Society* 68–69: 88–108.
Harbert, T. 2021. Tapping the power of unstructured data. *Sloan Management Review*.
Hemmer, T., and E. Labro. 2008. On the Optimal Relation between the Properties of Managerial and Financial Reporting Systems. *Journal of Accounting Research* 46 (5).
Huang, X., S. H. Teoh, and Y. Zhang. 2014. Tone Management. *The Accounting Review* 89 (3): 1083–1113.
Ittner, C. D., and J. Michels. 2017. Risk-based forecasting and planning and management earnings forecasts. *Review of Accounting Studies* 22 (3): 1005–1047.
Jegadeesh, N., and D. Wu. 2013. Word power: A new approach for content analysis. *Journal of Financial Economics* 110 (3): 712–729.
Jensen, M. C., and W. H. Meckling. 1992. Specific and General Knowledge and Organizational Structure. In *Contract Economics*, 251–274. Oxford: Blackwell.
Johnson, J. 2020. What is Dark Data? The Basics & The Challenges. *BMC Blogs*.
Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
Khatri, N., and H. A. Ng. 2000. The Role of Intuition in Strategic Decision Making. *Human Relations* 53 (1): 57–86.
Kreft, I., and J. De Leeuw. 1998. *Introducing Multilevel Modeling*. London: Sage Publications.
Kroos, P., M. Schabus, and F. Verbeeten. 2021. The Relation between Internal Forecasting Sophistication and Accounting Misreporting. *Journal of Management Accounting Research* forthcoming.
Labro, E., M. Lang, and J. D. Omartian. 2023. Predictive Analytics and Centralization of Authority. *Journal of Accounting and Economics* 75 (1): Forthcoming.
Lapide, L. 2004. Sales and Operations Planning Part I: The Process. *Journal of Business Forecasting*: 17–19.
Larcker, D. F., and A. A. Zakolyukina. 2012. Detecting Deceptive Discussions in Conference Calls. *Journal of Accounting Research* 50 (2): 495–540.
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.
Li, H., and W. Suen. 2004. Delegating Decisions to Experts. *Journal of Political Economy* 112 (S1): S311–S335.
Liberti, J. M., and A. R. Mian. 2009. Estimating the Effect of Hierarchies on Information Use. *Review of Financial Studies* 22 (10): 4057–4090.
Liberti, J. M., and M. A. Petersen. 2019. Information: Hard and Soft. *The Review of Corporate Finance Studies* 8 (1): 1–41.
Loughran, T., and B. McDonald. 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance* 66 (1): 35–65.
Marr, B. 2019. What is Unstructured Data and Why Is It So Important to Businesses? An Easy Explanation for Anyone. *Forbes*, October 16.
Mazmanian, M., W. J. Orlikowski, and J. Yates. 2006. Ubiquitous email: individual experiences and organizational consequences of blackberry use. *Academy of Management Annual Meeting Proceedings*: D1–D6.
McConnell, M. M., and G. Perez-Quiros. 2000. Output Fluctuations in the United States: What Has Changed Since the Early 198O's? *The American Economic Review* 90 (5): 1464–1476.
Mintzberg, H. 1994. *The Rise and Fall of Strategic Planning*. New York: The Free Press.

O'Leary, D., and V. C. Storey. 2020. Discovering and Transforming Exhaust Data to Realize Managerial Value. *Communications of the Association for Information Systems* 47: 315–337.

Oliva, R., and N. Watson. 2011. Cross-functional alignment in supply chain planning: A case study of sales and operations planning. *Journal of Operations Management* 29 (5): 434–448.

Prendergast, C. 1993. A Theory of "Yes Men." *The American Economic Review* 83 (3): 757–770.

Radner, R. 1993. The Organization of Decentralized Information Processing. *Econometrica* 61 (5): 1109–1146.

Raudenbush, S. W., and A. S. Byrk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Thousand Oaks, California: Sage Publications.

Rights, J. D., and S. K. Sterba. 2019. New Recommendations on the Use of R-Squared Differences in Multilevel Model Comparisons. *Multivariate Behavioral Research* 55 (4): 568-599.

Sedor, L. M. 2002. An Explanation for Unintentional Optimism in Analysts' Earnings Forecasts. *The Accounting Review* 77 (4): 731–753.

Splunk. 2019. The State of Dark Data.

Stein, J. C. 2002. Information Production and Capital Allocation: Decentralized versus Hierarchical Firms. *The Journal of Finance* 57 (5): 1891–1921.

Surowiecki, J. 2004. *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*. New York, NY: Anchor Books.

Tang, V. W. 2018. Wisdom of Crowds: Cross-Sectional Variation in the Informativeness of Third-Party-Generated Product Information on Twitter. *Journal of Accounting Research* 56 (3): 989–1034.

Taulli, T. 2019. What You Need to Know About Dark Data. *Forbes*.

Vickery, S. K., J. Jayaram, C. Droge, and R. Calantone. 2003. The effects of an integrative supply chain strategy on customer service and financial performance: an analysis of direct versus indirect relationships. *Journal of Operations Management* 21 (5): 523–539.

Wasiak, J., B. Hicks, L. Newnes, C. Loftus, A. Dong, and L. Burrow. 2011. Managing by E-Mail: What E-mail Can Do for Engineering Project Management. *IEEE Transactions on Engineering Management* 58 (3): 445–456.

Watson, R. B. 1987. The Effects of Demand-Forecast Fluctuations on Customer Service and Inventory Cost When Demand is Lumpy. *The Journal of the Operational Research Society* 38 (1): 75–82.

Wen, Q., P. A. Gloor, A. Fronzetti Colladon, P. Tickoo, and T. Joshi. 2020. Finding top performers through email patterns analysis. *Journal of Information Science* 46 (4): 508–527.

Whyte, G. 1989. Groupthink Reconsidered. *Academy of Management Review* 14 (1): 40–56.

Zhang, X. F. 2006. Information Uncertainty and Stock Returns. *The Journal of Finance* 61 (1): 105–137.

Figure 1. Sales and Operations Planning (S&OP) Process

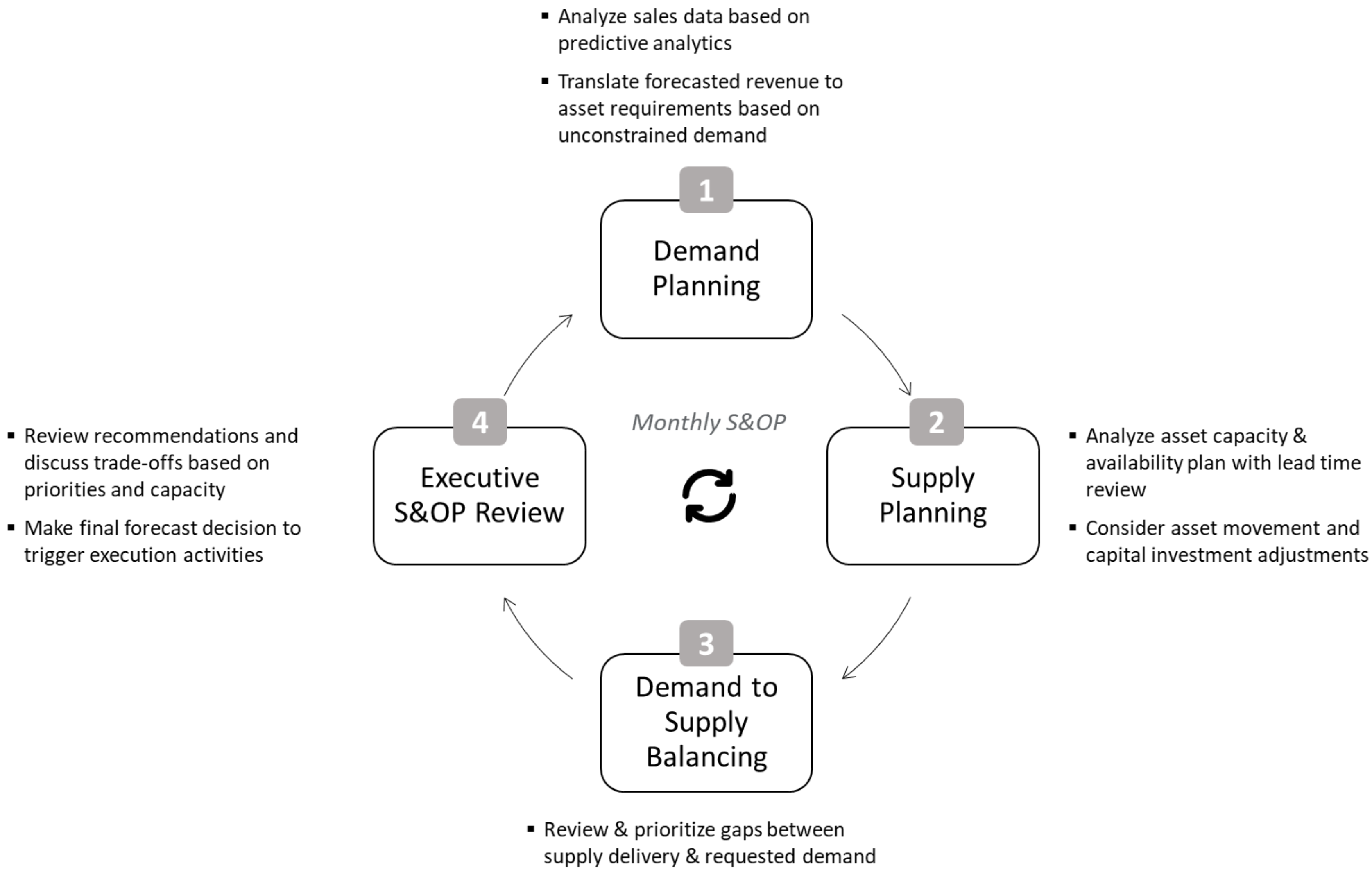

Figure 1 documents the broad process steps in the monthly S&OP process adapted from company documents.

Figure 2. Nested Data Structure

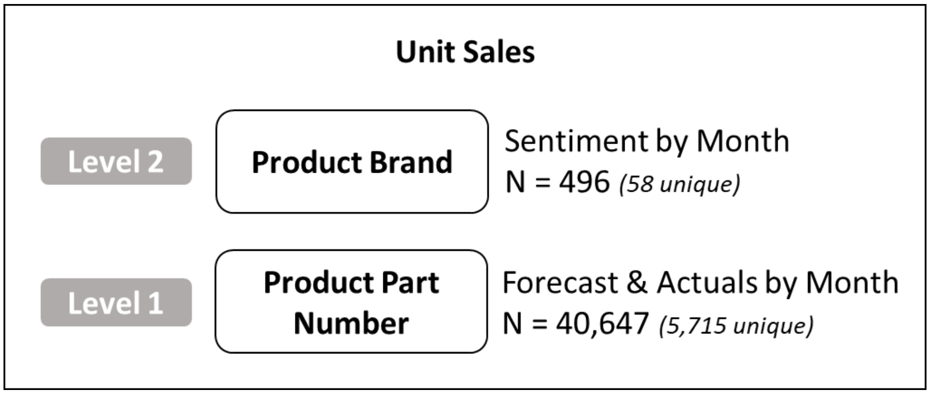

Figure 2 presents the multi-level, nested data structure. The unit sales data is characterized by a two-level structure where monthly forecast and actual unit sales are collected at the product part number level while monthly aggregate email sentiment is collected at the brand level.

Figure 3. S&OP Timeline

Figure 3 presents the timeline for the rolling, monthly S&OP process. Actual unit sales are collected at month *m* while forecast values for month *m* are collected at month *m-3*. Aggregate email sentiment is collected between monthly forecast decisions for the entire month *m-3*.

Table 1. Email Characteristics

Attribute	N	Mean	SD	Min	P25	P50	P75	Max
Emails per month	219,198	27,594	3,607	616	26,129	26,636	28,759	34,723
Emails per sender-month	219,198	282	208	1	118	229	399	985
No. of words per email	219,198	34	80	0	8	15	29	2,368
Email size (KB)	219,198	335	1,610	7	38	60	128	37,000
High Importance	219,198	0.01	0.07	0	0	0	0	1
No. of To Recipients	219,198	1.65	1.33	0	1	1	2	25
No. of CC Recipients	219,198	0.75	1.30	0	0	0	1	24
No. of BCC Recipients	219,198	0.01	0.24	0	0	0	0	25
Brand Tag	219,198	0.58	0.49	0	0	1	1	1
Disseminated Message	219,198	0.74	0.44	0	0	1	1	1
Sentiment	219,198	0.58	0.21	0	0.44	0.58	0.72	1
GenSentiment	91,851	0.59	0.21	0	0.45	0.59	0.73	1

Table 1 reports descriptive statistics for several email attributes based on email messages sent between July 2021 and March 2022. The sample size for *GenSentiment* only considers messages without a brand tag.

Table 2. User Characteristics by Hierarchical Level and Function

Panel A. Message Counts

		Function			
Level	**Acct**	**Ops**	**Sales**	**Total**	**% Total**
-2	3,125	1,591	5,749	10,465	5%
-3	8,613	15,124	13,233	36,970	17%
-4	3,426	12,323	49,124	64,873	30%
-5		43,339	43,459	86,798	40%
-6		20,092		20,092	9%
Total	15,164	92,469	111,565	219,198	
% Total	7%	42%	51%		

Panel B. Unique Senders

		Function			
Level	**Acct**	**Ops**	**Sales**	**Total**	**% Total**
-2	2	1	2	5	2%
-3	8	4	9	21	10%
-4	7	7	54	68	31%
-5		32	79	111	50%
-6		15		15	7%
Total	17	59	144	220	
% Total	8%	27%	65%		

(Table 2 continued)

Panel C. Mean *Sentiment* & *GenSentiment*

Level	Function: Acct	Ops	Sales	Overall Mean
-2	0.61	0.47	0.60	0.59
	0.62	0.47	0.61	0.59
-3	0.59	0.57	0.60	0.58
	0.60	0.58	0.59	0.59
-4	0.58	0.58	0.60	0.60
	0.58	0.59	0.60	0.60
-5		0.57	0.58	0.58
		0.58	0.58	0.58
-6		0.52		0.52
		0.56		0.56
Overall Mean	0.59	0.56	0.59	
	0.60	0.58	0.59	

Comparison between Functions

	Difference	p-value
Sales - Ops	0.037	0.000
	0.016	0.000
Acct - Ops	0.032	0.000
	0.022	0.000
Sales - Acct	0.005	0.003
	-0.005	0.031

Table 2 reports email sender characteristics by hierarchical level and function. Panel A shows the email message counts along each dimension. Panel B displays the number of unique senders by level and function. Panel C provides the mean *Sentiment* (top value) and mean *GenSentiment* (bottom value) in each group, with a comparison of functions based on the t-test comparing the difference in means between each pair.

Table 3. Descriptive Statistics for Key Analysis Variables

Variable	N	Mean	SD	Min	P25	P50	P75	Max
ActualSales	40,647	17.59	48.95	0	0	3	11	352
FcstSales	40,647	19.52	52.14	0	1	3	12	373
Sentiment	496	0.54	0.06	0.13	0.52	0.55	0.58	0.70
GenSentiment	9	0.58	0.02	0.54	0.58	0.59	0.59	0.59
Growth	40,647	0.26	0.44	0	0	0	1	1
Decline	40,647	0.004	0.06	0	0	0	0	1
DemandUncert	38,934	0.61	0.33	0	0.37	0.58	0.82	2.22
Q4	40,647	0.32	0.47	0	0	0	1	1
PPI	40,647	144.19	0.69	143.50	143.71	143.85	144.85	145.29

Table 3 reports the descriptive statistics for the key analysis variables. Actual unit sales are collected from October 2021 through June 2022 and the corresponding forecast values are taken from July 2021 through March 2022. *Sentiment* is measured at the brand-month level while *GenSentiment* is measured by month with repeated observations at the part number level based on the nested data structure.

Table 4. Correlation Matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) *ActualSales*	1							
(2) *FcstSales*	**0.783**	1						
(3) *Sentiment*	**0.038**	**0.054**	1					
(4) *GenSentiment*	0.004	0.000	**0.524**	1				
(5) *Growth*	**0.053**	**0.137**	**0.027**	-0.002	1			
(6) *Decline*	**-0.037**	**-0.059**	**0.019**	**0.010**	**-0.037**	1		
(7) *DemandUncert*	**-0.277**	**-0.286**	**0.058**	**0.012**	-0.008	0.007	1	
(8) *Q4*	0.001	0.001	**-0.186**	**-0.492**	-0.002	**-0.010**	-0.007	1
(9) *PPI*	**0.010**	-0.008	**0.055**	**0.237**	0.000	0.000	0.006	**-0.553**

Bold denotes significance at the 0.05 level.

Table 5. Aggregate Email Sentiment and Future Firm Performance

		DV = ***ln(ActualSales)***		
Variable		(1)	(2)	(3)
Sentiment	H1 (+)		0.570**	0.811**
			(0.234)	(0.347)
ln(FcstSales)		0.812***	0.812***	0.812***
		(0.034)	(0.034)	(0.034)
GenSentiment				-0.622
				(0.677)
Q4				0.042***
				(0.016)
PPI				0.053***
				(0.013)
Constant		0.264***	-0.047	-7.436***
		(0.090)	(0.109)	(1.900)
Observations		40,647	40,647	40,647
R-Squared		0.652	0.653	0.654
Number of clusters		58	58	58

Table 5 examines the predictive ability of aggregate email sentiment for future firm performance, reporting the estimation results from maximum likelihood regressions of *ln(ActualSales)* on S*entiment* based on the two-level specification from Equation (1). Model (1) serves as a base model including conventional information sources with *ln(FcstSales).* Model (2) adds *Sentiment* while Model (3) includes additional controls. Robust standard errors in parentheses are clustered by brand. ***, **, and * denote significance at the 1 percent, 5 percent, and 10 percent level (two-tailed), respectively. Rights and Sterba (2019) R^2 values are shown. See Appendix A for variable definitions.

Table 6. Role of Information Uncertainty

Panel A. Cross-Sectional Variation in the Predictive ability of Aggregate Email Sentiment

		DV = *ln(ActualSales)*			
Variable		(1)	(2)	(3)	(4)
Sentiment		0.490	0.785**	0.459	0.520***
		(0.413)	(0.347)	(0.416)	(0.190)
*Sentiment * Growth*	H2 (+)	1.742***		1.758***	
		(0.607)		(0.607)	
*Sentiment * Decline*	H2 (+)		3.715***	4.039***	
			(0.889)	(0.939)	
*Sentiment * DemandUncert*	H2 (+)				-0.871
					(1.227)
Growth		-0.238***		-0.238***	
		(0.091)		(0.091)	
Decline			-0.041	-0.053	
			(0.038)	(0.036)	
DemandUncert					-0.226***
					(0.078)
ln(FcstSales)		0.816***	0.812***	0.816***	0.810***
		(0.034)	(0.034)	(0.034)	(0.032)
GenSentiment		-0.846	-0.583	-0.806	-0.521
		(0.721)	(0.673)	(0.714)	(0.460)
Q4		0.040**	0.042***	0.040**	0.048***
		(0.016)	(0.016)	(0.016)	(0.013)
PPI		0.054***	0.053***	0.054***	0.050***
		(0.013)	(0.013)	(0.013)	(0.012)
Constant		-6.972***	-7.000***	-6.979***	-6.708***
		(1.915)	(1.939)	(1.916)	(1.772)
Observations		40,647	40,647	40,647	38,934
R-Squared		0.653	0.653	0.653	0.674
Number of clusters		58	58	58	57

(Table 6 Continued)

Panel B. Cross-Sectional Variation in the Predictive ability of Traditional Information Sources

	DV = *ln(ActualSales)*			
Variable	(1)	(2)	(3)	(4)
Sentiment	0.485	0.799**	0.466	0.487***
	(0.419)	(0.345)	(0.420)	(0.181)
*Sentiment * Growth*	1.750***		1.770***	
	(0.616)		(0.617)	
*Sentiment * Decline*		1.361**	1.703**	
		(0.650)	(0.686)	
*Sentiment * DemandUncert*				-0.244
				(1.197)
*ln(FcstSales) * Growth*	-0.050		-0.051	
	(0.081)		(0.081)	
*ln(FcstSales) * Decline*		-1.565***	-1.583***	
		(0.107)	(0.109)	
*ln(FcstSales) * DemandUncert*				-0.667***
				(0.039)
Growth	-0.225**		-0.225**	
	(0.088)		(0.088)	
Decline		-2.078***	-2.102***	
		(0.150)	(0.150)	
DemandUncert				-0.273***
				(0.086)
ln(FcstSales)	0.830***	0.813***	0.831***	0.696***
	(0.043)	(0.034)	(0.042)	(0.033)
GenSentiment	-0.841	-0.636	-0.854	-0.502
	(0.720)	(0.659)	(0.699)	(0.440)
Q4	0.040**	0.041***	0.039**	0.044***
	(0.016)	(0.016)	(0.017)	(0.012)
PPI	0.054***	0.052***	0.054***	0.045***
	(0.013)	(0.013)	(0.013)	(0.012)
Constant	-5.600***	-5.428***	-5.498***	-4.599***
	(1.910)	(1.922)	(1.901)	(1.785)
Observations	40,647	40,647	40,647	38,934
R-Squared	0.653	0.655	0.655	0.701
Number of groups	58	58	58	57

Table 6 examines cross-sectional variation in information uncertainty and its effect on the predictive ability of aggregate email sentiment for future firm performance. Panel A reports the maximum likelihood estimation results of regressing *ln(ActualSales)* on *Sentiment* based on interaction models using three proxies for information uncertainty. Model (1) includes an interaction term between *Sentiment* and *Growth*. Model (2) includes an interaction term between *Sentiment* and *Decline*. Model(3) includes both interactions with *Growth* and *Decline*. Model (4) includes an interaction term between *Sentiment* and *DemandUncert*. Panel B reports the estimation results of another set of interaction models that consider cross-sectional variation in the predictive ability of traditional information sources. Model (1) adds an interaction term between *ln(FcstSales)* and *Growth.* Model (2) adds an interaction term between *ln(FcstSales)* and *Decline.* Model (3) includes interactions with *Growth* and *Decline*. Model (4) adds an interaction term between *ln(FcstSales)* and *DemandUncert.* Continuous interaction term variables used in estimation are mean-centered for ease of interpretation. Robust standard errors in parentheses are clustered by brand. ***, **, and * denote significance at the 1 percent, 5 percent, and 10 percent level (two-tailed), respectively. See Appendix A for variable definitions.

Table 7. Role of Employee Hierarchical Distance

	DV = ***ln(ActualSales)***			
Variable	(1)	(2)	(3)	(4)
SentOrg-2	-0.075			
	(0.090)			
SentOrg-3		-0.035		
		(0.146)		
SentOrg-4			0.302*	
			(0.179)	
SentOrg-5				0.141
				(0.349)
ln(FcstSales)	0.821***	0.816***	0.812***	0.812***
	(0.033)	(0.034)	(0.034)	(0.034)
GenSentiment	-0.391	0.905	-0.080	0.500
	(0.634)	(0.694)	(0.656)	(0.528)
Q4	0.050***	0.050***	0.043***	0.048***
	(0.013)	(0.015)	(0.015)	(0.014)
PPI	0.054***	0.050***	0.049***	0.050***
	(0.015)	(0.013)	(0.013)	(0.013)
Constant	-7.312***	-7.527***	-6.900***	-7.361***
	(2.135)	(1.888)	(1.838)	(1.931)
Comparison between Levels				
	z-test	**p-value**		
SentOrg-2 - SentOrg-3	-0.233	0.408		
SentOrg-2 - SentOrg-4	-1.882	0.030		
SentOrg-2 - SentOrg-5	-0.599	0.274		
SentOrg-3 - SentOrg-4	-1.459	0.072		
SentOrg-3 - SentOrg-5	-0.465	0.321		
SentOrg-4 - SentOrg-5	0.410	0.659		
Observations	35,638	38,723	40,212	40,596
R-Squared	0.671	0.659	0.653	0.653
Number of clusters	49	56	57	57

Table 7 examines cross-sectional variation in employee hierarchical distance and its effect on the predictive ability of aggregate email sentiment for future firm performance. The table reports the maximum likelihood estimation results of regressing *ln(ActualSales)* on *Sentiment*, where sentiment is aggregated based on a different employee hierarchical cross-section (Org-2, Org-3, Org-4, or Org-5) for each model within the panel. The bottom of the table reports the results of post-estimation tests with which I analyze the equality of coefficients for each hierarchical distance pair. I report the z-test statistic and p-value for each comparison. Robust standard errors in parentheses are clustered by brand. ***, **, and * denote significance at the 1 percent, 5 percent, and 10 percent level (two-tailed), respectively. See Appendix A for variable definitions.

Table 8. Robustness to Alternative Sentiment Measures

Panel A. Descriptive Statistics

Variable	N	Mean	SD	Min	P25	P50	P75	Max
SentBlob	496	0.07	0.04	-0.10	0.05	0.07	0.09	0.38
SentH4	496	0.05	0.02	-0.03	0.04	0.05	0.05	0.22
SentLM	496	0.00	0.01	-0.08	0.00	0.00	0.00	0.03
GenSentBlob	9	-0.01	0.02	-0.04	-0.01	-0.01	0.00	0.01
GenSentH4	9	-0.01	0.01	-0.02	-0.01	-0.01	-0.01	0.01
GenSentLM	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Panel B. Aggregate Email Sentiment and Future Firm Performance

	DV = *ln(ActualSales)*		
Variable	(1)	(2)	(3)
SentBlob	0.892*		
	(0.488)		
SentH4		1.779***	
		(0.663)	
SentLM			3.575
			(2.758)
ln(FcstSales)	0.812***	0.812***	0.812***
	(0.034)	(0.034)	(0.034)
GenSentBlob	-0.334		
	(0.528)		
GenSentH4		-0.198	
		(0.620)	
GenSentLM			-2.151
			(1.710)
Q4	0.052***	0.032**	0.048***
	(0.014)	(0.015)	(0.013)
PPI	0.053***	0.052***	0.052***
	(0.013)	(0.014)	(0.013)
Constant	-7.498***	-7.332***	-7.283***
	(1.892)	(1.971)	(1.941)
Observations	40,647	40,647	40,647
R-Squared	0.653	0.653	0.652
Number of clusters	58	58	58

Table 8 examines the predictive ability of aggregate email sentiment for future firm performance using alternative measures for sentiment. Panel A presents the descriptive statistics, where sentiment is measured at the brand-month level and general sentiment is measured at the month level with repeated observations at the part number level based on the nested data structure. Panel B reports the maximum likelihood estimation results of regressing *ln(ActualSales)* on three different sentiment proxies. Model (1) uses *SentBlob*, Model (2) uses *SentH4*, and Model (3) uses *SentLM* for sentiment. Robust standard errors in parentheses are clustered by brand. ***, **, and * denote significance at the 1 percent, 5 percent, and 10 percent level (two-tailed), respectively. See Appendix A for variable definitions.

Table 9. Aggregate Email Sentiment and Forecast Error

		DV = *FcstError*	
Variable	(1)	(2)	(3)
Sentiment		-1.099*	-1.454**
		(0.666)	(0.726)
FcstErrorLag1	0.257***	0.257***	0.257***
	(0.020)	(0.020)	(0.020)
FcstErrorLag2	0.147***	0.147***	0.146***
	(0.011)	(0.011)	(0.011)
FcstErrorLag3	0.134***	0.134***	0.135***
	(0.008)	(0.008)	(0.008)
GenSentiment			-17.510**
			(7.283)
PPI			0.027*
			(0.016)
Constant	-0.132***	0.467	10.923**
	(0.027)	(0.362)	(4.257)
Observations	13,572	13,572	13,572
R-Squared	0.199	0.201	0.203
Number of clusters	57	57	57

Table 9 examines the predictive ability of aggregate email sentiment for future firm performance using an alternative dependent variable, reporting the maximum likelihood estimation results of regressing *FcstError* on *Sentiment.* Model (1) includes the lagged forecast error from the previous three months. Model (2) adds *Sentiment* and Model (3) includes additional controls. Robust standard errors in parentheses are clustered by brand. ***, **, and * denote significance at the 1 percent, 5 percent, and 10 percent level (two-tailed), respectively. See Appendix A for variable definitions.

Appendix A. Variable Definitions

Variable	Description
Dependent Variable	
ActualSales	Actual unit sales by part number in month *m*. Values are winsorized at the top and bottom one percent.
Independent Variable	
Sentiment	Measure of email sentiment based on naïve Bayes algorithm that assigns a sentiment score ranging from 0 (negative) to 1 (positive) to each email. Individual email sentiment is then aggregated at the brand-month level to derive a product-specific measure. Value is measured at *m-3*.
Control Variables	
FcstSales	Forecasted unit sales by part number for month *m* at time *m-3*. Values are winsorized at the top and bottom one percent.
GenSentiment	Measure of email sentiment based on *Sentiment*. However, individual email sentiment is aggregated at the month level for any emails that are not tagged with a brand to derive a general sentiment measure unrelated to product sentiment. Value is measured at *m-3*.
Q4	Indicator variable that takes the value of 1 if a forecast is completed for a month in the fourth quarter and 0 otherwise.
PPI	Monthly Producer Price Index for the surgical and medical instrument manufacturing industry: measures the average change over time in selling prices received by domestic producers for their output. Value is captured at the time of the forecast decision.
Cross-Sectional Variables	
Growth	Indicator variable that takes the value of 1 for Growth stage products and 0 otherwise. This measure is defined and used by the organization.
Decline	Indicator variable that takes the value of 1 for Decline stage products and 0 otherwise. This measure is defined and used by the organization.
DemandUncert	Measure of the unpredictable portion of demand uncertainty. Calculated as the standard deviation of the residuals of a first-order autoregression of log-changes in unit sales at month *m* on log-changes in unit sales at month *m-1*.
SentOrg-2/3/4/5	Measure of email sentiment based on variable *Sentiment* but aggregated by brand, month, and senders with a hierarchical distance of -2/3/4/5. Hierarchical distance measures the number of reporting levels a sender is removed (i.e., lower) from the CEO of the firm. Value is measured at *m-3*.

(Appendix A continued)

Variable	Description
Alternative Variables	
SentBlob	Measure of email sentiment based on the difference in counts of positive and negative words scaled by the total words in a message, using word lists from the TextBlob Python library. Individual email sentiment is aggregated at the brand-month level. Value is measured at *m-3*.
GenSentBlob	Measure of email sentiment based on *SentBlob*. However, individual email sentiment is aggregated at the month level for any emails that are not tagged with a brand to derive a general sentiment measure unrelated to product sentiment. Value is measured at *m-3*.
SentH4	Measure of email sentiment based on the difference in counts of positive and negative words scaled by the total words in a message, using word lists from the Harvard IV-4 dictionary. Individual email sentiment is aggregated at the brand-month level. Value is measured at *m-3*.
GenSentH4	Measure of email sentiment based on *SentH4*. However, individual email sentiment is aggregated at the month level for any emails that are not tagged with a brand to derive a general sentiment measure unrelated to product sentiment. Value is measured at *m-3*.
SentLM	Measure of email sentiment based on the difference in counts of positive and negative words scaled by the total words in a message, using word lists from the Loughran-McDonald dictionary. Individual email sentiment is aggregated at the brand-month level. Value is measured at *m-3*.
GenSentLM	Measure of email sentiment based on *SentLM*. However, individual email sentiment is aggregated at the month level for any emails that are not tagged with a brand to derive a general sentiment measure unrelated to product sentiment. Value is measured at *m-3*.
SentBERT	Measure of email sentiment based on the Google BERT algorithm and fine-tuned on the SST2 dataset. Sentiment scores range from -1 (negative) to 1 (positive) and are aggregated at the brand-month level. Value is measured at *m-3*.
GenSentBERT	Measure of email sentiment based on *SentBERT*. However, individual email sentiment is aggregated at the month level for any emails that are not tagged with a brand to derive a general sentiment measure unrelated to product sentiment. Value is measured at *m-3*.
FcstError	Signed unit sales forecast error, measured as the actual monthly unit sales by part number less the forecasted monthly unit sales, and scaled by the actual unit sales. Values are winsorized at the top and bottom one percent.
FcstErrorLag	Lagged unit sales forecast error for months *m-1*, *m-2*, and *m-3*.

Appendix B. Example Email Sentiment Classification from Condor Naïve Bayes Algorithm

Negative Sentiment
Sentiment Score = 0.180
Hi Team - in our tax construct, we cannot ship [Product Name] from {} countries to {} due to these countries tax set up. This is why they need to go through {} before transferring to these countries. I checked with {} and finance and unfortunately there isn't another option. Usually we keep these sort of transfers to a minimum because of this. Thanks, {}

Positive Sentiment
Sentiment Score = 0.936
Awesome, thanks for passing along. If there is any support he needs for [Product Name], let's make sure to provide it. Excellent opportunity.

Neutral Sentiment
Sentiment score = 0.521
{}/{}, Attached has the Q122 revenue forecast for the US by product. I spot checked a few and they match [Planning System Name], as expected. I'm passing this along for your reference in case you find the territory tab breakdown useful in your discussions with {}, as it may provide you more insight on each of the specific regions/businesses. Note: this was an outcome from a forecasting discussion for the US with the finance team. I had sat in on all the region finance calls mostly to ensure alignment with our S&OP process. Thanks, {}