A Textual Measure of Strategy

Rajiv Banker
Fox School of Business
Temple University
banker@temple.edu

Yi Liang
Fox School of Business
Temple University
yiliang@temple.edu

Xinjie Ma
Fox School of Business
Temple University
xinjie.ma@temple.edu

Acknowledgement: We thank Ram Mudambi, Paige Patrick (discussant), Charlotte Ren, and workshop participants at 2018 AAA Management Accounting Section Midyear Meeting, University of Calgary and Temple University for helpful comments and suggestions. We thank Fox School of Business and its Young Scholars Award Funding for providing financial support.

A Textual Measure of Strategy

Research summary: To gain competitive advantages, a firm may choose to orient its strategy to differentiating its products and services or reducing its costs. Most prior studies measure strategy using data from executive surveys, proprietary databases, or panelist evaluation. These methods are costly to implement and, consequently, restrict the scope of research. This paper develops a textual measure of firms' strategy using the annual 10-K filings of U.S. public firms. We validate our measure by showing its properties and its associations with operating activities are consistent with those suggested or documented in the prior literature. We also detect that strategy impacts different measures of firm performance differently and that much of the changes in earnings properties over time can be explained by firm strategy. Thus, our approach can be useful in many potential studies in strategic management and other business disciplines.

Keywords: strategy, competitive advantage, text analysis, operating activities, firm performance

Introduction

Prior measures do not provide a rigorous way to compare the intended strategy over time and over firms in different industries, relying primarily proprietary data such as survey responses from executives (e.g., Montemayor, 1996; Slater and Olson, 2001; Vorhies, Morgan, and Autry, 2009; Zatzick, Moliterno, and Fang, 2012). To overcome this challenge, in this paper we present a text analysis approach using readily available data to measure the strategy of firms. Specifically, we leverage the statutory-required documents of U.S. public firms that are filed annually in a prescribed format and develop a strategy index that characterizes firms based on their strategic emphases (i.e., either delivering unique products and services or lowering costs). Our approach is easy to execute and can be applied to large samples across different industries and years. We validate our strategy index by performing a variety of tests on its characteristics and verifying its associations with various variables suggested or documented in the prior literature (e.g., Hambrick, 1983a; Porter, 1985). Moreover, we explore both time-series and cross-sectional variations in firms' strategy and investigate how strategy impacts the short-term accounting-based performance, return on assets (ROA), and long-term market-based performance, stock return. We find that while ROA varies systematically with strategy, the stock return does not. Finally, we examine the changes in the volatility of earnings and the degree of matching between concurrent revenues and expenses in the past two decades and show the changes in earnings properties can be explained by the real economic changes in strategy, especially the differences of strategy among different cohorts of newly listed firms.

Strategic management involves a long-term vision to obtain a sustainable competitive advantage over competitors (Porter, 1996). A coherent strategy guides resource allocation decisions and influences operating activities and performance (Nag, Hambrick, and Chen, 2007;

Porter, 1980). Many prior strategic management studies have examined how firms' strategy explain the differences in firm performance (e.g., Hambrick, 1983a; Nair and Filer, 2003; Rumelt, Schendel, and Teece, 1991; Zott and Amit, 2008). Research in other business disciplines, however, often focuses on specific activities or behavior and examines how strategy impacts, for example, executive reporting structure (Banker *et al.*, 2010), executive compensation (Ittner, Larcker, and Rajan, 1997), financial reporting irregularity (Bentley, Omer, and Sharp, 2013), human resource systems (Arthur, 1992), purchasing management design (David *et al.*, 2002), and work-life balance programs (Wang and Verma, 2012). Given business researchers' wide range of interests, it is essential to develop a reliable and convenient approach to measure strategy to facilitate future academic research.

Prior studies have primarily created measures of strategy with survey responses from executives (e.g., Slater and Olson, 2001), proprietary databases (e.g., Hambrick, 1983a, 1983b; Thornhill and White, 2007), evaluations by panelists (e.g., Dess and Davis, 1984; Kabanoff and Brown, 2008; Zott and Amit, 2008), or accounting numbers (e.g., Berman *et al.*, 1999; David *et al.*, 2002; Kotha and Nair, 1995). While the strategy measures based on survey or proprietary databases may reflect the perceptions of executives who formulate and implement the firm-level strategy, they may also be distorted by executives' own incentives and may be subject to common-method bias and nonresponse bias (Armstrong and Overton, 1977; Podsakoff *et al.*, 2003). Measuring strategy with the help of external panelists is a popular alternative as outside experts can provide a detached and informative rating of relative strategies (Snow and Hambrick, 1980), but such a measure may not necessarily reflect managers' true beliefs due to misinterpretation and human bias caused by panelists' heterogeneous backgrounds. More importantly, the difficulty and cost of obtaining data using these methods limit the scope, power,

and replicability of the empirical studies. For example, many studies focus on a single industry and confine their empirical analyses to firms that responded to the surveys, which may impair the generalizability of the findings. Kabanoff and Brown (2008) utilize a Naïve Bayesian approach of text analysis to classify strategy for a sample of Australian firms. However, to implement the Naïve Bayesian approach, it still requires *ex ante* expert assessment for a training set of firms to help the algorithm learn the classification rules (McCallum and Nigam, 1998). Because the optimal size of the training set usually increases with sample size, this method can still involve a significant amount of expert evaluation and, as a result, excessive costs. Moreover, the impact of any small mistake or human bias in classifying the training set would potentially be magnified when the algorithm is applied to the full sample. In contrast to survey and proprietary databases, accounting numbers are often publicly available and easy to obtain. However, the strategy inferred from such data often suffers from various factors that influence the accounting numbers but are not necessarily related to the strategy such as a recession (Mintzberg, 1978). Finally, it is also worth noting that, to avoid the concern of circular reasoning, empirical studies may not use strategy measures created based on accounting numbers to examine the impact of strategy on financial outcomes such as firm performance.

The goal of this paper is to present a textual strategy index for a large sample of firms, utilizing publicly available management narratives. In particular, we draw on the prior economics and strategic management literature and consider two main strategies of firms, lowering costs and differentiating products and services. On the one hand, as argued in the early industrial organization literature, one way for firms to succeed is to provide similar products and services as their rivals do but reap profits by lowering costs and prices as much as possible (Hotelling, 1929). On the other hand, firms may pursue maximal differentiation by offering

unique products and services in order to soften price competition and attract customers with a high willingness to pay for such uniqueness (D'Aspremont, Gabszewicz, and Thisse, 1979). This classification of strategy is also widely supported by a variety of strategic management theories, all of which share broad similarities regarding how firms can (1) offer differentiated products and services through innovation or (2) reduce costs and improve operational efficiency to achieve superior performance (e.g., Miles and Snow, 1978; Porter, 1985; Treacy and Wiersema, 1995). For simplicity, we borrow the terms "differentiation" and "cost leadership" from Porter (1985) to label the two strategies that we study.

Our textual analysis takes the following three steps to generate the strategy index. First, guided by the prior strategic management literature, we develop keyword lists for differentiation and cost leadership orientations, which contain words and phrases describing how firms position themselves and how they implement their strategic positioning. Second, we parse all the Form 10-K that Securities and Exchange Commission (SEC) requires U.S. public firms to file annually and identify the Item 1 Business section which provides a general description of firm businesses including main products and services (Securities and Exchange Commission, 2011).[i] Finally, we use Python scripts to count the occurrences of each keyword in each Item 1 document and then sum the counts to generate the strategy index for each document.

We validate the strategy index using a sample of 46,032 firm-year observations. First, to ensure that our index makes intuitive sense, we list the top three differentiators and cost leaders of the Fortune 100 firms and verify that our classification is consistent with their corporation images. Second, consistent with "the tendency for strategies to persist over time," we find that our textual measure, *on average*, does not exhibit significant time-series variations (Ghemawat, 1991: 14). Third, we examine the construct validity of the measure by documenting a significant

association with realized operating activities, efficiency, and profit margin. Firms leaning toward differentiation spend more on research and development (R&D) and marketing functions than those leaning toward cost leadership do. They also have lower operation efficiency and higher profit margin. We further establish predictive validity by documenting the incremental predictive power of the strategy index for future operating activities and operating performance after controlling for the concurrent activities and performance.

Moreover, we explore the changes in strategy and show both time-series and cross-sectional variations in the relative importance of the strategies. We further explore the association between firm performance and strategy by contrasting a short-term accounting-based performance measure, ROA, with a long-term market-based performance measure, stock return. We find firms leaning toward the differentiation strategy have lower ROA, which may be partially attributable to the fact that R&D expenses are not capitalized as assets and hence reduce net income under the U.S. GAAP (Lev and Sougiannis, 1996). For stock return which incorporates investors' expectations for future performance, we do not find significant differences among firms with different strategies.

Finally, we examine how strategy explain the changes in earnings properties over time (matching and earnings volatility). Prior literature interprets the decrease in matching and the increase in earnings volatility over past decades as a decline in earnings quality (Collins, Maydew, and Weiss, 1997; Dechow, Ge, and Schrand, 2010; Dichev and Tang, 2008; Donelson, Jennings, and Mclnnis, 2011; Francis and Schipper, 1999; Lev and Zarowin, 1999). We show that the "decline in earnings quality" can be explained by real economic changes in strategy. We find that firms leanings toward the differentiation strategy have a lower degree of matching between revenue and concurrent expenses and higher volatility of earnings. Specifically, we find

that the changes in earnings properties are mainly driven by the differences in strategy among successive cohorts of newly listed firms. We document that the time-series variations of the average strategy of public firms are mainly driven by the differences among different cohorts of newly listed firms rather than within-firm variations. That is, successive cohorts for firms exhibit increasingly more emphasis on differentiation strategy. We also find an incremental explanatory power of strategy on earnings properties after controlling for the cohort effect.

This paper contributes to the literature in three ways. First, we develop a strategy index based on firms' public disclosures across different industries over the years. Because such public disclosures can be easily accessed and analyzed using contemporary information technology, our approach is easier to execute than those introduced in the prior literature. Also, our strategy index can be directly applied to future research.[ii] Second, our textual analysis approach involves little human intervention and thus largely reduces perceptual and interpretive bias. As noted, private survey responses may not reflect managers' true beliefs and panelists can be biased in classifying firms due to heterogeneous personal experiences (Snow and Hambrick, 1980). Our textual approach, to a large extent, avoids these human processing procedures and compliments these strategy measures. Third, we show that our strategy index is empirically associated with both the concurrent and subsequent financial outcomes of firms. These empirical findings serve as the cross-validation of strategy measures, which is an important step before applying the strategy measure to further studies and has been given limited attention in previous research (Zahra and Pearce, 1990). Moreover, because the strategy index has incremental predictive power for future operating activities and operating performance, it can assist shareholders in forecasting firm performance and making investment decisions.

This study also contributes to the policy debate on the SEC concept release No. 33-10064 (Securities and Exchange Commission, 2016), which solicits the public's views on financial disclosure issues. The release specifically asks for comments on whether to explicitly require disclosures of strategy in the annual reports. As our findings suggest that the disclosures in 10-K explain concurrent and subsequent firm financial outcomes, it is likely that explicit requirements on strategy disclosures would encourage managers to provide more reliable and precise information on their strategy and thus facilitate the analysis of strategies by stakeholders.

Measuring Strategy

To identify how firms *position* themselves, and how they *implement* the strategic positioning in their "self-introductions," we build annual firm-level strategy indices based on keyword counts found in their 10-K filings under Item 1 - Business. Specifically, we capture firms' emphases on two strategies, differentiation and cost leadership, which are discussed extensively in Porter's work and later studies. Firms adopting the differentiation strategy aim to differentiate themselves from their industry peers by constantly delivering unique products and services, with the uniqueness coming from innovative product features, advanced technology, customer intimacy, and brand image. Other firms adopting cost leadership attempt to achieve the low-cost position through economies of scale/scope, tight cost control, and operation efficiency.

We develop the strategy index using the textual components extracted from the Business section of firms' annual 10-K filings. The U.S. SEC requires domestic public firms to file annual reports on the Form 10-K which comprehensively discuss their business and financial condition (Securities and Exchange Commission, 2009). In addition, over a three-year period ending May 6, 1996, public firms were phased into filing on the Electronic Data Gathering, Analysis, and Retravel (EDGAR) system, a publicly accessible online system. Thus, any firm's 10-Ks can be

downloaded and analyzed by the public since then. Similar to other types of firm disclosure, such as a press release, conference call, or the company website, a firm's 10-K conveys private information and is usually written by the communication and legal departments under the instruction of management teams (Abrahamson and Hambrick, 1997). Unlike less formal disclosure, the language in a 10-K is often stricter, less colorful, and more comparable across firms, as it is filed following SEC regulations and is under close scrutiny of the public and auditors (Securities and Exchange Commission, 2011). As a result, researchers in various business disciplines can analyze the text of 10-K filings to extract qualitative information (e.g., Feldman *et al.*, 2010; Kothari, Shu, and Wysocki, 2009; Loughran and Mcdonald, 2016).

The focus of our analysis is on Item 1 of the 10-K Form, which SEC guidelines designate the Business section. Item 1 is "a good place to start to understand how the company operates" and is also a relatively standardized overview of firms' general operations (Securities and Exchange Commission, 2011). We restrict our text analysis to Item 1 only because it is where firm managers discuss firm products, services, and strategies (Hoberg and Phillips, 2016). Additionally, Regulation S-K requires updated disclosure on business development in Item 1 to provide investors timely information about the competitive position of a firm (Securities and Exchange Commission, 2016). Therefore, Item 1 effectively serves as firms' formal "self-introduction" to the public. Furthermore, since each item in a 10-K discusses a unique topic, focusing on Item 1 provides a relatively uniform context for our analysis and reduces the noise introduced by words and phrases which might be used in different ways across different items. For example, in other items of 10-K, the word "cost" is likely be used to describe accounting items such as "cost of goods sold," which may not necessarily outline the strategy. Therefore,

analyzing Item 1 instead of the complete 10-K can decrease measurement errors and improve the explanatory power of the textual strategy index.

After retrieving the text from Item 1, we deploy a word-count approach to measure strategy. That is, we create keyword lists for differentiation and cost leadership strategies, and then count the number of keywords that appear in Item 1 for each 10-K filing. The key assumption implicit in this approach is that firms will use more keywords associated with a certain strategy to describe themselves and their operations when the firm leans more toward that strategy. This is consistent with the prior linguistics literature, where frequency counts often represent the intensity of concern in a given document (Weber, 1990). Similar assumptions have been made in prior studies in the strategy literature that utilized textual analysis techniques where the word-count approach was applied (e.g, Abrahamson and Hambrick, 1997; Huff, 1990; Kaplan, 2008).

To implement the word-count approach, we develop strategy keyword lists that pinpoint the strategic positioning of a firm and its implementations in their 10-K discussions. That is, our keyword lists cover not only the words and phrases that are used to conceptually describe the strategy but also those that describe the value-creating activities that facilitate strategy implementations. For example, for the differentiation strategy, besides keywords such as "differentiation" and "unique" that are used to describe the strategy itself, keywords about activities that are often taken by the differentiators such as "research and development" and "innovation" are also considered. Including keywords that describe strategic activities can provide additional information to the strategy index because "a firm's strategy is manifested in the way in which it configures and links the many activities in its value chain relative to competitors" (Porter, 1991).

To ensure that the keywords are selected with a solid theoretical foundation, we draw the lists from Porter's work and subsequent studies. In Appendix A.1, we provide citations for each keyword to ensure that our keyword lists are supported by and can be traced back to the prior literature. Thus, this process is similar to establishing the content validity, which relies on experts to assess the usefulness of each keyword (Lawshe, 1975). Additionally, we randomly select 50 annual reports to assess the precision of the keywords in capturing strategy and revise the list accordingly to mitigate ambiguity. We discuss the keyword lists and their theoretical foundation in detail below.

Differentiators usually obtain uniqueness through product innovation and/or customer intimacy (Treacy and Wiersema, 1995). R&D investments help firms pioneer product innovation, which creates greater tangible customer value by delivering new and unique products that their competitors cannot offer (Hambrick, 1983a; Nair and Filer, 2003; Spanos, Zaralis, and Lioukas, 2004; Thornhill and White, 2007). Marketing functionality builds customer intimacy through anticipating and satisfying customer needs (Hambrick, 1983a; Narver and Slater, 1990; Spanos *et al.*, 2004). Customer intimacy, in turn, delivers greater intangible customer value that differentiates the firm from its industry peers. Also, to support the R&D function and marketing function, the human resource (HR) function is of vital importance since it helps acquire and maintain valuable human capital (Porter, 1985). Therefore, we develop the keyword list for differentiation by including keywords and phrases related to the differentiated position (Porter, 1985: 14) and how firms obtain the unique position through product innovation and customer intimacy (Campbell-Hunt, 2000; Hambrick, 1983a; Kotha and Vadlamani, 1995; Miller and Friesen, 1986a, 1986b; Nayyar, 1993; Shinkle, Kriauciunas, and Hundley, 2013a; Thornhill and White, 2007; Treacy and Wiersema, 1995; White, 1986; Zahra and Covin, 1993; Zatzick *et al.*,

2012). Drawing on the prior literature, we focus on the following words and phrases to discern the emphasis on differentiation. Here, an asterisk ("*") denotes a wildcard which matches alphabet letters. For example, "unique*" represents both "unique" and "uniqueness."

- Strategic positioning: *differenti*, unique*, superior*, premium*, high* price*, high* margin*, high* end*, inelasticity**
- Product: *new* product*, innovate*, creativ*, research and development, R&D, techni*, technolog*, patent**
- Quality management: *quality*, reliab*, durable**
- Marketing: *marketing*, advertis*, brand*, reputation*, trademark**
- Customer: *customer* service*, consumer* service*, customer* need*, sales support*, post-purchase service*, customer* preference*, consumer* preference*, consumer* relation*, consumer* experience*, consumer* support*, loyalty*, customiz*, tailor*, personaliz*, responsive**
- Human resources: *talent*, train*, skill*, intellectual propert*, human capital**

Cost leaders compete on price or cost. To achieve the low-cost position within an industry, firms can improve upon cost control, operation efficiency, and economies of scale/scope. Therefore, the list developed for cost leadership includes keywords and phrases describing the overall low-cost position (Porter, 1985: 12) and the operation activities that contribute to the low-cost position: cost control, operation efficiency, and economy of scale (Campbell-Hunt, 2000; Hambrick, 1983a; Kotha and Vadlamani, 1995; Miller and Friesen, 1986a, 1986b; Nayyar, 1993; Shinkle *et al.*, 2013a; Thornhill and White, 2007; Treacy and Wiersema, 1995; White, 1986; Zahra and Covin, 1993; Zatzick *et al.*, 2012). Consequently, we search for the occurrences of the following words and phrases to discern emphasis on cost leadership.

- Strategic positioning: *cost leader**, *low* pric**, *low* cost**, *cost advantage**, *competitive pric**, *aggressive pric**
- Cost control: *control* cost*, control* expense*, control* overhead*, minimiz* cost*, minimiz* expense*, minimiz* overhead*, reduce* cost*, reduce* expense*, reduce* cost*, cut* cost*, cut* expense*, cut* overhead*, decreas* cost*, decreas* expense*, decreas* overhead*, monitor* cost*, monitor* expense*, monitor* overhead*, sav* cost*, sav* expense*, sav* overhead*, cost* control*, cost* minimization*, cost* reduction*, cost* saving*, cost* improvement*, expense* control*, expense* minimization*, expense**

reduction, *expense** *saving**, *expense** *improvement**, *overhead** *control**, *overhead** minimization*, overhead* reduction*, overhead* *saving**, *overhead** *improvement**

- Operation efficiency: *efficien**, *high** *yield**, *process** *improvement**, *asset** *utilization**, *capacity** *utilization**
- Scale economy: *scope**, *scale**, *breath**, *broad*, *mass*, *high** *volume**, *large** *volume**

It is important to allow for flexibility when searching the 10-K Item 1 for keyword occurrences. First, we use wildcards to capture keyword variations. Second, for two-word phrases, we consider both the form with a hyphen and the form without a hyphen between the two words (e.g., low cost and low-cost). Third, for phrases that contain a verb and its object, we count the instances in which the verb and its object exist in the same sentence without requiring them to be adjacent to each other (for example, both "minimize cost" and "minimize the direct labor cost" are counted similarly). To illustrate how our measure captures competitive strategies, Appendix A.2 provides the discussions of strategy using the 2016 annual reports of Apple Inc. and Wal-Mart Stores Inc. as illustrative examples.

We then aggregate individual keyword frequencies into categories of differentiation and cost leadership by summing the counts and scaling them by the total number of words (in thousands) in Item 1. The resulting raw frequency measures for competitive strategies, *Diff* and *Cost*, are defined as follows:

$$Diff = \frac{Ndiff}{Textlen} \times 1000 \text{ and } Cost = \frac{Ncost}{Textlen} \times 1000 \quad (1)$$

where *Ndiff* and *Ncost* are the frequencies of differentiation keywords and cost leadership keywords, respectively, and *Textlen* is the total number of words in Item 1 in a 10-K filing. Therefore, greater *Diff* indicates a firm's greater emphasis on the differentiation, and greater *Cost* indicates its greater emphasis on the cost leadership.

Instrument Validation

Following the prior literature, we conduct a battery of validity checks for the strategy index (Shadish, Cook, and Campbell, 2002). First, we validate that our selection of keywords is supported by the prior literature and the categorization outputs. Second, we examine the persistence of the strategy index. We expect our strategy index to be persistent as firm-level strategy often persists over time. Finally, we test construct validity by replicating the empirical tests from the prior literature, and demonstrate that similar results can be generated.

Sample selection and descriptive statistics

For our empirical analyses, we construct the sample based on the universe of firms in the Compustat database from fiscal years 1995 to 2015 because a majority of firms started filing their 10-Ks to the EDGAR database after 1994. We use web-crawling techniques to download all the filings that appear as 10-K, 10-K405, 10-KSB, or 10-KSB40 released on the SEC online EDGAR database for the firms based on their Central Index Key (CIK), a unique identifier assigned by the SEC to identify firms. Then, we use a Python script to parse each filing to identify the Business section. For filings released before 2005, we extract the content between "Item 1 Business" and "Item 2 Properties." For filings released later, we extract the content between "Item 1 Business" and "Item 1A Risk Factors," since the SEC mandated the risk factor disclosures and introduced Item 1A in 2005 (Bao and Datta, 2014). Finally, we process the extracted texts and construct the strategy index for each firm-year observation. The initial sample of textual data contains 94,981 observations.

To assess the construct validity of the strategy index, we show that our strategy index is associated with concurrent and future firm operating activities and performance in a predictable way. Thus, we further merge our strategy index with the Compustat database, requiring sufficient

financial data for each observation to perform the regressions except that, for the observations with missing R&D, marketing, and SG&A expenses, we treat missing values as zero. Moreover, we eliminate financial firms (SIC code 6000-6999) and utility firms (SIC code 4900-4999) because they are regulated and the financial statements of the financial firms are substantially different from those of other firms, which makes it impossible to construct some of the measures that we use in the empirical tests. Finally, we follow Banker *et al.* (2016) and eliminate observations with the ratio of current period sales to prior period sales being less than 0.5 or greater than 2.0 to exclude the observations that potentially experienced non-articulating transactions such as mergers and acquisitions or significant divestitures. The remaining sample contains 46,032 observations for 4,985 unique firms between 1995 and 2015. In Appendix A.3, we summarize our sample selection procedure.

Table 1 provides the descriptive statistics of our sample. The mean of *Diff* and *Cost* are 11.02 and 1.27, representing strategy keyword occurrences scaled by the length of the focal document (per 1,000 words). Specifically, the median raw count of differentiation keywords (*Ndiff*) and cost leadership keywords (*Ncost*) are 59 and 6. The median number of total words in Item 1 Business (*Textlen*) of our sample is 5,804. One explanation for the noticeable difference between *Diff* and *Cost* lies with the nature of the strategies. That is, differentiation strategy tends to require longer discussion since firms usually explain in detail how they differ from others along multiple dimensions, while cost leadership discussion about prices and costs are usually less complicated.

Descriptive statistics of other variables depict the entrepreneurial nature of our sample firms. The mean market capitalization of sample firms is $3.267 billion, and the mean age of a sample firm is approximately 17 years. The mean gross margin ratio is 0.33 and the mean asset

turnover ratio is 1.19. The means for R&D expenses, marketing expense, and selling, general and administrative (SG&A) expense are 12 percent, 1 percent, and 29 percent of the sales revenue, respectively.

Insert Table 1 here

Keyword validation

As mentioned, our keywords are selected based on the discussion of strategy typologies in Porter's work and related studies. Our keyword lists aim to reflect both the strategic positioning and the operating activities that support the strategic positioning and, thus, comprehensively capture the strategy. Moreover, to ensure that the strategy index makes intuitive sense, we select a subset of firms and manually compare their strategy indices with the strategic information disclosed through other communication channels, such as a firm's corporate website and its press releases. In Appendix A.4, we provide a list of Fortune 100 firms that rank Top 3 of the average value of *Diff* and *Cost* from 2011 to 2015, respectively. We find that our index accurately represents the strategy delivered through firms' public disclosures. For example, Intel Corp is one of the top differentiators in our list (*Diff* =19.89), which is consistent with their well-received brands and their commitment to "invents at the boundaries of technology to make amazing experiences possible for business and society" (Intel Company, 2017).

Strategy persistence

The strategy of a firm provides "a long-term context within which constituents can attribute meaning to specific investments and projects" (Rindova and Fombrun, 1999). Specifically, strategic planning involves establishing a long-term plan to obtain competitive advantages and translating that plan into daily operations that are coherent and executable. A strategy is typically not changed frequently, as abrupt changes may blur the corporate image and confuse customers

(Porter, 1996). Moreover, past activities may also create organizational and environmental constraints that limit the managers' ability to modify the strategy (Kabanoff and Brown, 2008). Following this argument, for our strategy index to be valid, we expect the strategy to be persistent. That is, we expect the strategy indices of the same firm from different years to be largely consistent with each other.[iii]

To test this insight, we first compute Cronbach's alpha, which ranges from 0 to 1 and measures how consistent a set of items are as a group (Cronbach, 1951). We find that the Cronbach's alpha of *Diff* and *Cost* are 0.966 and 0.964, respectively, which are higher than 0.9, the commonly accepted threshold for high consistency. We find that some of the firms make incremental changes in their 10-K Item 1, which can lead to high persistence for those firms mechanically. However, as long as the 10-K still accurately represents the business conditions of the firm, our measure is not biased.

Insert Table 2 here

Table 2 presents another way to examine strategy persistence. We sort firms into quintiles for each year based on *Diff* and *Cost* and track the top and bottom quintiles over the subsequent three years. If the strategy index is persistent, most of the firms in those two quintiles should remain in the quintiles. The results in Table 2 are consistent with our expectation. For example, there are 9,206 firm-years in the fifth quintile of *Diff* (top differentiators) in Year 0, which is defined as the fiscal year when the observations are sorted into the fifth quintile.[iv] The percentages of these observations remain in the fifth quintile after one, two, and three years are 85.59 percent, 80.04 percent, and 75.41 percent, respectively. Overall, the descriptive data in Table 2 suggests that the textual strategy index is persistent while some time-series variations exist as well, again consistent with the strategy being a long-term concept.

Construct validity

To assess the construct validity of the textual strategy index, we examine how well it captures the strategy concept. A valid index should exhibit a systematic and predictable relationship with many other variables discussed and examined in the prior literature. To do so, we replicate and extend some of the prior studies to show that similar results can be generated using our textual measure. That is, we show that the textual strategy index can perform at least as well as those measures created based on more costly methods such as executive surveys in various empirical settings. Specifically, we examine the concurrent and intertemporal associations between our textual strategy index and several sets of financial variables, including operating activities, operation efficiency, and profit margin.

Operating activities

Firms that adopt a differentiation strategy often focus on product innovation and/or customer intimacy (Treacy and Wiersema, 1995). Firms with a product innovation focus would invest more in R&D to improve product designs and features, and firms with a customer intimacy focus would develop and maintain a close relationship with customers through branding, customized service, and skilled sales forces (Campbell-Hunt, 2000; Hambrick, 1983a; Miller and Friesen, 1986a). Moreover, to support their innovation and marketing activities, recruiting talented employees is critical and demands special attention from the management. By contrast, cost leaders constantly contain their SG&A costs to achieve the low-cost position in the industry. Thus, we test whether the textual index is associated with concurrent operating activities such as R&D, marketing, and HR functions. We predict that differentiators allocate more resources to these functions.

Additionally, our textual measure has the potential to predict future operating activities as well because 10-K disclosures may reflect firms' commitment to those functions with the

ambition to sustain the competitive advantages (Ghemawat, 1991; Porter, 1985). Thus, we test the *incremental* predictive power of our textual measure for future activities when the concurrent activities are already controlled for.

We estimate the following ordinary least squares (OLS) models with year- and industry-fixed effects to test the association between the strategy index and concurrent and future operating activities.

$$Operating_t = \gamma_{1,0} + \gamma_{1,1} Diff_t + \gamma_{1,2} Cost_t + \sum \rho_{1,i} Control_{i,t} + \xi_{1,t} \quad (2)$$

$$Operating_{t+1} = \gamma_{2,0} + \gamma_{2,1} Diff_t + \gamma_{2,2} Cost_t + \gamma_{2.3} Operating_t + \sum \rho_{2,i} Control_{i,t} + \xi_{2,t} \quad (3)$$

where the dependent variable, $Operating$, is $R\&D$, Adv, or $SG\&A$. $R\&D$ is defined as R&D expenses scaled by sales revenue and serves as a proxy for R&D function. Adv is defined as marketing expenses scaled by sales revenue and serves as a proxy for marketing function. $SG\&A$ is defined as selling, general and administrative expenses scaled by sales revenue and serves as an aggregated proxy for firms' orientation toward R&D, marketing, and HR functions. We expect positive $\gamma_{1,1}$ and $\gamma_{2,1}$ and negative $\gamma_{1,2}$ and $\gamma_{2,2}$ in the above models.

In the analysis that follows, we use two-way clustered standard errors by firm and year to correct for the potential correlations of the regression error terms between different firms in the same year and different years of the same firm (Thompson, 2011). The control variables include the size (*Ln(Mktcap)*) and firm age (*Ln(Age)*). Size can be viewed as a proxy for market power that affects the incentive for innovation investment and firm age as a proxy for past experience that affects productivity in innovation investment (Henderson, 1993). Both market power and past experience limit managerial choices in innovation and other operating activities and, thus, the inclusion of these variables strengthens our empirical tests. We also include industry- and

year-fixed effects to control for the similarities between observations from the same industry and the same year. Industry-fixed effects are defined based on Fama-French twelve-industry classification. The detailed definition of each variable is provided in Appendix A.5.

Insert Table 3 here

Table 3 shows the results of regressing operating activity variables on strategy variables. We find the coefficients on *Diff* are positive and significant at the one percent level in almost all specifications except for predicting R&D in the subsequent period, indicating differentiators incur more R&D, marketing, and SG&A expenses in both year t and year t+1. We find the coefficient on *Cost* is negative and significant at the one percent level in almost all specifications except for predicting R&D in the subsequent period, indicating cost leaders incur less R&D, marketing, and SG&A expenses in both year t and year t+1. Results in Table 3 also suggest that the impact of strategy on operating activities is economically significant. For example, the coefficient of *Diff* in Column I is 0.0037, indicating that, when the value of *Diff* increases by one standard deviation, the ratio of R&D expenses to sales revenue increases by 2.31 percent. Such an increase is equivalent to a rise of $67 million in R&D expenses on average (the mean sales revenue of our sample is $2,884 million). The coefficient of *Diff* and *Cost* in Column II is smaller in magnitude because of the inclusion of the lagged dependent variable in the model. Thus, the coefficients on *Diff* and *Cost* in Column II, IV, and VI capture the predictability of our strategy index for expenses over-and-above the first-order autoregression.

Operation efficiency

Firm strategy also affects operation efficiency (Campbell-Hunt, 2000; Dess and Davis, 1984; Hambrick, 1983a; Porter, 1985). Beyond cost reduction initiatives, another channel for firms to achieve the low-cost position is asset parsimony, i.e. reducing the assets needed for each

unit of output (Hambrick, 1983a). Cost leaders can commit to employing automated machinery to improve efficiency and productivity, showing a lower level of asset parsimony. Specifically, capital expenditure indicates such commitments to technology efficiency (Berman *et al.*, 1999; Kotha and Nair, 1995; Spanos *et al.*, 2004). As a result, cost leaders utilize assets more efficiently and generate greater revenue with given tangible resources. Therefore, we expect differentiators to have lower capital intensity and asset efficiency in both concurrent and subsequent periods, and validate our expectation by estimating the following OLS models:

$$Effi_t = \alpha_{1,0} + \alpha_{1,1} Diff_t + \alpha_{1,2} Cost_t + \sum \varphi_{1,i} Control_{i,t} + \mu_{1,t} \quad (4)$$

$$Effi_{t+1} = \alpha_{2,0} + \alpha_{2,1} Diff_t + \alpha_{2,2} Cost_t + \alpha_{2,3} Effi_t + \sum \varphi_{2,i} Control_{i,t} + \mu_{2,t} \quad (5)$$

where the dependent variable, $Effi$, is CE or ATO. We measure capital expenditure CE as capital expenditure scaled by beginning-of-year total assets. ATO is asset turnover ratio, defined as sales revenue scaled by beginning-of-year total assets, which is a proxy for asset efficiency (i.e., the inverse of asset parsimony). Thus, we expect negative $\alpha_{1,1}$ and $\alpha_{2,1}$ and positive $\alpha_{1,2}$ and $\alpha_{2,2}$.

Insert Table 4 here

Table 4 presents the results of regressing efficiency variables on strategy variables. Again, we examine both the concurrent and intertemporal associations between strategy variables and efficiency. We find the coefficients on *Diff* are negative and significant at the one percent level in Column I and II, suggesting that firms leaning toward differentiation have lower capital expenditure. We also find the coefficients on *Cost* are positive and significant at the one percent level in Column III and IV, suggesting that firms leaning toward cost leadership operate more efficiently. Moreover, the significant coefficients of *Diff* and *Cost* in Column II, IV, and VI

capture the predictability of strategy on efficiency after controlling the lagged efficiency measures.

Profit margins

The difference in operating activities further leads to the difference in financial outcomes (Selling and Stickney, 1989). Differentiators can often charge a premium price because customers enjoy the greater value of their products and services (Porter 1985). This unique position also helps differentiators obtain successive market power, which further influences the price-cost margins (Waterson, 1980). In other words, firms that emphasize a differentiation strategy are more likely to hold a unique position relative to their peers and command a greater profit margin (e.g., David *et al.*, 2002; Porter, 1985). Thus, we expect differentiators to have a higher margin and estimate the following OLS models:

$$Margin_t = \beta_{1,0} + \beta_{1,1} Diff_t + \beta_{1,2} Cost_t + \sum \eta_{1,i} Control_{i,t} + \varepsilon_t \quad (6)$$

$$Margin_{t+1} = \beta_{2,0} + \beta_{2,1} Diff_t + \beta_{2,2} Cost_t + \beta_{2,3} Margin_t + \sum \eta_{2,i} Control_{i,t} + \varepsilon_t \quad (7)$$

where the dependent variable, $Margin$, is $GM, OM, or\ OM\_adj$. GM is defined as gross profit (i.e., sales revenue minus cost of goods sold) scaled by sales revenue, and OM is defined as operating profit (i.e., sales revenue minus cost of goods sold and SG&A expenses) scaled by sales revenue. We also use adjusted operating margin OM_adj which excludes R&D and marketing expenses from SG&A expenses because R&D and marketing are often viewed as investments by shareholders and, hence, may not always be considered detrimental to profits (Ball *et al.*, 2015; Lev and Sougiannis, 1996). Overall, we expect $\beta_{1,1}$ and $\beta_{2,1}$ to be positive and $\beta_{1,2}$ and $\beta_{2,2}$ to be negative in model (6) and model (7).

Insert Table 5 here

Table 5 reports the results of regressing margins on strategy variables, testing both concurrent and intertemporal associations. Column I and II report the results for the gross margin model, and the coefficients on *Diff* are positive and significant at the five percent level and coefficients on *Cost* are negative and significant at the one percent level. The results suggest that firms that emphasize differentiation strategy charge a greater price premium and continue to have greater gross margin in the subsequent period even when the current gross margin is controlled for in the regression. Moreover, we find that the coefficient of *Cost* is positive and significant for operating margin in the concurrent period. However, the coefficient becomes negative when the operating margin measure is adjusted, indicating that the positive association between operating margin and *Cost* is driven by the investments in R&D and marketing functions. Thus, interpreting the results of the operating margin requires extra caution due to the current accounting treatment for R&D and marketing expenses.

Robustness tests

To ensure the robustness of our results, we perform several additional tests. First, we estimate the R&D, marketing, and SG&A models with Tobit regressions and find consistent results. Moreover, we re-estimate each dependent variable without setting missing values to zero but, instead, drop those observations. The coefficients of *Diff* and *Cost* are still statistically significant at the one percent level with the expected sign.

Moreover, we estimate the regression models using a modified Fama-MacBeth approach. Unlike the pooled OLS estimator, our modified Fama-MacBeth approach estimates the coefficients with the following three steps (Fama and MacBeth, 1973). First, we split the full sample into subsamples based on industry and estimate the regressions for each subsample. Second, we take the coefficients from the subsamples and calculate the average to obtain the

estimates for the full sample. Finally, we estimate the variance-covariance matrix of the estimates based on the variations in the industry-level coefficients and compute the t-statistics for the estimates. Therefore, the standard errors generated based on the modified Fama-MacBeth approach are robust to the correlations of the error terms among firms within the same industry. We find that our results are again consistent with those from the main tests.

Implications

In this section, we focus on the variation in strategy across industries and years and further explore its implication on firm performance and financial reporting.

Strategy and firm performance

We first examine the relationship between strategy and firm performance. Among all performance measures, accounting profitability and stock return are considered to be most comprehensive and, therefore, are widely studied in the prior literature (e.g., Desarbo *et al.*, 2005; Kabanoff and Brown, 2008; Rajagopalan, 1997). In particular, accounting profitability measures short-term performance, and the stock return reflects the long-term performance of firms (Ball and Brown, 1968). Perhaps because prior studies often use data from a single industry, they document limited and mixed evidence on how strategy impacts performance (e.g. Dess and Davis, 1984; Hambrick, 1983a; Miller and Friesen, 1986a, 1986b; Robinson and Pearce, 1988). With our textual measure, we can examine the association between strategy and performance using a sample that covers multiple years and multiple industries.

We employ ROA as the accounting-based measure of firm performance, which assesses how well a firm can utilize its assets to generate current-period profit. And stock return (RET) is a market-based measure that also captures the investors' expectation about firms' future performance.

Insert Figure 3 here

Figure 3 compares the time trend of performance measures between two strategic groups, differentiators and cost leaders. We classify firms into strategic groups based on the strategy index. Specifically, for each year and industry combination, we divide the sample into three subsamples based on the strategy index terciles.[v] We label the observations in the third tercile of *Cost* as cost leaders and the observations in the third tercile of *Diff* as differentiators. As shown in the figure, the differentiators' average ROA is consistently lower than the cost leaders' average ROA, while the differentiators' average RET closely resembles the cost leaders' average RET.

Insert Table 6 here

We further examine the difference in performances by regressing performance measures on the continuous strategy index and explore whether the results shown in the figure hold for the multivariate analysis. We estimate the following OLS models. The number of observations decreases in these regressions because of data availability.

$$ROA_t = \theta_0 + \theta_1 Diff_t + \theta_2 Cost_t + \sum \kappa_{1,i} Control_{i,t} + \epsilon_t \quad (8)$$

$$RET_t = \zeta_0 + \zeta_1 Diff_t + \zeta_2 Cost_t + \zeta_3 Ln(Mktcap)_{t-1} + \zeta_4 BTM_{t-1} + \zeta_5 Beta_{t-1} + \zeta_6 RET_{t-1} + \iota_t \quad (9)$$

where, for the RET regression, we control for the four risk factors that are commonly used in empirical asset pricing literature (Carhart, 1997), namely, firm size *Ln(Mktcap)*, book-to-market ratio *BTM*, Capital Asset Pricing Model (CAPM) beta *Beta*, and prior-year stock return RET_{t-1}.

As shown in Table 6, we find a negative and significant association between ROA and *Ratio*, indicating that differentiators have lower ROA. This observation may be attributable to the inclusion of R&D and marketing investment in operating expenses. Because firms' inputs

toward R&D and marketing activities are likely to generate future long-term cash flows, such inputs may not always be viewed as pure costs. Therefore, because differentiators tend to invest more in R&D and marketing functions, ROA, although widely used in the accounting and strategy literature, may not fully represent their performance.

For the market-based performance measure RET, we find that the coefficient of *Ratio* is statistically insignificant, suggesting that strategy does not affect stock returns. While *prima facie* differentiators with a riskier business would deliver greater stock returns, traditional CAPM approach argues that only systematic risk should be incorporated into asset prices and command a risk premium. Therefore, this result is consistent with that strategy being an idiosyncratic risk, which can be diversified and hence does not affect stock return.

Economic condition and changes in strategy

Next, we explore whether firms adjust their strategy as a response to economic conditions such as recessions and booms by calculating the average *Ratio* of firms for each year. For these analyses, we measure the relative importance of the strategies of a firm by computing a new variable *Ratio* as *Diff/(Diff+Cost)*. Therefore, a higher *Ratio* indicates a greater emphasis on differentiation. To ensure that we document the firms' adjustment in strategy rather the changes in market composition, we focus on a balanced sample, where all the firms survived the entire sample period, 1995 to 2015. The balanced sample comprises 242 firms.

Insert Figure 1 here

Figure 1 plots the time trend of strategy ratio of the balanced sample. The average ratio shows an overall upward trend and especially during economic booms, showing firms' growing emphasis on differentiation strategy. This may be partly driven by the growing emphasis on technology development over the past two decades. Moreover, the average ratio dropped

noticeably around both the burst of the dot-com bubble and the financial crisis. Thus, it is likely that managers emphasize more on improving cost structures and less on expanding intangible investment during the economic downturns.

Insert Figure 2 here

We further explore the changes in average strategy ratio by industry. Since the industry trends are subject to industry-specific economic factors, we expect the time-series variations defer by industry. In figure 2, we compare the trends of manufacturing industry and chemical industry which are, again, classified according to Fama-French 12-industry classification. The manufacturing firms produce machinery, trucks, and plans, and the chemical firms produce chemicals and allied products. We find the two industries present differing trends in the recent years. The greater increase of strategy ratio in chemicals highlights the trend that the industry has become a high-tech industry where the competitive advantages rely on technological intensity (Ahuja and Katila, 2004).

Financial reporting

Finally, we examine whether firm strategy has a real impact on financial reporting. Prior studies documented the changes in earnings properties over the past fifty years, namely a decline in earnings relevance and the matching of concurrent revenues and expenses and an increase in earnings volatility (Collins *et al.*, 1997; Dechow *et al.*, 2010; Dichev and Tang, 2008; Donelson *et al.*, 2011; Francis and Schipper, 1999; Lev and Zarowin, 1999). In this paper, we take a closer look at the data over past two decades and study whether firm strategy is at least partially driving the trend in earnings properties.

During the past two decades, an accelerated shift to a knowledge-based economy led to a closer focus of firms on information, technology, and learning. As we show in the previous

section, seasoned firms exhibit a heavier emphasis on differentiation strategy. In addition, more "idea" firms are established and listed over time and these new cohorts of firms lean toward differentiation to a greater extent. We estimate the following regression to observe the drivers of the changes in strategy and earnings properties over time.

$$Dep = \alpha_0 + \alpha_1 Recency + \alpha_2 Fyear \quad (10)$$

where the dependent variable, Dep, is $Ratio$, *R&D*, *Matching*, and *Vol(ROA)*. *Ratio* represents the relative important of the strategy within firm. *R&D* is R&D expenditures scaled by sales revenue, a proxy for intangible intensity. *Matching* is the estimates of concurrent revenue and expense matching for each cohort-year. And *Vol(ROA)* is the four-year rolling standard deviation of *ROA*. *Recency* is a categorical variable for different cohorts, with values of 1,2,3,4, and 5 assigning to firms listed before 1995, firms listed between 1995 and 2000, firms listed between 2000 and 2005, firms listed between 2006 and 2010, and firms listed between 2010 and 2015. Fyear is added to control for overall time trends.

Table 7 shows that the coefficients of *Recency* are negatively significant in column (I) and (II), indicating that firms listed in more recent years lean toward differentiation strategy and have greater intangible intensity. In addition, firms listed in more recent years have lower matching and higher earnings volatility. In the untabulated ANOVA tests, *Recency* rather than *fyear* explains more variations in strategy and earnings properties. That is, the changes in market composition explain more variations in strategy measures in the sample than the changes within firms, consistent with that strategy is a relatively stable attribute of a firm.

The differences in firm strategy lead to different activities in daily operating and, in turn, varying patterns in earnings properties in financial reporting. Differentiators' investments in product design, quality control, marketing and branding, and customer service usually reap

profits after more than a fiscal year while the investments are expensed immediately under the current GAAP. On the other hand, cost leaders control costs and improve efficiency and, thus, get immediate payback. Therefore, firms lean toward differentiation tend to have poor matching and higher earnings volatility. We estimate the following regression for matching and earnings volatility, respectively.

$$Revenue_{\mathrm{t}} = \vartheta_1 + \vartheta_2 Expense_{t-1} + \vartheta_3 Expense_t + \vartheta_4 Expense_{t+1} + \vartheta_5 \mathrm{Ratio}_{\mathrm{t}} + \vartheta_6 Expense_{t-1} * Ratio_t + \vartheta_7 Expense_t * Ratio_t + \vartheta_8 Expense_{t+1} * Ratio_t + \vartheta_9 Recency_t + \vartheta_{10} Expense_{t-1} * Recency_t + \vartheta_{11} Expense_t * Recency_t + \vartheta_{12} Expense_{t+1} * Recency_t + \tau'_1 Control + \tau'_2 Expense_{t-1} * Control + \tau'_3 Expense_t * Control + \tau'_4 * Expense_{t+1} * Control + \ni_{\mathrm{t}} \quad (11)$$

$$Vol(ROA_{t-2-t+1}) = \pi_0 + \pi_1 Ratio_t + \pi_2 Recency_t + \sigma' Control + \epsilon_t \quad (12)$$

where $Revenue_{\mathrm{t}}$ is the total revenue scaled by average assets and $Expense_t$ is the total expense scaled by average assets. *Vol(ROA)* is the four-year rolling standard deviation of *ROA*. $Ratio_t$ is the relative importance of strategy and $Recency_t$ is a categorical variable for different cohorts. We control for other reasons for poor matching and high volatility such as intangible intensity, loss, special item, operating cash flow, firm size (Collins *et al.*, 1997; Dichev and Tang, 2008; Donelson *et al.*, 2011; Givoly and Hayn, 2000; Srivastava, 2014).

Table 8 shows that the coefficients of $Recency_t$ are significant, indicating that newly listed firms have poorer matching and high earnings volatility. We also find an incremental explanatory power of $Ratio_t$ on earnings properties after controlling for the cohort effect, consistent with that firm strategy lead to varying patterns in financial reporting due to different operating activities.

Conclusion And Discussion

In this paper, we develop a firm-level strategy index based on the texts of firm's public disclosure. We first create comprehensive keyword lists for differentiation and cost leadership orientations and then compute the keyword frequency in the Item 1 Business of 10-K filings to generate the strategy index. Using a sample of 46,032 U.S. public firm-year observations, we show that our strategy index is persistent across years, which is consistent with the prediction that, on average, firms' strategy is long-term in nature. We further assess the construct validity of the index by replicating the prior studies and estimating the association between our strategy index and outcome variables including both financial variables such as asset turnover ratio and gross margin and content-based operating activity measures such as R&D expenses and marketing expenses. We can reliably reproduce the results documented in the prior literature using the textual strategy index. Furthermore, we show the time-series and cross-sectional variations of the strategy index and use the strategy index to examine how firm performance varies with strategy. We find that while cost leaders and differentiators exhibit systematic differences in ROA, their stock returns are similar.

Compared to the other approaches that measure strategy, our text analysis approach is easier to implement. Computer software (e.g. WordStat and QDA Miner) and programming languages such as Perl and Python can handle large-scale textual documents and produce reliable and consistent measurements for a large sample with a low cost. Thus, our measure has the unique advantage of enabling researchers to comprehensively analyze firms in multiple industries and for multiple years. Moreover, our strategy index is associated with not only concurrent but also subsequent firm performance, which suggests that summarizing the strategic

disclosures in 10-Ks with our approach can provide useful information to various stakeholders such as debt holders, shareholders, and business researchers.

Implementing or extending our approach also creates many interesting opportunities for future research. First, because our strategy index covers a large sample of multiple industries and multiple years, researchers can use it to explore whether results documented in the prior literature based on data from a single industry and a short period also hold for other industries and longer periods. It is likely that, because of the rapid change in information technology in recent decades, some of the findings in the prior literature may no longer apply to today's firms. Also, for the research that focuses on a specific industry, we encourage researchers to configure the keyword list to include industry-specific keywords. For example, for the airline industry, the keywords such as "on-time" and "convenient" may be included. Second, our measure draws inferences from firm disclosures, which is, hence, an "average" measure of generic strategies at the strategic business unit (SBU) level. Future research may apply our textual analysis approach to SBUs for a subgroup of the firms that discuss each segment in their annual reports and further explore how the strategy at SBU level interact with each other to generate synergy at the firm level. The challenge lies in the difficulty in identifying such firms and separating segment discussions. Finally, another interesting opportunity for future extension is to develop a textual measure of the focus strategy, which is not covered in this study. However, it is possible that an approach other than the word-count approach is needed to properly evaluate the focus strategy. Developing a focus strategy measure can contribute to, for example, the debate of pure strategy vs. hybrid strategy by showing how the tradeoff changes when the breath of the targeted market changes.

REFERENCES

Abrahamson E, Hambrick DC. 1997. Attentional homogeneity in industries: The effect of discretion. *Journal of Organizational Behavior* **18**(1997): 513–532.

Ahuja G, Katila R. 2004. Where do resources come from? The role of idiosyncratic situations. *Strategic Management Journal* **25**(89): 887–907.

Amazon.com Inc. 2017. Form 10-K. Available at: https://www.sec.gov/Archives/edgar/data/1018724/000101872417000011/amzn-20161231x10k.htm [20 December 2017].

AmerisourceBergen Corporation. 2016. Form 10-K. Available at: https://www.sec.gov/Archives/edgar/data/1140859/000104746915008939/a2226704z10-k.htm [7 October 2017].

Armstrong JS, Overton TS. 1977. Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research* **14**(3): 396.

Arthur JB. 1992. The link between business strategy and industrial relations systems in American steel. *ILR Review* **45**(3): 488–506.

Ball R, Brown P. 1968. An empirical evaluation of income numbers. *Journal of Accounting Research* **6**(2): 159–178.

Ball R, Gerakos J, Linnainmaa JT, Nikolaev VV. 2015. Deflating profitability. *Journal of Financial Economics* **117**(2): 225–248.

Banker RD, Basu S, Byzalov D, Chen JYS. 2016. The confounding effect of cost stickiness on conservatism estimates. *Journal of Accounting and Economics* **61**(1): 203–220.

Banker RD, Hu N, Luftman J, Pavlou PA. 2010. CIO reporting structure, strategic positioning, and firm performance: To whom should the CIO report? *MIS Quarterly* **35**(2): 487–504.

Bao Y, Datta A. 2014. Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science* **60**(6): 1371–1391.

Bentley KA, Omer TC, Sharp NY. 2013. Business strategy, financial reporting irregularities, and audit effort. *Contemporary Accounting Research* **30**(2): 780–817.

Berman SL, Wicks AC, Kotha S, Jones TM. 1999. Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance. *Academy of Management Journal* **42**(5): 488–506.

Campbell-Hunt C. 2000. What have we learned about generic competitive strategy? A meta-analysis. *Strategic Management Journal* **21**(2): 127–154.

Carhart MM. 1997. On persistence in mutual fund performance. *The Journal of Finance* **52**(1): 57.

Collins DW, Maydew EL, Weiss IS. 1997. Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics* **24**(1): 39–67.

Cronbach LJ. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* **16**(3): 297–334.

D'Aspremont C, Gabszewicz JJ, Thisse JF. 1979. On Hotelling's 'Stability in competition'. *Econometrica* **47**(5): 1145–1150.

David JS, Hwang Y, Pei BKW, Reneau JH. 2002. The performance effects of congruence between product competitive strategies and purchasing management design. *Management Science* **48**(7): 866–885.

Dechow P, Ge W, Schrand C. 2010. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*. Elsevier

50(2–3): 344–401.
Desarbo WS, Di Benedetto CA, Song M, Sinha I. 2005. Revisiting the Miles and Snow strategic framework: Uncovering interrelationships between strategic types, capabilities, environmental uncertainty, and firm performance. *Strategic Management Journal* **26**(1): 47–74.
Dess GG, Davis PS. 1984. Porter's (1980) generic gtrategies as determinants of strategic group membership and organizational performance. *Academy of Management Journal* **27**(3): 467–488.
Dichev ID, Tang VW. 2008. Matching and the changing properties of accounting earnings over the last 40 years. *Accounting Review* **83**(6): 1425–1460.
Donelson DC, Jennings R, Mclnnis J. 2011. Changes over time in the revenue-expense relation: Accounting or economics? *Accounting Review* **86**(3): 945–974.
DuPont. 2017a. Form 10-K. Available at: https://www.sec.gov/Archives/edgar/data/30554/000003055417000004/dd-12312016x10k.htm [20 December 2017].
DuPont. 2017b. Our Approach. Available at: http://www.dupont.com/corporate-functions/our-approach.html [20 December 2017].
Fama EF, MacBeth JD. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* **81**(3): 607–636.
Feldman R, Govindaraj S, Livnat J, Segal B. 2010. Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies* **15**(4): 915–953.
Francis J, Schipper K. 1999. Have Financial Statements Lost Their Relevance ? *Journal of Accounting Research* **37**(2): 319–352.
Ghemawat P. 1991. *Commitment: The Dynamic of Strategy*. Free Press: New York, NY.
Givoly D, Hayn C. 2000. The changing time-series properties of earnings,cash flows and accruals:Has financial reporting become more conservative? *Journal of Accounting & Economics* **29**: 287–320.
Hambrick DC. 1983a. High profit strategies in mature capital goods industries: A contingency approach. *Academy of Management Journal* **26**(4): 687–707.
Hambrick DC. 1983b. Some tests of the effectiveness and functional attributes of Miles and Snow's strategic types. *Academy of Management Journal* **26**(1): 5–26.
Henderson R. 1993. Underinvestment and incompetence as responses to radical innovation: Evidence from the photolithographic alignment equipment industry. *RAND Journal of Economics* **24**(2): 248–270.
Hoberg G, Phillips G. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* **124**(5): 1423–1465.
Hotelling H. 1929. Stability in competition. *The Economic Journal* **39**(153): 41.
Huff AS. 1990. *Mapping Strategic Thought*. John Wiley: Chichester, UK.
Intel Company. 2017. Intel Company Overview. Available at: https://www.intel.com/content/www/us/en/company-overview/company-overview.html [20 December 2017].
Ittner CD, Larcker DF, Rajan M V. 1997. The choice of performance measures in annual bonus contracts. *Accounting Review* **72**(2): 231–255.
Kabanoff B, Brown S. 2008. Knowledge structures of prospectors, analyzers, and defenders: Content, structure, stability, and performance. *Strategic Management Journal* **29**(2): 149–171.

Kaplan S. 2008. Cognition, capabilities, and incentives: Assessing firm response to the Fiber-Optic revolution. *Academy of Management Journal* **51**(4): 672–695.
Kotha S, Nair A. 1995. Strategy and environment as determinants of performance: evidence from the Japanese machine tool industry. *Strategic Management Journal* **16**(7): 497–518.
Kotha S, Vadlamani BL. 1995. Assessing generic strategies: An empirical investigation of two competing typologies in discrete manufacturing industries. *Strategic Management Journal* **16**(1): 75–83.
Kothari SP, Shu S, Wysocki PD. 2009. Do managers withhold bad news. *Journal of Accounting Research* **47**(1): 241–276.
Lawshe CH. 1975. A quantitative approach to content validity. *Personnel Psychology* **28**(4): 563–575.
Lev B, Sougiannis T. 1996. The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics* **21**(1): 107–138.
Lev B, Zarowin P. 1999. The Boundaries of Financial Reporting and How to Extend Them. *Journal of Accounting Research* **37**(2): 353.
Loughran T, Mcdonald B. 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* **54**(4): 1187–1230.
McCallum A, Nigam K. 1998. A comparison of event models for naive bayes text classification. *AAAI/ICML-98 Workshop on Learning for Text Categorization* : 41–48.
Miles RE, Snow CC. 1978. *Organizational Strategy, Structure, and Process*. McGraw-Hill: New York, NY.
Miller D, Friesen PH. 1986a. Porter's (1980) generic strategies and performance: An empirical examination with American data. Part I: Testing Porter. *Organization Studies* **7**(1): 37–55.
Miller D, Friesen PH. 1986b. Porter's (1980) generic strategies and performance: An empirical examination with American data. Part II: Performance implications. *Organization Studies* **7**(3): 255–261.
Mintzberg H. 1978. Patterns in Strategy Formation. *Management Science* **24**(9): 934–948.
Montemayor EF. 1996. Congruence between pay policy and competitive strategy in high-performing firms. *Journal of Management* **22**(6): 889–908.
Nag R, Hambrick DC, Chen M-J. 2007. What is strategic management, really? Inductive derivation of a consensus definition of the field. *Strategic Management Journal* **28**(9): 935–955.
Nair A, Filer L. 2003. Cointegration of firm strategies within groups: A long-run analysis of firm behavior in the Japanese steel industry. *Strategic Management Journal* **24**(2): 145–159.
Narver JC, Slater SF. 1990. The effect of a market orientation on business profitability. *Journal of Marketing* **54**(4): 20.
Nayyar PR. 1993. On the measurement of competitive strategy: Evidence from a large multiproduct U.S. firm. *Academy of Management Journal* **36**(6): 1652.
Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology* **88**(5): 879–903.
Porter ME. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Companies*. Free Press: New York, NY.
Porter ME. 1985. *Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press: New York, NY.
Porter ME. 1991. Towards a dynamic theory of strategy. *Strategic Management Journal* **12**(S2):

95–117.
Porter ME. 1996. What is strategy? *Harvard Business Review* **74**(6): 61–78.
Procter & Gamble. 2017. Core Strengths - Who We Are. Available at: http://us.pg.com/who-we-are/our-approach/core-strengths [24 September 2017].
Rajagopalan N. 1997. Strategic orientations, incentive plan adoptions, and firm performance: evidence from electric utility firms. *Strategic Management Journal* **18**(10): 761–785.
Rindova VP, Fombrun CJ. 1999. Constructing competitive advantage: the role of firm-constituent interactions. *Strategic Management Journal* **20**(8): 691–710.
Robinson RB, Pearce JA. 1988. Planned patterns of strategic behavior and their relationship to business-unit performance. *Strategic Management Journal* **9**(1): 43–60.
Rumelt RP, Schendel D, Teece D. 1991. Strategic management and economics. *Strategic Management Journal* **12**(S2): 5–29.
Securities and Exchange Commission. 2009. Form 10-K. Available at: https://www.sec.gov/fast-answers/answers-form10khtm.html [9 September 2017].
Securities and Exchange Commission. 2011. How to Read a 10-K. Available at: https://www.sec.gov/fast-answers/answersreada10khtm.html [9 September 2017].
Securities and Exchange Commission. 2016. Concept release on business and financial disclosure required by Regulation S-K.
Selling TI, Stickney CP. 1989. The effects of business environment and strategy on a firm's rate of return on assets. *Financial Analysts Journal* **45**(1): 43–52+68.
Shadish J, Cook T, Campbell D. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth Cengage learning: Boston, MA.
Shinkle GA, Kriauciunas AP, Hundley G. 2013a. Why pure strategies may be wrong for transition economy firms. *Strategic Management Journal* **34**(10): 1244–1254.
Shinkle GA, Kriauciunas AP, Hundley G. 2013b. Why pure strategies may be wrong for transition economy firms. *Strategic Management Journal* **34**(10): 1244–1254.
Slater SF, Olson EM. 2001. Marketing's contribution to the implementation of business strategy: An empirical analysis. *Strategic Management Journal* **22**(11): 1055–1067.
Snow CC, Hambrick DC. 1980. Measuring organizational strategies : Some theoretical and methodological problem. *Academy of Management Review* **5**(4): 527–538.
Spanos YE, Zaralis G, Lioukas S. 2004. Strategy and industry effects on profitability: Evidence from Greece. *Strategic Management Journal* **25**(2): 139–165.
Srivastava A. 2014. Why have measures of earnings quality changed over time? *Journal of Accounting and Economics*. Elsevier **57**(2–3): 196–217. Available at: http://dx.doi.org/10.1016/j.jacceco.2014.04.001.
Thompson SB. 2011. Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics* **99**(1): 1–10.
Thornhill S, White RE. 2007. Strategic purity: A multi-industry evaluation of pure vs. hybrid business strategies. *Strategic Management Journal* **28**(5): 553–561.
Treacy M, Wiersema F. 1995. *The discipline of market leaders: choose your customers, narrow your focus, dominate your market*. Addison-Wesley: Boston, MA.
Vorhies DW, Morgan RE, Autry CW. 2009. Product-market strategy and the marketing capabilities of the firm: Impact on market effectiveness and cash flow performance. *Strategic Management Journal* **30**(12): 1310–1334.
Wal-Mart. 2017. Our Business. Available at: https://corporate.walmart.com/our-story/our-business [4 October 2017].

Wang J, Verma A. 2012. Explaining organizational responsiveness to work-life balance issues: The role of business strategy and high-performance work systems. *Human Resource Management* **51**(3): 407–432.
Waterson M. 1980. Price-cost margins and successive market power. *Quarterly Journal of Economics* **94**(1): 135–150.
Weber RP. 1990. *Basic content analysis: Second edition*. Sage.
White RE. 1986. Generic business strategies, organizational context and performance: An empirical investigation. *Strategic Management Journal* **7**(3): 217–231.
WPP, Kantar Millward Brown. 2017. 2017 BrandZ Top 100 Global Brands. Available at: http://www.millwardbrown.com/brandz/top-global-brands/2017 [20 December 2017].
Zahra SA, Covin JG. 1993. Business strategy, technology policy and firm performance **14**(6): 451–478.
Zahra SA, Pearce JA. 1990. Research evidence on the Miles-Snow typology. *Journal of Management* **16**(4): 751–768.
Zatzick CD, Moliterno TP, Fang T. 2012. Strategic (MIS) fit: The implementation of TQM in manufacturing organizations. *Strategic Management Journal* **33**(11): 1321–1330.
Zott C, Amit R. 2008. The fit between product market strategy and business model: implications for firm performance. *Strategic Management Journal* **29**(1): 1–26.

Table 1
Summary statistics

Variable	N	Mean	Std. Dev.	P25	Median	P75
Ndiff	46,021	73.02	58.07	30	59	100
Ncost	46,021	8.38	7.66	3	6	12
Textlen	46,021	6869.15	4649.54	3753	5804	8649
Diff	46,021	11.02	5.75	6.72	10.80	14.60
Cost	46,021	1.27	0.93	0.6	1.07	1.72
R&D	46,021	0.12	0.47	0	0	0.07
SG&A	46,021	0.29	0.31	0.11	0.22	0.37
Adv	46,021	0.01	0.03	0	0	0.01
CE	46,021	6.04	7.06	1.87	3.70	7.27
ATO	46,021	1.19	0.79	0.64	1.03	1.53
GM	46,021	0.33	0.53	0.23	0.36	0.53
OM	46,021	0	0.79	0.05	0.11	0.19
OM_adj	46,021	0.13	0.35	0.07	0.14	0.25
ROA	46,021	0.01	0.17	-0.01	0.04	0.09
RET	42,450	0.16	0.60	-0.20	0.07	0.38
Ln(Mktcap)	49,057	6.11	1.99	4.71	6.06	7.40
Ln(Age)	49,057	2.82	0.71	2.30	2.83	3.37
BTM	49,057	0.56	0.52	0.25	0.45	0.74
Beta	39,296	1.24	0.79	0.69	1.12	1.64
Recency	46,021	1.69	1.06	1	1	2
Vol(ROA)	43,992	0.08	0.37	0.02	0.04	0.08
Revenue	46,021	1.19	0.8	0.64	1.03	1.53
Expense	46,021	1.2	0.82	0.65	1.02	1.51
Matching	46,021	0.73	0.21	0.66	0.77	0.84
SPI	46,021	0.29	0.46	0	0	1
Loss	46,021	0.28	0.45	0	0	1
CFO	45,995	0.07	0.17	0.03	0.09	0.14

This table reports the sample descriptive statistics and distribution. Variables are defined as in Appendix A.5. All continuous variables are winsorized for the extreme one percentiles.

Table 2
Stability of Strategy

Strategy	**Quintile 1**	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**	**N**
Panel A Diff						
Year 0					100	9,206
Year 1	0.08	0.32	1.06	12.95	85.59	7,542
Year 2	0.14	0.61	2.46	16.75	80.04	6,508
Year 3	0.18	0.82	3.74	19.85	75.41	5,702
Panel B Cost						
Year 0					100	9,207
Year 1	0.46	0.84	2.50	15.14	81.07	7,770
Year 2	0.97	1.82	4.26	19.20	73.75	7,357
Year 3	1.51	2.41	5.94	21.23	68.91	6,147

This table reports the percentage of observations fall into each quintile of *Strategy* in a given year.

Table 3
Strategy and operating activities

VARIABLES	Predicted sign	(I) $R\&D_t$	(II) $R\&D_{t+1}$	(III) Adv_t	(IV) Adv_{t+1}	(V) $SG\&A_t$	(VI) $SG\&A_{t+1}$
$\boldsymbol{Diff_t}$	**+**	**0.0037**	**0.0001**	**0.0009**	**0.0001**	**0.0092**	**0.0010**
		(0.000)	(0.095)	(0.000)	(0.000)	(0.000)	(0.000)
$\boldsymbol{Cost_t}$	**-**	**-0.0172**	**-0.0009**	**-0.0027**	**-0.0002**	**-0.0236**	**-0.0023**
		(0.000)	(0.021)	(0.000)	(0.000)	(0.000)	(0.000)
$Ln(Mktcap)_t$		-0.0020	-0.0001	0.0011	0.0001	-0.0142	-0.0010
		(0.337)	(0.644)	(0.000)	(0.001)	(0.000)	(0.001)
$Ln(Age)_t$		-0.0718	-0.0019	-0.0024	-0.0002	-0.0475	-0.0036
		(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
$R\&D_t$			0.9652				
			(0.000)				
Adv_t					0.9304		
					(0.000)		
$SG\&A_t$							0.9023
							(0.000)
Constant		0.2609	0.0086	0.0059	0.0007	0.3975	0.0344
		(0.000)	(0.001)	(0.037)	(0.012)	(0.000)	(0.000)
Adjusted R^2		0.1716	0.9039	0.0953	0.8872	0.2041	0.8561
Observations		46,021	38,408	46,021	38,408	46,021	38,408
Industry & Year FE		YES	YES	YES	YES	YES	YES

P-values are reported in parentheses and are based on clustering standard errors at the firm and year level.

Table 4
Strategy and operation efficiency

VARIABLES	Predicted sign	(I) CE_t	(II) CE_{t+1}	(III) ATO_t	(IV) ATO_{t+1}
***Diff*$_t$**	-	**-0.083**	**-0.031**	**0.003**	**0.000**
		(0.000)	(0.000)	(0.154)	(0.344)
***Cost*$_t$**	+	**0.056**	**0.011**	**0.094**	**0.006**
		(0.445)	(0.644)	(0.000)	(0.000)
Ln(Mktcap)$_t$		0.274	-0.024	-0.070	-0.009
		(0.000)	(0.017)	(0.000)	(0.000)
Ln(Age)$_t$		-1.307	-0.133	0.091	0.013
		(0.000)	(0.009)	(0.000)	(0.000)
CE_t			0.666		
			(0.000)		
ATO_t					0.937
					(0.000)
Constant		8.340	2.318	1.185	0.049
		(0.000)	(0.000)	(0.000)	(0.001)
Adjusted R^2		0.226	0.605	0.305	0.939
Observations		46,021	38,408	46,021	38,408
Industry & Year FE		YES	YES	YES	YES

P-values are reported in parentheses and are based on clustering standard errors at the firm and year level.

Table 5
Strategy and margins

VARIABLES	Predicted sign	(I) GM_t	(II) GM_{t+1}	(III) OM_t	(IV) OM_{t+1}	(V) OM_adj_t	(VI) OM_adj_{t+1}
***Diff*$_t$**	+	**0.006**	**0.001**	**-0.004**	**-0.000**	**-0.000**	**0.000**
		(0.000)	(0.001)	(0.019)	(0.559)	(0.557)	(0.764)
***Cost*$_t$**	-	**-0.018**	**-0.001**	**0.013**	**0.001**	**-0.007**	**-0.001**
		(0.000)	(0.045)	(0.008)	(0.161)	(0.003)	(0.092)
Ln(Mktcap)$_t$		0.024	0.001	0.044	0.001	0.040	0.005
		(0.000)	(0.003)	(0.000)	(0.024)	(0.000)	(0.000)
Ln(Age)$_t$		0.017	0.001	0.099	0.005	0.016	0.001
		(0.023)	(0.629)	(0.000)	(0.038)	(0.029)	(0.479)
GM_t			0.904				
			(0.000)				
OM_t					0.941		
					(0.000)		
OM_adj_t							0.845
							(0.000)
Constant		0.033	-0.006	-0.558	-0.045	-0.220	-0.039
		(0.298)	(0.362)	(0.000)	(0.000)	(0.000)	(0.000)
Adjusted R^2		0.070	0.791	0.117	0.844	0.088	0.697
Observations		46,021	38,408	46,021	38,408	46,021	38,408
Industry & Year FE		YES	YES	YES	YES	YES	YES

P-values are reported in parentheses and are based on clustering standard errors at the firm and year level.

Table 6
Strategy, performance, and uncertainty

VARIABLES	(I) ROA_t	(II) RET_t
***Diff*$_t$**	**-0.001**	**0.001**
	(0.018)	(0.247)
***Cost*$_t$**	**0.000**	**-0.001**
	(0.951)	(0.732)
$Ln(Age)_t$	0.024	
	(0.000)	
$Ln(Mktcap)_t$	0.019	-0.012
	(0.000)	(0.011)
BTM_{t-1}		0.100
		(0.002)
$Beta_{t-1}$		0.010
		(0.676)
RET_{t-1}		-0.034
		(0.130)
Constant	-0.195	-0.029
	(0.000)	(0.669)
Adjusted R^2	0.137	0.160
Observations	46,021	32,987
Industry &Year FE	YES	YES

P-values are reported in parentheses and are based on clustering standard errors at the firm and year level.

Table 7
Strategy and earnings properties over time

VARIABLES	(I) *Ratio*	(II) *R&D*	(III) *Matching*	(IV) *Vol(ROA)*
Recency	0.001	0.062	-0.041	0.023
	(0.010)	(0.000)	(0.000)	(0.000)
Fyear	-0.000	-0.004	0.002	-0.001
	(0.001)	(0.000)	(0.000)	(0.002)
Constant	1.455	7.556	-3.582	1.616
	(0.000)	(0.000)	(0.000)	(0.001)
Adjusted R^2	0.000	0.016	0.037	0.003
Observations	46,021	46,021	46,021	43,992

P-values are reported in parentheses and are based on robust standard errors.

Table 8
Strategy and earnings properties

VARIABLES	(I) $Revenue_t$	(II) $Revenue_t$	(III) *Vol(ROA)*	(IV) *Vol(ROA)*
$Expense_{t-1}$	-0.129	-0.023		
	(0.130)	(0.730)		
$Expense_t$	1.230	1.148		
	(0.000)	(0.000)		
$Expense_{t+1}$	-0.008	-0.081		
	(0.914)	(0.103)		
Ratio$_t$	0.158	0.072	**0.033**	**0.030**
	(0.000)	(0.000)	(0.000)	(0.000)
$Ratio_t$* $Expense_{t-1}$	0.365	0.102		
	(0.000)	(0.157)		
Ratio$_t$* Expense$_t$	**-0.534**	**-0.157**		
	(0.000)	(0.086)		
$Ratio_t$* $Expense_{t+1}$	0.029	0.010		
	(0.746)	(0.845)		
$Recency_t$		0.004		0.005
		(0.068)		(0.000)
$Recency_t$* $Expense_{t-1}$		0.008		
		(0.318)		
$Recency_t$* $Expense_t$		-0.020		
		(0.042)		
$Recency_t$* $Expense_{t+1}$		0.004		
		(0.546)		
$R\&D_t$		0.066		0.017
		(0.000)		(0.000)
$R\&D_t$* $Expense_{t-1}$		-0.051		
		(0.000)		
$R\&D_t$* $Expense_t$		-0.147		
		(0.000)		
$R\&D_t$* $Expense_{t+1}$		-0.030		
		(0.004)		
SPI_t		-0.023		0.011
		(0.000)		(0.000)
SPI_t* $Expense_{t-1}$		0.072		
		(0.000)		
SPI_t* $Expense_t$		-0.102		
		(0.000)		
SPI_t* $Expense_{t+1}$		0.029		
		(0.022)		
$Ln(Mktcap)_t$		0.002		-0.006
		(0.002)		(0.000)
$Ln(Mktcap)_t$* $Expense_{t-1}$		-0.004		
		(0.256)		
$Ln(Mktcap)_t$* $Expense_t$		-0.007		
		(0.110)		
$Ln(Mktcap)_t$* $Expense_{t+1}$		0.010		
		(0.001)		

$Loss_t$		-0.108		0.033
		(0.000)		(0.000)
$Loss_t$ * $Expense_{t-1}$		0.062		
		(0.000)		
$Loss_t$ * $Expense_t$		-0.153		
		(0.000)		
$Loss_t$* $Expense_{t+1}$		0.083		
		(0.000)		
CFO_t		0.397		-0.108
		(0.000)		(0.000)
CFO_t* $Expense_{t-1}$		-0.046		
		(0.530)		
CFO_t* $Expense_t$		0.265		
		(0.005)		
CFO_t* $Expense_{t+1}$		-0.078		
		(0.200)		
Adjusted R^2	0.947	0.986	0.091	0.270
Observations	42,524	42,500	43,992	43,968
Industry &Year FE	YES	YES	YES	YES

P-values are reported in parentheses and are based on clustering standard errors at the firm and year level.

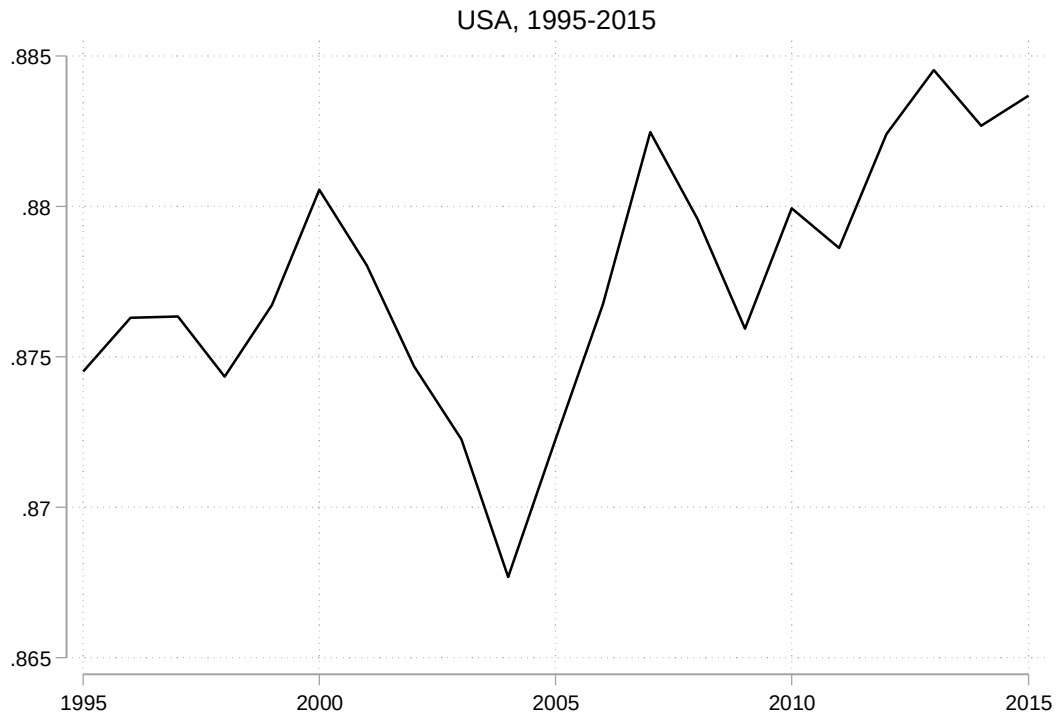

Figure 1. Changes in average strategy

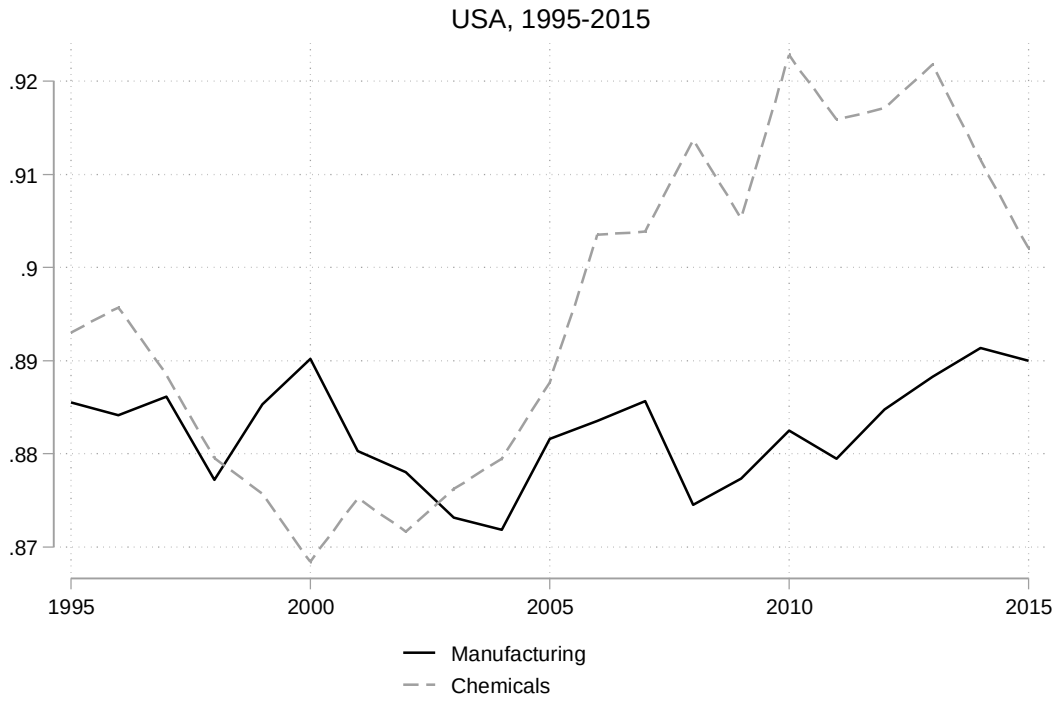

Figure 2. Changes in average strategy by industry

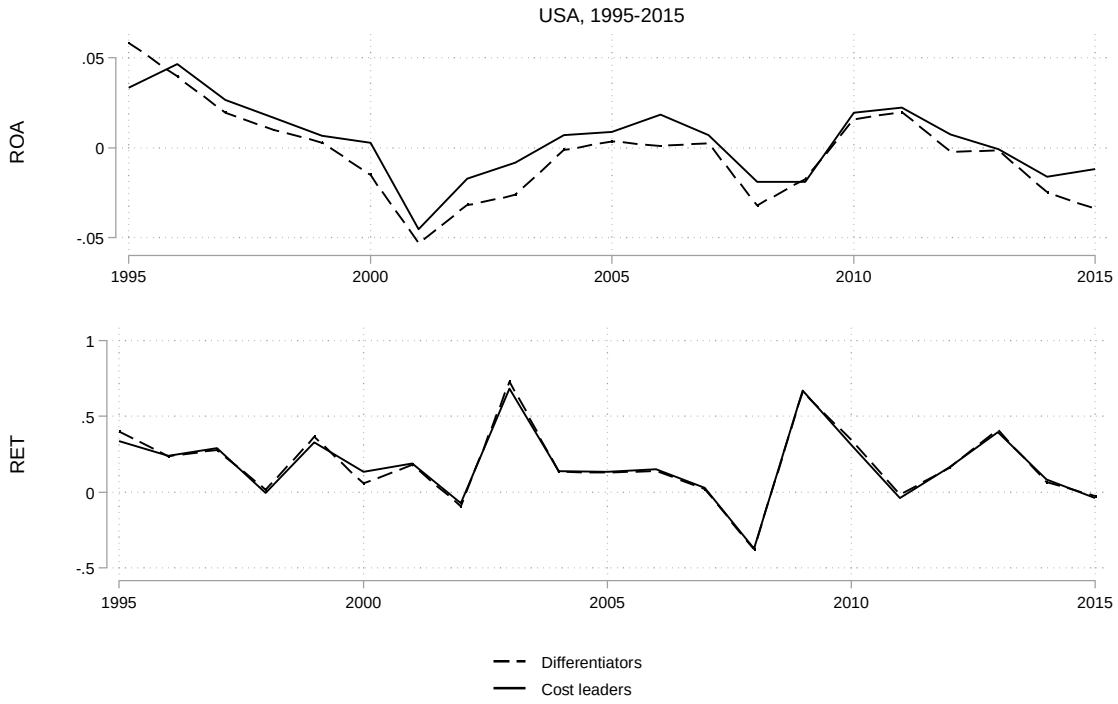

Figure 3. Changes in average performance

Methodological Appendix A.1
Keyword lists

A.1.1 Differentiation
I. Strategic positioning

1) differenti*
   (Campbell-Hunt, 2000; Porter, 1985; Thornhill and White, 2007)
2) unique*
   (Dess and Davis, 1984; Hambrick, 1983a; Kotha and Vadlamani, 1995; Porter, 1985; White, 1986; Zahra and Covin, 1993)
3) superior*
   (Campbell-Hunt, 2000)
4) premium*
   (Nayyar, 1993; Porter, 1985)
5) high* price*
   (Dess and Davis, 1984; Nayyar, 1993; White, 1986; Zahra and Covin, 1993)
6) high* margin*
   (Porter, 1985)
7) high* end*
8) inelasticity*
   (Miller and Friesen, 1986a, 1986b)

II. Product innovation

9) new* product*
   (Nayyar, 1993; Porter, 1985; Zatzick *et al.*, 2012)
10) innovat*
    (Campbell-Hunt, 2000; Miller and Friesen, 1986a, 1986b; Shinkle, Kriauciunas, and Hundley, 2013b; Thornhill and White, 2007; White, 1986; Zatzick *et al.*, 2012)
11) creativ*
    (White, 1986)
12) research and development
13) R&D
    (Zatzick *et al.*, 2012)
14) techni*
    (White, 1986; Thornhill and White, 2007; Zatzick *et al.*, 2012)
15) technolog*
    (Miller and Friesen, 1986a, 1986b)
16) patent*
17) quality*
    (Campbell-Hunt, 2000; Miller and Friesen, 1986a, 1986b; Nayyar, 1993; Porter, 1985; Thornhill and White, 2007; White, 1986)
18) reliab*
    (White, 1986)
19) durable*

III. Customer intimacy

20) marketing*

(Miller and Friesen, 1986a, 1986b; Nayyar, 1993; Porter, 1985; White, 1986; Zahra and Covin, 1993)

21) advertis*
(Nayyar, 1993; Kotha and Vadlamani, 1995; Zahra and Covin, 1993)

22) brand*
(Hambrick, 1983a; Miller and Friesen, 1986a, 1986b; Nayyar, 1993; Porter, 1985; Zahra and Covin, 1993)

23) reputation*
(Nayyar, 1993)

24) trademark*

25) customer* service*
(Hambrick, 1983a; Miller and Friesen, 1986a, 1986b; Nayyar, 1993; Porter, 1985)

26) consumer* service*
(Nayyar, 1993)

27) customer* need*
(White, 1986)

28) sales support*
(White, 1986)

29) post-purchase service*
(White, 1986)

30) customer preference*
(Campbell-Hunt, 2000; Zatzick *et al.*, 2012)

31) c preference*
(Campbell-Hunt, 2000; Zatzick *et al.*, 2012)

32) loyalty*
(Miller and Friesen, 1986a, 1986b)

33) customiz*

34) tailor*
(White, 1986)

35) personaliz*

36) responsive*
(Shinkle *et al.*, 2013; Thornhill and White, 2007; Zatzick *et al.*, 2012)

IV. Human resources (Porter, 1985)

37) talent*

38) train*

39) skill*

40) intellectual* propert*

41) human capital*

A.1.2 Cost leadership

I. Strategic positioning

1) cost leader*
(Porter, 1985)

2) low* pric*

3) low* cost*

(Dess and Davis, 1984; Miller and Friesen, 1986a, 1986b; Porter, 1985; White, 1986)

4) cost advantage*
   (Porter, 1985)
5) competitive pric*
   (White, 1986; Zahra and Covin, 1993)
6) aggressive pric*
   (Miller and Friesen, 1986a, 1986b)

II. Cost control (Dess and Davis, 1984; Hambrick, 1983a; Kotha and Vadlamani, 1995; Thornhill and White, 2007)

7) control* cost* or expense* or control* overhead*
   (Hambrick, 1983a; Miller and Friesen, 1986a, 1986b)
8) minimiz* cost* or expense* or overhead*
   (Miller and Friesen, 1986a, 1986b)
9) reduc* cost* or expense* or overhead*
   (Campbell-Hunt, 2000; Miller and Friesen, 1986a, 1986b; Nayyar, 1993)
10) sav* cost*
11) improv* cost*
12) cut* cost*or expense*or overhead*
13) decrease* cost* or expense* or overhead*
14) monitor* cost* or expense* or overhead*

III. Operation efficiency (Hambrick, 1983a; Kotha and Vadlamani, 1995)

15) efficien*
   (Miller and Friesen, 1986a, 1986b; Nayyar, 1993; Kotha and Vadlamani, 1995; Shinkle *et al.*, 2013; White, 1986; Zahra and Covin, 1993; Zatzick *et al.*, 2012)
16) high* yield*
   (White, 1986)
17) process* improvement*
   (Nayyar, 1993)
18) asset* utilization*
19) capacity* utilization*
   (Zahra and Covin, 1993)

IV. Economies of scale

20) (broad) scope*
   (Porter, 1985)
21) (large) scale*
   (Campbell-Hunt, 2000; Miller and Friesen, 1986a, 1986b; Porter, 1985)
22) breath*
   (Porter, 1985)
23) broad
24) mass
25) high* volume*
26) large* volume*

* denotes wildcards.

Appendix A.2
Illustrative examples

In these illustrative examples, we show the occurrences of the keywords we capture and count using Python script. The first example contains sentences from the 2016 10-K Item 1 of Apple Inc. and the second example contains sentences from the 2016 10-K Item 1 of **Wal-Mart Stores Inc.** As noted, we search with flexibility. Therefore, we search for the occurrences of single keywords, two-word phrases, and verb-noun combinations. We highlight the occurrences of differentiation keywords in **Bold** and the occurrences of cost leadership keywords in *Italic*.

1) Apple Inc. 2016

"…The Company is committed to bringing the best user experience to its customers through its **innovative** hardware, software and services. The Company's business strategy leverages its **unique** ability to design and develop its own operating systems, hardware, application software and services to provide its customers products and solutions with **innovative** design, **superior** ease-of-use and seamless integration…The Company believes ongoing investment in **research and development** ("**R&D**"), **marketing** and **advertising** is critical to the development and sale of **innovative** products and **technologies**…The Company believes a **high-quality** buying experience with knowledgeable salespersons who can convey the value of the Company's products and services greatly enhances its ability to attract and retain customers. Therefore, the Company's strategy also includes building and expanding its own retail and online stores and its third-party distribution network to effectively reach more customers and provide them with a **high-quality** sales and post-sales support experience…"

For the whole 2016 10-K Item 1 for Apple Inc., our program shows cost leadership keywords occur twice and differentiation keywords occur 69 times.

2) Wal-Mart Stores Inc.

"Our strategy is to lead on price, **differentiate** on access, be competitive on assortment and deliver a great experience. Leading on price is designed to earn the trust of our customers every day by providing a *broad* assortment of **quality** merchandise and services at everyday *low prices* ("EDLP"). EDLP is our pricing philosophy under which we price items at a *low price* every day so our customers trust that our prices will not change under frequent promotional activity. Price leadership is core to who we are. Everyday *low cost* ("EDLC") is our commitment to *control expenses* so those *cost savings* can be passed along to our customers. Our digital and physical presence provides customers access to our *broad* assortment anytime and anywhere. We strive to give our customers and members a great digital and physical shopping experience."

For the whole 2016 10-K Item 1 for Wal-Mart Stores Inc., our program shows cost leadership keywords occur 22 times and differentiation keywords occur 26 times.

Appendix A.3
Sample selection

	No. of obs
All available 10-K fillings on EDGAR database released from 1995 to 2016	101,808
Observations with extracted textual data from 1995 to 2016 (Calendar year)	94,981
All firm year observations with financial data required for our dependent variables and control variables from Compustat from 1994 to 2015 (Fiscal year)	77,709
Less:	
Observations in financial and utility industry	(18,984)
Observations with stock price smaller than $1	(6,103)
Observations with sales change greater than 100% or smaller than -50%	(5,653)
Observations in fiscal year 1994	(579)
Observations in fiscal year 2016	(338)
Observations with sufficient data for construct validity tests	46,032
Less:	
Observations without sufficient stock return data from CRSP	(3,582)
Observations for performance and uncertainty tests	42,450

Appendix A.4
Examples of differentiators and cost leaders

We compare our strategy classification to strategic information from corporate websites and annual reports. We picked the examples from Fortune 100 firms based on the average value of *Diff* and *Cost* from 2010 to 2015. The top 3 differentiation-oriented firms and cost-leadership-oriented firms are selected based on the average value of *Diff* and *Cost*, respectively.

Top 3 differentiation-oriented firms:

1. E. I. DU PONT DE NEMOURS AND CO (*Diff=20.12*)

E. I. DU PONT DE NEMOURS AND CO (commonly referred to as DuPont) is a conglomerate company and one of the largest chemical producers. They aim to be a world leader in science and innovation across a broad range of disciplines (DuPont, 2017a). DuPont manages to turn scientific breakthroughs into commercial breakouts and claims to be the world's most innovative science company (DuPont, 2017b).

2. INTEL CORP (*Diff*=19.89)

Intel Corporation is a multinational technology company headquartered in the Silicon Valley. Their brands are strongly recognized and ranked as #44 among the 2017 BrandZ Top 100 most valuable global brands (WPP and Kantar Millward Brown, 2017). In addition, Intel embraces technology intensity and "invents at the boundaries of technology to make amazing experiences possible for business and society" (Intel Company, 2017).

3. PROCTER & GAMBLE CO (*Diff*=18.17)

The Procter & Gamble Company, as a global manufacturer of household and industrial consumer products, builds competitive advantages around product innovation and customer intimacy. They recognize "six core strengths: Consumer Understanding, Innovation, Branding, Go-to-Market Capabilities, Scale and Productivity" and "is built to lead in these areas" (Procter & Gamble, 2017).

Top 3 cost-leadership-oriented firms:

1. AMAZON.COM INC (*Cost*=3.51)

Amazon is the world's largest internet company by revenue. As mentioned in their annual report, Amazon "strives to offer our customers the lowest price possible through low everyday product pricing and shipping offers" and aims to improve operational efficiency in order to make the low prices possible (Amazon.com Inc, 2017).

2. WAL-MART STORES INC (*Cost*=3.41)

Wal-Mart is another leading retailer in the U.S. It is known be an example of cost leaders while Wal-Mart is also dedicated to meet customers' needs. They provide customers with "every day low prices on a broad assortment", which facilitates a convenient one-stop shopping experience. As they claimed, "Every Day Low Price (EDLP) is the cornerstone of our strategy, and our price focus has never been stronger"(Wal-Mart, 2017).

3. AMERISOURCEBERGEN CORP (*Cost*=3.27)

AmerisourceBergen Corporation is one of the largest global pharmaceutical sourcing and distribution services companies. Perhaps due to the economies of scale, AmerisourceBergen has "one of the lowest cost operating structure among all pharmaceutical distributors" (AmerisourceBergen Corporation, 2016).

Appendix A.5
Variable definition

Variable	Definition
Textual Variables	
Ndiff	Number of differentiation-related words in 10-K Item 1
Ncost	Number of cost-leadership-related words in 10-K Item 1
Textlen	Total number of words in 10-K Item 1
Diff	Number of differentiation-related words per 1,000 total words in 10-K Item 1
Cost	Number of cost leadership-related words per 1,000 total words in 10-K Item 1
Ratio	The relative importance of strategy, defined as *Diff*/(*Diff*+*Cost*)
Other Variables	
R&D	Ratio of research and development expenses to sales
Adv	Ratio of marketing expense to sales
SG&A	Ratio of selling, general and administrative (SG&A) expenses to sales
CE	Ratio of capital expenditure (times 100) to beginning-of-year total assets
ATO	Asset turnover, defined as sales revenue scaled by average total assets
GM	Gross margin, defined as revenue minus cost of goods sold (COGS) scaled by sales revenue
OM	Operating margin, defined as revenue minus COGS and SG&A scaled by sales revenue
OM_adj	Adjusted operating margin, defined as revenue minus COGS and SG&A (excluding R&D and marketing expenses) scaled by sales revenue
ROA	Return on Assets, defined as the earnings before extraordinary items scaled by average total assets
Vol(ROA)	Five-year rolling standard deviation of annual return on assets

RET	Total return, defined as 12-month buy-and-hold stock return over the fiscal year
Vol(RET)	Five-year rolling standard deviation of monthly buy-and-hold stock return
Ln(Mktcap)	The Natural logarithm of market capitalization at the beginning of the year, defined as shares outstanding times closing stock price
Ln(Age)	The Natural logarithm of firm age as the number of years since the firm's initial appearance in Compustat annual file with non-missing financial information
BTM	Book-to-market ratio, defined as the book value of common stock scaled by market capitalization
Beta	The rolling slope coefficient from a model regressing monthly returns on the value weighted market index for last five fiscal periods (minimum of 24 observations are required)
Recency	Indicators for different cohorts of newly listed firms. Values of 1,2,3,4, and 5 indicate firms listed before 1995, firms listed between 1995 and 2000, firms listed between 2000 and 2005, firms listed between 2006 and 2010, and firms listed between 2010 and 2015
Vol(ROA)	Four-year rolling standard deviation of ROA from year t-2 to year t+1
Revenue	Sales revenue scaled by average total assets
Expense	Sales revenue mines earnings before extraordinary item, scaled by average total assets
Matching	Annual estimate (ϑ_2) of the following cross-sectional regression for each cohort-year $Revenue_{i,t} = \vartheta_1 + \vartheta_2 * Expense_{i,t-1} + \vartheta_3 * Expense_{i,t} + \vartheta_4 * Expense_{i,t+1} + \ni_{i,t}$
SPI	Indicator variable for large special item (greater than 1% of the total assets)
Loss	Indicator variable for negative earnings
CFO	Operating cash flow scaled by average total assets

REFERENCES FOR APPENDIXES

Abrahamson E, Hambrick DC. 1997. Attentional homogeneity in industries: The effect of discretion. *Journal of Organizational Behavior* **18**(1997): 513–532.

Ahuja G, Katila R. 2004. Where do resources come from? The role of idiosyncratic situations. *Strategic Management Journal* **25**(89): 887–907.

Amazon.com Inc. 2017. Form 10-K. Available at: https://www.sec.gov/Archives/edgar/data/1018724/000101872417000011/amzn-20161231x10k.htm [20 December 2017].

AmerisourceBergen Corporation. 2016. Form 10-K. Available at: https://www.sec.gov/Archives/edgar/data/1140859/000104746915008939/a2226704z10-k.htm [7 October 2017].

Armstrong JS, Overton TS. 1977. Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research* **14**(3): 396.

Arthur JB. 1992. The link between business strategy and industrial relations systems in American steel. *ILR Review* **45**(3): 488–506.

Ball R, Brown P. 1968. An empirical evaluation of income numbers. *Journal of Accounting Research* **6**(2): 159–178.

Ball R, Gerakos J, Linnainmaa JT, Nikolaev VV. 2015. Deflating profitability. *Journal of Financial Economics* **117**(2): 225–248.

Banker RD, Basu S, Byzalov D, Chen JYS. 2016. The confounding effect of cost stickiness on conservatism estimates. *Journal of Accounting and Economics* **61**(1): 203–220.

Banker RD, Hu N, Luftman J, Pavlou PA. 2010. CIO reporting structure, strategic positioning, and firm performance: To whom should the CIO report? *MIS Quarterly* **35**(2): 487–504.

Bao Y, Datta A. 2014. Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science* **60**(6): 1371–1391.

Bentley KA, Omer TC, Sharp NY. 2013. Business strategy, financial reporting irregularities, and audit effort. *Contemporary Accounting Research* **30**(2): 780–817.

Berman SL, Wicks AC, Kotha S, Jones TM. 1999. Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance. *Academy of Management Journal* **42**(5): 488–506.

Campbell-Hunt C. 2000. What have we learned about generic competitive strategy? A meta-analysis. *Strategic Management Journal* **21**(2): 127–154.

Carhart MM. 1997. On persistence in mutual fund performance. *The Journal of Finance* **52**(1): 57.

Collins DW, Maydew EL, Weiss IS. 1997. Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics* **24**(1): 39–67.

Cronbach LJ. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* **16**(3): 297–334.

D'Aspremont C, Gabszewicz JJ, Thisse JF. 1979. On Hotelling's 'Stability in competition'. *Econometrica* **47**(5): 1145–1150.

David JS, Hwang Y, Pei BKW, Reneau JH. 2002. The performance effects of congruence between product competitive strategies and purchasing management design. *Management Science* **48**(7): 866–885.

Dechow P, Ge W, Schrand C. 2010. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*. Elsevier

50(2–3): 344–401.
Desarbo WS, Di Benedetto CA, Song M, Sinha I. 2005. Revisiting the Miles and Snow strategic framework: Uncovering interrelationships between strategic types, capabilities, environmental uncertainty, and firm performance. *Strategic Management Journal* **26**(1): 47–74.
Dess GG, Davis PS. 1984. Porter's (1980) generic gtrategies as determinants of strategic group membership and organizational performance. *Academy of Management Journal* **27**(3): 467–488.
Dichev ID, Tang VW. 2008. Matching and the changing properties of accounting earnings over the last 40 years. *Accounting Review* **83**(6): 1425–1460.
Donelson DC, Jennings R, Mclnnis J. 2011. Changes over time in the revenue-expense relation: Accounting or economics? *Accounting Review* **86**(3): 945–974.
DuPont. 2017a. Form 10-K. Available at: https://www.sec.gov/Archives/edgar/data/30554/000003055417000004/dd-12312016x10k.htm [20 December 2017].
DuPont. 2017b. Our Approach. Available at: http://www.dupont.com/corporate-functions/our-approach.html [20 December 2017].
Fama EF, MacBeth JD. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* **81**(3): 607–636.
Feldman R, Govindaraj S, Livnat J, Segal B. 2010. Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies* **15**(4): 915–953.
Francis J, Schipper K. 1999. Have Financial Statements Lost Their Relevance ? *Journal of Accounting Research* **37**(2): 319–352.
Ghemawat P. 1991. *Commitment: The Dynamic of Strategy*. Free Press: New York, NY.
Givoly D, Hayn C. 2000. The changing time-series properties of earnings,cash flows and accruals:Has financial reporting become more conservative? *Journal of Accounting & Economics* **29**: 287–320.
Hambrick DC. 1983a. High profit strategies in mature capital goods industries: A contingency approach. *Academy of Management Journal* **26**(4): 687–707.
Hambrick DC. 1983b. Some tests of the effectiveness and functional attributes of Miles and Snow's strategic types. *Academy of Management Journal* **26**(1): 5–26.
Henderson R. 1993. Underinvestment and incompetence as responses to radical innovation: Evidence from the photolithographic alignment equipment industry. *RAND Journal of Economics* **24**(2): 248–270.
Hoberg G, Phillips G. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* **124**(5): 1423–1465.
Hotelling H. 1929. Stability in competition. *The Economic Journal* **39**(153): 41.
Huff AS. 1990. *Mapping Strategic Thought*. John Wiley: Chichester, UK.
Intel Company. 2017. Intel Company Overview. Available at: https://www.intel.com/content/www/us/en/company-overview/company-overview.html [20 December 2017].
Ittner CD, Larcker DF, Rajan M V. 1997. The choice of performance measures in annual bonus contracts. *Accounting Review* **72**(2): 231–255.
Kabanoff B, Brown S. 2008. Knowledge structures of prospectors, analyzers, and defenders: Content, structure, stability, and performance. *Strategic Management Journal* **29**(2): 149–171.

Kaplan S. 2008. Cognition, capabilities, and incentives: Assessing firm response to the Fiber-Optic revolution. *Academy of Management Journal* **51**(4): 672–695.
Kotha S, Nair A. 1995. Strategy and environment as determinants of performance: evidence from the Japanese machine tool industry. *Strategic Management Journal* **16**(7): 497–518.
Kotha S, Vadlamani BL. 1995. Assessing generic strategies: An empirical investigation of two competing typologies in discrete manufacturing industries. *Strategic Management Journal* **16**(1): 75–83.
Kothari SP, Shu S, Wysocki PD. 2009. Do managers withhold bad news. *Journal of Accounting Research* **47**(1): 241–276.
Lawshe CH. 1975. A quantitative approach to content validity. *Personnel Psychology* **28**(4): 563–575.
Lev B, Sougiannis T. 1996. The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics* **21**(1): 107–138.
Lev B, Zarowin P. 1999. The Boundaries of Financial Reporting and How to Extend Them. *Journal of Accounting Research* **37**(2): 353.
Loughran T, Mcdonald B. 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* **54**(4): 1187–1230.
McCallum A, Nigam K. 1998. A comparison of event models for naive bayes text classification. *AAAI/ICML-98 Workshop on Learning for Text Categorization* : 41–48.
Miles RE, Snow CC. 1978. *Organizational Strategy, Structure, and Process*. McGraw-Hill: New York, NY.
Miller D, Friesen PH. 1986a. Porter's (1980) generic strategies and performance: An empirical examination with American data. Part I: Testing Porter. *Organization Studies* **7**(1): 37–55.
Miller D, Friesen PH. 1986b. Porter's (1980) generic strategies and performance: An empirical examination with American data. Part II: Performance implications. *Organization Studies* **7**(3): 255–261.
Mintzberg H. 1978. Patterns in Strategy Formation. *Management Science* **24**(9): 934–948.
Montemayor EF. 1996. Congruence between pay policy and competitive strategy in high-performing firms. *Journal of Management* **22**(6): 889–908.
Nag R, Hambrick DC, Chen M-J. 2007. What is strategic management, really? Inductive derivation of a consensus definition of the field. *Strategic Management Journal* **28**(9): 935–955.
Nair A, Filer L. 2003. Cointegration of firm strategies within groups: A long-run analysis of firm behavior in the Japanese steel industry. *Strategic Management Journal* **24**(2): 145–159.
Narver JC, Slater SF. 1990. The effect of a market orientation on business profitability. *Journal of Marketing* **54**(4): 20.
Nayyar PR. 1993. On the measurement of competitive strategy: Evidence from a large multiproduct U.S. firm. *Academy of Management Journal* **36**(6): 1652.
Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology* **88**(5): 879–903.
Porter ME. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Companies*. Free Press: New York, NY.
Porter ME. 1985. *Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press: New York, NY.
Porter ME. 1991. Towards a dynamic theory of strategy. *Strategic Management Journal* **12**(S2):

95–117.
Porter ME. 1996. What is strategy? *Harvard Business Review* **74**(6): 61–78.
Procter & Gamble. 2017. Core Strengths - Who We Are. Available at: http://us.pg.com/who-we-are/our-approach/core-strengths [24 September 2017].
Rajagopalan N. 1997. Strategic orientations, incentive plan adoptions, and firm performance: evidence from electric utility firms. *Strategic Management Journal* **18**(10): 761–785.
Rindova VP, Fombrun CJ. 1999. Constructing competitive advantage: the role of firm-constituent interactions. *Strategic Management Journal* **20**(8): 691–710.
Robinson RB, Pearce JA. 1988. Planned patterns of strategic behavior and their relationship to business-unit performance. *Strategic Management Journal* **9**(1): 43–60.
Rumelt RP, Schendel D, Teece D. 1991. Strategic management and economics. *Strategic Management Journal* **12**(S2): 5–29.
Securities and Exchange Commission. 2009. Form 10-K. Available at: https://www.sec.gov/fast-answers/answers-form10khtm.html [9 September 2017].
Securities and Exchange Commission. 2011. How to Read a 10-K. Available at: https://www.sec.gov/fast-answers/answersreada10khtm.html [9 September 2017].
Securities and Exchange Commission. 2016. Concept release on business and financial disclosure required by Regulation S-K.
Selling TI, Stickney CP. 1989. The effects of business environment and strategy on a firm's rate of return on assets. *Financial Analysts Journal* **45**(1): 43–52+68.
Shadish J, Cook T, Campbell D. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth Cengage learning: Boston, MA.
Shinkle GA, Kriauciunas AP, Hundley G. 2013a. Why pure strategies may be wrong for transition economy firms. *Strategic Management Journal* **34**(10): 1244–1254.
Shinkle GA, Kriauciunas AP, Hundley G. 2013b. Why pure strategies may be wrong for transition economy firms. *Strategic Management Journal* **34**(10): 1244–1254.
Slater SF, Olson EM. 2001. Marketing's contribution to the implementation of business strategy: An empirical analysis. *Strategic Management Journal* **22**(11): 1055–1067.
Snow CC, Hambrick DC. 1980. Measuring organizational strategies : Some theoretical and methodological problem. *Academy of Management Review* **5**(4): 527–538.
Spanos YE, Zaralis G, Lioukas S. 2004. Strategy and industry effects on profitability: Evidence from Greece. *Strategic Management Journal* **25**(2): 139–165.
Srivastava A. 2014. Why have measures of earnings quality changed over time? *Journal of Accounting and Economics*. Elsevier **57**(2–3): 196–217. Available at: http://dx.doi.org/10.1016/j.jacceco.2014.04.001.
Thompson SB. 2011. Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics* **99**(1): 1–10.
Thornhill S, White RE. 2007. Strategic purity: A multi-industry evaluation of pure vs. hybrid business strategies. *Strategic Management Journal* **28**(5): 553–561.
Treacy M, Wiersema F. 1995. *The discipline of market leaders: choose your customers, narrow your focus, dominate your market*. Addison-Wesley: Boston, MA.
Vorhies DW, Morgan RE, Autry CW. 2009. Product-market strategy and the marketing capabilities of the firm: Impact on market effectiveness and cash flow performance. *Strategic Management Journal* **30**(12): 1310–1334.
Wal-Mart. 2017. Our Business. Available at: https://corporate.walmart.com/our-story/our-business [4 October 2017].

Wang J, Verma A. 2012. Explaining organizational responsiveness to work-life balance issues: The role of business strategy and high-performance work systems. *Human Resource Management* **51**(3): 407–432.
Waterson M. 1980. Price-cost margins and successive market power. *Quarterly Journal of Economics* **94**(1): 135–150.
Weber RP. 1990. *Basic content analysis: Second edition*. Sage.
White RE. 1986. Generic business strategies, organizational context and performance: An empirical investigation. *Strategic Management Journal* **7**(3): 217–231.
WPP, Kantar Millward Brown. 2017. 2017 BrandZ Top 100 Global Brands. Available at: http://www.millwardbrown.com/brandz/top-global-brands/2017 [20 December 2017].
Zahra SA, Covin JG. 1993. Business strategy, technology policy and firm performance **14**(6): 451–478.
Zahra SA, Pearce JA. 1990. Research evidence on the Miles-Snow typology. *Journal of Management* **16**(4): 751–768.
Zatzick CD, Moliterno TP, Fang T. 2012. Strategic (MIS) fit: The implementation of TQM in manufacturing organizations. *Strategic Management Journal* **33**(11): 1321–1330.
Zott C, Amit R. 2008. The fit between product market strategy and business model: implications for firm performance. *Strategic Management Journal* **29**(1): 1–26.

[i] We download the 10-K filings released from 1995 to 2016 on EDGAR database. Normally, firms release 10-K filings after the fiscal year end. Thus, 10-K filings released from 1995 to 2016 correspond to fiscal years between 1994 and 2016.
[ii] We will use this endnote to list the website where our textual strategy index is made freely available.
[iii] We find that some of the firms only make small changes in the Item 1 of 10-K, which thus can lead to high stability for those firms mechanically. However, as long as the 10-K still truthfully represent the business conditions of the firm, our measure is not biased.
[iv] Here, Year 0 refers to the fiscal year when the observations are sorted into the fifth quintile and Year 1 refers to the subsequent fiscal year
[v] We find consistent results if we divide the sample into quartiles.