PEAD.txt: Post-Earnings-Announcement Drift Using Text

Vitaly Meursault, Pierre Jinghong Liang, Bryan R. Routledge,
Madeline Marco Scanlon

December 13, 2021

Abstract

We construct a new numerical measure of earnings announcement surprises, standardized unexpected earnings call text (SUE.txt), that does not explicitly incorporate the reported earnings value. SUE.txt generates a text-based post-earnings-announcement drift (PEAD.txt) larger than the classic PEAD. The magnitude of PEAD.txt is considerable even in recent years when the classic PEAD is close to zero. We explore our text-based empirical model to show that the calls' news content is about details behind the earnings number and the fundamentals of the firm.

Author information: Vitaly Meursault (corresponding author), Federal Reserve Bank of Philadelphia, vitaly.meursault@phil.frb.org; Pierre Jinghong Liang, Carnegie Mellon University, liangj@andrew.cmu.edu; Bryan R. Routledge, Carnegie Mellon University, routledge@cmu.edu; Madeline Marco Scanlon, University of Pittsburgh, mms238@pitt.edu.

1. Introduction

Publicly traded firms in the United States announce earnings and related financial statement information quarterly. When reported earnings are high relative to expectations, stock prices tend to rise for over 60 trading days. Conversely, when earnings are low, prices continuously fall. This post-earnings-announcement drift (PEAD), first documented by Ball and Brown (1968) and so named by Bernard and Thomas (1989), is a long-standing robust market anomaly commonly attributed to investor underreaction, among other factors (Fink (2021) is a recent large scale review of the PEAD literature). Computation of earnings surprises underlying PEAD typically uses either the history of earnings or analysts' expectations as a benchmark (Livnat and Mendenhall, 2006), leading to what is called standardized unexpected earnings (SUE).

In this paper, we propose a new numerical earnings surprise measure based on the text of earnings calls without explicitly incorporating the earnings number. This measure, labeled SUE.txt, is calculated using output from a prediction model based on a regularized logistic text regression that extracts "good news" and "bad news" from earnings call text. The prediction model is trained using past earnings calls and associated one-day abnormal returns; its parameters are dynamically calibrated. We document a drift phenomenon associated with SUE.txt, which we label as PEAD.txt.

We report that the text-based post-earnings-announcement drift (PEAD.txt) is larger than PEAD at every calendar quarter mark within a year for our sample of 2010 to 2019 and quintile split portfolios (Figure 1). The difference is growing each quarter following the release of the earnings call text: 2.87% to 1.54%, 4.61% to

2

2.7%, 6.51% to 3.87%, and 8.01% to 4.63%. These magnitudes deepen the existing PEAD puzzle.

Using panel regressions, we find the association between SUE.txt and abnormal returns is more than twice as strong as between SUE and abnormal returns. The relationship persists across specifications with different controls and firm and year-quarter fixed effects.

Using the model's predictions, we construct a profitable trading strategy that goes long in companies whose calls contained the best news and shorts the companies with the worst news. The generated alpha is significant within the Fama–French five factors plus momentum framework (Fama and French 2015; Carhart 1997) and q5 factor framework (Hou et al., 2020). The PEAD.txt portfolio held for a quarter generates a larger alpha than the PEAD portfolio.

While our main contribution is the documentation of the PEAD.txt phenomenon, we also contribute by offering analytic tools to examine the nature of PEAD.txt. The goal of the literature since Bernard and Thomas (1989) has been to explain why PEAD is happening. PEAD.txt is based on a wide range of information, raising more questions. While this paper does not answer these questions directly, we propose two research tools for testing old and new hypotheses. These tools leverage the predictive model's output (regression coefficients) and the cross-section of earnings call content at the paragraph level. The first one is paragraph-level SUE.txt – a measure that reflects how important individual paragraphs are for our model (document-level SUE.txt is the sum of paragraph-level SUE.txt values plus a quarter-level constant). The second tool is a domain-knowledge-based paragraph classification scheme using

3

keywords related to the business curriculum.

With these two tools, we document the uneven distribution of words and phrases that mark news in the cross-section of paragraph content. We consider a wide range of paragraph groups and show that surprising information can appear in all of them, but with a lot of variation. Discussions of bottom line results, foreign exchange factors, operational interruptions, weather, and seasonality are most surprising on average, but surprises in paragraphs mentioning non-bottom line financial metrics contribute the most to SUE.txt overall due to their ubiquitousness.

Overall, our paper suggests that SUE.txt flexibly summarizes good news and bad news about the firm and its environment contained in earnings calls. In this sense, it is similar to the summary measure of earnings surprise. Our prediction model and empirical results confirm that earnings call texts share much of numerical earnings' communication capabilities in terms of expressing hierarchies and ordinality. These capabilities allow text to flexibly reflect the underlying firm economic activities. In this light, our results suggest that a more meaningful distinction between textual information and earnings might be its form (unstructured compared to structured) rather than substance (tone compared to facts, for example).

The magnitude of PEAD.txt relative to PEAD and text surprises' composition becomes apparent only after an empirical investigation, but the text's importance is fundamentally grounded. At the core, numerical earnings communicate a vast amount of primitive data via an imperfect summary statistic. This paper's foundational idea is that earnings call transcripts are designed to noisily communicate the *same* vast amount of primitive data, which numerical earnings are designed to

4

imperfectly summarize. Text and numbers compress primitive data in different ways and are not completely orthogonal nor completely identical. This heterogeneity in how text about earnings and earnings numbers aggregate underlying data and how market participants react to text and numbers motivates us to explore the parallel PEAD.txt phenomenon.

Insights gained by the analyses in this paper contribute to our understanding of two related well-developed literatures: PEAD anomaly and fundamental analysis. We next briefly describe the connections of this paper to these two literatures, especially the rapidly developing work incorporating machine learning techniques (such as textual analysis).[1]

First, text analysis has been used in the literature to study the cross-section of PEAD. For example, research has shown that interaction between earnings surprises and negative tone (Engelberg, 2008) or readability (Lee, 2012) produces a larger drift. These text analysis studies add to a list of determinants of PEAD's cross-section that includes the proportion of institutional investors (Bartov et al., 2000), arbitrage risk (Mendenhall, 2004), and revenue surprises (Jegadeesh and Livnat, 2006). Our study shows that text surprises on their own can produce a larger drift than earnings surprises. Our findings also have implications for the recent debate about the potential disappearance of PEAD. Several studies, including Chordia et al. (2014), Milian (2015) and Martineau (2021) argue that PEAD has recently shrunk to the point of disappearance. However, other recent studies, like Ali et al. (2020) and Cox (2020) find that PEAD persists. We document that while both PEAD and

---

[1]For a comprehensive and more historical review of the literature see Richardson et al. (2010).

5

PEAD.txt decrease in the second half in our sample, the shrinkage of PEAD.txt is smaller, and it is far from disappearing.

Second, this paper contributes to the long fundamental analysis literature recently invigorated by data mining and AI techniques. Classic work, like Ou and Penman (1989) and its modern extensions such as Yan and Zheng (2017), focus almost entirely on accounting numbers to explain current and predict firm-level future outcome variables, like earnings and stock returns. A more recent model built by Cao et al. (2021) incorporates corporate financial information, qualitative disclosure, and macroeconomic indicators. The paper shows that this comprehensive AI ensemble model outperforms human analysts as a whole, while human analysts perform better when firms are subject to more information asymmetry (e.g., more illiquid or more intangible assets). In this context, our paper identifies a potentially valuable avenue for future AI analysts to process textual data to improve prediction tasks on future earnings and prices. Further along this line of thoughts, our work has implications to the recent literature on robo-analysts (Coleman et al. (2020) and Grennan and Michaely (2020)) and the effect of AI-readership on corporate disclosure (Cao et al. 2020).[2]

---

[2]For example, Cao et al. 2020 show a potential feedback mechanism: higher AI-readership causes disclosure to be more catered to machine readers (than human readers) by avoiding words that are known to be perceived negatively by computational algorithms. In our paper, the market perception of word impact, positive or negative, is dynamically updated rather than frozen in time (such as the Loughran-McDonald dictionary), which makes reactive disclosure strategy potentially more challenging.

2. Documentation of PEAD.txt

In this section, we describe the process of generating PEAD.txt starting from data. We begin with the data description, followed by the machine learning-based methodology to develop the SUE.txt measure, the abnormal returns calculation procedure, and finish with a statistical comparison of PEAD.txt and PEAD phenomena.

2.1. Datasets

We construct the corpus of earnings call transcripts using the Capital IQ Transcripts database, which is available through the Wharton Research Data Services (WRDS) platform. Various numerical variables are constructed based on Center for Research in Security Prices (CRSP), Compustat, and Institutional Brokers' Estimate System (IBES) datasets available through the WRDS platform.[3] The details about dataset construction, merging, abnormal returns calculation, and returns timing are in Online Appendix A.

The dataset used to construct SUE.txt contains 108,704 observations between 2008Q1 and 2019Q4. The final dataset after the construction of surprises contains 85,160 observations 2010Q1 and 2019Q4. There are 4,701 unique firms in the dataset.

2.2. Construction of SUE.txt

We create a measure of earnings call text surprises, standardized unexpected earnings text (SUE.txt). Our measure reflects the following intuition: if certain content

---

[3]Transcripts and Compustat are provided by S&P Global Market Intelligence, CRSP is provided by University of Chicago Booth School of Business, IBES is provided by Refinitiv.

Electronic copy available at: https://ssrn.com/abstract=4007361

predicts abnormal returns around the call, that content reflects unexpected information. We compute SUE.txt using a regularized logistic text regression that connects the text of earnings call transcripts to one-day abnormal returns. We reestimate the model for every quarter using only information from the past eight quarters as the training set. This procedure ensures that our model is applicable in a dynamic setting. In the description below we focus on how our model and estimation procedure allows us to robustly capture unexpected textual content. A technical description of the model is provided in Online Appendix B.

2.2.1. Predictive model

Our approach to identifying unexpected information is returns-based. We assume that abnormal announcement returns are generated by unexpected information and that an earnings call with zero announcement returns was entirely expected by the market. We identify words and two-word combinations associated with positive or negative return surprises using a flexible machine learning model. We consider these words unexpected because they are associated with abnormal market reactions. The cumulative impact of these unexpected words is SUE.txt.

Our model is regularized logistic regression with elastic net regularization (Zou and Hastie, 2005).[4] Because textual data is high-dimensional, overfitting is a concern. To ensure that our model produces robust measures of surprises, we use standard

---

[4]This approach is most similar to work like Kogan et al. (2009) and Frankel et al. (2016) who use support vector regressions, Li (2010) who uses naive bayesian model, Brown et al. (2020) who use a combination of a topic model and supervised regression, Ke et al. (2020) who use a multistep procedure involving a supervised model and Garcia et al. (2021), who use multinomial inverse regression.

8

machine learning approaches of regularization and cross-validation and use only out-of-sample predictions for the main analyses.

Regularization is a technique aimed at improving out-of-sample performance by constraining in-sample error minimization to prefer solutions with smaller norms of coefficients. In our case, an unregularized model would load much more on individual words (especially rare words that appear in a few documents with large market reactions) and would capture chance co-occurrence of words and returns rather than true textual surprises. We use cross-validation to produce an optimally regularized model by splitting the sample and evaluating how strong regularization needs to be to predict announcement returns well out-of-sample. A model that doesn't regularize enough will overfit on chance associations between text and returns and will not produce robust predictions out-of-sample. Likewise, a model that regularizes too much would not be able to capture even robust associations between text and returns, which would also result in bad-out-sample predictions. Cross-validation ensures that we pick the right regularization values for our task.

To further ensure that our model identifies robust surprises, we only use one-quarter ahead predictions of the model for all our analyses. This means that when we compute SUE.txt for a specific earnings call, we use a model that has never seen that earnings call during estimation (see Table 2 for a visual representation of our rolling window approach). That further ensures that the results we obtain are due to robust measures of surprises rather than chance associations between text and returns. To ensure that we have a large panel of out-of-sample SUE.txt, we re-estimate our model every quarter using the data for two previous years. Therefore,

9

we only lose two years from our original sample of earnings calls (2008 and 2009).

Our target variable is one-day abnormal returns split into *high*, *flat*, and *low* categories (see Online Appendix B for the details about their construction). The model outputs the log-odds of a given earnings call being associated with high, flat and low returns.

2.2.2. Variables and model training

The regularized logistic text regression uses log-frequencies of individual words (unigrams) and two-word combinations (bigrams) in documents as independent variables (the bag-of-words approach). Let $\text{freq}(j, n)$ denote the frequency of the term j in the document n. The associated independent variable is $x_{n,j} = \log\left(1 + \text{freq}(j, n)\right)$. The specification includes the 1,000 most common unigrams and 1,000 most common bigrams in the presentation and the Q&A sections separately (total of 4,000 variables).[5] We use Snowball stemmer's stopword list to remove some ubiquitous English words like "the."[6] The numerical part of all terms containing numbers is replaced with #, so that "$1000.00" becomes "$#" and "Q3" becomes "Q#." We also render all words lower case but do not perform any other word processing. Most common tokens are selected using the training set and so vary across time.

Summary statistics for the dataset used to construct SUE.txt are presented in Table 1. The number of documents across all years is approximately 117,000. Management presentation sections of earnings calls are large documents; the median one

---

[5]For example, the log frequencies of the word "revenue" in the presentation and the Q&A sections have different variables associated with them.

[6]https://snowballstem.org/. Last accessed: 08/26/2020.

10

is approximately 3,000 words long. Q&A sections are even larger; the median one is about 4,000 words long. The median abnormal return is very close to zero, and the split into the three categories is even.

We also experiment with introducing an array of numerical variables to the model and compare text-only and text-and-numeric models in Online Appendix B.

2.2.3. Computation of SUE.txt

We construct our measure of earnings call text surprises based on the text-based model's log-odds ratio output: To stress the analogy with classic earnings surprises (SUE), we call our measure standardized unexpected earnings <call> text (SUE.txt):

$$\text{SUE.txt} = \text{log-odds}(H) - \text{log-odds}(L).$$

Our measure is standardized in the sense that it is directly comparable between different companies. Like classic SUE, positive and negative values of SUE.txt correspond to good and bad earnings announcement news, respectively, and zero value indicates no unexpected information.

Intuitively, SUE.txt is high if the call contains many words and phrases associated with high returns and few words and phrases associated with low returns, according to the model's predictions. As shown in later sections, these words and phrases are general markers of "good news" or "bad news." They appear in paragraphs discussing widely varying content types, from firm financial performance to general economic conditions. We can think about segments containing the news markers as unexpected text, and the segments containing no news markers as expected text.

11

We further discuss analogy between SUE.txt and SUE in Section 3.2, the words and phrases driving the SUE.txt in Section 4.1, and the context in which they appear in Section 4.2.

2.3. Construction of PEAD.txt based on SUE.txt

To demonstrate PEAD.txt and compare it to PEAD, we compute the cumulative abnormal returns for a spread portfolio formed on the day following the earnings call that buys the stocks that we estimate to be in the top quintile of SUE.txt or SUE in a given quarter and shorts the stocks in the bottom quintile:

$$\text{CAR}_t^S = \prod_{t=S}^{E} \left(\text{AR}_t^S \right),$$

$$\text{AR}_t^S = \frac{1}{|T|} \sum_{\{f,q,t\} \in T} \text{AR}_{f,q,t} - \frac{1}{|B|} \sum_{\{f,q,t\} \in B} \text{AR}_{f,q,t},$$

$$\text{AR}_{f,q,t} = \text{R}_{f,q,t} - \text{R}_{f,q,t}^b,$$

where f, q, t are the firm, quarter, and event time indices; S, E indicate the start and end times of the calculation; T and B are sets of observations belonging to the top and bottom quintiles of SUE.txt or SUE;[7] $|T|$ and $|B|$ are the sizes or respective sets; R is the firm stock return; and R^b is the benchmark return of one of the six size and book-to-market matched portfolios. [8]

---

[7]In the case of SUE.txt, we estimate the quintile an observation will belong to in its quarter by using training set SUE.txt quintile cutoffs. In the case of SUE, we use the previous quarter's SUE quintile cutoffs.

[8]The cutoffs used to match stocks to their benchmark portfolios and the portfolio returns are from Kenneth R. French's data library at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Last accessed: 08/12/2020.

SUE.txt generates a much larger drift than classic SUE, deepening the PEAD puzzle. Figure 1 compares PEAD.txt and PEAD over the 252 trading days horizon (one calendar year). At every calendar quarter mark, PEAD.txt is much larger and growing: 2.87% to 1.54% on trading day 63, 4.61% to 2.7% on trading day 126, 6.51% to 3.87% on trading day 189, and 8.01% to 4.63% on trading day 252. PEAD is only larger at the very beginning of the window.

2.4. Comparing statistical properties of PEAD.txt with traditional PEAD

Tables 3 and 4 provide some further diagnostics for the first 63 trading days. The PEAD.txt based only on call transcripts is larger than the drift based on a regularized logistic regression with both the text and numerical variables (see Online Appendix C). The larger magnitude of PEAD.txt relative to PEAD comes from both the top and bottom quintiles, but mostly from the top one (1.31% compared to 0.16% for the first 63 trading days). As a comparison, we find that using quintiles of percentages of negative words in the transcripts (similarly to Engelberg 2008) produces much a smaller drift than PEAD.txt, 1.11% to 2.87%.[9] The quintile spread of the earnings call day abnormal returns also produces a smaller drift, 1.65% (using only abnormal returns to generate the drift is the approach of Brandt et al. 2008).

SUE.txt has stronger associations with CAR than classic SUE in a panel regression setting with fixed effects, as Table 5 shows. We compute CAR at the stock-quarter level using the returns of the six size and book-to-market portfolios as a benchmark. One standard deviation increase in earnings call surprise is associated

---

[9]We use Loughran and McDonald's financial domain sentiment dictionary to identify negative words (Loughran and McDonald, 2011).

with 3% to 6% of a standard deviation increase in 63 trading days CAR, depending on specification. This result is robust to including firm and year-quarter fixed effects and clustering by firm and year-quarter. Earnings surprises have a weaker association with CAR, with normalized coefficients ranging between 1% and 2% across specifications and only significant at the 5% or 10% level in some specifications (since the last column includes interactions between earnings surprises and other variables, the coefficient size there is not comparable with other columns). Qualitatively similar results hold at days 1 to 32 and 33 to 63 (Table 6).

Finally, a trading strategy that utilizes PEAD.txt produces alpha. We consider a portfolio that buys the stocks that we estimate to be in the top quintile of SUE.txt in a given quarter, and shorts the stocks in the bottom quintile. The portfolio is equally weighted, opens the position at the first close after the earnings call, and holds it for 63 trading days. We regress the daily portfolio returns minus the risk-free rate on the five Fama-French factors and momentum. The spread portfolio earns a statistically significant daily alpha of 3.9 basis points, as Table 7 shows. Both top and bottom quintile portfolios have statistically significant alphas (positive and negative, respectively). In contrast to that, alpha generated by the classic SUE is lower (2.6 basis points), as shown in Table 8. While the SUE spread alpha has high statistical significance, alpha for the bottom quintile is only significant at the 5% level, and alpha for the top quintile is not significant at the 5% level.

Additional comparison of spread portfolio alphas is presented in Table 9. A strategy that equally weights SUE.txt and SUE signals is the best performing strategy overall with an alpha of 4.2 basis points. Strategy based on a regularized logistic

14

regression with both the text and numerical variables (see Online Appendix C) underperforms the SUE.txt strategy. Strategy based on the percentages of negative words from the financial sentiment dictionary does not produce alpha in our setting.

Table 10 compares strategies with a shorter holding period, trading days 1 to 32. In this case, the classic SUE spread strategy comes ahead of the SUE.txt strategy with 4.3 to 3.4 basis points alpha. The best performing strategy overall is a mix of the two with an alpha of 5.2 basis points. Table 11 presents the portfolio performance results (63 days holding period) using the q5 factors (Hou et al., 2020).[10] The results are very similar to the results obtained using Fama–French factors.

2.5. PEAD.txt and PEAD over time

Figure 2 demonstrates PEAD.txt and PEAD across the years. PEAD.txt is larger than PEAD in eight out of ten years, except in 2012 and 2013. Both PEAD.txt and PEAD are smaller and plateau sooner in the second half of the sample. However, PEAD.txt never falls below 3.4% at the calendar year mark. We also see signs of a large resurgence in PEAD.txt in 2019 at the end of our sample. These results suggest that PEAD.txt has been more robust to forces that are reducing PEAD potentially to the point of disappearance as discussed in Chordia et al. (2014), Milian (2015) and Martineau (2021).

---

[10]We obtain the factor returns data at http://global-q.org/index.html. Last accessed: 08/12/2020.

15

3. Economic interpretation of PEAD.txt and comparison with PEAD

3.1. Economic interpretation of SUE.txt and PEAD.txt

SUE.txt is a summary statistic that reflects the sign and magnitude of news about a firm's economic activity based on the text of earnings calls. SUE.txt does not explicitly incorporate the various numbers mentioned in earnings calls but still reflects them through correlations between word choice and numbers. The intuition behind PEAD.txt is similar to PEAD: firms with positive surprises tend to have upward price drift, and firms with negative surprises tend to drift downward. The difference is an expanded definition of surprise.

Economic activities occur in the physical world. They involve the circulation of goods, money, and information; contracts; physical and mental activities; and environmental and societal factors, to name just a few aspects. The accounting system economically summarizes these activities using numerical disclosures consisting of financial statement figures such as net income. Natural language disclosure, such as earnings calls, performs a similar summarization function.

One interesting aspect of SUE.txt is its relationship to the numbers contained in earnings calls. The numbers are not incorporated in our measure directly (we replace every number with "number token"). However, the relationship between reported numbers and firm value can often be inferred from language ("our EPS improved from numtoken to numtoken," "we experienced a loss of numtoken"), and SUE.txt heavily utilizes this (see Section 4.1).

While the content of numerical and natural language disclosure is similar, the form is naturally very different. Numbers come in an easy-to-process manner, with

16

clear hierarchies and ordinality. Hierarchy and ordinality are also present in the language but are harder to process mathematically. For example, a human reader sees that "great earnings this quarter" is better news for firm value than "ok earnings," and that "increase in total revenues" has higher importance than "loss of one of many contracts," but a computer algorithm needs more processing.

Our algorithm to create the SUE.txt measure is one way to process and summarize the surprising content of earnings calls by relying on regularized logistic text regression and one-day abnormal returns for calibration. Both the model and the way to calibrate it can be tweaked in future work to produce better measures.

3.2. Comparing the economic meaning of SUE.txt and SUE

SUE.txt is a text-based analog of SUE because, like SUE, it reflects the difference between the firm's reported results and the market expectations. However, how the two measures incorporate the results and expectations is notably different.

SUE incorporates firm results and market expectations directly in the form of reported earnings and analyst earnings forecasts. The beauty of this measure is in the fact that we have direct access to analyst expectations measured in the same units as the firm results.

SUE.txt identifies what is news in text and quantifies it. The challenge is to separate the new (and relevant) content in earnings calls from old (or irrelevant) content.[11]

---

[11]To validate that our model separates unexpected textual information that drives market response from expected information that generates no market reaction, we performed a human annotation study. We produce a dataset of 100 paragraphs, 50 from the top decile of unexpected paragraphs identified by our model and 50 from the bottom decile (most expected paragraphs have

In a perfect world, we would task professional analysts with highlighting the new information in earnings calls.[12] While using analysts in such a way is unfeasible at scale, one can use a statistical model and some external measure of information relevance to infer the impact of new text content, expressed in a numerical form such as log-odds of high or low return. We use one-day abnormal stock returns to discipline the regularized logistic text regression. Returns work under the assumption that prices incorporate publicly available information and that earnings calls contain a significant portion of the information released that day. The result is a model that finds words and phrases marking new information.

Proposing SUE.txt as a stand-alone measure similar to SUE and generating larger drift without explicitly utilizing the earnings number distinguishes the present paper from papers using text to study PEAD's cross-section. Engelberg (2008) and Lee (2012) use negative tone and readability, respectively, to further refine SUE-based sorting. The papers contrast their language measures with numbers and argue that they are associated with higher information processing costs. The present paper focuses on text as a reflection of economic activity, similar to earnings number in content, but different in form. Naturally, this view does not contradict the results

SUE.txt close to zero, and most surprising paragraphs have large positive or negative SUE.txt). A human annotator (an RA who didn't participate in the project in any other capacity) was asked whether each of the paragraphs was "likely to contain unexpected good or bad news about the firm that is likely to cause a large market reaction". In 68% of the cases, the annotator identified as paragraphs in the top (bottom) decile of the absolute value of SUE.txt as unexpected (expected) by investors. That is better than chance performance (50%) at the 1% level of statistical significance.

[12]Asking analysts to write hypothetical earnings calls based on all available prior information would achieve a similar end. Importantly, this would allow us to study both what is new in the actual earnings call and what is omitted. The economic cost of doing this at scale would be prohibitive, but one could perhaps design a conditional text generation model that does that.

Engelberg (2008) and Lee (2012). Instead, we focus on the aspects of language that are more similar to numerical disclosure and argue that earnings call text reflects objective information about firms' value, not just as much as the earnings numbers do, but to a more considerable extent than earnings.

3.3. Examples

To help build intuition about the SUE.txt measure, we provide two example paragraphs below. The first example is identified as expected text by the model, and the second one contains a large positive surprise. We highlight the words that were assigned non-zero coefficients by the model and normalize the coefficients by the largest absolute in the paragraph for tractability.

No Surprise **Smart Technologies Inc., 2015Q3**	Positive Suprise **WD-40 Company, 2017Q2**
"Following our prepared remarks, we will open the call for questions. Please note that some of the information you'll hear during our discussion today will consist of forward-looking statements within the meaning of applicable U.S. and Canadian securities laws. These statements, which are further discussed in the important cautionary statement found on Page # of our presentation include, without limitations, statements regarding our sales and performance outlook for the fourth quarter and full year fiscal #, including adjusted revenue, adjusted EBITDA, adjusted gross margin and cash operating expense; our market expectations, future sales of our new and existing products, including SMART kapp and our interactive flat panels; the addressable market for certain of our products and our future business product and other plans and strategies."	"*Additionally(-4%)*, you heard that the reduction in sales was significantly offset by $4.3 million in *transaction(-5%)*-related impacts in EMEA due to the strengthening of the euro and the U.S. dollar against the pound sterling. You heard that our sales was *strong(+100%)* in Canada and that we believe the market will continue to see *growth(-1%)* in the coming quarters. You heard that our sales was *strong(+58%)* in Asia with a 21% sales *growth(-1%)* in our distributor markets and a 17% sales *growth(-1%)* in China. You heard we're maintaining our net income and EPS guidance for the fiscal year, but we revised a couple of other components of our fiscal year guidance."

The first paragraph is boilerplate and the model correctly identifies that as no words have non-zero coefficients associated with them. In the second paragraph, the management is conveying positive news which is correctly identified by the model. SUE.txt of this paragraph is high and positive mainly because the word "strong" is used twice, in this case, in the context of sales. The first time the coefficient is 100% since it is the largest coefficient in this paragraph, the second time the word strong appears the coefficient is smaller because our model operates on log word counts (all other coefficients are scaled relative to the first instance of the word "strong"). Our interpretation is that managers use the word "strong" to highlight results exceeding expectations, which is consistent with the content of this paragraph. The rest of the coefficients are negative but small in value and the paragraph doesn't provide much context to understand why these words are statistically more likely to be used in earnings calls associated with low returns. It is expected for the coefficients to not be completely interpretable because SUE.txt is an output of a supervised model that is optimized for explaining returns out-of-sample rather than closely following human judgment about the polarity of individual words.

4. Analytic tools for explaining PEAD.txt

PEAD.txt is larger than PEAD. That deepens the PEAD puzzle. But earnings calls also allow us to have a more detailed look into the drift's driving forces, which is the ultimate goal of PEAD literature. While the present paper does not provide answers about PEAD's drivers, we propose new tools to examine SUE.txt and PEAD.txt (and potentially other text-based measures). New tools are needed be-

20

cause regularized logistic text regression is a complex model, and earnings call text is a complex environment. The proposed tools help make sense of the measure by reducing the complexity to a more manageable level.

The first step to explain PEAD.txt is to understand how SUE.txt aggregates information from words and paragraphs. Understanding how a text classification model arrives at its conclusions is not an easy task. The difficulty is due to the interaction of two already complex components – a text classification model and the diverse content of earnings calls. Furthermore, the model interacts with textual content at different levels – directly through word counts in individual documents, and indirectly through the context in which individual words appear. For example, it could be helpful to know that "increasing" is associated with high returns, but it is also essential to understand what the companies report as increasing.

We start making sense of SUE.txt using the concept of word impact (Yano et al., 2012). Building on word impact, we propose two new tools: paragraph-level SUE.txt and a domain-knowledge-based paragraph classification scheme reflecting the business curriculum.

4.1. Word-level impact and news directionality

At the word level, we focus on three quantities of interest: model coefficients, average word counts (or log counts) per document, and their product, called impact (Yano et al., 2012).[13] The trained model has parameters, β coefficients, associated with individual tokens that tell us which words and phrases drive the model's pre-

---

[13]Note that there is a different definition of word impact (Routledge et al., 2013).

dictions. In this case, words like "improvement" and "strong" shift the prediction of the model to the high return category, and words like "lower" and "impacted" shift it to the low return category. The primary way the model interacts with the content of the documents is through the word frequencies (more specifically, log frequencies of tokens that are the x variables of the model). Impact I of term j is defined as the product between the model coefficient and the mean log frequencies across all observations:

$$I_j = (\beta_j^H - \beta_j^L)\frac{1}{N}\sum_{i=1}^{N} x_{ij},$$

where β^H and β^L are the coefficients in the parts of the logistic regression that predict *high* and *low* returns, respectively.

Model coefficients and mean log frequencies define a two-dimensional space. Figure 3 plots fifteen tokens with the largest positive impact and fifteen tokens with the largest negative impact. Many of these coefficients' signs are consistent with the intuition that good news about firm value correlates with positive returns, like "favorable," "strong" or "improvement." Among the words with a negative sign, we also see tokens confirming that intuition, like "issue," "loss," or "decline." It is also clear that words can be highly impactful in two different ways: uncommon words like "nice" or "issue" are influential when they do appear, whereas words like "good" and "not" are much more ubiquitous and influence the model prediction through x rather than β.

Overall, the coefficients support the intuition that favorable news is associated with an increase in firm value (and the opposite for bad news). However, the model

22

picks up positive or negative news signals in various ways, some of which are more straightforward than others. Notable types of signals include:

- Tokens semantically indicating directionality of news, like "numtoken increase" or "lower."

- Tokens implying directionality of effect, like "benefited" or "impacted."

- The implied polarity of "income" and "loss."

- Markers of analyst behavior. Analysts can either acknowledge good results ("great <quarter>") or satisfactory answers ("<ok, > good"), or raise "issues" and ask for clarifications to help them "understand" something. That also provides signals for the model.

4.2. Descriptive patterns of paragraph-level SUE.txt

The model also interacts with document content on a deeper level, through the context in which words and phrases with non-zero β coefficients appear. If the coefficients indicate good or bad news, what tends to be the subject of the news? To answer that question, we propose a domain-knowledge-based paragraph classification scheme reflecting the business curriculum, calculate paragraph-level SUE.txt (SUE.txt^P), and analyze how SUE.txt^P differs across different paragraph groups and subgroups. A multitude of possible paragraph classification schemes would reflect the goals and preferences of various domain experts. We propose a business curriculum-based scheme as a starting point because it allows us to cover the vast majority of earnings call paragraphs and because this scheme seems reasonable for texts produced to a large extent by people with business school degrees.

23

We focus on paragraphs as units of text unified by a single theme. Capital IQ Transcripts database conveniently provides paragraph splits.

4.2.1. Paragraph-level SUE.txt and paragraph groups based on business curriculum

We modify the measure of impact discussed above to apply at the paragraph level. Paragraph-level SUE.txt (SUE.txt$^\text{P}$) aggregates the coefficients of words present in the paragraph with necessary log frequency adjustments. Document-level SUE.txt is the sum of paragraph-level SUE.txt values plus a quarter-level constant. We define paragraph-level SUE.txt as follows:

$$\text{SUE.txt}^\text{P} = \sum_{w=1}^{W}(\beta_w^H - \beta_w^L)\Delta_w,$$

$$\Delta_w = \log{(2 + b_w)} - \log{(1 + b_w)},$$

where β^H and β^L are the coefficients in the parts of the logistic regression that predict *high* and *low* returns, respectively, and b is the number of times a given n-gram appeared in the document before (we use this weighting because our bag-of-words model operates on log word counts).

Further, we split paragraphs into groups using the following keyword-based scheme consisting of (potentially overlapping) paragraph groups, subgroups, and keywords. The groups and subgroups are the following (see Online Appendix D for keywords and paragraph examples):

1. **Financial accounting**: bottom line, metrics, adjustments, lending, financing.
2. **Operations management and marketing**: operational and marketing metrics, segments, supply chain, production, interruptions, marketing.

3. **Global economics**: foreign exchange, seasonality and weather, general global economics.

4. **Strategy**: competition, expansion, contraction, partners, deals, government, restructuring, general strategy.

5. **Forward-looking**: paragraphs including forward-looking phrases that are identified following Muslu et al. (2015).

Here, we examine absolute paragraph-level SUE.txt for different groups. Mean absolute SUE.txt$^{\mathrm{P}}$ for paragraph group g is defined straightforwardly:

$$\left| \mathrm{SUE.txt}^{\mathrm{P}}{}_{g} \right| = \frac{1}{|G|} \sum_{k \in G} \mathrm{SUE.txt}^{\mathrm{P}}{}_{k},$$

where G is the set of paragraphs belonging to a specific group.

The absolute value of SUE.txt$^{\mathrm{P}}$ shows us where the information is, without specifying whether the information is good or bad for firm value. In a world where good and bad news about firm revenue is equally likely, and all firms report revenue news, the related paragraphs would likely have an average SUE.txt$^{\mathrm{P}}$ of zero even if the revenue news is significant. Looking at the absolute value of SUE.txt$^{\mathrm{P}}$ allows us to see what the big news is about without worrying that good and bad news cancel out.

All groups of paragraphs are informative, but there is a lot of variation between and within the groups. As Figure 4 shows, the bottom line, forex, interruption, and seasons group have the highest mean absolute SUE.txt$^{\mathrm{P}}$ (within 5% of the top subgroup, bottom line). However, these groups are rare and, overall, the financial accounting metrics subgroup dominates as the most prevalent (around 37%) and

quite impactful (0.8 of the absolute SUE.txt^P of the most impactful group). General strategy and segment subgroups, as well as the forward-looking group, fall somewhat in the middle as being quite prevalent but not as impactful as some other subgroups.

Overall, the results in this section show that SUE.txt reflects a wide range of information about the firm and its environment. Naturally for financial disclosure, discussions of financial metrics dominate overall. Nevertheless, when certain rare topics, such as operational interruptions or foreign exchange are brought up, they drive up our surprise measure in extreme directions.

5. Autocorrelation of SUE.txt

A subset of PEAD literature, including Narayanamoorthy (2006) and Cao and Narayanamoorthy (2012) discusses the cross-sectional differences in autocorrelations of SUE and links them to possible causes of investor underreactions. Within the SUE.txt setting, we explore the parallel association analysis between autocorrelation and contemporaneous earnings characteristics such as loss and ex-ante (expected) earnings characteristics such as volatility.

Table 12 shows the SUE.txt autocorrelation results following Narayanamoorthy (2006) and Cao and Narayanamoorthy (2012). We construct deciles of SUE.txt and run regressions with decile in the current period on the left-hand side. On the right-hand side, the regressions include the decile of SUE.txt in the previous period, earnings volatility, market value, and loss indicator (all for the previous period) as well as interactions between the lag of SUE.txt decile and the other variables. We

26

see that SUE.txt has positive autocorrelation.[14] The presence of autocorrelation is important because cross-sectional differences in autocorrelation coefficients offer potential explanations for PEAD. For example, Narayanamoorthy (2006) links autocorrelations to accounting conservatism and shows, amongst other results, that SUE is more mean-reverting for loss firms. We confirm that the same holds for SUE.txt.[15] Examining the relationship between earnings volatility and the autocorrelation of SUE.txt, we find that unlike SUE (see Cao and Narayanamoorthy 2012), SUE.txt is more persistent when earnings volatility is higher. Overall, we confirm that there are predictable patterns in the cross-section of autocorrelation of SUE.txt similar to patterns in autocorrelations of SUE. Naturally, there are some differences in the cross-sections of earnings and analyst-based SUE and SUE.txt based on textual information. Understanding the impact of these differences on the autocorrelations of SUE and SUE.txt can be an important direction for future research that further explores how emerging technologies can be connected and contribute to capital markets research.

6. Conclusions

We develop a measure of earnings call text surprise, SUE.txt. We compute it using a regularized logistic text regression that links the text to the market reaction around the call. We find that in our sample period of 2010 to 2019, PEAD.txt, the

---

[14]In an untabulated analysis we also confirm that the positive autocorrelation extends to at least four lags, the same as SUE in our sample.

[15]In untabulated analyses we find that the interaction is significant and negative for the first three lags of SUE.txt decile and loss indicator, and insignificant for the fourth lag.

post-earnings-announcement drift based on SUE.txt alone, without directly incorporating earnings numbers, is much larger than the classic SUE-based PEAD. Panel regressions of cumulative abnormal returns on SUE.txt and SUE and trading strategy alpha tests confirm these results. Since earnings calls contain a wide range of information regarding the firm's performance, this indicates that investor underreaction to earnings announcements goes far beyond the headline number. In this way, we deepen the PEAD puzzle.

While extracting information from the unstructured text can be profitable, understanding how markets process information is a more important goal academically. We propose a new tool that helps understand what kinds of earnings call content drives the market reaction, paragraph-level SUE.txt. Using paragraph-level SUE.txt in conjunction with a keyword-based paragraph classification scheme reflecting the business curriculum, we show that paragraphs related to financial accounting are significant drivers of SUE.txt. This suggests that a more meaningful distinction between textual information and earnings might be its form (unstructured compared to structured), rather than substance (objective compared to subjective, tone compared to facts). Questions regarding how text and numbers interact with each other to help investors understand the state of the firm and cross-sectional differences in SUE.txt and its autocorrelations call for future theoretical, structural, and empirical research.

28

Online Appendix A. Dataset construction details

We construct the corpus of earnings call transcripts using the Capital IQ Transcripts database, which is available through the Wharton Research Data Services (WRDS) platform. Various numerical variables are constructed based on the Center for Research in Security Prices (CRSP), Compustat, and Institutional Brokers' Estimate System (IBES) datasets available through the WRDS platform. All continuous variables are winsorized at the 1% and 99% levels.

To merge the Transcripts, CRSP, and IBES datasets, we use the identifier crosswalk provided with the Transcripts dataset, as well the CRSP to IBES linking table constructed using the script by Qingyi (Freda) Song Drechsler.[16] We link Compustat to other datasets using the Compustat / CRSP merged dataset available at WRDS.

To compute announcement day abnormal returns used to create the ML-based earnings call surprise measure, we use the WRDS Event Studies tool to compute one-day abnormal returns using the Fama–French plus momentum risk model with default estimation window, number of valid returns, and gap parameters. In the Fama-French three-factor model, proposed in Fama and French (1993), stock returns are adjusted for excess returns on the market portfolio of stocks, the size factor, and the book-to-market factor. These factors adjust for some major correlations between stock returns. The excess return on the market portfolio of stocks captures the common variation in stock returns. The size factor relates to profitability, as small firms tend to have lower earnings on assets than large firms. The book-to-

---

[16]https://www.fredasongdrechsler.com/full-python-code/iclink. Last accessed: 08/12/2020.

29

market factor follows from the fact that firms with a low (high) stock price relative to book value tend to have persistently low (high) earnings. The Carhart momentum factor, proposed in Carhart (1997), extends the Fama-French three-factor model to account for the profitable trading strategy associated with stock price momentum. Momentum is the tendency of stock price to continue increasing following a rise and to continue decreasing following a decline. By using the Fama–French three-factor model with the Carhart momentum factor in the calculation of abnormal returns, we extract the "unexpected" portion of returns that we hypothesize to be associated with the information contained in the earnings call.

For the econometric analysis of the earnings call surprise measure, we compute long-run cumulative abnormal returns based on the Fama–French six size and book-to-market matched portfolios.[17]

We compute classic earnings surprises, following Livnat and Mendenhall (2006).[18] We compute three versions of standardized unexpected earnings (SUE). The first version (*SUE1*) is based on earnings reported in Compustat and measures the surprise as the difference between the current earnings and the earnings in the same quarter the previous year. The second version (*SUE2*) is computed in the same way as (*SUE1*) but uses earnings adjusted for special items. The main specification, *SUE3*, uses analyst forecasts to measure deviations in reported earnings from expectations. Prior to firm earnings announcements, financial analysts assigned to the firm issue

---

[17]The cutoffs used to match stocks to their benchmark portfolios and the portfolio returns are from Kenneth R. French's data library at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Last accessed: 08/12/2020.

[18]We use a python script available at https://www.fredasongdrechsler.com/full-python-code/pead. Last accessed: 08/12/2020.

predictions of future firm performance based on firm fundamentals, prior industry experience, and other information. These forecasts are then compared to reported earnings to calculate unexpected earnings (also called earnings surprises) and adjusted by share price. Earnings surprises are summary statistics for new information contained in earnings announcements. The prices tend to move in the direction of the surprise, and as a result, buying (selling) firms with positive (negative) SUE can generate excess returns.

Earnings calls are often conducted either before market open or after market close. This influences which daily return should be used as the earnings call day return. For calls that happened in the morning of day t, the return on day t (price change between the market close on day $t-1$ and the market close on day t) is appropriate. But if the call happens on day t after the market close, the return on day $t+1$ (price change between the market close on day t and the market close on day $t+1$) is the one that covers the time of the call. The Capital IQ Transcripts database does not include the time of the call, but it can be estimated using the transcript creation time. We use the following heuristic. Given that the date of the earnings call is t,

- If the first version of the transcript is marked as preliminary and was uploaded before 3:00 PM Eastern Time on the day t, we use the returns between $t-1$ and t as the call day returns.

- If the first version of the transcript is marked as final (edited) and was uploaded sometime during the day t, we use the return between $t-1$ and t as the call day return.

31

- Otherwise, we use the return between t and $t+1$ as the call day return.

Note that this approach picks the later return as earnings call day return in cases when we might be unsure about when the call happened, safeguarding against using information from the future.

32

Online Appendix B. Regularized logistic regression

We use a regularized logistic text regression to predict short-term market reactions based on earnings transcript text. Then we use model predictions to construct a measure of earnings call text surprises. We consider an earnings call to have a positive SUE.txt if the model assigns high log-odds to one-day abnormal returns being positive and low log-odds to one-day abnormal returns being negative.

We use the elastic net model developed by Zou and Hastie (2005) and implemented in R package *glmnet* (Friedman et al., 2010).[19] It is a multinomial logistic regression model that combines $L1$ and $L2$ penalties of the lasso and ridge methods. The model, in log-odds form, is the following:

$$\text{for } r \in \{H, F, L\},$$

$$\text{log-odds}(r) = \log \frac{\Pr\left(R_{t=0} = r | X = x\right)}{\Pr\left(R_{t=0} \neq r | X = x\right)} = \beta_{0r} + \beta_r^T x,$$

where $R_{t=0}$ is the earnings call day return split into categories $r \in \{H, F, L\}$, *high*, *flat*, and *low* (more on the category split below); x is a vector of predictor variables (word frequencies or standard numerical variables); β_{0r} is the intercept; and β_r is the vector of regression coefficients.

---

[19] While the model we use is distinct from support vector regression used in Frankel et al. (2016), the general approach of using a regularized linear model is the same. Our approaches also share a feature of reestimating the model over time.

The objective function is the following:

$$L\left(\{\beta_{0r}, \beta_r\}_{r\in\{H,F,L\}}\right) =$$
$$-\left[\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{r\in\{H,F,L\}} y_{ir}\left(\beta_{0r} + x_i^T\beta_r\right) - \log\left(\sum_{r\in\{H,F,L\}} e^{\beta_{0r}+x_i^T\beta_r}\right)\right)\right] +$$
$$\lambda\left[(1-\alpha)\|\beta\|_2^2/2 + \alpha\sum_{j=1}^{p}\|\beta_j\|_1\right],$$

where y_{ir} is an indicator variable equal to one if the one-day return for an earnings call i falls into category r, α is the mixing hyperparameter that controls the relative strength of $L1$ and $L2$ regularization, λ is the hyperparameter that controls overall strength of regularization, and p is the total number of coefficients associated with all categories. To avoid excessive subscripts, we denote the combination of vectors β_H, β_F and β_L as β, and individual coefficients as β_j. We set α to 0.5 and choose λ using 10-fold cross-validation on the training sets.

Our target variable is one-day abnormal returns split into *high*, *flat*, and *low* categories in the following way:

1. *Flat*: 33% of observations closest to zero (first tercile of the absolute value of abnormal returns). The cutoffs for training and test sets are based on the terciles of the training set.

2. *High*: large positive abnormal returns (positive returns not falling into the *flat* category).

3. *Low*: large negative abnormal returns (negative returns not falling into the *flat* category).

34

The SUE.txt is an a priori measure in the sense that one can compute it without access to returns associated with a specific call. We achieve this by setting up our problem as a sliding window prediction task, as demonstrated in Table 2. At each iteration, we use eight consecutive quarters of observations as the training set, and one subsequent quarter as the test set. We slide the window one quarter at a time. We perform 40 iterations, with test quarters ranging from the first quarter of 2010 to the fourth quarter of 2019. We only use test set predictions to calculate SUE.txt, therefore ending up with 40 quarters of a priori measurement. Within each training set, we perform 10-fold cross-validation to find the best value of the regularization strength hyperparameter λ.

Online Appendix C. Combining text and numbers in the regularized logistic text regression

In this appendix, we investigate interactions between text and numerical variables for explaining announcement returns and for producing drift. We estimate a model that includes both text and an array of numerical variables reflecting the firm's earnings, fundamentals, and market responses to firm-specific information before the earnings call (Text + Num model).[20] To hone in further on potential interactions between text and numbers, we also experiment with a novel model setup that in addition to an array of numerical variables includes interactions between word/phrase counts and terciles of earnings surprises (Text × Num model). We find complementarities between text and numbers that suggest the two media work together to help investors understand the value of the firm.

Interacting text and numbers in ML model

We begin by examining announcement returns generated by the Text, Text + Num, and Text × Num models. We find that text and numbers together classify announcement day returns better than text or numbers alone. The difference in announcement returns of firms classified as "high return" and "low return" is 4.3% for the Text + Num model, compared to 3.8% for the Num model and 2.5% for the Text model (see Table 13). Text × Num model generates slightly lower returns

---

[20]The variables are: SUE, abnormal return on the day before the earnings call, abnormal return on the day two days before the earnings call, abnormal return for the earnings call day last quarter, firm size, share turnover, idiosyncratic volatility, number of analysts following the firm, Fama–French 49 industries indicator, and the interactions between SUE and firm size, turnover, idiosyncratic volatility, and the analyst coverage.

than Text + Num model (4.2% vs 4.3%). We interpret the underperformance of the Text × Num model as indicating that while management language changes with numerical results, the same word means the same thing no matter what numerical results are. The relative magnitudes of returns generated by Text, Num, and Text + Num suggest that announcement day information incorporation is based on both numbers and text, but on numbers to a larger degree than text.

Interacting text and numbers in generating drift

Then we use the log-odds of three different models to generate surprises in the same manner as we did with SUE.txt and use the surprises to generate separate drifts. The results are presented in Figure 5. We find that PEAD.txt produced by the text-only model is the largest purely ML model-based drift, 8.01% for the period starting one day after the earnings call and ending at the calendar year mark. Despite performing better at classifying announcement day returns, the Text + Num model generates a lower drift of 6.15%. We speculate that it happens because the Text + Num model underweights text and overweights numbers because numbers are incorporated quickly and are more predictive of announcement day returns. In contrast, textual information takes longer for markets to incorporate (a common notion since at least Engelberg 2008), which makes pure text surprises most associated with the drift. Text × Num model performs slightly below Text + Num model (5.44% vs. 6.15%) and, unsurprisingly, model based only on the numbers produces a drift similar to classic SUE-based PEAD (4.11% vs 4.64%).

Interacting text and numbers in generating drift: alternative solution

While ML-based models struggle at incorporating text and numbers in a way that produces larger drift, we achieve better results with a simpler method – rank aggregation of surprises. We equally weight percentiles of SUE.txt and SUE and renormalize the result to fall between 0 and 1 to produce SUE.mix (for example, a stock in 100th percentile of SUE and 50th percentile of SUE.txt is assigned to 75th percentile of SUE.mix). Sorting the stocks by SUE.mix creates PEAD.mix, which is the largest drift we generate (8.87% vs 8.01% in the case of PEAD.txt). This analysis shows that text and numbers produce drifts that complement each other, but the magnitude of PEAD.mix is largely attributable to the textual information.

Conclusion

Overall, we show that text and numbers are complementary in helping investors uncover the firm value. Numbers provide more information on announcement day, but text produces larger subsequent drift. Some ways of combining numbers and text work better than others and the optimal way of doing so is an open research question. The connection between the predictive model and PEAD is interesting and has scope for a follow-up paper. Importantly, the PEAD.txt result is quite robust to including text and numbers.

Online Appendix D. Paragraph groups examples

1. **Financial accounting.** Discussions of financial results focused around metrics from various financial statements.

   (a) **Bottom line.** Net earnings / income, net losses or EPS. Keywords: *earnings, eps, net earnings, net income, net loss.*

   > Thank you, Jack. Good morning, everyone. For the third quarter, we generated net income of $118.9 million or $2.96 per fully diluted share compared with $100.4 million or $2.33 per diluted share in the third quarter last year. (The Hanover Insurance Group, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

   (b) **Metrics.** A wide range of financial metrics excluding the bottom line ones above. Keywords: *amortization, book-to-bill, capex, cash, cash flow, cashflow, charge, cogs, cost, current assets, current liabilities, ebit, ebitda, ebt, expenditure, expense, fee, goodwill, gross margin, income, inventory, liability, long - term debt, loss, margin, operational income, operational margin, payable, pp&e, profit, profit margin, profitability, receivable, retained earnings, revenue, sale, tax, treasury stock, working capital.*

   > Turning to some other key metrics. In addition to delivering robust constant currency revenue growth, which we were able to achieve even when facing a – even when faced with a difficult prior-year comparable, it was – I was also pleased to see the year-over-year and sequential expansion in both our adjusted gross and operating margin. And during the quarter, our adjusted gross margin reached 58.6%, which was an increase of 160 basis points as compared to the prior year and 90 basis points sequentially, while our adjusted operating margin totaled 27%, which was an increase of 100 basis points as compared to the prior year and 180 basis points sequentially. (Teleflex Incorporated, Q3 2019 Earnings Call, Oct 31, 2019.)

   (c) **Adjustments.** Non-GAAP adjusted metrics, specific write-offs etc. Keywords: *accrual, adjusted, adjusting for, adjustment, allowance, benefit, conversion, depreciation, excluding, gaap, gain from, gain, gain on, impairment, non - gaap, non -gaap, non gaap, non- gaap, non-gaap, write - down, write off, write offs, write- offs, writedown, writeoff.*

   > Our MMP segment delivered adjusted earnings of $210 million compared to $300 million in the same period last year. Weak

39

refining results affected MMP negatively this quarter. In addition, we also had a negative timing effect related to the valuation of gas in storage. According to accounting principles, we had a write-down of the gas inventories due to the drop in gas prices in the quarter. However, the gas inventory is sold forward at higher prices and expect to deliver is during winter. Without these timing effects, MMP would have delivered adjusted earnings within the guided range. (Equinor ASA, Q2 2019 Earnings Call, Jul 25, 2019.)

(d) **Lending.** Financial metrics specific for lending institutions. Keywords: *bond, borrowing, capital ratio, debt, deposit, financing, interest, interest income, loan, noninterest income, return on assets, return on equity, roa, roe.*

Both Medallion Bank and Medallion Capital continue to add shareholder value. The consumer and commercial business segments recorded $27.5 million of net interest income for the third quarter and $77 million for the first 9 months of the year. Our consumer segments continue to grow as demand for the bank's products remain strong. In the 2019 9 months, Medallion Bank originated $351.8 million of recreational and home improvement loans compared to $333.4 million in the same period last year, including $125.3 million in the 2019 third quarter. As of September 30, our net consumer portfolio stood at $919 million, a 24% increase from the 2018 third quarter and has grown 21% year-to-date. (Medallion Financial Corp., Q3 2019 Earnings Call, Oct 31, 2019.)

(e) **Financing.** Equity and debt raising, dividends and stock repurchases. Keywords: *dividend, equity raise, loan financing, pay interest, payback, payout, raise equity, repurchase, share purchase, stock purchase.*

Got it. And just for – and you can do this in multiple phases over the next 12 months? Or is it a 1/1 period you will get the full equity raise or one date, rather? (Regency Centers Corporation, Q3 2019 Earnings Call, Oct 31, 2019.)

2. **Operations management and marketing**. Production, delivery, supply and demand, and supply chain.

(a) **Operational and marketing metrics.** Metrics that are not reported in traditional financial statements, like orders, backlog, volume, subscrip-

40

tions etc. Keywords: *asset, backlog, barrel, booking, commission, orders, shipment, subscription, traffic, unit, volume.*

> Moving to Page 10 and ER. Q3 2019 revenue grew 7% to $64 million from $60 million in the prior year. The growth in ER revenue was driven by higher volume and pricing. Adjusted EBITDA in Q3 declined to a loss of $1.1 million from a profit of $0.6 million in the prior year. The decline was primarily the result of unfavorable mix, higher material cost and was partially offset by pricing and volume. Our backlog grew 11% to $195 million compared to $176 million in the prior year. (Spartan Motors, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

(b) **Segments.** Specific lines of business. Keywords: *application, business, category, group, product, project, segment, service, subsidiary.*

> Turning to our business highlights by segment. In Personal Lines, we grew our topline by approximately 6%, as a result of new business growth at mid-single-digit rate increases. Our deep agency partnerships and account focus remain important differentiators for us in the market, especially in the face of increasing competition. We continue to increase our penetration in targeted markets, executing on our strategy to be our agents carrier of choice for preferred account business. Account business now represents 85% of both new business and our overall portfolio. (The Hanover Insurance Group, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

(c) **Supply chain.** Supply and demand, and specifically supply chain. Keywords: *demand, supplier, supply, supply chain.*

> As discussed in prior calls, we expect our largest OEM customer for ligands to diversify their supply chain next year, which will reduce overall ligand demand for Repligen in 2020. To limit the impact on our overall proteins business, we continue to invest in a Repligen-owned ligand portfolio, and we are very encouraged by reports that the Protein A resin utilizing our NGL-Impact A ligand is gaining traction in the marketplace, with a number of key accounts expected to scale up over the next few years. (Repligen Corporation, Q3 2019 Earnings Call, Oct 31, 2019.)

(d) **Production.** Production cycle, including development, launch, production etc. Keywords: *channel, completion, construct, develop, downstream, exploration, franchise, install, inventory management, open, out - source,*

outsource, overhaul, produce, production, relocate, remodel, repair, upgrage, upstream.

> And then, James, if I could, you're guiding up next quarter led by Defense and you called out GaN and X-band. Are you looking at production now with some of these large systems like SPY-60 or Gator? Or is it more development work? And if it's the former, can you give some kind of color on the run? I know this stuff has been in development for many years now, but some of these are very large systems with big unit volumes in the long term, just trying to get a feel for how Defense will play out over the next – well, long term actually, the next 12 months or so? And then I have a follow up. (Qorvo, Inc., Q2 2020 Earnings Call, Oct 31, 2019.)

(e) **Interruptions.** Disruptions of firm operations including interruptions and shutdowns. Keywords: *delay, interruption, shutdown.*

> As you know, we did encounter some challenges to begin the new year in early January during a time of extreme winter weather conditions. The Oil Sands base plant incurred a power interruption, which resulted in a controlled shutdown of extraction and upgrading. We executed on a very disciplined recovery process and returned the assets to service, with no lasting impacts to overall operations. We are now back at full production rates, and we remain on track to meet our guidance commitments for the year. (Suncor Energy Inc., Q4 2017 Earnings Call, Feb 08, 2018).

(f) **Marketing.** Customers, prices and promotions. Keywords: *ad, check, client, customer, launch, population, price, promotion.*

> Okay. Great. And you mentioned that ESG dealers place some orders early in the quarter. I wonder why that timing was there. Was there a price increase that they were getting in front of? Or why those orders came in, was there sell-through to end customers? Was there a customer attached to those orders? (Federal Signal Corporation, Q3 2019 Earnings Call, Oct 31, 2019.)

3. **Global economics.** Discussions of global, national and regional economy as well as factors such as weather.

(a) **Foreign exchange.** Foreign exchange and currency environment. Keywords: *currency, foreign exchange, forex, fx.*

> As a result of prevailing foreign exchange conditions, we now expect that foreign exchange will result in a 225-basis point head-

42

wind to full year revenue as compared to our previous expectation of a 150-basis point headwind. As such, we are lowering our as-reported revenue growth guidance from a range of between 6% and 6.5% to a revised range of between 5.75% and 6%. (Teleflex Incorporated, Q3 2019 Earnings Call, Oct 31, 2019.)

(b) **Seasonality and weather.** Seasonal factors including weather. Keywords: *season, seasonality, seasonally, weather.*

Even with the late season weather slowdown across much of the West, revenue, profitability and earnings improved significantly in 2018, and we expect 2019 to be another year of significant top and bottom line expansion. (Granite Construction Incorporated, Q4 2018 Earnings Call, Feb 20, 2019.)

(c) **General global economics.** General economic environment or conditions in which the firm operates. Keywords: *conditions, country, economic, economy, environment, export, fluctuation, geographic, import, inflation, international, macro, macroeconomy, region, stability, trend, uncertainty, world.*

We continue to meet competitive challenges by substantially reducing component and manufacturing costs, while introducing new products that improve productivity and increase flexibility for our customers. As widely reported, the macro economy has softened and the geopolitical climate has become more unstable over the last few months. As a result, we again found ourselves in a more uncertain position with limited near-term visibility to improved – improving business conditions. Despite the volatile geopolitical and business environment, we continue to invest in new products and application to substantially enhance our competitive position. (IPG Photonics Corporation, Q2 2019 Earnings Call, Jul 30, 2019.)

4. **Strategy.** Discussions of other parties with which the companies engage in strategic interactions, including competitors, partners and governments. Also discussions of mergers and restructuring, and general corporate strategy.

(a) **Competition.** Firm's competitors and competitive advantage. Keywords: *compete, competition, competitive, industry, market.*

You'll have to excuse me, Anthony, I don't go into the details on the pricing. I'm sure our competitors listen to the call, same as anyone else. So I don't want to share that for competitive reasons.

43

But I can tell you, Anthony, that we feel incredibly confident in our pricing strategy and we believe that it will not be a barrier to adoption. (Teleflex Incorporated, Q3 2019 Earnings Call, Oct 31, 2019.)

(b) **Expansion.** Discussions of M&A, integrations and investments. Keywords: *acquisition, integration, investment, m a, m&a, merger, reinvestment.*

Now let me provide some more background on our latest acquisition. In late August, we purchased the assets of Backflow Direct, a California-based company that designs and manufactures large-diameter stainless steel backflows, primarily used in commercial fire protection applications. Backflow Direct was established about 7 years ago, and the founder is well known in the backflow prevention industries and to Watts as well. Strategically, Backflow Direct provides some innovative products in fire protection applications that broaden our offerings to meet customers' requirements. We are also excited that the former owner, a well-respected professional in backflow prevention, is joining our team to aid in future backflow development opportunities. We are excited to add Backflow Direct to our portfolio. (Watts Water Technologies, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

(c) **Contraction.** Divestitures and divestments. Keywords: *divestiture, divestment.*

Yes. So about half of the revenue change is due to the divestitures, and a good chunk is due to FX. When I look at the range, I also hedge the range based on the lumpiness that we tend to see in the recall business, and we continue to see. But predominantly, it's FX and divestiture, and divestiture is about half of that. When I look at – on an EBITDA basis, the divestitures is about 1/3 of that, again, reflecting that these are lower-margin businesses. (Stericycle, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

(d) **Partners.** Partners and joint ventures. Keywords: *joint venture, jv, partner.*

The first and largest is our agreement to form a 90/10 joint venture with a local real estate operator, we're the 90, for an initial 40-plus individual street retail properties in Hoboken, New Jersey. Our share of the investment approximates $185 million. The

44

properties, mostly apartments over Street retail, our prime retail – prime real estate sites on either Washington Street or 14th Street, two of Hoboken's main commercial thoroughfares. We're very bullish on Hoboken and its access to the increasingly important west side of Manhattan, including the $20-billion-plus Hudson Yards development. That access is easier than in any areas of Manhattan through the path, Ferry and the bus through the immediately adjacent Lincoln Tunnel. One or more transportation choices of which is walkable from the buildings we're buying. (Federal Realty Investment Trust, Q3 2019 Earnings Call, Oct 31, 2019.)

(e) **Deals.** Contracts, purchases and payments etc. Keywords: *award, bid, close, contract, deal, payment, purchase, transaction, win.*

Importantly, if we conclude the transaction, which we believe is likely, Noble will receive a payout of its share of the drilling contract and be free to market the Bully II. Given the effect of the advanced stages of the negotiations, we recognized a gross impairment on the Bully II in the third quarter of $596 million or $331 million net of noncontrolling interest. We would expect to receive the lump-sum payment in the fourth quarter of this year. (Noble Corporation plc, Q3 2019 Earnings Call, Oct 31, 2019.)

(f) **Government.** Government actions, regulations and approvals. Keywords: *action, approval, compliance, government, law, legal resolution, legal settlement, license, patent, regulation, regulator, regulatory, requirement, rule, trial.*

Additionally, the controlled rollout in Japan continues to gain traction as we generated $9.6 million of sales in the quarter due to new sites and patient utilization. We opened 17 sites in Q2 and now are in approximately 100 hospitals, with 135 hospitals currently approved by the government. It is worth noting that we do expect to open fewer sites in the second half of the year as the team closes out the post-approval study and prioritizes a broader CP launch, which will require additional training at existing sites. (Abiomed, Inc., Q2 2020 Earnings Call, Oct 31, 2019.)

(g) **Restructuring.** Restructuring, reorganization and lay-offs. Keywords: *lay - off, lay - offs, layoff, reorganization, restructuring.*

45

Then in addition, we put in an additional restructuring program in 2019, and that program doesn't generate savings until 2021. And there's also cash outflows that we have to invest in the project in 2019. There's also CapEx associated with it. (Teleflex Incorporated, Q3 2019 Earnings Call, Oct 31, 2019.)

(h) **General strategy.** Strategy, initiatives and growth. Keywords: *efficiency, growth, initiative, opportunity, strategic, strategically, strategy.*

You guys have been very consistent with your strategy of organic investment and driving growth and margin expansion. So I'm going to focus first on growth. Your comments on the macro were helpful. I was wondering though if you could talk about some of your strategies of penetrating some of your international markets like the Middle East. What does that look going forward? Are you seeing traction there? And then also, if you could talk about some of your connected solutions? And maybe an update there. (Watts Water Technologies, Inc., Q3 2019 Earnings Call, Oct 31, 2019.)

5. **Forward-looking.** Paragraphs including forward-looking phrases that are identified following Muslu et al. (2015).

As a result of these changes, our revenues from new products introduced in the past 5 years is now at 15% compared to approximately 11% in 2016. We aim to increase this number to above 20% by 2024. (Luxfer Holdings PLC, Q3 2019 Earnings Call, Oct 31, 2019.)

46

References

Ali, A., Chen, X., Yao, T., Yu, T., 2020. Can mutual funds profit from post earnings announcement drift? The role of competition. Journal of Banking & Finance 114, 105774.

Ball, R., Brown, P., 1968. An empirical evaluation of accounting income numbers. Journal of Accounting Research 6, 159.

Bartov, E., Radhakrishnan, S., Krinsky, I., 2000. Investor sophistication and patterns in stock returns after earnings announcements. The Accounting Review 75, 43–63.

Bernard, V.L., Thomas, J.K., 1989. Post-earnings-announcement drift: Delayed price response or risk premium? Journal of Accounting Research 27, 1.

Brandt, M.W., Kishore, R., Santa-Clara, P., Venkatachalam, M., 2008. Earnings announcements are full of surprises. Working paper.

Brown, N.C., Crowley, R.M., Elliott, W.B., 2020. What are you saying? Using topic to detect financial misreporting. Journal of Accounting Research 58, 237–291.

Cao, S., Jiang, W., Wang, J.L., Yang, B., 2021. From man vs. machine to man+ machine: The art and AI of stock analyses. Technical Report. National Bureau of Economic Research.

Cao, S., Jiang, W., Yang, B., Zhang, A.L., 2020. How to talk when a machine is listening: Corporate disclosure in the age of AI. Technical Report. National Bureau of Economic Research.

Cao, S.S., Narayanamoorthy, G., 2012. Earnings volatility, post–earnings announcement drift, and trading frictions. Journal of Accounting Research 50, 41–74.

Carhart, M.M., 1997. On persistence in mutual fund performance. The Journal of Finance 52, 57–82.

Chordia, T., Subrahmanyam, A., Tong, Q., 2014. Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? Journal of Accounting and Economics 58, 41–58.

Coleman, B., Merkley, K.J., Pacelli, J., 2020. Man versus machine: A comparison of robo-analyst and traditional research analyst investment recommendations. Available at SSRN 3514879 .

Cox, J., 2020. Market fragmentation and post-earnings announcement drift. Journal of Economics and Finance 44, 587–610.

Engelberg, J., 2008. Costly information processing: Evidence from earnings announcements. Working paper.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Fink, J., 2021. A review of the post-earnings-announcement drift. Journal of Behavioral and Experimental Finance 29, 100446.

Frankel, R., Jennings, J., Lee, J., 2016. Using unstructured and qualitative disclosures to explain accruals. Journal of Accounting and Economics 62, 209–227.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33, 1–22.

Garcia, D., Hu, X., Rohrer, M., 2021. The colour of finance words. Working paper.

Grennan, J., Michaely, R., 2020. Artificial intelligence and high-skilled work: Evidence from analysts. Available at SSRN .

Hou, K., Mo, H., Xue, C., Zhang, L., 2020. An augmented q-factor model with expected growth. Review of Finance 25, 1–41.

Jegadeesh, N., Livnat, J., 2006. Revenue surprises and stock returns. Journal of Accounting and Economics 41, 147–171.

Ke, Z.T., Kelly, B., Xiu, D., 2020. Predicting returns with text data. Working paper.

Kogan, S., Levin, D., Routledge, B.R., Sagi, J.S., Noah Smith, O.A., 2009. Predicting risk from financial reports with regression. Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL, Boulder, Colorado, 272–280.

Lee, Y.J., 2012. The effect of quarterly report readability on information efficiency of stock prices. Contemporary Accounting Research 29, 1137–1170.

Li, F., 2010. The information content of forward-looking statements in corporate filings—a naïve bayesian machine learning approach. Journal of Accounting Research 48, 1049–1102.

Livnat, J., Mendenhall, R.R., 2006. Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts. Journal of Accounting Research 44, 177–205.

Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. The Journal of Finance 66, 35–65.

Martineau, C., 2021. Are there (still) post-earnings announcement drifts? Working paper.

Mendenhall, R.R., 2004. Arbitrage risk and post-earnings-announcement drift. The Journal of Business 77, 875–894.

Milian, J.A., 2015. Unsophisticated arbitrageurs and market efficiency: Overreacting to a history of underreaction? Journal of Accounting Research 53, 175–220.

Muslu, V., Radhakrishnan, S., Subramanyam, K.R., Lim, D., 2015. Forward-looking MD&A disclosures and the information environment. Management Science 61.

Narayanamoorthy, G., 2006. Conservatism and cross-sectional variation in the post–earnings announcement drift. Journal of Accounting Research 44, 763–789.

Ou, J.A., Penman, S.H., 1989. Financial statement analysis and the prediction of stock returns. Journal of accounting and economics 11, 295–329.

Richardson, S., Tuna, I., Wysocki, P., 2010. Accounting anomalies and fundamental analysis: A review of recent research advances. Journal of Accounting and Economics 50, 410–454.

Routledge, B.R., Sacchetto, S., Smith, N.A., 2013. Predicting merger targets and acquirers from text. Working paper.

Yan, X., Zheng, L., 2017. Fundamental analysis and the cross-section of stock returns: A data-mining approach. The Review of Financial Studies 30, 1382–1423.

Yano, T., Smith, N.A., Wilkerson, J.D., 2012. Textual predictors of bill survival in congressional committees, in: Proceedings of the 2012 Conference of the North

49

American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Montréal, Canada. pp. 793–802.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B 67, 301–320.

Figures

Figure 1: Comparison of text-based post-earnings-announcement drift (PEAD.txt) and classic PEAD. The lines represent cumulative abnormal returns (CAR) of spread portfolios formed on the day following the earnings call that buy the stocks that we estimate to be in the top quintile of SUE.txt or SUE in a given quarter and short the stocks in the bottom quintile. We calculate the abnormal returns using the returns on the matched six size and book-to-market portfolios. The starting point is the day after the earnings call. The sample period is 2010 to 2019.

51
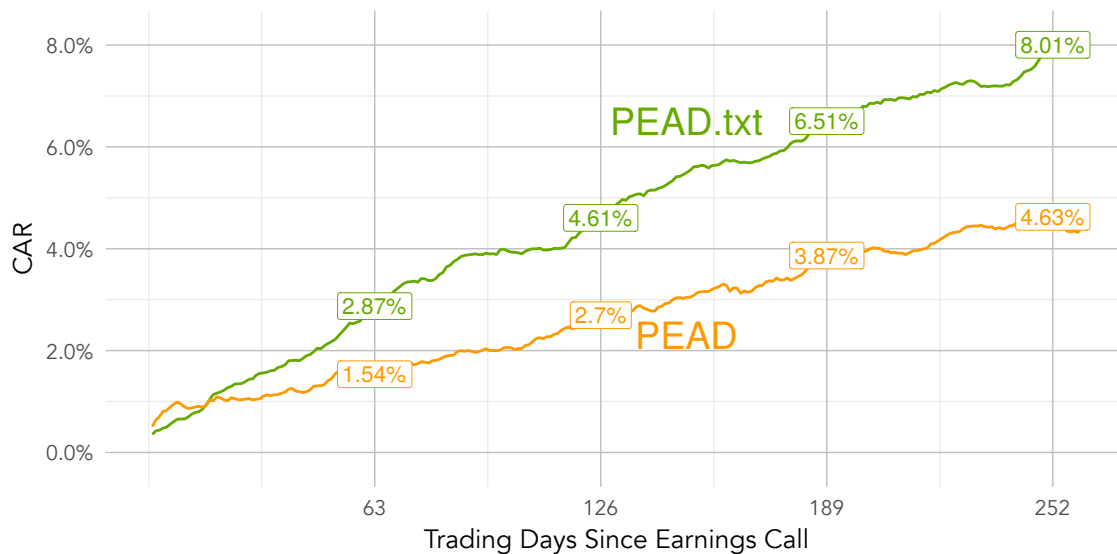
Figure 2: PEAD.txt and PEAD across years, part 1. The lines represent cumulative abnormal returns (CAR) of spread portfolios formed on the day following the earnings call that buy the stocks that we estimate to be in the top quintile of SUE.txt or SUE in a given quarter and short the stocks in the bottom quintile. We calculate the abnormal returns using the returns on the matched six size and book-to-market portfolios. The starting point is the day after the earnings call. The labels correspond to PEAD.txt and PEAD at the 252 tradings days mark (one calendar year).
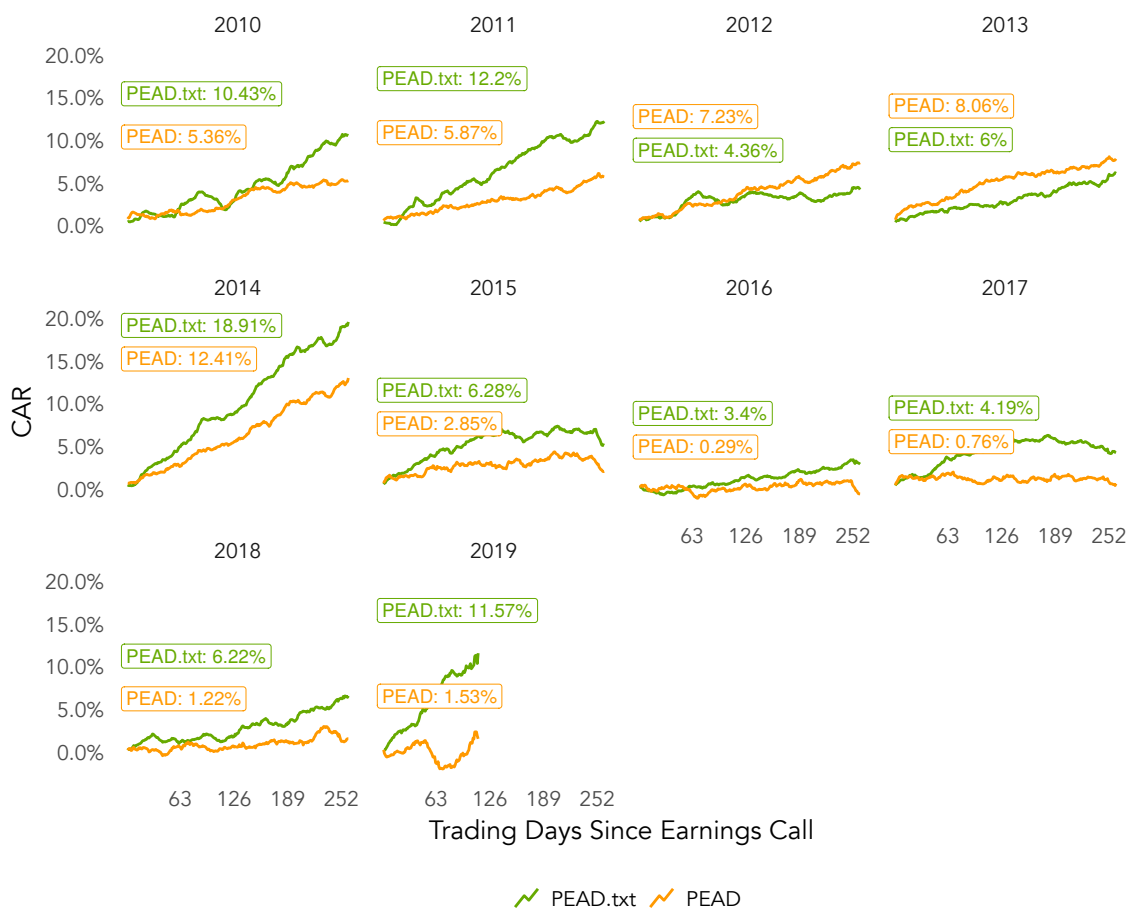
Figure 3: Tokens with largest positive and negative impact. The tokens above zero are positively associated with high returns and / or negatively associated with low returns. Coefficients are normalized by the largest absolute value. The x-axis is average log frequency of tokens across all documents. Impact is the product of $\beta^H - \beta^L$ and the mean log frequency. High impact values are associated with high returns ("good news"), and low impact values are associated with low returns ("bad news"). We display the top fifteen tokens with the largest positive impact and the top fifteen tokens with the largest negative impact. PR indicates presentation section, QA indicates Q&A section.
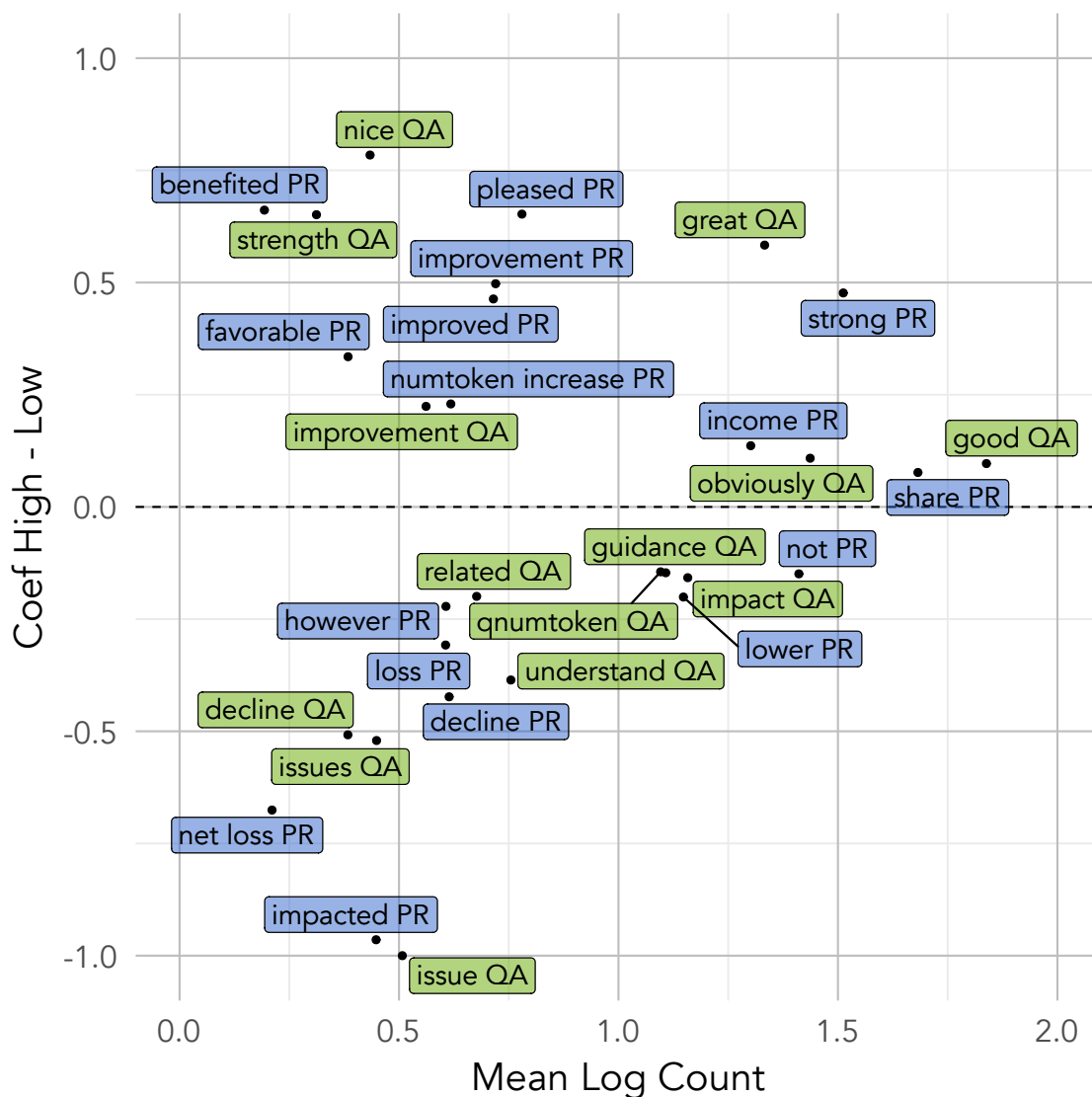
Figure 4: Absolute value of paragraph-level SUE.txt (SUE.txt^P) and prevalence of paragraph subgroups. The three-letter abbreviations refer to paragraph groups based on the business curriculum. Y-axis represents mean absolute value of SUE.txt^P normalized by the largest absolute value. SUE.txt^P aggregates the coefficients of words in the paragraph with log frequency adjustments. High impact values are associated with high returns ("good news"); low impact values are associated with low returns ("bad news"). X-axis represents the percentage of paragraphs belonging to a given subgroup. The dotted lines represent the x- and y-axis values whose product is equal to (right to left) 100%, 75%, 50% and 25% of the largest product among the subgroups. A paragraph can belong to more than one subgroup.
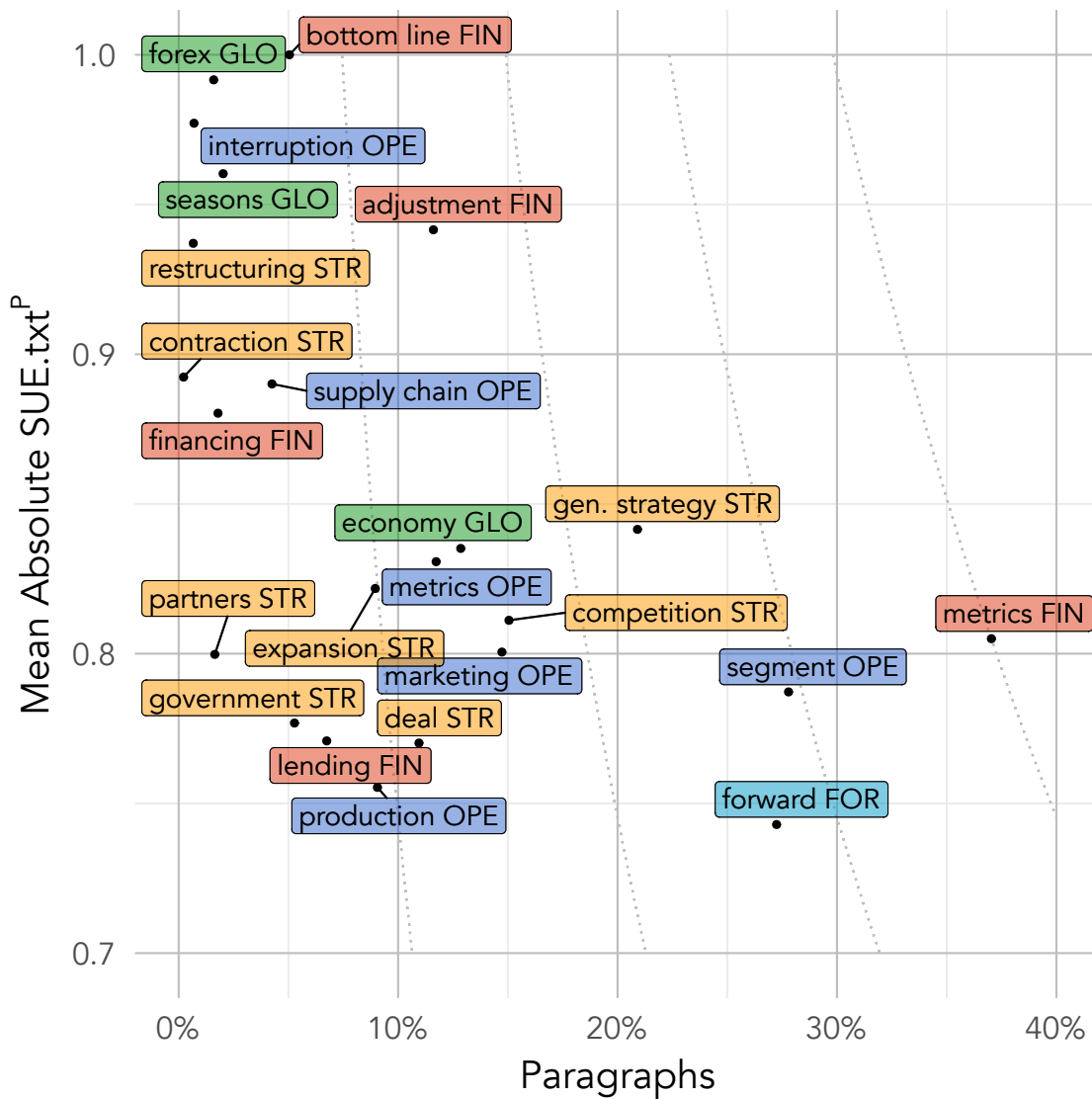
Figure 5: Comparison of drifts created using different methods.
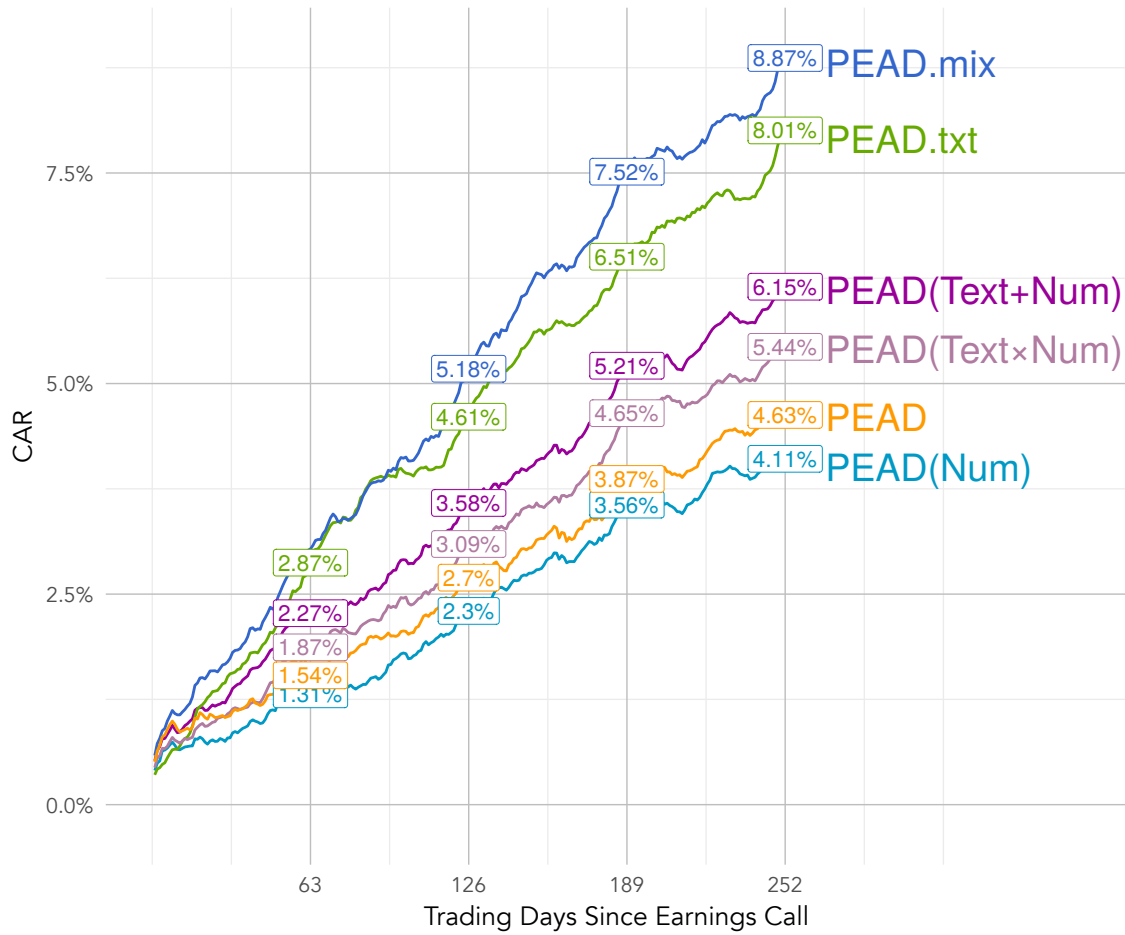
Tables

Table 1: Earnings call text surprise construction: summary statistics for the combined dataset.

	Value
Total Obs	108,704
Median Tokens Pres.	2,833
Median Tokens Q&A	4,018
Median AR	0.02%
Median AR Cutoff	+-1.87%
AR Split	33%/34%/33%

Table 2: Sliding window setup. At each iteration, we use eight consecutive quarters of observations as the training set, and one subsequent quarter as the test set. Number of observations is given in parentheses.

	Train Quarters	Test Quarter
1	2008Q1 – 2009Q4 (14,748 obs.)	2010Q1 (2022 obs.)
2	2008Q2 – 2010Q1 (16,584 obs.)	2010Q2 (2487 obs.)
...	...	...
40	2018Q4 – 2019Q3 (21,378 obs.)	2019Q4 (2,219 obs.)

56

Table 3: PEAD.txt and PEAD comparison. We calculate earnings call text surprises using the output of a regularized logistic text regression that predicts one-day return. Earnings surprises are standardized unexpected earnings calculated using the analyst forecasts. We calculate AR and CAR using the returns on the matched six size and book-to-market portfolios.

Quintile	AR(0)	CAR(1,63)	CAR(1,32)	CAR(33,63)
SUE.txt (PEAD.txt)				
Q 1	-0.0288	-0.0152	-0.0089	-0.0064
Q 2	-0.0075	-0.0102	-0.0057	-0.0045
Q 3	0.0022	0.0003	-0.0002	0.0005
Q 4	0.0089	0.0041	0.0021	0.0020
Q 5	0.0201	0.0131	0.0066	0.0064
Spread	**0.0489**	**0.0287**	**0.0156**	**0.0129**
SUE (PEAD)				
Q 1	-0.0325	-0.0136	-0.0093	-0.0043
Q 2	-0.0146	-0.0020	-0.0011	-0.0008
Q 3	0.0024	0.0032	0.0015	0.0017
Q 4	0.0156	0.0043	0.0018	0.0025
Q 5	0.0285	0.0016	0.0017	-0.0001
Spread	**0.0610**	**0.0154**	**0.0111**	**0.0042**

Table 4: PEAD.txt and other drifts comparison. We calculate earnings call text surprises using the output of a regularized logistic text regression that predicts one-day return based only on earnings call text. Numeric and text split uses the output of a regularized logistic regression that predicts one-day return based on earnings call text and an array of numerical variables. Sentiment dictionary (negative) split uses percentage of negative words identified using the financial domain sentiment dictionary (Loughran and McDonald, 2011). AR(0) splits on abnormal returns on the earnings call day. We calculate AR and CAR using the returns on the matched six size and book-to-market portfolios.

Quintile	AR(0)	CAR(1,63)	CAR(1,32)	CAR(33,63)
SUE.txt (PEAD.txt)				
Q 1	-0.0288	-0.0152	-0.0089	-0.0064
Q 3	0.0022	0.0003	-0.0002	0.0005
Q 5	0.0201	0.0131	0.0066	0.0064
Spread	**0.0489**	**0.0287**	**0.0156**	**0.0129**
Numeric and Text				
Q 1	-0.0374	-0.0120	-0.0084	-0.0036
Q 3	0.0002	-0.0016	0.0003	-0.0019
Q 5	0.0334	0.0104	0.0053	0.0051
Spread	**0.0707**	**0.0227**	**0.0138**	**0.0088**
Sent. Dict. (Neg.)				
Q 1	-0.0119	-0.0065	-0.0036	-0.0029
Q 3	-0.0008	-0.0026	-0.0010	-0.0016
Q 5	0.0111	0.0047	0.0019	0.0028
Spread	**0.0231**	**0.0113**	**0.0055**	**0.0058**
AR(0)				
Q 1	-0.0949	-0.0099	-0.0058	-0.0040
Q 3	0.0003	0.0012	-0.0011	0.0023
Q 5	0.0927	0.0065	0.0064	0.0002
Spread	**0.1876**	**0.0165**	**0.0122**	**0.0042**

Table 5: Earnings call text surprise and cumulative abnormal returns regression, specification comparison. We calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts one-day return. We calculate CAR using the returns on the matched six size and book-to-market portfolios. Errors are clustered at the firm and year-quarter level.

	CAR(1,63)				
	Model 1	Model 2	Model 3	Model 4	Model 5
SUE.txt	0.06***	0.06***	0.03**	0.05***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SUE	0.02***	0.02***	0.01	0.01*	0.18
	(0.01)	(0.01)	(0.01)	(0.01)	(0.10)
Sent. Dict. (Neg.)				−0.01	−0.01
				(0.01)	(0.01)
AR(0)				−0.01	−0.01
				(0.01)	(0.01)
CAR(-31,-1)				−0.05***	−0.05***
				(0.01)	(0.01)
Size				−0.70***	−0.70***
				(0.07)	(0.07)
Turnover				0.03*	0.03*
				(0.01)	(0.01)
I. Vol.				−0.06**	−0.06**
				(0.02)	(0.02)
Coverage				−0.00	−0.00
				(0.00)	(0.00)
SUE * Size					−0.12
					(0.09)
SUE * Turnover					0.01
					(0.01)
SUE * I. Vol.					−0.05**
					(0.02)
SUE * Coverage					−0.01
					(0.01)
Num. obs.	85160	85160	85160	85160	85160
Fixed Effects	None	Ind, YQ	Firm, YQ	Firm, YQ	Firm, YQ
Adj. R^2	0.00	0.02	0.05	0.08	0.08

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

59

Table 6: Earnings call text surprise and cumulative abnormal returns regression, timing comparison. We calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts one-day return. We calculate CAR using the returns on the matched six size and book-to-market portfolios. Errors are clustered at the firm and year-quarter level.

	CAR(1,63)	CAR(1,32)	CAR(33,63)
SUE.txt	0.05***	0.04***	0.03**
	(0.01)	(0.01)	(0.01)
SUE	0.01*	0.03***	−0.01
	(0.01)	(0.01)	(0.01)
Sent. Dict. (Neg.)	−0.01	−0.01	−0.01
	(0.01)	(0.01)	(0.01)
AR(0)	−0.01	0.01	−0.02***
	(0.01)	(0.01)	(0.00)
CAR(-31,-1)	−0.05***	−0.01	−0.05***
	(0.01)	(0.02)	(0.01)
Size	−0.70***	−0.51***	−0.46***
	(0.07)	(0.08)	(0.05)
Turnover	0.03*	−0.01	0.04**
	(0.01)	(0.01)	(0.01)
I. Vol.	−0.06**	−0.04	−0.03*
	(0.02)	(0.02)	(0.01)
Coverage	−0.00	0.00	−0.01**
	(0.00)	(0.00)	(0.00)
Num. obs.	85160	85160	85160
Fixed Effects	Firm, YQ	Firm, YQ	Firm, YQ
Adj. R^2	0.08	0.05	0.05

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

60

Table 7: Alpha for different quintiles of earnings call text surprise, 63 trading days, Fama–French five factors plus momentum. We calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts one-day return.

	SUE.txt Portfolios					
	Spread	Q 1	Q 2	Q 3	Q 4	Q 5
Alpha	0.039***	−0.020***	−0.008	0.001	0.009**	0.020***
	(0.007)	(0.006)	(0.005)	(0.004)	(0.003)	(0.004)
MKT	−0.003	1.022***	0.996***	0.997***	1.001***	1.020***
	(0.010)	(0.010)	(0.008)	(0.006)	(0.005)	(0.005)
SMB	−0.117***	0.684***	0.660***	0.633***	0.613***	0.567***
	(0.016)	(0.013)	(0.011)	(0.010)	(0.008)	(0.010)
HML	−0.134***	0.133***	0.035	0.062***	0.024	−0.003
	(0.027)	(0.023)	(0.018)	(0.015)	(0.013)	(0.013)
RMW	0.056	−0.124***	−0.159***	−0.125***	−0.100***	−0.067***
	(0.029)	(0.025)	(0.019)	(0.017)	(0.014)	(0.018)
CMA	0.002	0.053	0.107***	0.092***	0.074***	0.057***
	(0.033)	(0.030)	(0.021)	(0.021)	(0.017)	(0.017)
UMD	27.218***	−24.771***	−18.879***	−11.292***	−7.103***	2.468***
	(1.472)	(1.325)	(1.139)	(0.938)	(0.719)	(0.675)
Num. obs.	2379	2376	2369	2376	2377	2379
Adj. R^2	0.319	0.945	0.966	0.974	0.981	0.975

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

61

Table 8: Alpha for different quintiles of classic earnings surprise, 63 trading days, Fama–French five factors plus momentum. Earnings surprises are standardized unexpected earnings calculated using the analyst forecasts.

	SUE Portfolios					
	Spread	Q 1	Q 2	Q 3	Q 4	Q 5
Alpha	0.026***	−0.014*	−0.001	0.004	0.006*	0.013
	(0.007)	(0.006)	(0.003)	(0.003)	(0.003)	(0.007)
MKT	0.027***	1.000***	0.965***	0.999***	1.023***	1.028***
	(0.007)	(0.010)	(0.005)	(0.004)	(0.005)	(0.009)
SMB	−0.009	0.781***	0.552***	0.442***	0.604***	0.772***
	(0.011)	(0.016)	(0.008)	(0.007)	(0.008)	(0.014)
HML	−0.009	0.069**	0.038***	−0.019*	0.050***	0.058*
	(0.021)	(0.021)	(0.009)	(0.009)	(0.011)	(0.023)
RMW	0.053	−0.315***	−0.030*	0.029*	−0.048***	−0.259***
	(0.029)	(0.027)	(0.012)	(0.013)	(0.014)	(0.030)
CMA	0.016	0.137***	0.064***	0.037**	0.043**	0.153***
	(0.028)	(0.032)	(0.014)	(0.013)	(0.016)	(0.028)
UMD	12.942***	−31.316***	−9.406***	0.787	−4.755***	−18.328***
	(0.971)	(1.429)	(0.564)	(0.552)	(0.601)	(1.405)
Num. obs.	2379	2377	2378	2378	2378	2379
Adj. R^2	0.078	0.942	0.982	0.984	0.984	0.939

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Table 9: Alpha for different spread portfolios, 63 trading days, Fama–French five factors plus momentum. We calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts one-day return. Earnings surprises are standardized unexpected earnings calculated using the analyst forecasts. SUE.txt and SUE is a strategy that equally weights earnings call text surprises and earnings surprises signals. Numeric and text split uses the output of a regularized logistic regression model that predicts one-day return based on earnings call text and an array of numerical variables. Sentiment dictionary (negative) split uses percentage of negative words identified using the financial domain sentiment dictionary (Loughran and McDonald, 2011).

	Spread Portfolios				
	SUE.txt	SUE	SUE.txt and SUE	Numeric and Text	Sent. Dict.
Alpha	0.039***	0.026***	0.042***	0.035***	0.004
	(0.007)	(0.007)	(0.008)	(0.006)	(0.006)
MKT	−0.003	0.027***	0.036***	0.003	−0.013
	(0.010)	(0.007)	(0.009)	(0.008)	(0.007)
SMB	−0.117***	−0.009	−0.070***	−0.043**	−0.022
	(0.016)	(0.011)	(0.014)	(0.014)	(0.012)
HML	−0.134***	−0.009	−0.070**	−0.052*	−0.313***
	(0.027)	(0.021)	(0.023)	(0.023)	(0.019)
RMW	0.056	0.053	0.084**	0.017	−0.000
	(0.029)	(0.029)	(0.030)	(0.027)	(0.023)
CMA	0.002	0.016	0.011	0.055*	−0.052*
	(0.033)	(0.028)	(0.028)	(0.027)	(0.026)
UMD	27.218***	12.942***	26.358***	22.245***	23.252***
	(1.472)	(0.971)	(1.271)	(1.270)	(1.017)
Num. obs.	2379	2379	2379	2379	2379
Adj. R^2	0.319	0.078	0.247	0.225	0.490

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

63

Table 10: Alpha for different spread portfolios, 32 trading days, Fama–French five factors plus momentum. We calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts one-day return. Earnings surprises are standardized unexpected earnings calculated using the analyst forecasts. SUE.txt and SUE is a strategy that equally weights earnings call text surprises and earnings surprises signals. Numeric and text split uses the output of a regularized logistic regression model that predicts one-day return based on earnings call text and an array of numerical variables. Sentiment dictionary (negative) split uses percentage of negative words identified using the financial domain sentiment dictionary (Loughran and McDonald, 2011).

	Spread Portfolios				
	SUE.txt	SUE	SUE.txt and SUE	Numeric and Text	Sent. Dict.
Alpha	0.034***	0.043***	0.053***	0.040***	0.010
	(0.009)	(0.010)	(0.010)	(0.009)	(0.008)
MKT	0.005	0.046***	0.048***	0.013	−0.011
	(0.013)	(0.011)	(0.012)	(0.011)	(0.009)
SMB	−0.122***	−0.027	−0.076***	−0.042*	−0.045*
	(0.019)	(0.023)	(0.021)	(0.018)	(0.019)
HML	−0.126***	0.024	−0.055	−0.044	−0.302***
	(0.031)	(0.031)	(0.031)	(0.030)	(0.025)
RMW	0.079*	0.083*	0.103**	0.040	0.023
	(0.037)	(0.040)	(0.039)	(0.035)	(0.031)
CMA	0.031	0.036	0.046	0.071	−0.039
	(0.042)	(0.042)	(0.043)	(0.039)	(0.036)
UMD	29.475***	15.095***	27.905***	23.188***	26.441***
	(1.747)	(1.625)	(1.677)	(1.613)	(1.402)
Num. obs.	2379	2379	2379	2379	2379
Adj. R^2	0.226	0.048	0.158	0.136	0.345

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Table 11: Alpha for different spread portfolios, 63 trading days, q5 factors (Hou et al., 2020). We calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts one-day return. Earnings surprises are standardized unexpected earnings calculated using the analyst forecasts. SUE.txt and SUE is a strategy that equally weights earnings call text surprises and earnings surprises signals. Numeric and text split uses the output of a regularized logistic regression model that predicts one-day return based on earnings call text and an array of numerical variables. Sentiment dictionary (negative) split uses percentage of negative words identified using the financial domain sentiment dictionary (Loughran and McDonald, 2011).

	Spread Portfolios				
	SUE.txt	SUE	SUE.txt and SUE	Numeric and Text	Sent. Dict.
Alpha	0.0372***	0.0248***	0.0397***	0.0330***	0.0058
	(0.0079)	(0.0071)	(0.0081)	(0.0069)	(0.0073)
MKT	0.0363***	0.0445***	0.0765***	0.0371***	0.0058
	(0.0105)	(0.0073)	(0.0095)	(0.0090)	(0.0100)
ME	−0.0735***	0.0074	−0.0301*	−0.0059	−0.0073
	(0.0166)	(0.0117)	(0.0153)	(0.0147)	(0.0171)
IA	−0.0458	0.0514	0.0344	0.0509	−0.2625***
	(0.0339)	(0.0263)	(0.0316)	(0.0287)	(0.0325)
ROE	0.2906***	0.0944***	0.2422***	0.2342***	0.1407***
	(0.0326)	(0.0197)	(0.0299)	(0.0283)	(0.0294)
EG	0.2097***	0.1200***	0.2461***	0.1322***	0.2117***
	(0.0331)	(0.0279)	(0.0350)	(0.0293)	(0.0328)
Num. obs.	2379	2379	2379	2379	2379
Adj. R^2	0.1886	0.0403	0.1493	0.1226	0.1781

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 12: Autocorrelation of SUE.txt. *dSUE.txt* is the decile of *SUE.txt*. *EVol* is earnings volatility, *MktVal* is market value and *Loss* is an indicator variables equal to one if the firm has negative earnings in the quarter. The dependent variables are lagged by one quarter. The standard errors are clustered at the firm level.

	dSUE.txt			
	Model 1	Model 2	Model 3	Model 4
dSUE.txt1	0.28***	0.28***	0.28***	0.29***
	(0.00)	(0.00)	(0.00)	(0.00)
EVol1	−0.02***		−0.02***	−0.00
	(0.01)		(0.01)	(0.01)
dSUE.txt1×EVol1	0.04**		0.02	0.04***
	(0.01)		(0.01)	(0.01)
MktVal1		0.06***	0.06***	
		(0.01)	(0.01)	
dSUE.txt1×MktVal1		0.09***	0.08***	
		(0.01)	(0.01)	
Loss1				−0.10***
				(0.00)
dSUE.txt1×Loss1				−0.05***
				(0.01)
Adj. R^2	0.28	0.27	0.27	0.28
FE	Firm	Firm	Firm	Firm
Num. obs.	79337	78247	77232	79160

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

66

Table 13: Performance of the regularized logistic text regression model on the one-day return prediction task, all test sets combined. Naive benchmark is a "model" that always predicts the largest category in the training set. Text model is the main model we use to construct earnings call text surprises. SUE model predicts one-day returns using SUE, Num model includes an array of market and analyst following-based numerical variables, and Text + Num and Text × Num models uses both the text and numeric variables. Return spread is the difference between the announcement abnormal return of stocks classified as high return and the stocks classified as low return.

Model	Acc	F1 Macro	Return Spread
Naive	34.23%		
Text	46.95%	46.93%	2.49%
SUE	44.99%	44.08%	2.99%
Num	50.62%	50.55%	3.76%
Text + Num	52.03%	51.99%	4.28%
Text × Num	51.46%	51.44%	4.13%