JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 58, No. 6, Sept. 2023, pp. 2299–2326

doi:10.1017/S0022109022001181

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

Vitaly Meursault
Federal Reserve Bank of Philadelphia Research Department
vitaly.meursault@phil.frb.org (corresponding author)

Pierre Jinghong Liang
Carnegie Mellon University Tepper School of Business
liangj@andrew.cmu.edu

Bryan R. Routledge
Carnegie Mellon University Tepper School of Business
routledge@cmu.edu

Madeline Marco Scanlon
University of Pittsburgh Katz School of Business
mms238@pitt.edu

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 0. 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.

I. 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

We are grateful for the helpful comments of the conference and seminar participants at the 2021 SFS Cavalcade North America, 2021 FARS, ETH Zurich, Bocconi University, Indiana University, and Brown University. We thank Andrew Gross for his research assistance with the annotation task. Meursault is grateful for the 2020 William W. Cooper Doctoral Dissertation Award and PNC Presidential Fellowship for Outstanding Research on the Future of Financial Services that supported this project while he was a Ph.D. student at the Tepper School of Business, Carnegie Mellon University. We also thank an anonymous referee, Sean Cao (a referee), and Jarrad Harford (the editor) for their guidance throughout the revision and publication process. The views expressed in this article are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. Any errors or omissions are the responsibility of the authors. No statements here should be treated as legal advice.

(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 article, 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 1-day abnormal returns; its parameters are dynamically calibrated. We document a drift phenomenon associated with standardized unexpected earnings call text (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 2010–2019 and quintile split portfolios (Figure 1). The difference is growing each quarter following the release of the earnings call text: 2.87%–1.54%, 4.61%–2.7%, 6.51%–3.87%, and 8.01%–4.63%. These magnitudes deepen the existing PEAD puzzle.

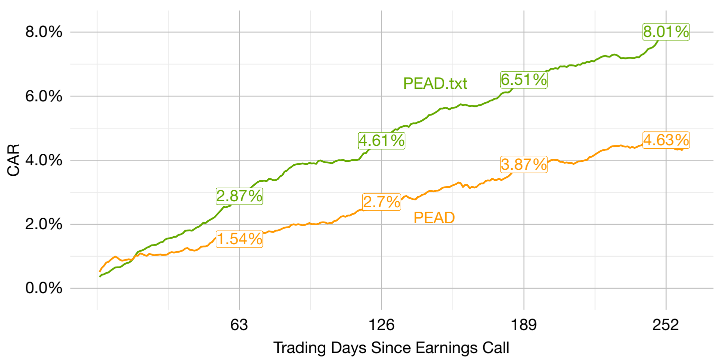
Using panel regressions, we find that the association between SUE.txt and abnormal returns is more than twice as strong as that between SUE and abnormal returns. The relationship persists across specifications with different controls and firm and year-quarter fixed effects.

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

FIGURE 1

Comparison of Text-Based Post-Earnings-Announcement Drift and Classic PEAD

In Figure 1, the lines represent cumulative abnormal returns 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 2010–2019.



the Fama–French 5 Factors Plus momentum framework (Carhart (1997), Fama and French (2015)) and q5 factor framework (Hou, Mo, Xue, and Zhang (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 article 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 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 nonbottom-line financial metrics contribute the most to SUE.txt overall due to their ubiquitousness.

Overall, our article 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's 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 (e.g., tone compared to facts).

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 article'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 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 article contribute to our understanding of two related well-developed literatures: PEAD anomaly and fundamental analysis. We next briefly describe the connections of this article to these two literatures,

https://doi.org/10.1017/S0022109022001181 Published online by Cambridge University Press

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, Radhakrishnan, and Krinsky (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, Subrahmanyam, and Tong (2014), Milian (2015), and Martineau (2021), argue that PEAD has recently shrunk to the point of disappearance. However, other recent studies, like Ali, Chen, Yao, and Yu (2020) and Cox (2020), find that PEAD persists. We document that while both PEAD and PEAD.txt decrease in the second half of our sample, the shrinkage of PEAD.txt is smaller, and it is far from disappearing.

Second, this article 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, Jiang, Wang, and Yang (2021a) incorporates corporate financial information, qualitative disclosure, and macroeconomic indicators. The article shows that this comprehensive AI ensemble model outperforms human analysts as a whole, although human analysts perform better when firms are subject to more information asymmetry (e.g., more illiquid or more intangible assets). In this context, our article 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 thought, our work has implications to the recent literature on robo-analysts (Coleman, Merkley, and Pacelli (2022), Grennan and Michaely (2020)) and the effect of AI readership on corporate disclosure (Cao, Jiang, Yang, and Zhang (2020)).[2]

II. 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

[1]For a comprehensive and more historical review of the literature, see Richardson, Tuna, and Wysocki (2010).

[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 article, 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.

calculation procedure, and finish with a statistical comparison of PEAD.txt and PEAD phenomena.

A. Data Sets

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 CRSP, Compustat, and IBES data sets available through the WRDS platform.[3] The details about data set construction, merging, abnormal returns calculation, and returns timing are in Appendix A of the Supplementary Material.

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

B. 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 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 1-day abnormal returns. We reestimate the model for every quarter using only information from the past 8 quarters as the training set. This procedure ensures that our model is applicable in a dynamic setting. In Section II.B.1, we focus on how our model and estimation procedure allow us to robustly capture unexpected textual content. A technical description of the model is provided in Appendix B of the Supplementary Material.

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 2-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)). Because textual data are high-dimensional, overfitting is a concern. To ensure that our model produces robust measures of surprises, we use standard 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

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

https://doi.org/10.1017/S0022109022001181 Published online by Cambridge University Press

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 does not 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 of 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 use only 1-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. 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 reestimate our model every quarter using the data for 2 previous years. Therefore, we lose only 2 years from our original sample of earnings calls (2008 and 2009).

Our target variable is 1-day abnormal returns split into *high*, *flat*, and *low* categories (see Appendix B of the Supplementary Material 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. Variables and Model Training

The regularized logistic text regression uses log frequencies of individual words (unigrams) and 2-word combinations (bigrams) in documents as independent variables (the bag-of-words approach). Let $\mathrm{freq}(j, n)$ denote the frequency of the term j in the document n. The associated independent variable is $x_{n,j} = \log(1 + \mathrm{freq}(j, \ n))$. 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).[4] We use Snowball stemmer's stopword list to remove some ubiquitous English words like "the" (https://snowballstem.org/ (last accessed Aug. 26, 2020)). 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 lowercase, 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 data set 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 is approximately 3,000 words long. Q&A sections are even larger; the median one is approximately 4,000 words long. The median abnormal return is very close to 0, and the split into the three categories is even.

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

TABLE 1

Earnings Call Text Surprise Construction: Summary Statistics for the Combined Data Set

In Table 1, we calculate some basic summary statistics of the combined data set used to train and evaluate the machine learning model, including the information about the size of the documents and abnormal return (AR) cutoffs used to create the categorical target variable.

	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%

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

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):

$$\text{SUE.txt} = \text{log-odds}(H) - \text{log-odds}(L) \tag{1}$$

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 0 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. We further discuss the analogy between SUE.txt and SUE in Section IV.B, the words and phrases driving the SUE.txt in Section IV.A, and the context in which they appear in Section IV.B.

C. 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:

$$\text{(2)} \qquad \text{CAR}_t^S = \prod_{t=S}^{E} \left(\text{AR}_t^S\right),$$
$$\text{AR}_t^S = \frac{1}{|T|} \sum_{\{f,q,t\} \in T} \text{AR}_{f,q,t} - \frac{1}{|B|} \sum_{\{f,q,t\} \in B} \text{AR}_{f,q,t},$$
$$\text{AR}_{f,q,t} = R_{f,q,t} - R^b_{f,q,t},$$

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;[5] $|T|$ 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.[6]

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 (1 calendar year). At every calendar quarter mark, PEAD.txt is much larger and growing: 2.87%–1.54% on trading day 63, 4.61%–2.7% on trading day 126, 6.51%–3.87% on trading day 189, and 8.01%–4.63% on trading day 252. PEAD is larger only at the very beginning of the window.

D. Comparing Statistical Properties of PEAD.txt with Traditional PEAD

Tables 2 and 3 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 Appendix C of the Supplementary Material). 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 a much smaller drift than PEAD.txt, 1.11%–2.87%.[7] 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, Kishore, Santa-Clara, and Venkatachalam (2008)).

SUE.txt has stronger associations with CAR than classic SUE in a panel regression setting with fixed effects, as Table 4 reports. 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 with 3%–6% of a standard deviation increase in 63 trading days CAR, depending on specification. This result is robust to including firm and year-quarter

[5]In the case of SUE.txt, we estimate that the quintile of 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.

[6]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 Dec. 8, 2020).

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

https://doi.org/10.1017/S0022109022001181 Published online by Cambridge University Press

TABLE 2

PEAD.txt and PEAD Comparison

In Table 2, we calculate earnings call text surprises using the output of a regularized logistic text regression that predicts 1-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)				
Q1	−0.0288	−0.0152	−0.0089	−0.0064
Q2	−0.0075	−0.0102	−0.0057	−0.0045
Q3	0.0022	0.0003	−0.0002	0.0005
Q4	0.0089	0.0041	0.0021	0.0020
Q5	0.0201	0.0131	0.0066	0.0064
Spread	**0.0489**	**0.0287**	**0.0156**	**0.0129**
SUE (PEAD)				
Q1	−0.0325	−0.0136	−0.0093	−0.0043
Q2	−0.0146	−0.0020	−0.0011	−0.0008
Q3	0.0024	0.0032	0.0015	0.0017
Q4	0.0156	0.0043	0.0018	0.0025
Q5	0.0285	0.0016	0.0017	−0.0001
Spread	**0.0610**	**0.0154**	**0.0111**	**0.0042**

TABLE 3

PEAD.txt and Other Drifts Comparison

In Table 3, we calculate earnings call text surprises using the output of a regularized logistic text regression that predicts 1-day return based only on earnings call text. Numeric and text split uses the output of a regularized logistic regression that predicts 1-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)				
Q1	−0.0288	−0.0152	−0.0089	−0.0064
Q3	0.0022	0.0003	−0.0002	0.0005
Q5	0.0201	0.0131	0.0066	0.0064
Spread	**0.0489**	**0.0287**	**0.0156**	**0.0129**
Numeric and text				
Q1	−0.0374	−0.0120	−0.0084	−0.0036
Q3	0.0002	−0.0016	0.0003	−0.0019
Q5	0.0334	0.0104	0.0053	0.0051
Spread	**0.0707**	**0.0227**	**0.0138**	**0.0088**
Sent. dict. (neg.)				
Q1	−0.0119	−0.0065	−0.0036	−0.0029
Q3	−0.0008	−0.0026	−0.0010	−0.0016
Q5	0.0111	0.0047	0.0019	0.0028
Spread	**0.0231**	**0.0113**	**0.0055**	**0.0058**
AR(0)				
Q1	−0.0949	−0.0099	−0.0058	−0.0040
Q3	0.0003	0.0012	−0.0011	0.0023
Q5	0.0927	0.0065	0.0064	0.0002
Spread	**0.1876**	**0.0165**	**0.0122**	**0.0042**

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 1–32 and 33–63 (Table 5).

TABLE 4

Earnings Call Text Surprise and Cumulative Abnormal Returns Regression, Specification Comparison

In Table 4, we calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts 1-day return. We calculate CAR using the returns on the matched six size and book-to-market portfolios. The errors are clustered at the firm and year-quarter level. *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively.

	CAR(1,63)				
	1	2	3	4	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)
IVOL				−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 × IVOL					−0.05**
					(0.02)
SUE × COVERAGE					−0.01
					(0.01)
No. of obs.	85,160	85,160	85,160	85,160	85,160
Fixed effects	None	Ind, YQ	Firm, YQ	Firm, YQ	Firm, YQ
Adj. R^2	0.00	0.02	0.05	0.08	0.08

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 equal-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 5 Fama–French factors and momentum. The spread portfolio earns a statistically significant daily alpha of 3.9 BPS, as Table 6 reports. 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 BPS), as reported in Table 7. While the SUE spread alpha has high statistical significance, alpha for the bottom quintile is significant only 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 8. A strategy that equally weights SUE.txt and SUE signals is the best-performing strategy overall with an alpha of 4.2 BPS. Strategy based on a regularized logistic regression with both the text and numerical variables (see Appendix C of the

https://doi.org/10.1017/S0022109022001181 Published online by Cambridge University Press

TABLE 5

Earnings Call Text Surprise and Cumulative Abnormal Returns Regression, Timing Comparison

In Table 5, we calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts 1-day return. We calculate CAR using the returns on the matched six size and book-to-market portfolios. The errors are clustered at the firm and year-quarter level. *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively.

	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)
IVOL	−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)
No. of obs.	85,160	85,160	85,160
Fixed effects	Firm, YQ	Firm, YQ	Firm, YQ
Adj. R^2	0.08	0.05	0.05

TABLE 6

Alpha for Different Quintiles of Earnings Call Text Surprise, 63 Trading Days, Fama–French 5 Factors Plus Momentum

In Table 6, we calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts 1-day return. *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively.

	SUE.txt Portfolios					
	Spread	Q1	Q2	Q3	Q4	Q5
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)
No. of obs.	2,379	2,376	2,369	2,376	2,377	2,379
Adj. R^2	0.319	0.945	0.966	0.974	0.981	0.975

TABLE 7

Alpha for Different Quintiles of Classic Earnings Surprise, 63 Trading Days, Fama–French 5 Factors Plus Momentum

In Table 7, earnings surprises are standardized unexpected earnings calculated using the analyst forecasts. *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively.

	SUE Portfolios					
	Spread	Q1	Q2	Q3	Q4	Q5
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)
No. of obs.	2,379	2,377	2,378	2,378	2,378	2,379
Adj. R^2	0.078	0.942	0.982	0.984	0.984	0.939

TABLE 8

Alpha for Different Spread Portfolios, 63 Trading Days, Fama–French 5 Factors Plus Momentum

In Table 8, we calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts 1-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 1-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)). *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively.

	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)
No. of obs.	2,379	2,379	2,379	2,379	2,379
Adj. R^2	0.319	0.078	0.247	0.225	0.490

TABLE 9

Alpha for Different Spread Portfolios, 32 Trading Days, Fama–French 5 Factors Plus Momentum

In Table 9, we calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts 1-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 1-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)). *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively.

	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)
No. of obs.	2,379	2,379	2,379	2,379	2,379
Adj. R^2	0.226	0.048	0.158	0.136	0.345

Supplementary Material) 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 9 compares strategies with a shorter holding period, trading days 1–32. In this case, the classic SUE spread strategy comes ahead of the SUE.txt strategy with 4.3–3.4 BPS alpha. The best-performing strategy overall is a mix of the two with an alpha of 5.2 BPS. Table 10 presents the portfolio performance results (63 days holding period) using the q5 factors (Hou et al. (2020)). We obtain the factor returns data at http://global-q.org/index.html (last accessed Dec. 8, 2020). The results are very similar to the results obtained using Fama–French factors.

E. PEAD.txt and PEAD over Time

Figure 2 demonstrates PEAD.txt and PEAD across the years. PEAD.txt is larger than PEAD in 8 out of 10 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).

TABLE 10

Alpha for Different Spread Portfolios, 63 Trading Days, q5 Factors (Hou et al. (2020))

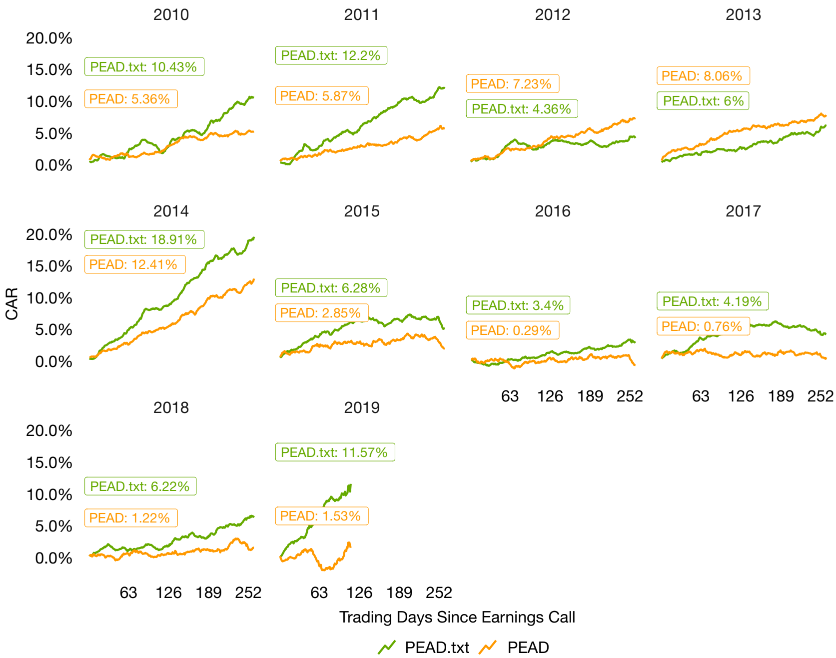
In Table 10, we calculate earnings call text surprises (SUE.txt) using the output of a regularized logistic text regression that predicts 1-day return. Earnings surprises are standardized unexpected earnings calculated using the analyst forecasts. SUE.txt and SUE is a strategy that equally weights earnings call text surprises and earnings surprises signals. Numeric and text split uses the output of a regularized logistic regression model that predicts 1-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)). *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively.

	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)
No. of obs.	2,379	2,379	2,379	2,379	2,379
Adj. R^2	0.1886	0.0403	0.1493	0.1226	0.1781

FIGURE 2

PEAD.txt and PEAD Across Years, Part 1

In Figure 2, the lines represent cumulative abnormal returns 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 (1 calendar year).



F. Why Linear Model?

We choose regularized logistic regression with log word count inputs as the main model because of its interpretability and the relative ease of computation while potentially trading-off explanatory power and the realism of the underlying model of disclosure language. Our approach is most similar to other work that uses word count models like Kogan, Levin, Routledge, Sagi, and Noah Smith (2009) and Frankel, Jennings, and Lee (2016), who use support vector regressions, Li (2010), who uses naive Bayesian model, Brown, Crowley, and Elliott (2020), who use a combination of a topic model and supervised regression, Ke, Kelly, and Xiu (2020), who use a multistep procedure involving a supervised model, and Garcia, Hu, and Rohrer (2021), who use multinomial inverse regression.

Deep learning models present an important alternative to word count models because of their ability to take into account word context.[8] The current state-of-the-art models for text classification are typically based on some deep learning architecture (e.g., Bidirectional Encoder Representations from Transformers (BERT; Devlin, Chang, Lee, and Toutanova (2019))). Such models treat text as an ordered sequence of words, allowing the relationship between the LHS variable and an individual word in the document to depend on the context in which the word appears. While context is undoubtedly important, such models are hard to interpret,[9] are computationally expensive,[10] require specific hardware (graphics processing units), and do not straightforwardly extend to very long documents such as earning calls.[11] The ability of deep learning to model word context opens exciting opportunities for academic Finance (see, e.g., Meursault (2019), Cao, Kim, Wang, and Xiao (2021b), Cao, Yang, and Zhang (2021c), and Huang, Wang, and Yang (2021)). However, it presents researchers with trade-offs that need to be taken into account when choosing an appropriate model for the specific task at hand.

III. Combining Text and Numbers in the Regularized Logistic Text Regression

In this section, 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

[8]Deep learning is a subset of machine learning methods that includes neural networks with multiple layers (a series of function compositions).

[9]There are methods to post-process such modes to gain word-in-context level coefficients that sum to model outputs at the document level (e.g., Lundberg and Lee (2017)). However, the complexity of the models still hinders clear interpretation. Additionally, these attribution methods require significant extra computation.

[10]The regularized logistic regression used in this paper has 4,000 parameters. A deep neural net with convolutional neural network with gated recurrent units (CNN-GRU) architecture used in Meursault (2019) to predict absolute abnormal returns around earnings press releases has approximately 6 million parameters, whereas BERT Large has 340 million (Devlin et al. (2019)).

[11]In case of BERT, the document has a 512 token limit and potentially lossy partitioning is required to handle longer documents. Adapting these models to longer documents is an area of ongoing research (see, e.g., Beltagy, Peters, and Cohan (2020)).

https://doi.org/10.1017/S0022109022001181 Published online by Cambridge University Press

before the earnings call (Text + Num model).[12] We also generate PEAD.mix by rank aggregating SUE.txt and SUE. We find complementarities between text and numbers that suggest the two media work together to help investors understand the value of the firm.

A. Combining Text and Numbers in Machine Learning Model

We begin by examining announcement returns generated by the Text, Text + Num, and Num models (the Num model includes the same numerical variables as Text + Num model, but no text). 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 11). 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.

TABLE 11

Performance of the Regularized Logistic Text Regression Model on the 1-Day Return Prediction Task, All Test Sets Combined

In Table 11, 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 1-day returns using SUE, Num model includes an array of market and analyst following-based numerical variables, and Text + Num and Text × Num models use 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%

B. Combining 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 3. We find that PEAD.txt produced by the text-only model is the largest purely Machine Learning (ML) model-based drift, 8.01% for the period starting 1 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

[12]The variables are: SUE, abnormal return on the day before the earnings call, abnormal return on the day 2 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.

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. Unsurprisingly, the ML model based only on the numbers produces a drift similar to classic SUE-based PEAD (4.11% vs. 4.64%).

C. Combining Text and Numbers in Generating Drift: Alternative Solution

While the ML-based model struggles 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 (e.g., a stock in the 100th percentile of SUE and the 50th percentile of SUE.txt is assigned to the 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.

D. Text and Numbers are Complementary

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.

IV. Economic Interpretation of PEAD.txt and Comparison with PEAD

A. 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

https://doi.org/10.1017/S0022109022001181 Published online by Cambridge University Press

(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" and "we experienced a loss of numtoken"), and SUE.txt heavily utilizes this (see Section V.A).

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 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 1-day abnormal returns for calibration. Both the model and the way to calibrate it can be tweaked in future work to produce better measures.

B. 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.[13]

In a perfect world, we would task professional analysts with highlighting the new information in earnings calls.[14] 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

[13]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 data set 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 SUE.txt close to 0, and most surprising paragraphs have large positive or negative SUE.txt). A human annotator (a Research Assistant (RA) who did not 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" (note that this means we are asking the RA to annotate "unexpectedness" based on their judgment generally and not the full set of available information released before the earnings call specifically). In 68% of the cases, the annotator identified 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.

[14]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.

numerical form such as log odds of high or low return. We use 1-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 standalone measure similar to SUE and generating larger drift without explicitly utilizing the earnings number distinguishes the present article from articles 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 articles contrast their language measures with numbers and argue that they are associated with higher information processing costs. The present article focuses on text as a reflection of economic activity, similar to earnings numbers in content, but different in form. Naturally, this view does not contradict the results of 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.

C. Examples

To help build intuition about the SUE.txt measure, we provide 2 example paragraphs below. The first example is identified as expected text by the model, and the second one contains a large positive surprise. We italicize the words that were assigned nonzero coefficients by the model and normalize the coefficients by the largest absolute in the paragraph for tractability.

No Surprise Smart Technologies Inc., 2015Q3
"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."

Positive Surprise WD-40 Company, 2017Q2
"*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 are 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 no words have nonzero 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 coefficient of the first appearance of the word "strong" is 100%, 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 does not 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 not to 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.

V. Analytic Tools for Explaining PEAD.txt

PEAD.txt is larger than PEAD. That deepens the PEAD puzzle. However, 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 article 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 because 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, Smith, and Wilkerson (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.

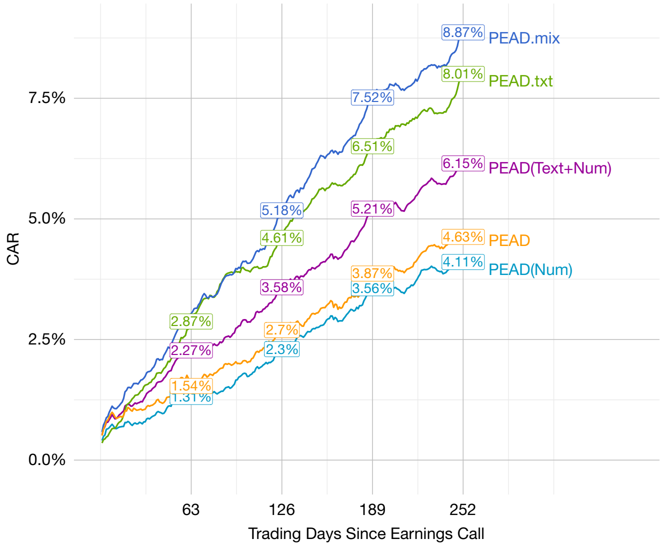
A. Word-Level Impact and News Directionality

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)).[15] The trained model has parameters, β coefficients, associated with individual tokens that tell us which words and phrases drive the model's predictions. 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:

[15] Note that there is a different definition of word impact (Routledge, Sacchetto, and Smith (2013)).

FIGURE 3

Comparison of Drifts Created Using Different Methods



$$I_j = \left(\beta_j^H - \beta_j^L\right) \frac{1}{N} \sum_{i=1}^{N} x_{ij}, \tag{3}$$

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 2-dimensional space. Figure 4 plots 15 tokens with the largest positive impact and 15 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 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"

https://doi.org/10.1017/S0022109022001181 Published online by Cambridge University Press

and ask for clarifications to help them "understand" something. That also provides signals for the model.

B. 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 nonzero 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$^{\mathrm{P}}$), and analyze how SUE.txt$^{\mathrm{P}}$ 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.

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

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$^{\mathrm{P}}$) 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:

$$\text{(4)} \qquad \mathrm{SUE.txt}^{\mathrm{P}} = \sum_{w=1}^{W} (\beta_w^H - \beta_w^L)\Delta_w,$$
$$\Delta_w = \log(2+b_w) - \log(1+b_w),$$

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).

Furthermore, 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 Appendix D of the Supplementary Material for keywords and paragraph examples):

- Financial accounting: bottom line, metrics, adjustments, and lending, financing.
- Operations management and marketing: operational and marketing metrics, segments, supply chain, production, interruptions, and marketing.
- Global economics: foreign exchange, seasonality and weather, and general global economics.
- Strategy: competition, expansion, contraction, partners, deals, government, restructuring, and general strategy.
- Forward-looking: paragraphs including forward-looking phrases that are identified following Muslu, Radhakrishnan, Subramanyam, and Lim (2015).

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

$$|\text{SUE.txt}_g^{\text{P}}| = \frac{1}{|G|}\sum_{k \in G} \text{SUE.txt}_k^{\text{P}}, \tag{5}$$

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

The absolute value of $\text{SUE.txt}^{\text{P}}$ 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 $\text{SUE.txt}^{\text{P}}$ of 0 even if the revenue news is significant. Looking at the absolute value of SUE. txt^{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 5 shows, the bottom-line, forex, interruption, and seasons groups have the highest mean absolute $\text{SUE.txt}^{\text{P}}$ (within 5% of the top subgroup, bottom line). However, these groups are rare and, overall, the financial

FIGURE 4

Tokens with Largest Positive and Negative Impact

In Figure 4, the tokens above 0 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 15 tokens with the largest positive impact and the top 15 tokens with the largest negative impact. PR indicates the presentation section and QA indicates the Q&A section.

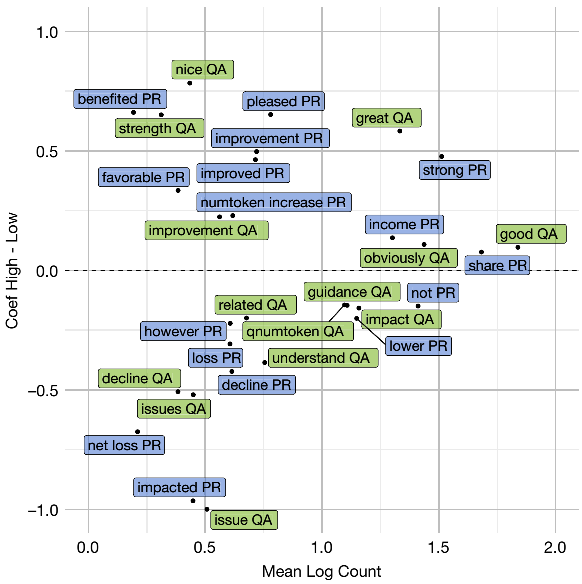
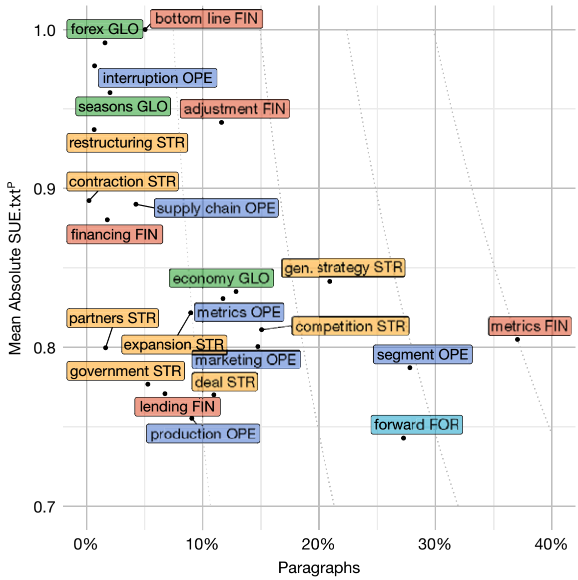


FIGURE 5

Absolute Value of Paragraph-Level SUE.txt ($SUE.txt^P$) and Prevalence of Paragraph Subgroups

Here in Figure 5, the 3-letter abbreviations refer to paragraph groups based on the business curriculum. The *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"). The *x*-axis represents the percentage of paragraphs belonging to a given subgroup. The dotted lines represent the *x*-axis 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.



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.

VI. 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

https://doi.org/10.1017/S0022109022001181 Published online by Cambridge University Press

TABLE 12

Autocorrelation of SUE.txt

In Table 12, dSUE.txt is the decile of SUE.txt. EVOL is earnings volatility, MKTVAL is market value, and LOSS is an indicator variables equal to 1 if the firm has negative earnings in the quarter. The dependent variables are lagged by 1 quarter. The standard errors are clustered at the firm level. *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively.

	dSUE.txt			
	1	2	3	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
No. of obs.	79,337	78,247	77,232	79,160

autocorrelation and contemporaneous earnings characteristics such as loss and ex ante (expected) earnings characteristics such as volatility.

Table 12 reports the SUE.txt autocorrelation results following Narayanamoorthy (2006) and Cao and Narayanamoorthy (2012). We construct the 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 the SUE.txt decile and the other variables. We see that SUE.txt has positive autocorrelation.[16] 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, among other results, that SUE is more mean-reverting for loss firms. We confirm that the same holds for SUE.txt.[17] 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

[16] In an untabulated analysis we also confirm that the positive autocorrelation extends to at least four lags, the same as SUE in our sample. This result is different from the findings of Narayanamoorthy (2006), who finds positive autocorrelation of SUE for up to three lags, followed by the negative autocorrelation with the fourth lag. We leave a detailed discussion of these differences to future work.

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

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.

VII. 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 PEAD 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 drive 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; and 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.

Supplementary Material

To view supplementary material for this article, please visit http://doi.org/10.1017/S0022109022001181.

References

Ali, A.; X. Chen; T. Yao; and T. Yu. "Can Mutual Funds Profit from Post Earnings Announcement Drift? The Role of Competition." *Journal of Banking & Finance*, 114 (2020), 105774.

Ball, R., and P. Brown. "An Empirical Evaluation of Accounting Income Numbers." *Journal of Accounting Research*, 6 (1968), 159.

Bartov, E.; S. Radhakrishnan; and I. Krinsky. "Investor Sophistication and Patterns in Stock Returns After Earnings Announcements." *Accounting Review*, 75 (2000), 43–63.

Beltagy, I.; M. E. Peters; and A. Cohan. "Longformer: The Long-Document Transformer." ArXiv Working Paper, available at https://arxiv.org/abs/2004.05150 (2020).

Bernard, V. L., and J. K. Thomas. "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?" *Journal of Accounting Research*, 27 (1989), 1.

Brandt, M. W.; R. Kishore; P. Santa-Clara; and M. Venkatachalam. “Earnings Announcements are Full of Surprises.” Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=909563 (2008).

Brown, N. C.; R. M. Crowley; and W. B. Elliott. “What are You Saying? Using Topic to Detect Financial Misreporting.” *Journal of Accounting Research*, 58 (2020), 237–291.

Cao, S.; W. Jiang; J. L. Wang; and B. Yang. “From Man vs. Machine to Man + Machine: The Art and AI of Stock Analyses.” Technical Report, National Bureau of Economic Research (2021a).

Cao, S.; W. Jiang; B. Yang; and A. L. Zhang. “How to Talk When a Machine is Listening: Corporate Disclosure in the Age of AI.” Technical Report, National Bureau of Economic Research (2020).

Cao, S.; Y. Kim; A. Wang; and H. Xiao. “From Words to Syntax: Identifying Context-Specific Information in Textual Analysis.” Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3568504 (2021b).

Cao, S.; B. Yang; and A. L. Zhang. “Deep Learning Mutual Fund Disclosure: Risk Sentiment, Risk Taking, and Performance.” Working Paper (2021c).

Cao, S. S., and G. Narayanamoorthy. “Earnings Volatility, Post-Earnings Announcement Drift, and Trading Frictions.” *Journal of Accounting Research*, 50 (2012), 41–74.

Carhart, M. M. “On Persistence in Mutual Fund Performance.” *Journal of Finance*, 52 (1997), 57–82.

Chordia, T.; A. Subrahmanyam; and Q. Tong. “Have Capital Market Anomalies Attenuated in the Recent Era of High Liquidity and Trading Activity?” *Journal of Accounting and Economics*, 58 (2014), 41–58.

Coleman, B.; K. J. Merkley; and J. Pacelli. “Human Versus Machine: A Comparison of Robo-Analyst and Traditional Research Analyst Investment Recommendations.” *Accounting Review*, 95 (2022), 221–244.

Cox, J. “Market Fragmentation and Post-Earnings Announcement Drift.” *Journal of Economics and Finance*, 44 (2020), 587–610.

Devlin, J.; M.-W. Chang; K. Lee; and K. Toutanova, “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, MN: Association for Computational Linguistics (2019), 4171–4186.

Engelberg, J. “Costly Information Processing: Evidence from Earnings Announcements.” Working Paper, University of California San Diego (2008).

Fama, E. F., and K. R. French. “A Five-Factor Asset Pricing Model.” *Journal of Financial Economics*, 116 (2015), 1–22.

Fink, J. “A Review of the Post-Earnings-Announcement Drift.” *Journal of Behavioral and Experimental Finance*, 29 (2021), 100446.

Frankel, R.; J. Jennings; and J. Lee. “Using Unstructured and Qualitative Disclosures to Explain Accruals.” *Journal of Accounting and Economics*, 62 (2016), 209–227.

Garcia, D.; X. Hu; and M. Rohrer. “The Colour of Finance Words.” Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3630898 (2021).

Grennan, J., and R. Michaely. “Artificial Intelligence and High-Skilled Work: Evidence from Analysts.” Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3681574 (2020).

Hou, K.; H. Mo; C. Xue; and L. Zhang. “An Augmented q-Factor Model with Expected Growth.” *Review of Finance*, 25 (2020), 1–41.

Huang, A.; H. Wang; and Y. Yang. “FinBERT – A Deep Learning Approach to Extracting Textual Information.” Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3910214 (2021).

Jegadeesh, N., and J. Livnat. “Revenue Surprises and Stock Returns.” *Journal of Accounting and Economics*, 41 (2006), 147–171.

Ke, Z. T.; B. Kelly; and D. Xiu. “Predicting Returns with Text Data.” NBER Working Paper No. 26186 (2020).

Kogan, S.; D. Levin; B. R. Routledge; J. S. Sagi; and N. A. Smith. “Predicting Risk from Financial Reports with Regression.” In *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Boulder, CO: Association for Computational Linguistics (2009), 272–280.

Lee, Y.-J. “The Effect of Quarterly Report Readability on Information Efficiency of Stock Prices.” *Contemporary Accounting Research*, 29 (2012), 1137–1170.

Li, F. “The Information Content of Forward-Looking Statements in Corporate Filings – A Naïve Bayesian Machine Learning Approach.” *Journal of Accounting Research*, 48 (2010), 1049–1102.

Livnat, J., and R. R. Mendenhall. “Comparing the Post-Earnings Announcement Drift for Surprises Calculated from Analyst and Time Series Forecasts.” *Journal of Accounting Research*, 44 (2006), 177–205.

Loughran, T., and B. McDonald. "When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance*, 66 (2011), 35–65.

Lundberg, S. M., and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions." In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (NIPS'17). Red Hook, NY: Curran Associates Inc. (2017), 4768–4777.

Martineau, C. "Are There (Still) Post-Earnings Announcement Drifts?" SSRN Working Paper (2021).

Mendenhall, R. R. "Arbitrage Risk and Post-Earnings-Announcement Drift." *Journal of Business*, 77 (2004), 875–894.

Meursault, V. "The Language of Earnings Announcements." SSRN Working Paper (2019).

Milian, J. A. "Unsophisticated Arbitrageurs and Market Efficiency: Overreacting to a History of Underreaction?" *Journal of Accounting Research*, 53 (2015), 175–220.

Muslu, V.; S. Radhakrishnan; K. R. Subramanyam; and D. Lim. "Forward-Looking MD&A Disclosures and the Information Environment." *Management Science*, 61 (2015), 931–948.

Narayanamoorthy, G. "Conservatism and Cross-Sectional Variation in the Post-Earnings Announcement Drift." *Journal of Accounting Research*, 44 (2006), 763–789.

Ou, J. A., and S. H. Penman. "Financial Statement Analysis and the Prediction of Stock Returns." *Journal of Accounting and Economics*, 11 (1989), 295–329.

Richardson, S.; I. Tuna; and P. Wysocki. "Accounting Anomalies and Fundamental Analysis: A Review of Recent Research Advances." *Journal of Accounting and Economics*, 50 (2010), 410–454.

Routledge, B. R.; S. Sacchetto; and N. A. Smith. "Predicting Merger Targets and Acquirers from Text." SSRN Working Paper (2013).

Yan, X., and L. Zheng. "Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach." *Review of Financial Studies*, 30 (2017), 1382–1423.

Yano, T.; N. A. Smith; and J. D. Wilkerson, "Textual Predictors of Bill Survival in Congressional Committees." In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Montréal, QC: Association for Computational Linguistics (2012), 793–802.

Zou, H., and T. Hastie. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society: Series B*, 67 (2005), 301–320.