Federal Reserve Bank of Philadelphia

Working Papers
Research Department

WP 21-40
November 2021
https://doi.org/10.21799/frbp.wp.2021.40

Corporate Disclosure: Facts or Opinions?

Shimon Kogan
IDC Herzliya and the University of Pennsylvania

Vitaly Meursault
Federal Reserve Bank of Philadelphia Research Department

ISSN: 1962-5361

Corporate Disclosure: Facts or Opinions?

Shimon Kogan[a], Vitaly Meursault[b,*]

[a] *Arison School of Business, IDC Herzliya and Wharton*
[b] *Federal Reserve Bank of Philadelphia*

November 26, 2021

Abstract

A large body of literature documents the link between textual communication (e.g., news articles, earnings calls) and firm fundamentals, either through pre-defined "sentiment" dictionaries or through machine learning approaches. Surprisingly, little is known about why textual communication matters. In this paper, we take a step in that direction by developing a new methodology to automatically classify statements into objective ("facts") and subjective ("opinions") and apply it to transcripts of earnings calls. The large scale estimation suggests several novel results: (1) Facts and opinions are both prominent parts of corporate disclosure, taking up roughly equal parts, (2) higher prevalence of opinions is associated with investor disagreement, (3) anomaly returns are realized around the disclosure of opinions rather than facts, and (4) facts have a much stronger correlation with contemporaneous financial performance but facts and opinions have an equally strong association with financial results for the next quarter.

Keywords:
Subjectivity, Machine Learning, NLP, Text Analysis
JEL: G14, G12, C00

[*] Corresponding author

Disclaimer: This Philadelphia Fed working paper represents preliminary research that is being circulated for discussion purposes. The views expressed in these papers 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. Philadelphia Fed working papers are free to download at https://philadelphiafed.org/research-and-data/publications/working-papers.

Author information: Vitaly Meursault (corresponding author), vitaly.meursault@phil.frb.org; Shimon Kogan, skogan@upenn.edu.

1. Introduction

A large and growing body of literature in finance and accounting establishes the usefulness of text data in predicting future performance and returns (e.g., Antweiler and Frank (2001), Tetlock et al. (2008), Ke et al. (2020)). These papers attempt to extract directional predictions from text using a lexicon or machine learning approaches, and, in general, interpret their measure as capturing "sentiment".

While it is clear that this "sentiment", however it is measured, is useful, it is not clear why. Tetlock et al. (2008) interprets the usefulness of text as capturing "otherwise hard-to-quantify aspects of firms' fundamentals". But what is the nature of these hard-to-quantify aspects of firms' fundamentals? After all, firms provide lengthy communications to investors through annual reports, earnings press releases, earnings calls, and special announcements, alongside their numerical communications.

An important distinction, in this case, would be between textual communications made by the company or its managers that convey facts (future or past) that are hard to convey numerically and opinions that provide an interpretation of facts. In principle, both can be informative to investors and both would affect firm fundamentals and returns. Ke et al. (2020) are careful in stating that their approach "does not differentiate between non-sentiment (i.e., objective information) and sentiment content of news per se."

In this paper, we contribute to this growing literature by suggesting a new methodology that allows us to automate the classification of sentences into "objective" and "subjective" in the context of earnings calls. The first category contains sentences stating facts that are, in principle, falsifiable, for example, “We opened a new store in Washington DC.” The second category contains sentences that cannot be falsified, for example, “We had a great quarter.” The last category contains sentences that are irrelevant for facts or opinions distinction, for example, “We now turn to slide 13.” We develop this methodology by manually tagging a large number of sentences from earnings calls and then training a machine learning model to capture the relevant text attributes associated with each.

It is important to note that this approach merges the benefits of human intuition advocated by, for example, Loughran and McDonald (2020) with the statistical, machine learning approach advocated by, for example, Gentzkow et al. (2019). That is, we identify the object of interest through human tagging of text and only then scale it by finding the associated word attributes. We do that precisely since there is no empirical outcome that can be measured and used to train a model. Put differently, it is quite possible that both objective and subjective statements are informative to investors. Also, the standard sentiment lexicons in finance do not distinguish between statements that are positive in facts (“Revenues improved by 10%”) and positive in coverage (“Revenues improved substantially”).

We show that the separation of disclosure into objective and subjective is important in understanding the heterogeneity of disclosure across firms and executives and that it has important implications for the incorporating of information by investors. First, despite the potential view that only facts should matter, subjective sentences make up roughly 50% of earnings calls. This composition appears to be stable over the sample period but varies systematically with the complexity of the company –

growth and large firms use more subjective language than value and small firms, with a similar variation being observed across industries.

Second, executives' language varies systematically, with CEOs using more subjective language than CFOs and both using more subjective language during the Q&A section relative to the opening section. Using changes in executives over time, we find that CEO communication style, when it comes to the use of subjective language, is more a function of the CEO than of the company, while the opposite is true for CFOs. This is consistent with the idea that CEOs have more leeway in the way they communicate with investors than CFOs have.

Third, we turn to examine how the form of communication affects the incorporation of information. A large body of literature examined how the release of information impacts investors' disagreement. We contribute to that literature by showing that more subjective information is associated with higher disagreement as measured by standard proxies for disagreement (i.e., abnormal volume) or by more direct measures of retail traders' disagreement (using StockTwits data, see Cookson and Niessner 2020). Next, we link the form of communication to anomaly returns. Engelberg et al. (2018) find that anomaly returns are more pronounced around the release of earnings. Their intuition is that the release of public information pushes investors' biased beliefs to be closer to the truth, thereby sending the prices of under (over) valued firms up(down). We repeat their analysis while separating instances when the information is more and less subjective. We find strong results suggesting that the high anomaly returns are associated with the release of subjective information.

Finally, we apply machine learning to study how the language in objective and subjective sentences, separately, is linked to the accounting performance and returns of the announcing firms. We apply the model (out of sample) and find that the accounting performance is more closely linked to the text in objective sentences while the returns on and the day after the earnings calls are linked with similar strength to the language in both objective and subjective sentences.

2. Data

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.[1] Executive information is based on Capital IQ People Intelligence dataset (through WRDS). Various numerical variables are constructed based on the Center for Research in Security Prices (CRSP),[2] Compustat,[3] and Refinitiv's Institutional Brokers' Estimate System (IBES) datasets (through WRDS). We use measures of investor disagreement from Cookson and Niessner (2020).[4]

We compute abnormal returns based on the Fama–French six size and book-to-market matched portfolios.[5] We use the call start time variable in Capital IQ

[1]This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

[2]©2021 Center for Research in Security Prices (CRSP®), The University of Chicago Booth School of Business.

[3]©2021 S&P Global Market Intelligence (and its affiliates, as applicable). No further distribution and or reproduction permitted.

[4]We download the measures from the author's website, https://www.marinaniessner.com/data. Last accessed: 07/17/2021.

[5]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.

Transcripts database to determine the appropriate timing of returns associated with the call. 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.

We compute standardized unexpected earnings, SUE, following Livnat and Mendenhall (2006).[6]

3. Our approach to identifying objectivity and subjectivity

We define objectivity as the linguistic expression of facts, and subjectivity as the linguistic expression of opinion.[7] In the context of financial disclosure, specifically earnings calls, the facts reflect the state of the firm observable by the manager, whereas opinions reflect the state of the mind of the manager.

We take the definition of objectivity and subjectivity to data by manually annotating a dataset of 3,673 earnings calls sentences following the definition and a set of guidelines discussed below. To scale this approach, we train a classifier that links the content of the annotated sentences to the human-assigned labels and apply it to all sentences in the corpus.[8] Having a human-annotated dataset is highly advantageous because it follows specific definitions of relevant concepts and because

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

[7]We follow the linguistics literature that defines subjectivity as "the linguistic expression of belief, emotion, evaluation, or attitude." See Taboada (2015) and Wiebe (1994).

[8]A similar approach was used by Li (2010) for sentiment classification.

the performance of models trained on the dataset can be evaluated out-of-sample on the subset of sentences not seen during training. It is also very important that our annotated dataset is based on the same kind of text (financial disclosure in the form of earnings calls) as the text we apply the model to, since the linguistic features associated with objectivity and subjectivity do not necessarily stay constant across different contexts.

3.1. Operationalization

While the difference between facts and opinions might appear intuitive, classifying sentences as reflecting one or the other involves nuances and practical considerations that can result in inconsistent annotations if instructions are not provided. In this subsection, we discuss some important points that clarify what sentences are considered as facts and opinions under our definitions.

We choose sentence as the unit of analysis. Sentences contain a significant amount of contextual information that allows us to decide whether the sentence is subjective or objective with reasonable precision.

We consider a sentence objective if it expresses facts, like "our EPS is 10 cents." We consider something a fact if two informed people who have the same information set as the manager would definitely agree with the statement, as is the case with reporting of financial metrics. The following are examples of sentences we classify as objective:

- For the trailing 12-month period ending June 30, same-store occupancy was 88.3%, and same-store EBITDARM coverage was 1.18x.

- And you know these are primarily related to the planned closure of the 2 manufacturing facilities in Detroit, Michigan and Cheektowaga, New York.

We consider a sentence subjective if it expresses opinions, like "we had a great quarter." We consider an expression to be an opinion if two equally informed people might disagree about it. For example, we can reasonably expect an informed person to have a different threshold for "great earnings" than a manager. The following are examples of sentences we classify as subjective:

- The analysis has showed that the brand has been distorted to bottoms.
- Turning to the Wholesale side of our green equation on Slide 13, we have talked about the immense opportunity we see in the solar space.

A large number of sentences in earnings calls express neither facts nor opinions. Many of these sentences are procedural, like "welcome to the earnings call," "moving on to slide 10." We also classify short sentences like "okay," or "thank you" as irrelevant. We put all questions into the irrelevant category. The following are examples of sentences we classify as irrelevant:

- We caution listeners that during this call, Zogenix management will be making forward-looking statements.
- The last line relates to the impact of the timing of the Gas Transmission rate-case decision.

Other classes of instances include:

- *Treatment of causality*: We consider causal statements like "due to decreasing demand" as being generally subjective. A very common exception from that is statements about accounting identities like "our profits fell due to an increase in costs."

- *Treatment of forecasts*: We treat forecasts and other forward-looking sentences the same way we treat statements about the past. If a sentence is a statement of an expected fact, like "we forecast our earnings to be 10 cents per share," we consider it objective. If a sentence expresses a general opinion about the future, like "we expect next quarter to be great," we consider it subjective.

- *Treatment of accounting metrics*: We assume that an informed person would know the method used for calculating accounting metrics such as earnings, even when there are different ways to calculate them. For example, we treat the sentence "our core earnings increased relative to the previous quarter" as objective even though it is possible to disagree about the appropriate method of calculating "core earnings."

- *Sentences containing both facts and opinions*: It is common for sentences in earnings calls to express both facts and opinions. In this case, the annotators are asked to decide whether fact or opinion is the main point of the sentence. The following are examples of sentences we classify as objective because we consider the objective component to be the focus:

  - We've made strong progress, not only in reducing inventory on the balance sheet, but also in reducing unit inventory levels, which are down over 10%

from June 30 and are now below 2011 year end levels.

– Despite the recent macro uncertainty, positive core television advertising trends are continuing for Nexstar in the third quarter and we remain on track to continue growing all of our non-political revenue sources in the second half of 2011.

The following are examples of sentences we classify as subjective because we consider the subjective component to be the focus:

– And we're very pleased with our performance for 2014 as we continue to experience strong growth across our legacy, gathering and processing, contract compression and NGL logistics businesses.

– Really I think for the full year, it's probably the better way to look at it because there's always going to be lumpiness from quarter-to-quarter when you think about $4 billion cost base of cost of goods.

3.2. Annotation

We create the corpus for annotation using a stratified sampling procedure. We split transcripts into sentences using *tokenize_sentences* function in *quanteda* R library (Benoit et al., 2018) and select sentences that have 3 or more words. We stratify by various categories described below to ensure that our model uses a wide range of content for training and can better account for potential differences in objectivity and subjectivity markers across the groups. For example, if a large and a small industry both use industry-specific accounting terms as objectivity markers, using stratified sampling aims to let the model see a more equal number of terms specific for each industry and make model performance more even across the two groups.

We use the following sentence categories for stratification:

- Fama-French 12 industries (12 categories). We want to cover both larger and smaller industries.

- Year (12 categories for years 2008–2019). We want to cover earlier years for which we have fewer earnings calls.

- Earnings calls with negative earnings surprises and negative abnormal returns as opposed to all others (2 categories). Calls presenting negative results are less common in our sample years but we want to represent them in the corpus.

- First five paragraphs of the first and second presentation in the call (2 categories). The beginning of the presentation tends to use more diverse language and we want that to be represented in the corpus.

We assign each sentence to a compound category which is a cartesian product of the above categories. The dimensionality of the compound category is $12\times12\times2\times2 = 576$. We assign a sampling factor to each of the categories that is equal to 2 if the category is associated with bad news and/or the beginning of the presentation and 1 otherwise. We sample a large number of sentences from each category according to the sampling factor, shuffle them, and assign them to annotators in the resulting order.

The annotation is performed using *prodigy*[9] and *spaCy*.[10] The annotation is performed by the two coauthors and two research assistants. The instructions were developed during several preliminary runs that are not included in the dataset. One of the coauthors was in charge of developing the instructions. After the instructions were developed, the other annotators were asked to read them and follow them during annotation as closely as possible, exercising judgment where appropriate. Separate practice sessions were conducted by the coauthor who developed the instructions for other annotators that involved annotating 100 sentences with explicit explanations of how the decision was made in ambiguous cases. The dataset was collected over several weeks in the Winter of 2020.

We annotate a total of 3,673 sentences, out of which 1,871 are manually classified as objective, 1,298 as subjective and 504 as irrelevant.

3.3. Subjectivity model

We train the sentence classifier using out-of-the-box text classification architectures from the NLP python library *spaCy*. We opt to rely on a widely used practical NLP library to emphasize that statistical text classification models can be almost as simple to use as dictionaries. Our sentence classification models are integrated with *spaCy* and can be used as a part of the spaCy pipeline. We use the CNN (convolutional neural network) model with default settings.

We randomly select 300 objective, 300 subjective and 200 irrelevant sentences for the test set (800 sentences total). The model achieves the F1 score, precision and

[9]https://prodi.gy/. Last accessed: 04/22/2021.
[10]https://spacy.io/. Last accessed: 04/22/2021.

recall of 79%/83%/76% for the objective category, 76%/78%/73% for the subjective category, and 79%/69%/91% for the irrelevant category.

Figure 1 reports F1 scores across different industries years and days with bad news. The performance of the model is stable across splits.

3.4. Characteristic words of objective and subjective content

In this section, we explore the words characteristic of objective and subjective sentences. Listing tokens that contribute most to model predictions and providing text together with model predictions is a common way for validating text-based models. While the models used in this paper do not generate word lists directly, characteristic words can be determined post factum based on which sentences were classified as objective or subjective. We use *scattertext* Python library described in Kessler (2017) as it provides a convenient interface to explore the characteristic words, look up example sentences and examine relative word frequencies. This tool can be used to explore predictions of a large set of models where two-way splits of categories are applicable.[11]

Figure 2 shows the relative frequency of unigrams and bigrams in objective and subjective sentences (we sample 1,000 sentences categorized as objective and 1,000 sentences categorized as subjective for performance reasons). The tokens on the diagonal appear equally often in objective and subjective sentences, whereas off-diagonal terms are characteristic of one group.[12]

[11] This is more general than it may seem. With many categories, you can do one versus all; with continuous variables you can do a median split. The categorization model can be anything including a word list.

[12] Concretely, the x- and y- values are dense ranks of frequencies. Dense rank is a function that

Overall, the distribution of terms is fairly disperse suggesting that objective and subjective sentences are well separated. The terms in the objective category lean heavily on numeric expressions ("numtoken," "million") and financial terms ("revenue," "EPS"). In the subjective category, we naturally see expressions of personal belief ("believe," "think," "feel"), adjectives like "hard" or "strong," and also words that appear in general discussions of company performance, ("environment," "economic," "business"). The five most characteristic subjective words are "focus," "focused," "environment," "opportunities," and "sort of." The five most characteristic objective words are "million in," "approximately," "GAAP," "numtoken per," and "million of." It is important to note that while there are some clear patterns in relative distributions of certain terms, the concept of subjectivity in earnings calls can't be reduced to any one of them, because of how diverse the content of earnings calls is.

4. Results

4.1. Summary statistics and cross-section of facts and opinions

We apply the trained model to the dataset of earnings calls, which includes 12 years, 4,346 firms, and 85,840 firm-quarter observations (Table 1). As Table 2 shows, we classify 37% of sentences as objective, 48% as subjective, and the rest as irrelevant.

Figure 3 shows summary statistics across years, industries, size and book-to-

---

assigns ranks to data dealing with ties in the following way: equal numbers are assigned equal rank and the next highest element is assigned the rank immediately after those assigned to the tied elements. See documentation for *scipy.stats.rankdata* function, https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.rankdata.html. Last accessed: 05/19/2021.

market portfolios. The sentence split is very stable across years. Industries vary more: utilities has the highest percentage of objective sentences (43%/42%/16% objective/subjective/irrelevant split) and business equipment has the lowest percentage of objective sentences (35%/51%/14% split). Across size and book-to-market portfolios, small firms are more objective than big firms, value firms are more objective than growth firms. The biggest difference is between small value (39%/46%/15%) and big growth (33%/53%/13%).

The largest differences in subjectivity are within the earnings calls, between CEOs and CFOs on one hand, and between the presenter speech and answer sections on the other. Overall, CEOs are far more subjective than CFOs with 60% of subjective sentences compared to 36%. Likewise, answer section is far more subjective than presentation section with 65% of subjective sentences compared to 34%. Figure 4 provides a more detailed view including a double split on executive and section (Panel A) and an illustration of how subjectivity evolves across sentence positions (Panel B). We see that the two tendencies described above are complementary and that CEO answers are the most subjective section (67% subjective sentences) and CFO presentation is the least subjective (18% subjective sentences). However, both types of executives and both sections generally become progressively more subjective as the call progresses, although there are some non-linearities.

4.2. Does subjectivity reflect the executive style or the firm's business activities?

Based on the discussion in the previous sections, subjectivity could reflect both the personal style of the executives (how often they are willing to talk about their opinions or express beliefs) and the business activities of the firm (growth/value, for example). In this section, we show that the style component is the primary one for CEOs and about as important as the firm component in the case of CFOs. Intuitively, we also find that the magnitude of the style component relative to the firm activity component is larger for the answer section.

To separate the executive style and the firm component, we utilize the event of executive turnover. Specifically, we examine firm–quarters when the executive changes **and** when past values of subjectivity both for the given firm and the given executive in a given position are available. That is, we only look at the cases when a CEO or a CFO at one firm in our sample transfers to the same position at another firm in our sample. We have 130 cases like that for CEOs and 499 for CFOs. This setup allows us to run the following regression (we do that separately for CEOs and CFOs, presentation and answer sections, four regressions total):

$$\%SubjSentFirm(q+1)_i = \beta_1 \%SubjSentFirm(q-2)_i + \beta_2 \%SubjSentExec(q-2)_i + \epsilon_i,$$

where $\%SubjSentFirm$ and $\%SubjSentExec$ are percentages of subjective sentences for a given firm or executive.

We are asking the following question: Is the subjectivity of the new executive at a firm associated more strongly with their subjectivity at the previous job, or with the subjectivity of their predecessor? We exclude the first quarter the executive is at the new job and the last quarter at the old job because the start and the end of tenure are typically reflected in the text of the calls through introductions and reflections on the history with the firm.

Figure 5 presents the results. For CEOs' presentations and answers alike (Panels

A and C), the subjectivity at the new firm is strongly associated with their subjectivity at the previous firm (β is 0.48 for presentation and 0.4 for answers, both p-values < 0.001). In contrast, the subjectivity of the previous CEO at the same position has no significant association with the LHS variable. For CFOs' presentations, both components are significant and the firm component matters more, while for CFOs' answers, both components are equally important. Comparing presentation regressions to the answer ones, we see that for both CEOs and CFOs, style's importance relative to the firm component is larger for the answers, as evidenced by the difference of coefficients' means.

4.3. Is subjective disclosure associated with higher disagreement?

A large body of literature has focused on disagreement as way to understand what otherwise appears to be excessive volume (e.g., Hong and Stein 2007). Some of these studies focused on the role of public information release as a laboratory to study disagreement. We contribute to this literature by showing that the *nature* of public information matters for investors disagreement.

To measure investor disagreement we use abnormal volume and three social media-based measures from Cookson and Niessner (2020). Abnormal volume is a standard measure in the literature often related to investor disagreement. The other three measures are calculated by Cookson and Niessner (2020) and measure investor disagreement directly using the StockTwits platform at the stock-day level.[13] The definitions of the measures are the following:

[13]We download the measures from the author's website, https://www.marinaniessner.com/data. Last accessed: 07/17/2021.

- $AbLogVol_{f,q}$, abnormal log volume, defined as the log volume on the date of the earnings call (t) minus the average log volume between trading days $t-140$ and $t-20$.

- $DisWithin_{f,q}$, average disagreement among investors with the same investment approach (Fundamental, Technical, Value, Momentum, Growth) measured using StockTwits platform.

- $DisAcross_{f,q}$, disagreement among investors with different investment approaches on StockTwits platform.

- $DisAll_{f,q}$, disagreement among all investors on StockTwits platform.

To evaluate the association between the degree of earnings call subjectivity and investor disagreement, we estimate the following regressions:

$$AbLogVol_{f,q} = \beta_1 \% SubjSent_{f,q} + \beta_2 DisMeasure_{f,q} + \beta_c C_{f,q} + FE_f + FE_q + \epsilon_{f,q},$$
$$DisMeasure_{f,q} = \beta_1 \% SubjSent_{f,q} + \beta_c C_{f,q} + FE_f + FE_q + \epsilon_{f,q},$$

where $\% SubjSent_{f,q}$ is the percentage of subjective sentences in the earnings call,[14] $DisMeasure_{f,q}$ is one of the three measures from Cookson and Niessner (2020) defined above, $C_{f,q}$ are controls (abnormal returns for days -30 to -6, abnormal returns for days -5 to -1, and return volatility for days -5 to -1, and standardized unexpected earnings, SUE), and FE_f and FE_q are firm and year-quarter fixed effects. Continu-

[14] We exclude irrelevant sentences from the calculation so that % of subjective sentences is 100% - % of objective sentences.

ous variables are normalized so the coefficients reflect changes in the LHS variables associated with $\beta\times$ one standard deviation changes in the RHS variable.

As Figure 6 shows, a one standard deviation change of $\%SubjSent$ is associated with 5% of one standard deviation change in abnormal volume, which is statistically and economically significant. The inclusion of any of the disagreement measures from Cookson and Niessner (2020) does not change the size of the coefficient, suggesting that the association between subjectivity and abnormal value is largely not captured by social-media-based measures of disagreement. However, when subjectivity is high, disagreement of StockTwits users within and across investment philosophies tends to be higher (β's for $DisWithin$ and $DisAcross$ are both around 1.5%, see Figure 7). The association between $\%SubjSent$ and $DisAll$ is not statistically significant.

4.4. Is subjective disclosure associated with higher anomaly returns?

One of the biggest debates in asset pricing is centered at the underlying drives of cross-sectional stock predictability (e.g., Cochrane 2011). To provide evidence for the idea that anomalies arise from investors' erroneous beliefs, Engelberg et al. (2018) study how anomaly returns accrue around earning announcements. They focus on earning announcements since these is a large set of events around which beliefs should converge toward fundaments. Engelberg et al. (2018) show that anomaly returns are largely realized on earnings announcement days, consistent with investors holding misguided beliefs prior to these announcements.

We expand their framework by asking whether the nature of information matters for belief updating. We do so by separating earnings announcements into more subjective and more objective ones, and we show that higher realization of anomaly returns is more strongly associated with subjective ones. We follow the methodology in Engelberg et al. (2018) – a stock's exposure to anomalies is captured with a variable Net constructed using the dataset from Green et al. (2017), which is similar to the dataset in Engelberg et al. (2018) but is more recent and covers years up to 2016. The dataset includes 94 stock characteristics associated with anomalies identified in the literature. The procedure for constructing Net is the following:

- Using the data from 1997 to 2007, we split stock characteristics into quintiles and compare the first and the fifth quintile returns to establish which quintile is the long quintile and which is the short one. For binary variables, we establish whether stocks with value zero or value one are long in the same way.

- For our sample period of 2008 to 2019, we calculate monthly quintiles of stock characteristics. Then we compute the number of stock characteristics that are long and short for a given stock–month.

- Net is computed as a number of long stock characteristics minus the number of short stock characteristics for a given stock–month.

To investigate the relationship between subjectivity and anomaly returns, we estimate the following regressions, the first of which follows the framework of Engelberg et al. (2018) and the second one introduces subjectivity:

$$\begin{aligned}
R_{f,t} =& \beta_1 Net_{f,t} + \beta_2 Eday_{f,t} + \beta_3 Net_{f,t} \times Eday_{f,t} + \\
& \sum_{i=1}^{10} \gamma_i R_{f,t-i} + \sum_{i=1}^{10} \delta_i R^2_{f,t-i} + \sum_{i=1}^{10} \rho_i LogVol_{f,t-i} + \epsilon_{f,t}, \\
R_{f,t} =& \beta_1 Net_{f,t} + \beta_2 Eday(Subj)_{f,t} + \beta_3 Eday(Obj)_{f,t} + \\
& \beta_4 Net_{f,t} \times Eday(Subj)_{f,t} + \beta_5 Net_{f,t} \times Eday(Obj)_{f,t} + \\
& \sum_{i=1}^{10} \gamma_i R_{f,t-i} + \sum_{i=1}^{10} \delta_i R^2_{f,t-i} + \sum_{i=1}^{10} \rho_i LogVol_{f,t-i} + \epsilon_{f,t}
\end{aligned}$$

where$R_{f,t}$ and $LogVol_{f,t}$ are daily stock returns and log volume, $Net_{f,t}$ is constructed as discussed above, $Eday_{f,t}$ is an indicator variable equal to one if the firm f has an earnings call on day t (or day $t-1$ but after market hours), and $Eday(Subj)_{f,t}$ and $Eday(Obj)_{f,t}$ are indicator variables equal to one if there is an earnings call and the call has above or below median percentage of subjective sentences, respectively. In order for coefficient magnitudes to be comparable with Engelberg et al. (2018), we follow their procedure of multiplying returns by 100, and dividing Net by 100.

Figure 8 presents the results. Panel A shows that we can qualitatively replicate the results of Engelberg et al. (2018): stocks that are long on anomalies tend to have higher returns (β of Net is 0.25, p-value < 0.001), returns are higher on earnings call days (β of $Eday$ is 0.15, p-value < 0.001) and, crucially, anomaly returns are much higher on earnings call days (β of the interaction $Net \times Eday$ is 1.11, p-value $<$ 0.05). Panel B presents our main result in this section: the higher anomaly returns manifest themselves only when the earnings call has above median subjectivity (β of

$Eday(Subj)_{f,t}$ is 0.24, p-value < 0.001 and β of $Net \times Eday(Subj)_{f,t}$ is 1.79, p-value < 0.01, while β's associated with $Eday(Subj)_{f,t}$ are not significant at 5% level).

To see whether the difference comes from the long or the short side anomaly splits, we follow Engelberg et al. (2018) and replace the Net variable with indicator variables $HighNet$ and $LowNet$ that are equal to one when Net is larger or smaller than zero, respectively. The results are presented in Figure 9. Panels A and C show that we qualitatively replicate results of Engelberg et al. (2018) about anomaly returns on earnings days being more positive on the long side and more negative on the short side. However, as we see on Panel B, only subjective earnings calls are associated with long side anomaly returns being more positive (β's of $LowNet \times Eday(Subj)_{f,t}$ is 0.16, p-value < 0.05), the coefficients associated with $Eday(Obj)$ are not statistically significant. On the short side (Panel D), however, subjective and objective calls have a very similar association with negative returns (β's of $LowNet \times Eday(Subj)_{f,t}$ and $LowNet \times Eday(Obj)_{f,t}$ are both -0.14, p-value < 0.05). These results suggest that larger anomaly returns are realized on the days of subjective earnings calls due to the long side of anomaly portfolios.

4.5. Subjectivity, fundamentals and returns

Finally, we turn to see if subjective and objective disclosure have different roles in explaining contemporaneous and subsequent measures of firm performance – stock returns and earnings. We do so by estimating separate machine learning models in which the text in the respective set of sentences (objective or subjective) is used as the set of explanatory variables. By doing so, we allow the set of relevant terms to differ based on the nature of disclosure and the performance variable.

Specifically, we use the following variables y:

1. **Δ_4 Net Income (Q0)**: the difference between net income in the current quarter and the net income in the same quarter last year, divided by total assets in the previous quarter.[15]
2. **Δ_4 Net Income (Q+1)**: the difference between net income in the next quarter and the net income in the same quarter last year, divided by total assets in the current quarter.
3. **Return (D0)**: abnormal return for the day of earnings call (nearest market close before to nearest market close after).[16]
4. **Return (D+1)**: abnormal return for the day after the earnings call.

For each of the variables y above we construct $\hat{y}(Obj)$ and $\hat{y}(Subj)$, the adjusted values of y implied by a bag-of-words model trained on the objective and subjective content of the earnings call, respectively. The adjustment of y consists of subtracting the mean of y for 8 preceding quarters. This reduces the amount of non-zero bag-of-words coefficients associated with firm activities that are stable over time (similar to firm fixed effects) rather than markers of news polarity, which we are primarily interested in.[17]

We estimate $\hat{y}$ using adaptive lasso model Zou (2006).[18] We reestimate the models

[15]We use variables IBQ (net income before extraordinary items) and ATQ (total assets) in Compustat.

[16]To compute abnormal returns, 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.

[17]This procedure does not materially affect the regressions using $\hat{y}$ which we examine later since those include firm fixed effects.

[18]The estimation involves two steps: ridge regression to obtain regularization weights, and lasso

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. The adaptive lasso model is linear log counts for words:

$$\hat{y}(Obj) = \beta_0 + \beta^T x(Obj),$$
$$\hat{y}(Subj) = \beta_0 + \beta^T x(Subj),$$

where β_0 is the intercept; and β is the vector of regression coefficients (the coefficients for objective and subjective regressions are estimated separately using the same procedure); $x(Obj)$ and $x(Subj)$ are vectors of log frequencies of common tokens from the objective and subjective sentences.

The objective function (same for objective and subjective content) is the following:

$$L\left(\{\beta_0, \beta\}\right) = -\left[\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \beta_0 + \beta^T x_i\right)\right] + \lambda \sum_{j=1}^{p} \hat{\omega}_j |\beta_j|,$$
$$\hat{\omega}_j = \frac{1}{\left|\hat{\beta}_j^{\text{ridge}}\right|},$$

λ is the hyperparameter that controls overall strength of lasso regularization; p is the total number of coefficients associated with all categories; and ω is a vector of weights estimated using a ridge regression. We choose λ using 10-fold cross-validation on the training sets.

We use log-frequencies of individual tokens (unigrams). Let $\text{freq}(j, n)$ denote the frequency of the term j in the document n. The associated independent variable

---

regression. We implement these steps using R library *glmnet* (Friedman et al., 2010).

is $x_{n,j} = \log(1 + \text{freq}(j, n))$. The specification includes the 1,000 most common unigrams. We use Snowball stemmer's stopword list to remove some ubiquitous English words like "the."[19] 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.

4.5.1. Regression setup

We assess the information content and complementarity of objective and subjective content using a fixed-effects regression framework. For $y \in \{\Delta_4$ Net Income (Q0), Return (D0)$\}$, we estimate

$$y = \beta_1 \hat{y}(Obj) + \beta_2 \hat{y}(Subj) + FE_{Firm} + FE_{year/quarter} + \epsilon,$$

and for $y \in \{\Delta_4$ Net Income (Q+1), Return (D+1)$\}$, we estimate

$$y = \beta_1 \hat{y}(Obj) + \beta_2 \hat{y}(Subj) + \beta_3 y_0 + FE_{Firm} + FE_{year/quarter} + \epsilon,$$

where $\hat{y}(Obj/Subj)$ is the out-of-sample output of the bag-of-words model using tokens from objective or subjective sentences; y_0 is Δ_4 Net Income (Q0) or Return (D0) depending on y; and FE_{Firm} and $FE_{year/quarter}$ are the fixed effects. We cluster the errors at the firm level.

We winsorize all variables at 1% and 99% levels and standardize them by subtract-

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

ing the mean and dividing by the standard deviation, so the regression coefficients measure % of standard deviation change in the left-hand-side variable associated with one standard deviation change in the right-hand-side variable.

4.5.2. Regression results

Objective and subjective sentences have a different association with current and future fundamentals and returns. Figure 10 demonstrates the relative explanatory power of objective and subjective content for current and next quarter's changes in net income relative to the same quarter a year ago (Panels A and B), the 1-day return around the earnings call (Panel C), and return for next day after the earnings call (Panel D). We see a diverging pattern: objective sentences have much higher exploratory power for current change net income, and subjective content has relatively little complementary information content. However, future changes in net income and returns are explained equally well by objective and subjective content; the two are highly complementary. These results suggest that investors actively trade both on objective content that reflects both current and future fundamentals and on subjective content that is as informative about the future as objective content, but less informative about the present. The information is largely incorporated into prices within the day of the call.

5. Relation to other prominent linguistic measures

5.1. Subjectivity vs numbers vs forward-looking sentences

In this section, we discuss to what extent objectivity coincides with other potentially related measures: the presence of numbers in disclosure and the presence of forward-looking markers. Numbers are a common way to represent many possible facts about the firm and are much easier to identify than linguistic markers of objectivity or subjectivity. It is also natural to think that subjectivity is more prevalent in discussions about the future because a degree of speculation is necessarily involved. Therefore, it is important to establish that our measure brings something to the table in addition to the simple presence of numbers or markers of forward-looking statements identified in previous research (for example, Muslu et al. 2015). We show that our measure identifies a large number of objective sentences that do not contain numbers, and that forward-lookingness is not a substitute for a subjectivity measure.

Panels A and B of Figure 11 explore the relationship between subjectivity and the presence of numbers at the sentence level. We identify a sentence as containing numbers if it contains a digit that is not a part of a year or date, phone number, or slide number. That leaves mostly numbers representing units, percentages, or dollar amounts. Panel A shows the percentage of sentences with and without numbers among objective, subjective and irrelevant sentences. Objective sentences are split between sentences with and without numbers, 59% to 41%. In contrast, the vast majority of subjective sentences do not contain numbers (93%). Looking at the prevalence of objectivity and subjectivity among sentences with and without numbers (Panel B), we see 85% to 14% and 20% to 62% split. These results show that while the presence of numbers is strongly and negatively associated with subjectivity, a dedicated measure of objectivity is required to be able to identify numerous objective sentences that do not contain numbers.[20]

[20]Some generic examples: "we signed a contract with company X," "we opened a new store," "we

Panels C and D of Figure 11 explore the relationship between subjectivity and the presence of forward-looking markers (identified following Muslu et al. (2015)). Note that we do not automatically consider any statement about the future to be subjective. At the annotation stage, we choose to treat forward-looking sentences in the same way as sentences about the past, meaning that we treat a sentence "we expect EPS of 10 cents" as objective and a sentence "we expect great EPS" as subjective (see Section 3.1). This allows us to see what proportion of forward-looking statements is framed as a statement that would be objectively true or false post hoc. As Panel D shows, that proportion is large, as 79% of forward-looking sentences are classified as objective and 19% as subjective. In contrast, non-forward-looking sentences are split more evenly, 35% to 51%. Additionally, Panel C demonstrates that forward-looking sentences are relatively rare (6% of all objective sentences and 1% of all subjective sentences). This shows that forward-lookingness and subjectivity are not conceptual substitutes.

5.2. Sentiment dictionaries

While the usefulness of sentiment measures is undisputable, we still don't understand what exactly they measure. While the name "sentiment" implies a strong connection with emotions and opinions, financial disclosure is based on facts, and applying standard sentiment measures to factual content can result in mixing the polarities of opinions and facts. In this section, we use our measures of objectivity and subjectivity to show that negative sentiment words are slightly more likely to be

experienced an outage."

associated with facts than opinions, and only positive sentiment words are more likely to be associated with opinions. Therefore, we argue that "text polarity" is a better term than "sentiment" for the common dictionary- and machine learning-based text measures.

The term "sentiment" in common usage is inseparable from feelings and opinions. The first two definitions in the Merriam-Webster dictionary are "an attitude, thought, or judgment prompted by feeling (synonym: predilection)" and "a specific view or notion (synonym: opinion)."[21] Likewise, computational linguistics considers sentiment to be a property of subjective but not objective text. For example, Esuli and Sebastiani (2006) follow a three-step sentiment classification procedure, where steps one and two are "deciding whether a given text has a factual nature" and "deciding if a given Subjective text expresses a Positive or a Negative opinion on its subject matter."[22] In financial disclosure context, the distinction between facts and opinions in the text has not been previously studied, but its relevance for the concept of sentiment has been acknowledged, for example, by Ke et al. (2020) who state that their text-based return prediction approach "does not differentiate between non-sentiment (i.e., objective information) and sentiment content of news per se."

To see whether the common dictionary measure of sentiment (Loughran and McDonald's financial domain sentiment dictionary (Loughran and McDonald,

[21]Other definitions are related to emotion or opinion as well. See https://www.merriam-webster.com/dictionary/sentiment. Last accessed: 06/28/2021.

[22]Step three is about the magnitude of sentiment, "deciding e.g. whether the Positive opinion expressed by a text on its subject matter is Weakly Positive, Mildly Positive, or Strongly Positive."

2011)) captures opinions rather than facts, we state two hypotheses, separately for negative and positive sentiment dictionaries.[23]

Hypothesis 1: Sentences including words from the *negative* sentiment dictionary are more likely to be subjective.
Hypothesis 2: Sentences including words from the *positive* sentiment dictionary are more likely to be subjective.

To test these hypotheses, we estimate the following linear probability model:

$$\mathbb{1}\{Subj\}_{c,s} = \beta_1 \mathbb{1}\{Neg > 0\}_{c,s} + \beta_2 \mathbb{1}\{Pos > 0\} + FE_f + FE_q + \epsilon_{c,s},$$

where c and s are indices for the earnings call and for sentences within the earnings call, $\mathbb{1}\{Subj\}_{c,s}$ is the indicator variable equal to one when the sentence is classified as subjective, $\mathbb{1}\{Neg > 0\}_{c,s}$ and $\mathbb{1}\{Neg > 0\}_{c,s}$ are indicator variables equal to one when a sentence has at least one word from the negative or positive sentiment dictionary, and FE_f and FE_q are firm and year-quarter fixed effects. Sentences classified as irrelevant are excluded. Standard errors are clustered at the firm level.

For sentiment scores to adequately capture opinions rather than facts, the presence of negative and positive words should be positively associated with the probability of a sentence being subjective. Therefore we expect both β_1 and β_2 to be positive and large.

[23]The asymmetric importance of negative and positive word lists is well known and discussed, for example, in Loughran and McDonald (2020).

Figure 12 presents the results. In the full specification (Panel A), we cannot reject the null hypothesis of no association between the presence of negative words and subjectivity, meaning that words from the negative sentiment dictionary are as likely to appear in subjective as objective sentences. In contrast, the β for the presence of positive words is significant with the magnitude of 0.216, corresponding to a 22 percentage point increase in the probability of a sentence being subjective when the words from the positive sentence dictionary are present. In the specification without the fixed effects (Panel B) and univariate specifications (Panels C and D), the β for positive words stays approximately the same, while the β for negative words is either not statistically significant, or negative and very small in magnitude.

Figure 13 presents relevant descriptive evidence. Panel A shows that negative and neutral sentences are split almost evenly between facts and opinions (roughly 47% to 53%), while for positive sentences the split is skewed towards opinions (25% to 75%).[24] Aggregating data the other way (Panel B), we see that a similar number of objective sentences are negative and positive (12% and 15%), while the split of subjective sentences is skewed towards positive sentiment (10% to 32%).

Overall, the regressions show that the negative sentiment dictionary applied to earnings calls does not capture managerial opinions rather than facts, and instead captures an equal mixture of both. In contrast, the words from the positive sentiment dictionary are more likely to capture opinions rather than facts. However, positive words are still often used in factual sentences (25 % of cases). Therefore, we conclude

[24] We classify sentences with more negative than positive words as negative, sentences with more positive than negative words as positive, and sentences with an equal number of both (most commonly zero) as neutral.

that "polarity" is a more appropriate term than "sentiment" for text-based measures generated by applying sentiment dictionaries to financial disclosure.

6. Conclusion

We show that separating facts from opinions is important for understanding the nature of corporate textual communication. We propose a definition of subjectivity based on computational linguistic literature, operationalize it with respect to earnings calls, create an annotated dataset, and train a model that classifies sentences as predominantly objective, subjective or irrelevant. Applying this automated classification on a large corpus of earnings calls, we show that facts and opinions are both highly prevalent in earnings calls, across all years and firm groups, and that the usage of subjective language varies systematically across firm types, executive roles, and the section of the call. More importantly, we find evidence consistent with the idea that subjective and objective disclosure represents fundamentally different types of information or that investors process that information differently. This is evident by the differential role subjective and objective disclosure has on disagreement, anomaly returns, and firm fundamentals.

References

Antweiler, W., Frank, M.Z., 2001. Is all that talk just noise? The information content of internet stock message boards. The Journal of Finance 59, 1259–1294.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A., 2018. quanteda: An R package for the quantitative analysis of textual data. Journal of Open Source Software 3, 774.

Cochrane, J.H., 2011. Presidential address: Discount rates. The Journal of Finance 66, 1047–1108.

Cookson, J.A., Niessner, M., 2020. Why don't we agree? Evidence from a social network of investors. The Journal of Finance 75, 173–228.

Engelberg, J., McLean, R.D., Pontiff, J., 2018. Anomalies and news. The Journal of Finance 73, 1971–2001.

Esuli, F.A., Sebastiani, 2006. SentiWordNet: A publicly available lexical resource for opinion mining, in: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006).

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

Gentzkow, M., Kelly, B., Taddy, M., 2019. Text as data. The Journal of Economic Literature 57, 535–74.

Green, J., Hand, J.R.M., Zhang, X.F., 2017. The characteristics that provide independent information about average U.S. monthly stock returns. The Review of Financial Studies 30, 4389–4436.

Hong, H., Stein, J.C., 2007. Disagreement and the stock market. The Journal of Economic Perspectives 21, 109–128. URL: https://www.aeaweb.org/articles?id=10.1257/jep.21.2.109, doi:10.1257/jep.21.2.109.

Ke, Z., Kelly, B.T., Dacheng, X., 2020. Predicting returns with text data. Working paper.

Kessler, J.S., 2017. Scattertext: a browser-based tool for visualizing how corpora differ, in: Proceedings of ACL-2017 System Demonstrations, Association for Computational Linguistics, Vancouver, Canada.

Li, F., 2010. The information content of forward-looking statements in corporate filings—a naïve bayesian machine learning approach. The 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.

Loughran, T., McDonald, B., 2020. Textual analysis in finance. Working paper.

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

Taboada, M., 2015. Sentiment analysis: An overview from linguistics. Annual Review of Linguistics 2, 1–23.

Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: Quantifying language to measure firms' fundamentals. The Journal of Finance 63, 1437–1467.

Wiebe, J., 1994. Tracking point of view in narrative. Computational Linguistics 20, 233–287.

Zou, H., 2006. The adaptive lasso and its oracle properties. Journal of the American Statistical Association 101, 1418–1429.

Tables

Table 1: Descriptive statistics, Part 1. Number of earnings calls and firms, mean number of sentences in presentation and answer sections by year.

Year	Calls	Firms	Avg Sent Pres	Avg Sent Ans
2008	4,280	1,938	144	145
2009	6,377	2,096	137	138
2010	7,026	2,267	137	157
2011	7,887	2,479	139	158
2012	7,966	2,475	139	160
2013	8,048	2,494	137	158
2014	8,211	2,503	138	164
2015	8,048	2,496	142	170
2016	7,886	2,428	143	172
2017	8,033	2,516	136	167
2018	8,188	2,537	137	165
2019	3,890	2,261	136	164
Full Sample	85,840	4,346	139	161

Table 2: Descriptive statistics, Part 2. Mean, 25'th percentile and 75'th percentile of % of sentences classified as objective and subjective by year.

Year	Avg Sent Obj, %	Avg Sent Subj, %	P25 Sent Obj, %	P25 Sent Subj, %	P75 Sent Obj, %	P75 Sent Subj, %
2008	0.391	0.473	0.33	0.412	0.445	0.536
2009	0.395	0.479	0.332	0.417	0.452	0.546
2010	0.379	0.488	0.322	0.429	0.432	0.55
2011	0.379	0.478	0.321	0.418	0.433	0.543
2012	0.382	0.477	0.32	0.419	0.436	0.545
2013	0.376	0.48	0.316	0.422	0.431	0.544
2014	0.369	0.484	0.31	0.427	0.423	0.548
2015	0.366	0.485	0.31	0.428	0.417	0.549
2016	0.361	0.489	0.304	0.433	0.412	0.55
2017	0.36	0.489	0.302	0.43	0.413	0.553
2018	0.363	0.483	0.305	0.426	0.416	0.548
2019	0.36	0.483	0.303	0.425	0.415	0.547
Full Sample	0.373	0.483	0.314	0.424	0.427	0.547

Figures

Figure 1: Performance of the objectivity classifier on the test sample. Panel A shows the F1 scores for individual sentence classes and the whole test set, Panel B shows average F1 scores for each year, Panel C shows average F1 scores for the twelve Fama-French industries, Panel D shows F1 score for documents associated with negative earnings surprises accompanied by negative returns ("bad news") and all other documents ("not bad news").

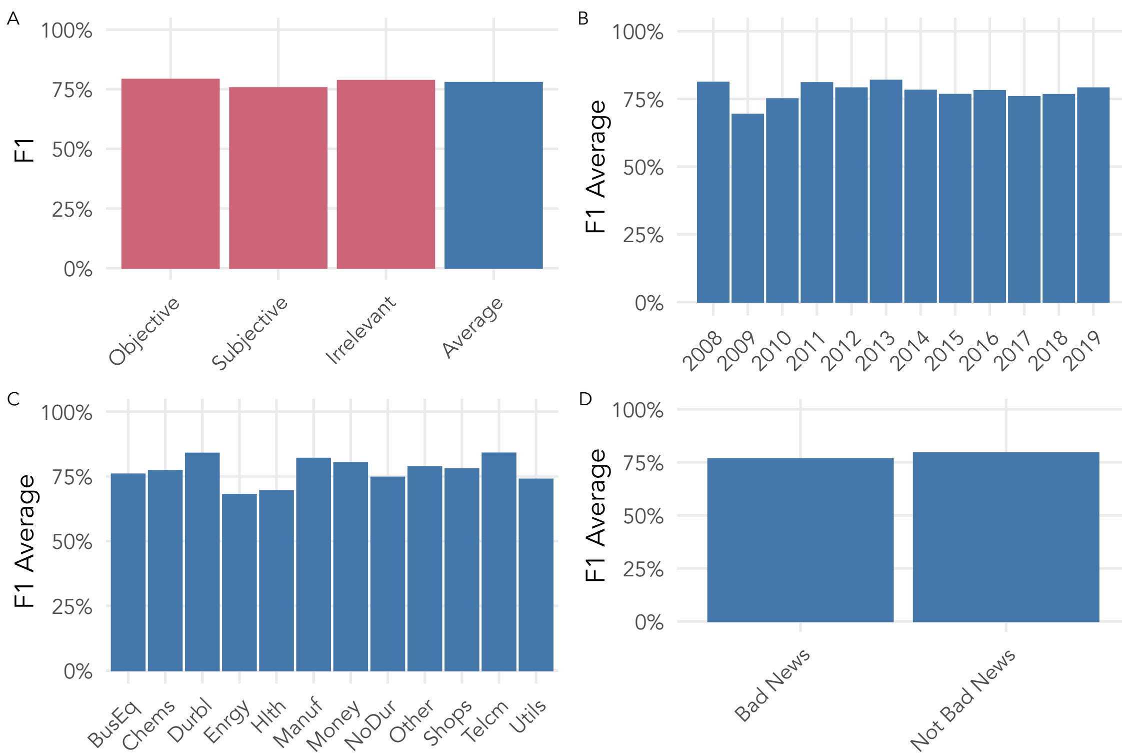
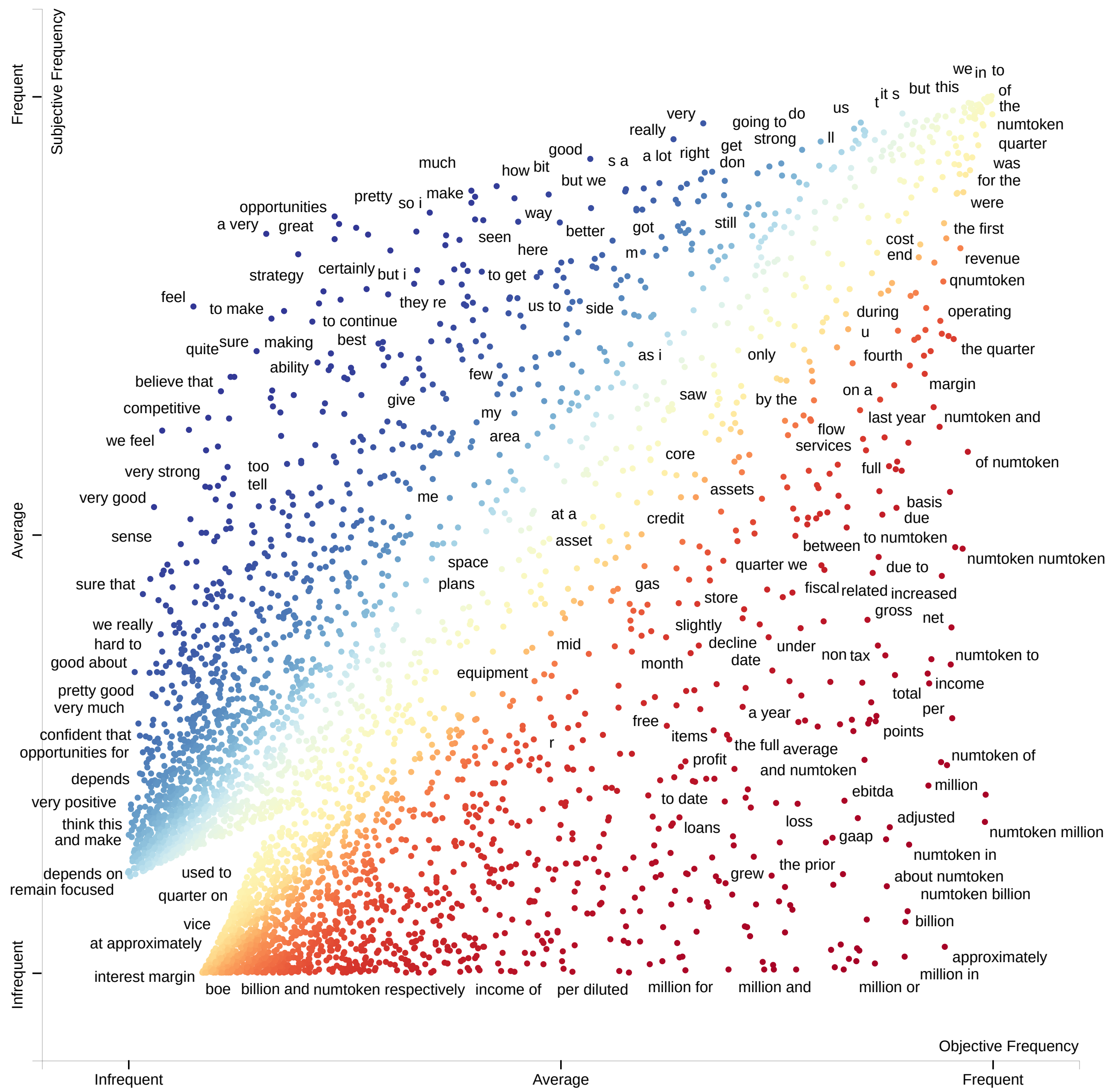


Figure 2: Relative unigram and bigram frequencies across objective and subjective sentences. X- and y- axes represent the dense ranks of tokens in the sentence class.



Search the chart

Figure 3: Prevalence of objective, subjective and irrelevant sentences across years (Panel A), eleven Fama-French industries (Panel B), and size and book-to-market portfolios (Panel C)

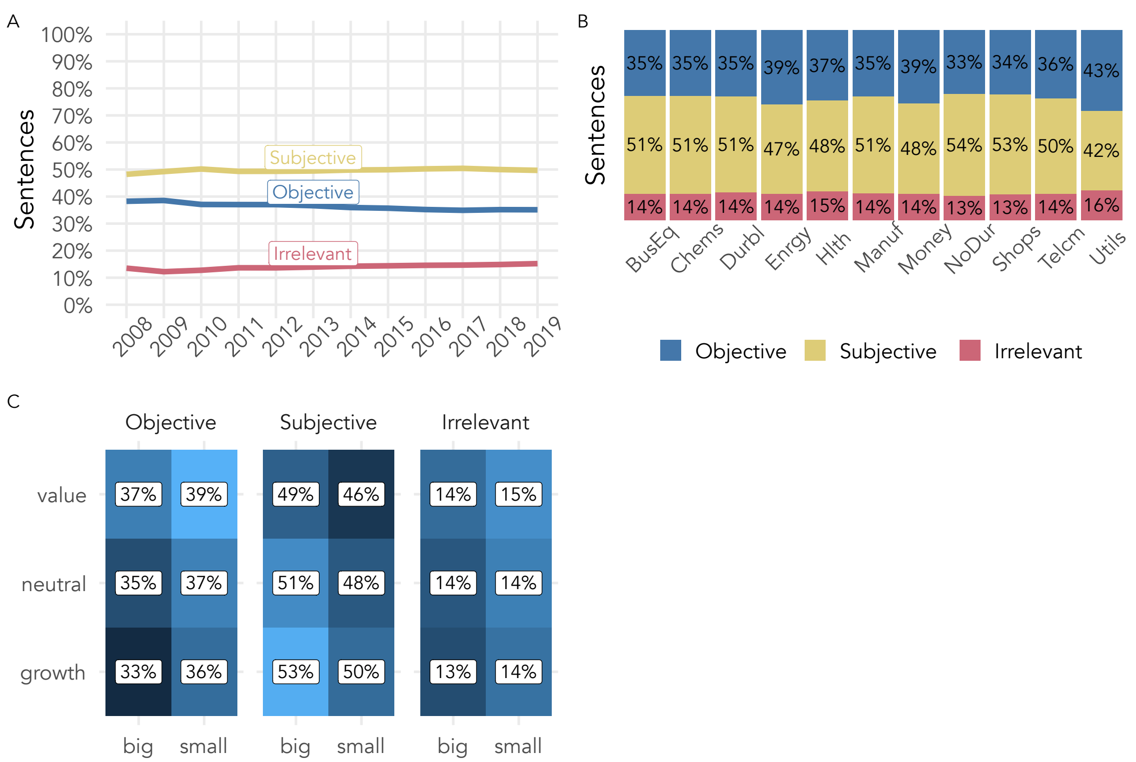


Figure 4: Prevalence of objective, subjective and irrelevant sentences across executive type (CEO or CFO) and earnings call section (Panel A), and sentence position by exectuive type and section (Panel B).

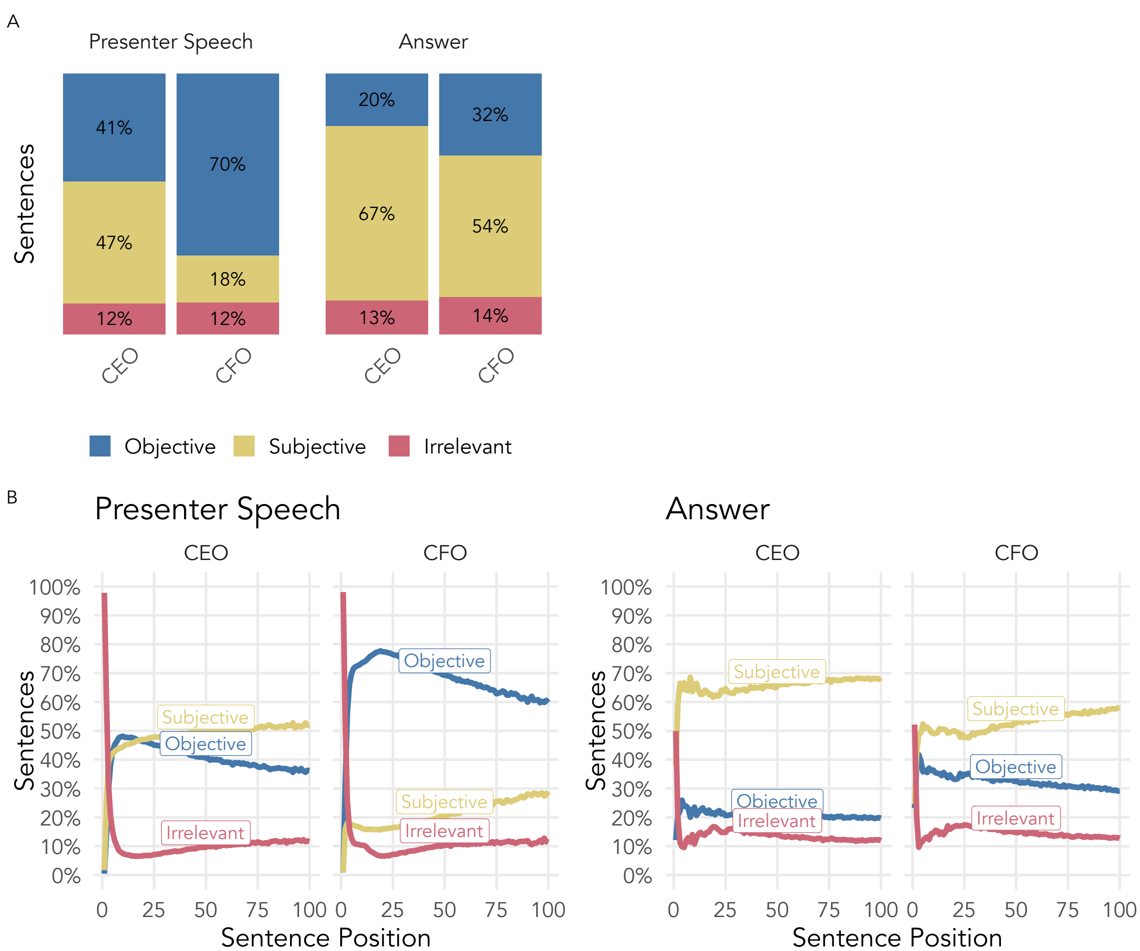


Figure 5: Subjectivity as a measure of executive style or firm activities, regression evidence. The observations represent cases when the executive of the firm is replaced by another executive of the same type from our sample. The y variable represents subjectivity of the new executive one quarter after the move. The right-hand-side variables represent subjectivity of the predecessor one quarter before the last one and the new executive's own subjectivity at their previous firm. P-values less than 0.05, 0.01 and 0.001 are indicated by *,** and ***.

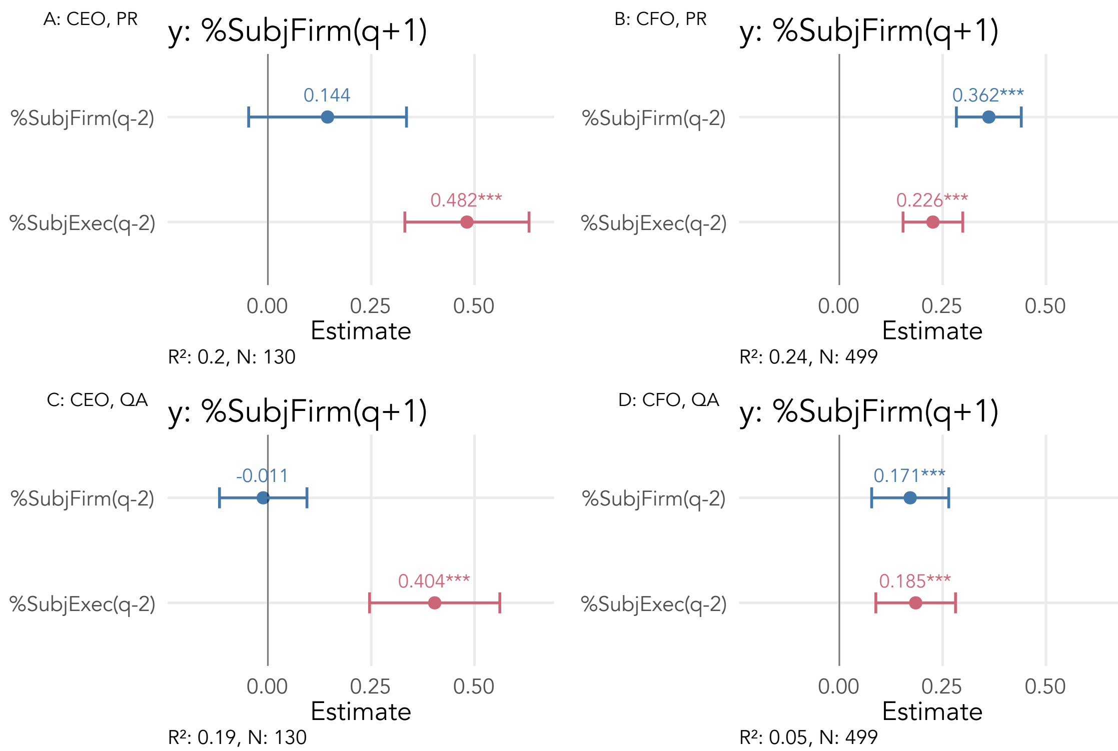


Figure 6: Subjectivity and investor disagreement, fixed effects regression evidence. Each panel represents one regression. The y variables are abnormal log volume (Panel A) and three social media-based measures of investor disagreement from Cookson and Niessner (2020) (Panels B, C, D). The right-hand-side variable, $\%SubjSent$ is percentage of subjective sentences in the earnings call. P-values less than 0.05, 0.01 and 0.001 are indicated by *,** and ***.

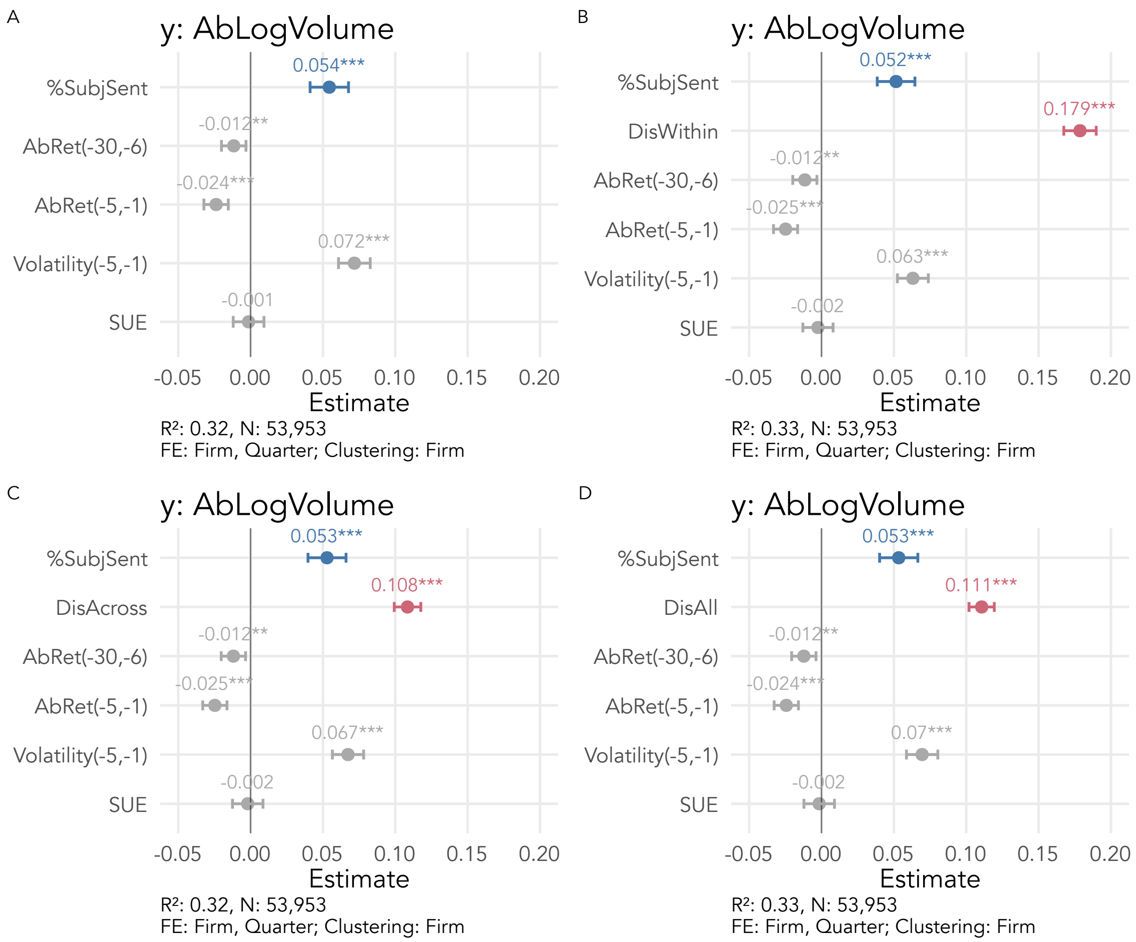


Figure 7: Subjectivity and investor disagreement, fixed effects regression evidence. Each panel represents one regression. The y variables are abnormal log volume (Panel A) and three social media-based measures of investor disagreement from Cookson and Niessner (2020) (Panels B and C). The right-hand-side variable, $\%SubjSent$ is percentage of subjective sentences in the earnings call. P-values less than 0.05, 0.01 and 0.001 are indicated by *,** and ***.

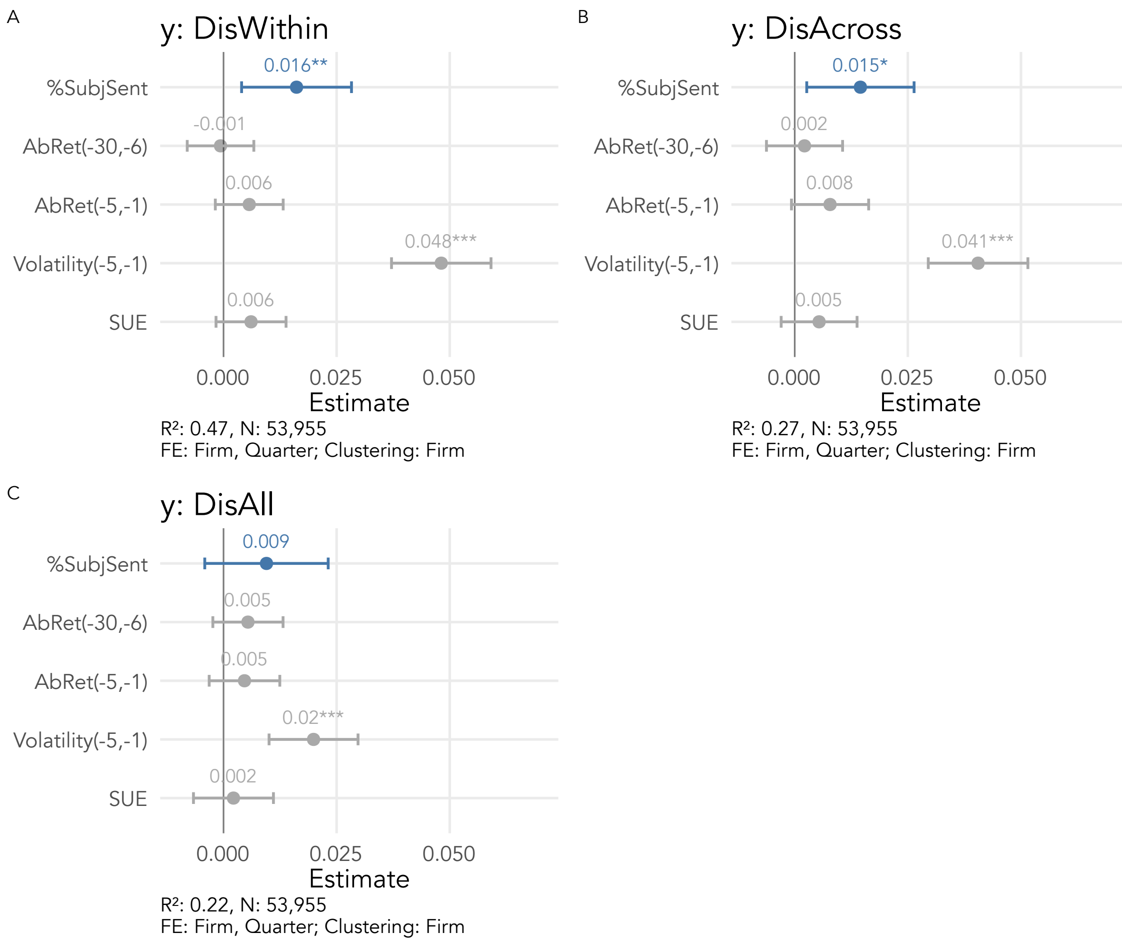


Figure 8: Subjectivity and anomaly returns, fixed effects regression evidence. The regression is run at the stock-day level. The Net variable is the difference between the number of long and short anomaly portfolios the stock is in during a given month, $Eday$ is an indicator variable equal to one if the day is associated with an earnings call for a given stock, and $Eday(Subj.call)$ and $Eday(Obj.call)$ are indicator variables equal to one if the day is associated with an earnings call for a given stock and the call has above or below median percentage of subjective sentences. P-values less than 0.05, 0.01 and 0.001 are indicated by *,** and ***.

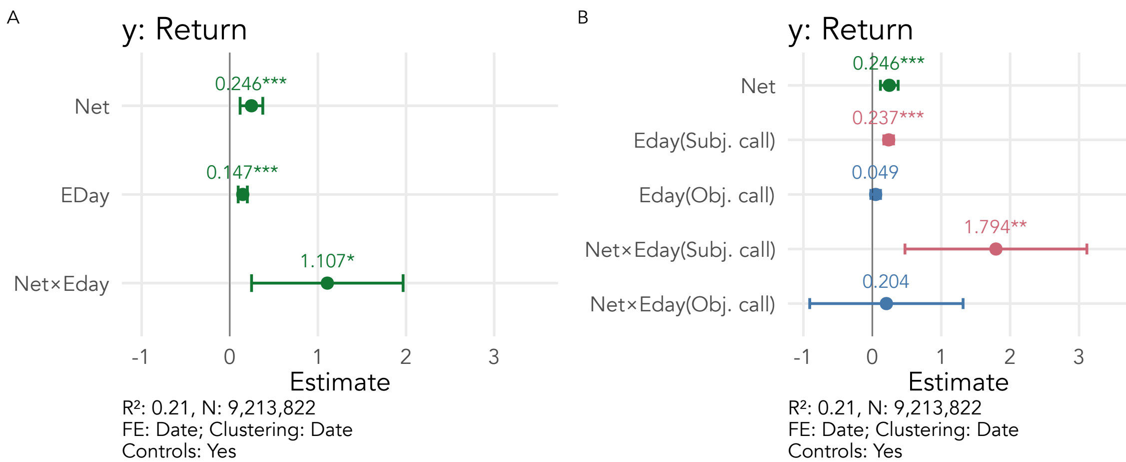


Figure 9: Subjectivity and anomaly returns, fixed effects regression evidence. The regression is run at the stock-day level. The $HighNet$ and $LowNet$ variables are indicators equal to one when the Net variable defined in the text is positive or negative, $Eday$ is an indicator variable equal to one if the day is associated with an earnings call for a given stock, and $Eday(Subj.call)$ and $Eday(Obj.call)$ are indicator variables equal to one if the day is associated with an earnings call for a given stock and the call has above or below median percentage of subjective sentences. P-values less than 0.05, 0.01 and 0.001 are indicated by *,** and ***.

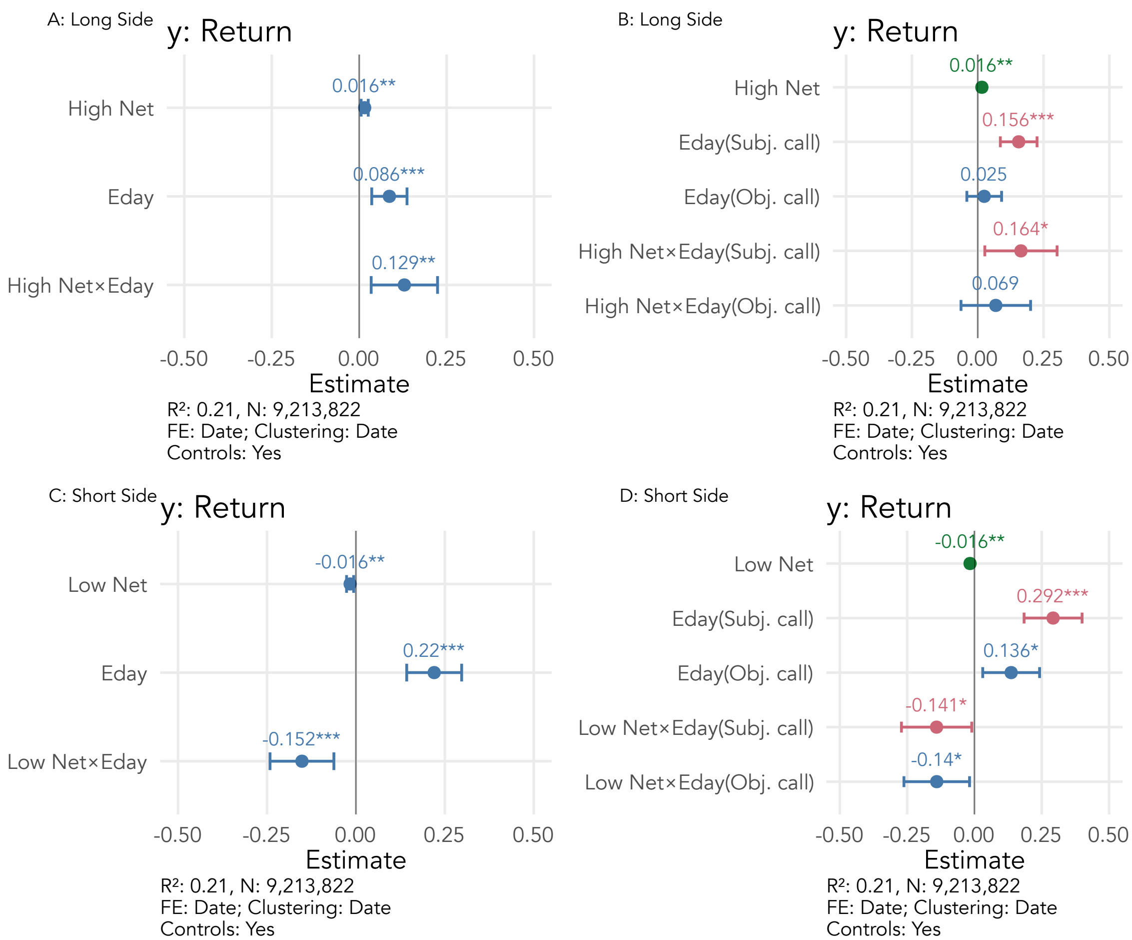


Figure 10: Ability of objective and subjective content to explain fundamentals and returns, fixed effects regression evidence. The y variable is change in net income relative to the same quarter last year (Panel A), one-quarter-ahead change in net income relative to the same quarter last year (Panel B), abnormal return for the day of the earnings call (Panel C) and abnormal return for the day after the earnings call (Panel D). The main right-hand-side variables are the $\hat{y}$'s implied by objective and subjective content respectively. The $\hat{y}$'s are rolling window out-of-sample outputs of an adaptive lasso bag-of-words model.

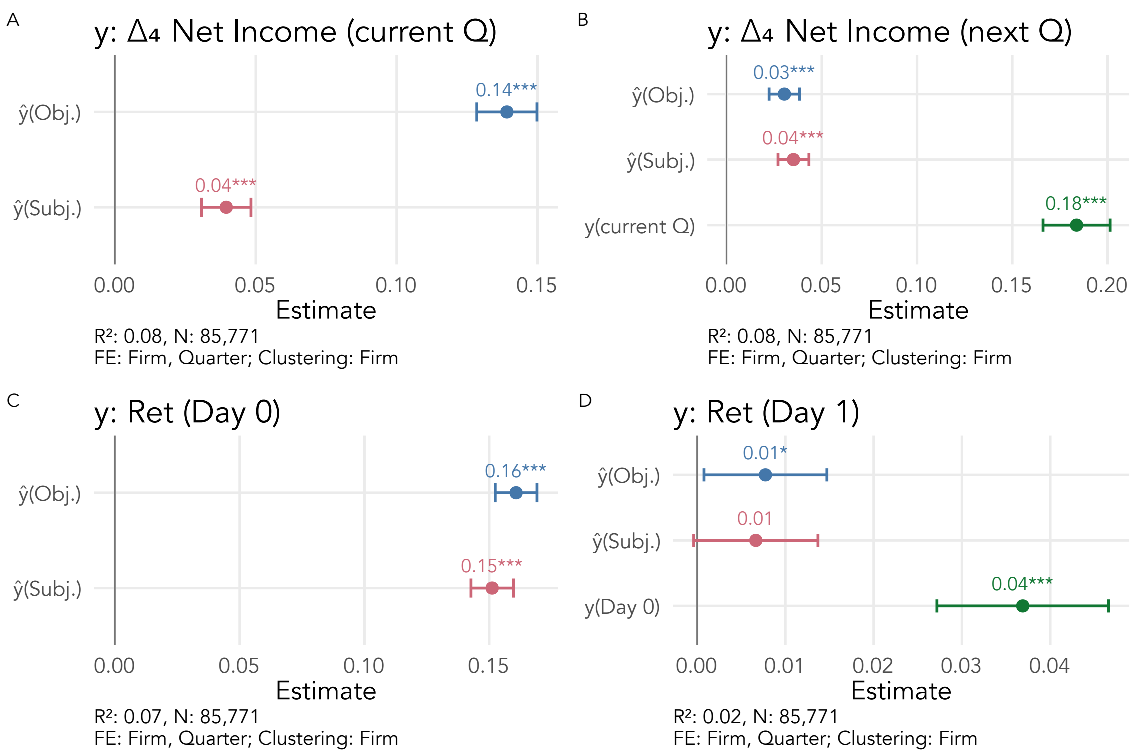


Figure 11: Objectivity and the presence of numbers (excluding years, dates and phone numbers) and forward-looking markers based on Muslu et al. (2015). Panel A shows percentage of objective, subjective and irrelevant sentences that contain or do not contain numbers. Panel B shows the percentage of sentences with and without numbers that are objective, subjective or irrelevant. Panel C shows percentage of objective, subjective and irrelevant sentences that contain or do not contain forward-looking markers. Panel D shows the percentage of sentences with and without forward-looking markers that are objective, subjective or irrelevant.

Figure 12: Sentiment scores and subjectivity, fixed effects regression evidence. The regression is run at the sentence level. The y variable is an indicator equal to one if the sentence is subjective. The right-hand-side variables are indicators equal to one when the sentence has at least one word marked as negative or positive in the Loughran and McDonald financial sentiment dictionary. Panel A presents fixed effect regression results, Panel B presents multivariate regression results without fixed effects, and Panels C and D present univariate regressions results. P-values less than 0.05, 0.01 and 0.001 are indicated by *,** and ***.

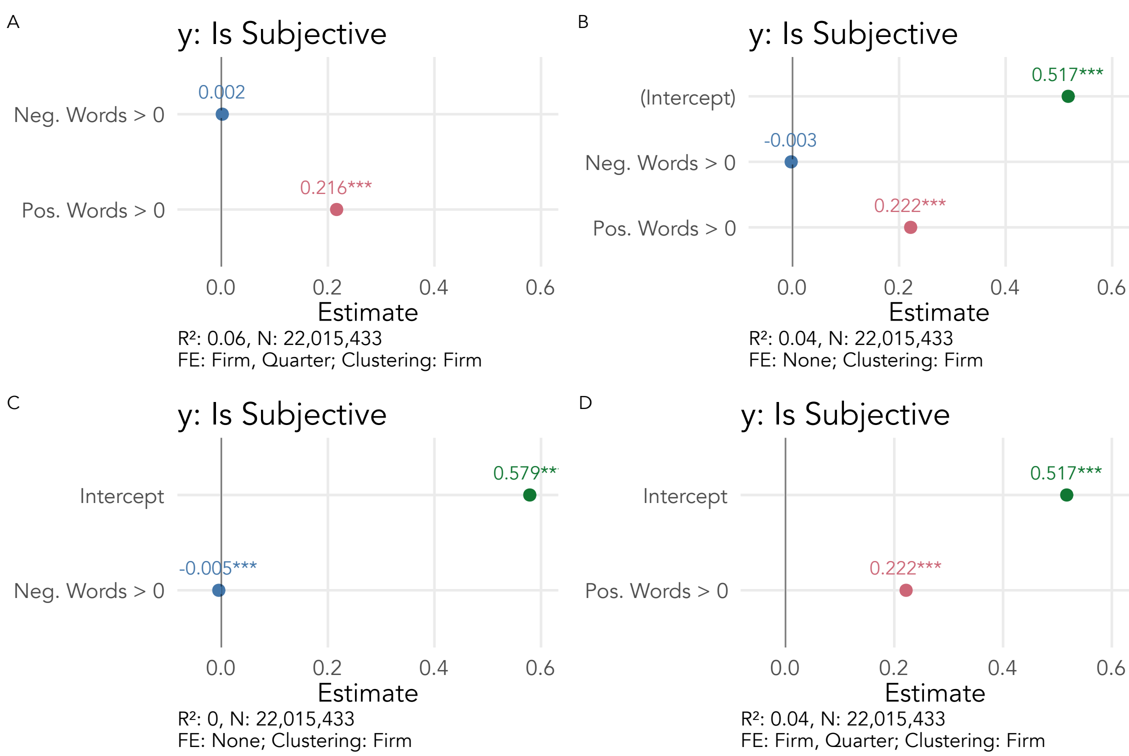


Figure 13: Prevalence of sentiment words from Loughran and McDonald dictionary across objective, subjective and irrelevant sentences. Panel A shows what percentages of tokens in a given sentence class appears in negative or positive sentiment dictionaries. Panel B shows the split between negative and positive words appearing in objective, subjective and irrelevant sentences.