Natural Language Processing – Part III: Feature Engineering

Applying NLP Using Domain Knowledge to Capture Alpha from Transcripts

Author
Frank Zhao
Quantamental Research
617-530-8107
fzhao@spglobal.com

Unstructured data[1] are largely underexplored in equity investing due to their higher costs[2]. The information content, as a result, remains largely untapped and offers an investment edge[3] to discerning investors who are adept at extracting those investment insights. One particularly valuable unstructured data set is S&P Global Market Intelligence's earnings call transcripts.

This publication, the third in the series (NLP I, NLP II), introduces new stock selection ideas in the areas of I) Topic Identification, II) Call Transparency and III) Call Sentiment using more advanced NLP techniques. The new signals complement the existing suite of signals that is offered in our Textual Data Analytics (TDA)[4] product, an off-the-shelf NLP solution for both quantitative and fundamental analysts. Our high-level U.S. findings are:

Topic Identification - Firms That Referenced the Most Positive Descriptors around Their Financials Outperformed Historically. Firms whose executives most frequently articulated references to growth- and expansion-related descriptors around I) revenue II) earnings or III) profitability topics outperformed their counterparts by 9.16%, 8.60% and 6.76% per year, respectively (Table 1). This group of signals has historically shown additive forecasting power above those from our NLP II paper (Table 4) and commonly used signals (Table 13).

Transparency - Firms That Provided Greater Call Transparency Exhibited by Executives' Behaviors and Decisions Outperformed Historically. Firms whose executives I) referenced the most instances of guidance or II) used the most similar language between calls outperformed their counterparts by 5.84% and 3.75% per year, respectively (Table 8 and 10). Firms whose executives III) introduced numerical values earlier or IV) blamed exogenous elements the least outperformed their counterparts by 3.97% and 2.39% per year, respectively (Table 11).

Weighted Average Sentiment - Quantifying Call Sentiment Using A Weighted Average Construct Relative to A Simple Average Led to Better Returns and Less Volatility Historically. Firms whose executives used the most positive language in the most recent call that was scarcely used in the previous calls outperformed their counterparts by 5.18% per year (Table 5). A weighted average sentiment, relative to a simple average, improved the annualized information ratio to 1.81 from 1.19 historically (Table 6).

Additive Forecasting Power - The Newly Introduced Signals Demonstrated Additive Forecasting Power above Commonly Used Alpha and Risk Signals Historically. The additive economic performance ranged from 2.02% to 6.69% per year after controlling for market, size, value, quality, momentum, earnings surprise and analyst revision (Table 13).

[1] Unstructured data sets are non-numerical data sets such as texts, audios and images that are of primary source. Primary source data in this context i) are furthest up the information chain containing the most relevant and timely information, ii) are easily mapped to publicly traded firms and iii) have good historical and cross-sectional coverage.
[2] The exploration of non-numerical data sets requires more advanced technical infrastructure and tools. The exploration requires more time from researchers to vet the data. The probability of finding additive signals is lower.
[3] Other examples may be more sophisticated modeling (e.g., non-linear) and execution efficiency.
[4] See Section 5 Data & Methodology for details.

1. Introduction

"*Coming up with features is difficult, time-consuming, requires expert knowledge. Applied machine learning is basically feature engineering*." – Andrew Ng[5]

Feature engineering is the process of applying domain expertise and knowledge to a data set to extract predictor variables for modeling. It is the second of the three sequential steps within the pipeline of every data science endeavor and sits in between the data cleaning and the analysis. The application of domain knowledge to generate the predictors is instrumental in allowing the algorithms to better understand the objectives at hand and improves the probability of a project's success. In the context of this paper, we are applying NLP on earnings call transcripts to create stock-level predictors with financial economic underpinnings to forecast stock returns at the forward one-month horizon. The predictors are grouped into three broad categories:

- **Market Moving Topics and Neighboring Descriptors:** Distilling earnings call transcripts down to a set of key financials-related topics (e.g., earnings) and neighboring descriptors to extract insights on the financial state and the trajectory of the firms.

  In our narrative, firms whose executives reference the most instances of a financials-related topic (e.g., earnings) in a directionally positive light should outperform. Our hypothesis is that directional tags around a topic may be a good proxy to gauge whether a firm is growing, stagnating or declining. The frequency of the mentions may be a good proxy to measure the magnitude of that growth and the confidence around it.

- **Weighted Average Sentiment Scores**: In our NLP II paper, we introduced sentiment measures during calls as simple averages. In this paper, we introduce weighted average constructions using two different weighting schemes.

  The first uses a concept from information retrieval called term frequency-inverse document frequency (TF-IDF). The weights (i.e., IDF) are determined by how commonly a word reference appears in a corpus, which in our context is a rolling window of a firm's five most recent earnings calls. A sentiment word that appears scarcely (abundantly) in such a corpus gets a bigger (smaller) weight.

  The second weighting scheme examines the language similarity (i.e., usage) of a firm's call relative to its historical ones. The similarity scores are the weights that are used to amplify or deflate changes in sentiment scores (e.g., year-over-year sentiment change). In our narrative, a low call similarity score (i.e., dissimilarity) is viewed as a flagged instance of oddity or discontinuity.

[5] Founder & CEO of Landing AI, Co-Cofounder and Chairman of Coursera and Adjunct Professor of Computer Science at Stanford University

- **III. Gauging Call Level Transparency via Executives' Behaviors and Decisions**: In our NLP II publication, we explored a number of measures to gauge and quantify the level of transparency that firms provide during earnings calls. A call with a higher level of transparency is deemed of a higher quality and is viewed favorably in our narrative.

  In this publication, we continue that exploration and introduce four additional, differentiated measures: I) Number of References to Guidance (II) Language Similarity between Calls III) The Decision of Timing via How Early Do Firms Introduce Numbers IV) Self-Attribution Bias via the Blame Game.

The following sections discuss the elements of an earnings call (see Appendix A.1), the intuition underlying our signal constructions, and the empirical results of our analysis.

2. Signals, Intuition & Construction

2.1 Market Moving Topics & Neighboring Descriptors

Introduction & Intuition

Since the calendar quarter Q1 2008, the average duration of U.S. earnings calls[6] is approximately 300 sentences and 7000 words where executives of the firms review and discuss a small number of material topics that are market moving in nature. The motivation for this section is to strip away the 'noise' of all those sentences and systematically identify, tag and categorize those material set of topics and their neighboring descriptors. We discuss the details of this process below.

Step 1: Identifying Market Moving Topics

One natural place to find those market moving topics is to identify and tag executives' discussions around the financials of a firm. The chief purpose of an earnings call is for executives to review the latest financials and provide additional visibility into those numbers. For example, a firm's articulation around its revenue and earnings performance are intuitive topics to tag.

Second, how does one come up with a comprehensive set of tags in a systematic way such that material topics are not left out? One solution is to empirically examine the most frequently discussed items in the prepared remarks section of the call transcripts.

Since many of the topics are comprised of two or more contiguous words (e.g., earnings per share), each transcript is initially parsed into eight sets of n-grams ranging from 1-gram (i.e., unigram) to 8-gram after stripping out the stop words[7]. Then the 1000 most frequently

[6] Point-in-time constituents in the Russell 3000 universe
[7] The list from the NLTK library has 170+ stop words. Available upon request.

appearing n-grams from all eight sets are manually vetted where the relevant (irrelevant) n-grams are kept (discarded). For example, see Exhibit 1 for the top 20 most frequently appearing 3-gram (i.e., trigram) from constituents in the S&P 100 investable universe between 2008 and 2017.

The topics that are highlighted in green are deemed as relevant and kept. Empirically, trigrams seem to be the best n-gram to capture the desired market moving topics.[8] Along with trigrams, four- and five-gram are useful in identifying neighboring descriptors.

Exhibit 1: Tri-Grams from the Prepared Remarks Section of S&P 100 January 2008 - December 2017

Description: Green-colored trigrams denote identified and kept topics.

('Trigram', Frequency Count)	
('earnings per share', 7572)	('quarter full year', 1753)
('forward looking statements', 5609)	('versus last year', 1744)
('free cash flow', 4824)	('second half year', 1655)
('good morning everyone', 3524)	('operating cash flow', 1646)
('quarter last year', 2990)	('versus prior year', 1579)
('thank good morning', 2250)	('chief executive officer', 1399)
('double digit growth', 2178)	('cause actual results', 1373)
('chief financial officer', 2088)	('thanks good morning', 1373)
('effective tax rate', 2009)	('actual results differ', 1368)
('constant currency basis', 1800)	('mid single digit', 1231)

Note: Stop words removed. We used the list of 170+ stop words from the NLTK library. Source: S&P Global Market Intelligence Quantamental Research. Data as of 11/01/2019.

Step 2: Creating Market Moving and Neighboring Descriptor Tags

After manually examining the most frequently appearing trigrams, related tags are grouped into a category (e.g., earnings) of tags (e.g., EPS, earnings per share, bottom line). For this paper, there are six categories of topic tags: revenue, operating income, earnings, cash flow, profitability and shareholder return. The finalized tags in the topic categories of revenue, earnings and profitability are shown below. See Appendix A.2 for the other three categories.

Exhibit 2: Revenue and Earnings Topic-Related Tags

Topic Category	Topic Tags
Revenue	{"sales", "revenue", 'top line', 'top bottom line', 'net revenue', 'organic revenue growth', 'organic sales growth', 'operational sales'}
Earnings	{"eps", "earnings", "earnings per share", "net income", "bottom line", "top bottom line"}
Profitability	{"margin", "gross margin", "operating margin", "return invested capital", "return capital"}

Source: S&P Global Market Intelligence Quantamental Research, as of 12/01/2019.

[8] Using unigrams and bigrams is too general and they are not adequate to capture topics that have three contiguous words (e.g. earnings per share). Using n-grams that are greater than five is too specific. They mainly capture legal disclosures (e.g., the most frequently appearing 8-gram is 'risks uncertainties cause actual results differ materially').

Neighboring Descriptor Tags: Directional & Guidance

Having topics is not enough. As we empirically examine the most frequently discussed 4-gram through 8-gram word phrases, many of the flagged topics have neighboring directional descriptors (e.g., increased, decreased). We proceed to create a second broad category of tags – directional tags – as directional indicators for the accompanying topics. For example, see the following sentence from Microsoft's call on October 26, 2017.

"*Our commercial cloud business had another quarter of robust **revenue growth** and …". -* the 113th sentence from Microsoft's call on October 26, 2017.

The word 'revenue' is highlighted in orange denoting a presence of a tag in the revenue topic category. The word 'growth' is highlighted in green denoting a presence of a neighboring tag in the category of directionally positive descriptors.

For simplicity sake, there are only two sets of directional tags for this publication: directionally positive or negative. The directional tags are empirically determined in the same way as the tags for the topics but we manually vet the most frequently appearing 4-gram through 8-gram phrases. See Exhibit 3 for the complete sets. Additionally, we also create one other broad category of descriptors: guidance (Exhibit 3).

Exhibit 3: Direction & Guidance Descriptor Tags

Directional Category	**Directional Descriptor Tags**
Positive Direction	{ "increase", "increased", "increases", "increasing", "increasingly", "expand", "expanded", "expanding", "expands", "expansion", "expansions", "grow", "grows", "grew", "growth", "growths", "improve", "improved", "improves", "improvement", "improvements", "strong", "stronger", "strongest", "strongly" }
Negative Direction	{"decline", "declined", "declines", declining", "deteriorate", "deteriorates", "deteriorated", "deteriorating", "compress", "compressed", "compresses", "compressing", "compressible", "compression", "reduce", "reduces", "reduced", "reducing", "reduction", "reductions", "weak", "weaker", "weakest", "weaken", "weakens", "weakened", "weakening", "weakness", "weaknesses"}

Guidance Category	**Guidance Tags**
Guidance	{"full year outlook", "full year expect", "guidance", "outlook", "forecast", "expect", "expects", "expected", "expecting", "expectation", "expectations"}

Source: S&P Global Market Intelligence Quantamental Research, as of 12/01/2019.

Step 3: Signal Construction - Using the Tags

First, each transcript is processed sentence-by-sentence.[9] In all, 13% of all sentences remain that have at least one instance of a flagged topic.[10]

Second, for each sentence, we create corresponding sentence-level binary flags[11] that denote the presence or the absence of a topic or a descriptor tag. In other words, each sentence becomes a numerical vector of binary flags where each cell represents a category of tags.

Third, using various combinations of the binary flags, we are able to identify systematically at a basic level "the content" of a sentence. For example, Microsoft's executives during sentence number 113 on its October 26, 2017 call discussed the revenue topic and used a neighboring directionally positive descriptor (Exhibit 4).

Exhibit 4: MSFT's October 26, 2017 Call Transcript - Executives' Prepared Remarks & Q&A

Note: The exhibit contains five columns of binary flags, a subset of all available binary flags in the analysis.

Sentence Sequence	**Text** (after some text preprocessing)	**Presence Of a Revenue Topic Tag (orange)**	**Presence of a Profitability Topic Tag (purple)**	**Presence of a Positive Directional Tag (green)**	**Presence of a Negative Directional Tag (red)**	**Presence of a Guidance Tag (blue)**
1 (Call Start)	thanks, jessie.	0	0	0	0	0
2	good afternoon, and thank you for joining us today.	0	0	0	0	0
…	…	…	…	…	…	…
113	our commercial cloud business had another quarter of robust **revenue** **growth** and …	1	0	1	0	0
…	…	…	…	…	…	…
137	gross **margin** percentage **declined** due to the impact of linkedin-related amortization.	0	1	0	1	0
…	…	…	…	…	…	…
193	we **expect** continued year-over-year **margin** **improvement** and sequential trends consistent with prior years.	0	1	1	0	1
…	…	…	…	…	…	…
385	thank you so much.	0	0	0	0	0
386 (Call End)	thank you all.	0	0	0	0	0

Source: S&P Global Market Intelligence Quantamental Research, as of 12/01/2019.

[9] Relative to a paragraph or a speaker block[8], the sentence-level processing has a much better probability of isolating the topic in question.
[10] Most of the sentences about 87% do not contain one or more of our defined topics.
[11] A binary variable takes two values: zero or one. One (zero) is indicative of presence (absence).

The set of signals in this section is binned into the following three sub-groups:

- Standalone Topic + Directional Tags
- Composite Topics + Directional Tags
- Standalone Topic + Directional Tags + Guidance Tags

Given a market moving topic (e.g., revenue), we sum up the number of sentences that contains at least one flagged instance of the topic and one flagged instance of a directionally positive (negative) tag in a call. We net out the two sums to create a call level net directionally positive sum (e.g., similar to the Net Positivity Score[12] from the NLP II publication) for the topic.[13] Lastly, we scale the topic's directionally net positive score by the total number of sentences in the call. See Exhibit 5 for the actual construction, the signals sort order and the intuition.

Exhibit 5: Signal Construction and Intuition - Textual Tagging & Categorization

Signal Groups	Construction	Sort Order	Intuition
Standalone Topic + Directional Tags (Go to Results)	Signal = A / B • A = # of sentences with net directionally positive instances given the presence of a topic • B = number of sentences in a call	D	• Firms with the highest net directionally positive score surrounding a market moving topic (e.g., earnings) are viewed as financially growing the most and are expected to outperform. • The number of sentences with net directionally positive references is viewed as a proxy for the magnitude of that growth and the executives' confidence level around it. • A sentence containing a guidance tag indicates that a firm's financial outlook is being discussed.
Composite Topics + Directional Tags (Go to Results)	Signal = A / B • A = # of sentences with net directionally positive instances given the presence of composite topics (i.e., 2+) • B = number of sentences in a call	D	
Standalone Topic + Directional Tags + Guidance Tags (Go to Results)	Signal = A / B • A = # of sentences with net directionally positive instances given the presence of a topic and the presence of a guidance tag • B = number of sentences in a call	D	

Note: D = descending sort; A = ascending sort
Source: S&P Global Market Intelligence Quantamental Research, as of 11/01/2019.

[12] Net Positivity Score = (# of Positive Sentiment Words - # of Negative Sentiment Words) / # of Total Words in a call using the Loughran & McDonald (2011) sentiment dictionary. The score and its components are readily available in our Textual Data Analytics (TDA) product.
[13] A directionally positive word that has a negation word within a neighborhood of three tokens is not included as a positive tally, but is included as a negative count.

2.2 Weighted Average Sentiment

In our previous publication (NLP II), we constructed four simple averages as proxies to gauge call sentiment. Each sentiment word in a call as defined by the L&M dictionary[14] has the same importance (i.e., equal weights). **In this section, we introduce the construction of weighted average sentiment using two different weighting schemes: term frequency-inverse document frequency (TF-IDF) and call similarity.**

2.2.1 Weighting Scheme One: Term Frequency-Inverse Document Frequency (TF-IDF)

Introduction, Motivation & Intuition

TF-IDF is a concept from information retrieval (e.g., internet searches). There are two main inputs into the metric: I) term frequency (TF) II) inverse document frequency (IDF). Each unique non-stop word in a call receives a TF-IDF score. In turn, the call level TF-IDF score is aggregated from the word level scores.

TF measures how often a word appears in a call. IDF measures whether a word appears scarcely or ubiquitously in a corpus, which in our case is a rolling window of five earnings calls for a given firm. The IDF term is the weighting scheme. A word that appears scarcely (ubiquitously) across a corpus of earnings calls gets a higher (lower) weight. **The intuition is that the scarce (ubiquitous) usage of a term in a corpus is deemed to have more (less) informational content.** See the equation below for the specific calculation that is used throughout this publication. Exhibit 6 has a TF-IDF example with real scores for visualization and for facilitating a more intuitive understanding.

TF-IDF Equation – Details & Intuition

Term	Calculation	Intuition
TF	TF = A / B • A = Frequency of a word appearing in an earnings call • B = Number of unique non-stop words in a corpus, which is a rolling window of five calls for a firm	• The numerator is how frequently a word appears in a call and the denominator is the number of unique non-stop words in a corpus such that the TF term and the IDF term are both consistently using a particular corpus as a reference point. • Since we normalize the numerical representation of a call to a unit vector, the call level TF-IDF score is independent of whether the denominator uses the number of unique words in a call or in a corpus as long as the selection is consistently being applied.
IDF	IDF = natural log((C+1) / (D+1)) + 1 • C = Number of calls in a corpus • D = Number of calls that a word appears in a corpus	• The weighting scheme • Places a bigger weight on a sentiment word that appears scarcely in a corpus
Normalizing	Euclidean Norm	• Normalize for calls of various length

[14] Loughran & McDonald (2011)

Exhibit 6: TF-IDF Example

Description: i) Five contiguous earnings calls where each call has four words. ii) Three of which is the same across the corpus and the fourth word highlighted in green is unique.

Earnings Calls for Stock i	Text
Call(t-4): Four Calls Ago	"W1 W2 W3 W8"
Call(t-3)	"W1 W2 W3 W7"
Call(t-2)	"W1 W2 W3 W6"
Call(t-1)	"W1 W2 W3 W5"
Call(t): Most Recent Call	"W1 W2 W3 W4"

			[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]		[9]
	Word Number		**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**		
	Sentiment		**Neutral**	**Neutral**	**Neutral**	**Positive**	**Positive**	**Positive**	**Positive**	**Positive**		
			TF-IDF (W1)	**TF-IDF (W2)**	**TF-IDF (W3)**	**TF-IDF (W4)**	**TF-IDF (W5)**	**TF-IDF (W6)**	**TF-IDF (W7)**	**TF-IDF (W8)**		**Call Level TF-IDF Weighted Avg Sentiment**
[1]	Call(t-4)		0.368	0.368	0.368	0.000	0.000	0.000	0.000	0.771		0.771
[2]	Call(t-3)		0.368	0.368	0.368	0.000	0.000	0.000	0.771	0.000		0.771
[3]	Call(t-2)		0.368	0.368	0.368	0.000	0.000	0.771	0.000	0.000		0.771
[4]	Call(t-1)		0.368	0.368	0.368	0.000	0.771	0.000	0.000	0.000		0.771
[5]	Call(t)		0.368	0.368	0.368	0.771	0.000	0.000	0.000	0.000		0.771

0.771 = TF-IDF Word W4 Score for Call(t) = TF * IDF / Normalizing = 0.125 * 2.0986 / 0.3401	
• TF	= (Frequency of word W4 in Call(t)) / (Number of Unique Words in the Corpus) = 1/8 = 0.125
• IDF	= natural log((# of calls in the corpus + 1) / (# of calls word w appears +1)) + 1 = natural log(6/2) + 1 = 2.0986
• Normalizing	= Euclidean Norm = √(Sum(X^2)) where X = TF-IDF(W1, W2, …, W8) for Call(t) = 0.3401

Source: S&P Global Market Intelligence Quantamental Research, as of 12/01/2019.

Signal Construction

For every firm, the corpus is a rolling window of its five most recent contiguous calls. Every non-stop word in a call has a TF-IDF word score. Then the scores of the words that are positive or negative in sentiment according to the L&M dictionary are summed to create two call level TF-IDF positive and negative scores (see column [9] in Exhibit 6 for the overall TF-IDF positive sum score for each of the five calls). The two sums are subsequently used as inputs to construct the various flavors of the weighted average sentiments. In Exhibit 7, we show two such examples using this weighting scheme.

Exhibit 7: Signal Construction and Intuition - TF-IDF Weighted Average Sentiment

Signal	Construction	Sort Order	Intuition
TF-IDF Weighted Average Sentiment (Go to Results)	Weighted Average = A/B or A - B • A = Call Level TF-IDF Positive Weighted Average Sentiment • B = Call Level TF-IDF Negative Weighted Average Sentiment where examples of sentiment are Positivity-to-Negativity or Net Positivity scores	D	• Takes into account the IDF weights for positive and negative words
TF-IDF Weighted Average Sentiment Per Word (Go to Results)	Weighted Average = A/B or A - B • A = TF-IDF Weight Average Sentiment Per Positive Word • B = TF-IDF Weighted Average Sentiment Per Negative Word where examples of sentiment are Positivity-to-Negativity or Net Positivity scores	D	• Takes into account the IDF weights for positive and negative words. The division by the number of positive and negative words is to assign a higher signal score to a call with the highest TF-IDF sum score yet achieved with the fewest infrequently appearing words in a corpus.

Note: D = descending sort; A = ascending sort
Source: S&P Global Market Intelligence Quantamental Research, as of 12/01/2019.

2.2.2 Weighting Scheme Two – Call Similarities

Introduction & Intuition

The second weighting scheme uses call similarity scores. Earnings calls are one of the main ways that company executives provide visibility and transparency into operational results. **When one or more customary norms in language usage change during an earnings call, the dissimilarity is (highly) noticeable and in our narrative serves as a flagged instance of oddity or discontinuity.[15]**

A good case in point is Tesla Inc.'s earnings call on May 2, 2018. Elon Musk, the founder and CEO of the firm, refused to answer cash flow and Model 3 related questions from the sell-side analysts and then abruptly took questions for the next 23 minutes (31% of the 74-minute call)

[15] Our thesis is that even when there is a change of personnel, the new leadership and writers would look to the language and the format of prior calls for guidance and emulation.

from a blogger with a YouTube channel.[16] The opaqueness, the dismissiveness and the defensiveness of the CEO exacerbated the fear of analysts and investors alike that the firm may not be able to ramp up Model 3 production as promised. In turn, the cash burn will only intensify going forward with no end in sight. According to the call similarity measure, this call has been flagged as one of the most dissimilar calls in terms of language since Tesla Inc. went public in June 2010.

In this section, we introduce the concept of call similarity as a proxy to identify an instance of oddity or discontinuity in an earnings call via language usage. **In our narrative, when a firm changes its language as measured by the call similarity scores (e.g., the presence and the frequency of wording and phrasing), this set of actions may contain rich and material information for the future financial outcomes of firms.**

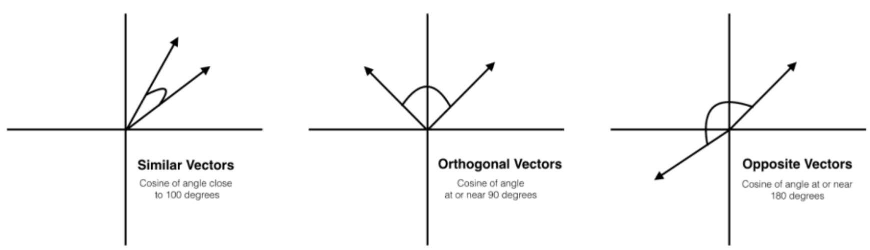
Measuring Call Similarity

How does one measure language similarity between two calls? There are a number of measures[17]. The one that is used throughout this paper is called cosine similarity. Specifically, the words within each call are represented as a numerical vector of word level TF-IDF scores (see above Section 2.2.1).

A similarity score between two calls is quantified by the geometric angle between two numerical vectors in an n-dimensional Euclidean space where the n is the number of unique words between the two calls. One of the advantages of a non-normalized cosine similarity measure over a non-normalized Euclidean distance measure is that the former does not get overly influenced by the length of calls.

The cosine similarity score is bounded between -1 and 1 where the first value denotes a pair of calls that is diametrically opposite and the latter value denotes a pair of calls that is exactly the same. Empirically, cosine similarity for a pair of calls, however, is bounded between 0 and 1.[18] See Exhibit 8.

Exhibit 8: Cosine Similarity



Source: S&P Global Market Intelligence Quantamental Research, as of 11/01/2019.

[16] https://money.cnn.com/2018/05/03/investing/elon-musk-tesla-earnings-call/index.html
[17] Others are Jaccard similarity, minimum edit distance and simple similarity.
[18] Zero is the lower bound because all the weights are strictly non-negative.

Signal Construction

In our narrative, the dissimilarity of a firm's call relative to its prior ones signifies a flagged instance of oddity or discontinuity for the firm. The oddity, however, does not speak to whether a firm's outlook is going to improve, deteriorate or stay neutral. **We use the call similarity scores as weights to amplify or deflate changes in sentiment scores of firms. A low (high) similarity score results in a higher (lower) weight** (Exhibit 9).

Exhibit 9: Signal Construction & Intuition - Call Similarity Weighted Average Sentiment

Signal	**Construction**	**Sort Order**	**Intuition**
YoY Change in Simple Sentiment Average * Call Similarity (t, t-4) (Go to Results)	Score = A / B • A = YoY change in simple sentiment average between call(t) and call(t-4) • B = Call similarity score between call(t) and call(t-4) where t = {earnings calls for firm i} where examples of sentiment are Positivity-to-Negativity or Net Positivity	D	Go long firms that have had the highest YoY change in sentiment and amplify that with the language similarity score between the two calls.
Demeaned Simple Sentiment Average * Average Call Similarity Scores (Go to Results)	Score = A / B • A = Simple sentiment average demeaned relative to the previous four calls (%) • B = Average call similarity scores between call(t) and call(t-1), call(t-2), …, call(t-4) where examples of sentiment are Positivity-to-Negativity or Net Positivity	D	Same as above except the demeaned calculation takes into account the previous four calls and is used as a proxy to measure call sentiment trend of a firm.

Note: D = descending sort; A = ascending sort

Source: S&P Global Market Intelligence Quantamental Research, as of 11/01/2019.

2.3 Gauging Call Level Transparency via Executives' Behaviors & Decisions:

In this section, we introduce four additional, differentiated signals[19] to measure the level of transparency that firms provide on earnings calls:

I. The Abundance or Scarcity of References to Guidance
II. Language Similarity between Calls
III. The Decision of Timing via How Early Do Firms Introduce Numbers
IV. Self-Attribution Bias via the Blame Game.

[19] See the previous publication NLP II for the prior set of signals in the call level transparency category.

2.3.1 The Abundance or Scarcity of References to Guidance

In Section 2.1, we introduced the guidance tags as a category of neighboring descriptors to complement the various categories of market moving topics. The discussion of guidance by definition provides a higher level of transparency as sell- and buy-side analysts are able to use the firm provided estimates as inputs in their modeling.

The discussion of guidance during calls could also be viewed as a (positive) signaling effect that is analogous to the act of declaring or increasing dividends or buybacks. Our hypothesis is that executives are more willing to discuss guidance when they have a higher degree of confidence and visibility into the future performance of the financials. **In our narrative, firms whose executives reference the most instances of guidance neighboring the six categories of market moving topics from Section 2.1 should outperform historically.**

Signal Construction

For each call, the number of sentences where one or more of the guidance tags is referenced is scaled by the total number of sentences in a call (Exhibit 10).

Exhibit 10: Number of References to Guidance - Signal Construction & Intuition

Signal	Construction	Sort Order	Intuition
Number of References to Guidance in a Call (Go to Results)	Sum the number of sentences containing one or more tags from the guidance category and one or more tags from the six categories of the market-moving topics from Section 2.1. Then the sum is scaled by the number of sentences in a call.	D	Our hypothesis is that executives are more willing to discuss guidance when they have a higher degree of confidence and visibility into the future performance of the financials.

Note: D = descending sort; A = ascending sort

Source: S&P Global Market Intelligence Quantamental Research, as of 02/01/2018.

2.3.2 Language Similarity between Calls

Introduction & Intuition

In Section 2.2.2, we introduced the example of how Elon Musk, the founder and CEO of Tesla Inc., during its May 2, 2018 earnings call refused to answer cash flow and Model 3 related questions from sell-side analysts. Then he abruptly took a dozen plus questions for 23 minutes from a blogger with a YouTube channel.[20] The dismissiveness and the defensiveness exacerbated the fear of analysts and investors alike that the firm may not be able to ramp up Model 3 production as projected. In turn, the cash burn would only intensify going forward. The stock proceeded to fall by -5.6% by the end of the call despite the firm reporting in-line numbers.

[20] See footnote 16.

In this section, we are proposing a second usage of call similarity scores. Our narrative here is that when call similarity scores are examined standalone[21], a low call similarity score (i.e., dissimilarity), which signifies a flagged instance of oddity or discontinuity, is viewed unfavorably.

Our hypothesis is that the path of least resistance is for firms to keep the language and the content of earnings calls similar especially within the prepared remarks section. Even when there is a change of personnel, the new leadership and writers would look to the language and the format of prior calls for guidance and emulation. The largest and the most drastic changes in the language and the content come about when firms are compelled to fulfill their fiduciary or legal responsibilities. Our hypothesis is that firms that have the most dissimilar (similar) calls relative to their historical ones should underperform (outperform).

Signal Construction

The signal is a simple average of four pairwise call similarity scores from the five most recent calls of a firm. Each of the four scores uses the most recent call as the reference point.

Exhibit 11: Language Similarity between Calls - Signal Construction & Intuition

Signal	Construction	Sort Order	Intuition
Average Call Similarity Score of the Most Recent Five Calls (Go to Results)	Simple Average of {A, B, C, D} • A = Similarity (Call(t), Call(t-4)) • B = Similarity (Call(t), Call(t-3)) • C = Similarity (Call(t), Call(t-2)) • D = Similarity (Call(t), Call(t-1)) where pairwise call similarity score = [0, 1]	D	Our hypothesis is that the language and the content tend to have the largest changes when firms are compelled to fulfill their fiduciary or legal responsibilities.

Note: D = descending sort; A = ascending sort
Source: S&P Global Market Intelligence Quantamental Research, as of 11/01/2019.

2.3.3 The Decision of Timing – How Early Do Firms Introduce Numbers?

Introduction, Motivation & Intuition

In our NLP II paper, we introduced the Numerical Transparency Score, which calculates the proportion of numbers relative to words in a call. An abundant usage of numbers is viewed favorably in our narrative and suggests the firm is exuding objectivity, transparency and confidence about their future prospects.

This score, however, does not speak to the element of timing: how early or late do the executives reference numbers in a call. **Our hypothesis is that firms delaying the introduction of numbers (relative to words) in an earnings call exhibit a lower level of**

[21] When call similarity scores are used as weights to amplify or deflate sentiment changes

transparency where the delay is viewed as a "stalling tactic". Our hypothesis is that these firms should underperform historically.

Although we only explore one signal with respect to 'the decision of timing', the idea could be extended to other facets of an earnings call (e.g., the order that bullish/bearish sell-side analysts are selected by firms to ask questions where the disproportionate selection of bullish analysts is viewed unfavorably).

Signal Construction

An earnings call is divided into 10 parts, each containing an equal number of tokens (i.e., numbers and words). The Numerical Transparency Score is calculated for each section where the earlier sections receive larger weights. Specifically, we use a linear decay where the first section receives a weight of 1.0 and then the weights for the subsequent nine sections decrement by one-tenth with the last section receiving a weight of 0.1. The overall signal is a weighted average of the 10 Numerical Transparency Scores (Exhibit 12).

Exhibit 12: The Decision of Timing via How Early Do Firms Reference Numbers – Signal Construction & Intuition

Signal	Construction	Sort Order	Intuition
The Decision of Timing – How Early Do Firms Reference Numbers (Go to Results)	• Divide a call into 10 sections by word and number token counts • Calculate the Numerical Transparency Score in each section • Overall score = Sum(w(i) * Score(i)) where i denotes the i^{th} section and w(i) = {1, 0.9, …, 0.1}	D	Our hypothesis is that firms delaying the introduction of numbers (relative to words) in an earnings call exhibit a lower level of transparency where the delay is viewed as a "stalling tactic".

Note: D = descending sort; A = ascending sort

Source: S&P Global Market Intelligence Quantamental Research, as of 11/01/2019.

2.3.4 Self-Attribution Bias - The Blame Game

Introduction & Intuition

When firms miss their expected numbers or have softness in their guidance, executives may attribute the weakness to exogenous elements that are outside of their control such as the economy or the industry. **Our hypothesis is that firms that most frequently blame the economy, the market or the industry during earnings calls should underperform historically.** Exhibit 13 has the set of identified tags that denotes a flagged instance when executives are discussing such topics.

Exhibit 13: Economy, Industry and Market-Related Tags

Exogenous Elements	Economy, Industry and Market-Related Tags
The economy, the industry and the market	{"industry", "industries", "industrywide", "industry-wide", "sector", "sectors", "market", "markets", "economy", "economies"}

Source: S&P Global Market Intelligence Quantamental Research, as of 02/01/2018.

Signal Construction

For every call, we identify sentences that contain at least one of the tags from Exhibit 13. Then we look for neighboring negative sentiment words according to the L&M dictionary in the sentences. We sum the number of the flagged sentences and then scale the sum by the total number of sentences in a call. See Exhibit 14.

Exhibit 14: Self-Attribution Bias via the Blame Game - Signal Construction & Intuition

Signal	Construction	Sort Order	Intuition
Self-Attribution Bias via The Blame Game (Go to Results)	• Identify the number of sentences in a call where one or more of the economy-related tags are referenced. • Sum up the number of sentences from above that have neighboring words that are negative in sentiment according to the L&M dictionary. • Scale the sum by the total number of sentences in a call.	A	Executives attribute the financial softness to exogenous elements like the economy, the market or the industry.

Note: D = descending sort; A = ascending sort.
Source: S&P Global Market Intelligence Quantamental Research, as of 11/01/2019.

3. Empirical Results: Standalone

Earnings calls are events[22] that occur on different dates. In order to achieve breadth in the cross-section, we apply a look-back window of four calendar months to construct each of the monthly rebalancing strategies. For example, at the end of April when we are rebalancing and constructing our portfolio, we look at earnings calls that have taken place since January of the same year and extract the information that is from a firm's most recent earnings call to construct the stock-level signal score. At the end of May, we construct our portfolio by looking back as far as February.

All long-only and long-short returns in our back-test are equal-weighted, are rebalanced monthly at month end and are binned into quintiles where the top (bottom) quintile or the long (short) portfolio contains the 20% of stocks with the highest (lowest) signal scores in the cross-

[22] Generally, event-driven signals lack a sufficient number of stocks for portfolio construction and the timing of the next event is unknown.

section.[23] All signals are industry-neutral (GICS Level 3) to mitigate industry tilts.[24] The definitions of columns in the exhibits containing back-test results are in Appendix A.3.

3.1 Empirical Results: Textual Tagging & Categorization

Results: Market Moving Topics & Directional Descriptors

Historically, firms whose executives referenced the most instances of i) revenue- ii) earnings- or iii) profitability-related topics in a directionally positive light outperformed their most extreme counterparts by 9.16%, 8.60% and 6.76% per year, respectively, with statistical significance at the 1% level since 2010 (see rows 1, 3 and 5 in Table 1).

Both the predictive power and the economic significance are especially strong for the two signals surrounding the tags in the revenue- and earnings-related topic categories. **In fact, the performance is among the highest of the signals that we have introduced so far including the prior sets from our NLP II publication.**

The second noteworthy observation is that the long-sides of these strategies contribute meaningfully and significantly (i.e., >50%) to the overall long-short strategies (compare the values in column 7 to the values in column 10 in Table 1). This is highly desirable since there are impediments against shorting stocks.

Table 1: Standalone Topic + Directional Tags
Russell 3000 May 2010 – December 2017

(Go to Signal Definitions - Exhibit 1) (Go to the Cover Page)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
		Signal Sort Order	Strategy Start Date	Average Firm Count in Each Quintile Bin	Average Monthly Spearman Correlation	Hit Rate Monthly Spearman Correlation	Annualized Average Monthly Long - Market	Annualized Monthly IR (Long - Market)	Hit Rate Average Monthly Long - Market	Annualized Average Monthly Long - Short	Annualized Monthly IR (Long - Short)	Hit Rate Average Monthly Long - Short
[1]	Revenue + Directionally Positive	D	201005	custom sort	0.030	66.3%	4.24%	1.27	67.4%	9.16%	1.09	65.2%
[2]	p-value	NaN	NaN	NaN	0.000	0.001	0.001	NaN	0.001	0.003	NaN	0.002
[3]	Earnings + Directionally Positive	D	201005	custom sort	0.036	71.7%	5.05%	1.15	60.9%	8.60%	1.23	64.1%
[4]	p-value	NaN	NaN	NaN	0.000	0.000	0.002	NaN	0.028	0.001	NaN	0.005
[5]	Profitability + Directionally Positive	D	201005	custom sort	0.027	68.5%	3.57%	1.18	64.1%	6.76%	1.34	67.4%
[6]	p-value	NaN	NaN	NaN	0.000	0.000	0.002	NaN	0.005	0.000	NaN	0.001

[23] The number of stocks in the quintile portfolios for signals in the textual tagging and categorization group is not even since there are a number of ties between the stock-level signal ranks. The bottom quintile has signal scores of zero or less. Actual portfolio counts for each of the signals are available upon request. This is denoted by the phrase 'custom sort'.

[24] The results for signals in the textual tagging and categorization group are not industry-neutral due to the fact that signal values have ties. The industry-neutral results for select signals are available upon request (e.g., revenue topic).

Note: Values that are shaded in green (red) are statistically significant at least at the 10% level and are consistent (inconsistent) with our ex-ante hypothesis. D = descending sort and A = ascending sort. Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 10/01/2019.

Results: Composite Market Moving Topics & Directional Descriptors

Historically, firms whose executives referenced the most instances of i) revenue- ii) earnings- or iii) profitability-related topics in combination with a directionally positive light outperformed the market and their most extreme counterparts by at least 4.49% and 11.30% per annum, respectively, at the 1% statistical significance level (see rows 1 and 3 in Table 2).[25]

Table 2: Composite Topics + Directional Tags

Russell 3000 May 2010 – December 2017

(Go to Signal Definitions - Exhibit 1)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
		Signal Sort Order	Strategy Start Date	Average Firm Count in Each Quintile	Average Monthly Spearman Correlation	Hit Rate Monthly Spearman Correlation	Annualized Average Monthly Long - Market	Annualized Monthly IR (Long - Market)	Hit Rate Average Monthly Long - Market	Annualized Average Monthly Long - Short	Annualized Monthly IR (Long - Short)	Hit Rate Average Monthly Long - Short
[1]	Revenue & Earnings + Directionally Positive	D	201005	494	0.037	72.8%	4.67%	1.38	67.4%	11.30%	1.07	66.3%
[2]	p-value	NaN	NaN	NaN	0.000	0.000	0.000	NaN	0.001	0.004	NaN	0.001
[3]	Revenue & Earnings & Profitability + Directionally Positive	D	201005	494	0.038	73.9%	4.49%	1.35	71.7%	11.93%	1.09	69.6%
[4]	p-value	NaN	NaN	NaN	0.000	0.000	0.000	NaN	0.000	0.003	NaN	0.000

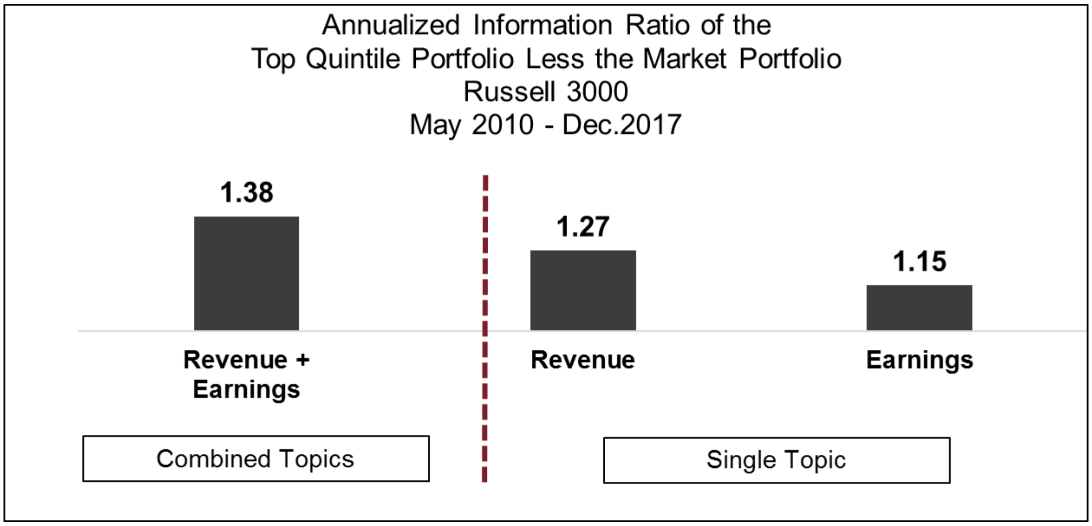
Note: Values that are shaded in green (red) are statistically significant at least at the 10% level and are consistent (inconsistent) with our ex-ante hypothesis. D = descending sort and A = ascending sort. Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 02/01/2018.

The second noteworthy observation is that the annualized information ratios from the active long side of these signals (i.e., with composite topics) are higher than the information ratios of the signals with a standalone topic. One plausible explanation is that the combined topics signify the reinforcing confirmation of the strength in the underlying business at the firms (See Exhibit 15).

[25] For empirical analysis, U.S. earnings call transcripts data starts in Q1 2008. In the body of the paper, we show results starting in May 2010 such that results for sentiment level, change in sentiment level and change in sentiment trend all consistently have the same sample period.

Exhibit 15: Annualized Information Ratio

Combined Topics versus Standalone Topics



Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 10/01/2019.

Thirdly, the performance of the signals with composite topics (i.e., 2+) are especially strong and the predictive power extends robustly to the larger caps using the Russell 1000 as a proxy (Table 3).

Table 3: Composite Topics + Directional Tags in the Large-Caps

Russell 1000 May 2010 – December 2017

(Go to Signal Definitions - Exhibit 1)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
		Signal Sort Order	Strategy Start Date	Average Firm Count in Each Quintile Bin	Average Monthly Spearman Correlation	Hit Rate Monthly Spearman Correlation	Annualized Average Monthly Long - Market	Annualized Monthly IR (Long - Market)	Hit Rate Average Monthly Long - Market	Annualized Average Monthly Long - Short	Annualized Monthly IR (Long - Short)	Hit Rate Average Monthly Long - Short
[1]	Revenue & Earnings + Directionally Positive	D	201005	188	0.033	60.9%	2.65%	0.81	65.2%	7.11%	0.64	63.0%
[2]	p-value	NaN	NaN	NaN	0.005	0.028	0.028	NaN	0.002	0.079	NaN	0.009
[3]	Revenue & Earnings & Profitability + Directionally Positive	D	201005	188	0.034	64.1%	2.32%	0.76	65.2%	7.41%	0.67	65.2%
[4]	p-value	NaN	NaN	NaN	0.002	0.005	0.038	NaN	0.002	0.068	NaN	0.002

Note: Values that are shaded in green (red) are statistically significant at least at the 10% level and are consistent (inconsistent) with our ex-ante hypothesis. D = descending sort and A = ascending sort. Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 10/01/2019.

Additive to Our Published Results from NLP II

The signals in this category are also additive to our published results in the NLP II paper, specifically after controlling for the Net Positivity Score and the Numerical Transparency Score, which are two of the strongest predictive signals. In Table 4, the annualized economic performance in excess of those two signals ranged from 4.59% to 6.86% per year with statistical significance at the 1% level.

Table 4: Additive Forecast Power above Net Positivity Score
& Numerical Transparency Score from the NLP II Publication
Russell 3000 May 2010 – Dec. 2017

(Go to the Cover Page)

Signal Name	Standalone Long-Short Return Spread	t-statistic	Adjusted Long-Short Return Spread	t-statistic
Revenue Topic + Directionally Positive	9.16%	3.02	6.86%	2.76
Net Income Topic + Directionally Positive	8.60%	3.41	6.24%	2.87
Profitability + Directionally Positive	6.76%	3.71	4.59%	3.04

Note: Values that are shaded in green (red) are statistically significant at least at the 10% level and are consistent (inconsistent) with our ex-ante hypothesis. Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 10/01/2019.

3.2 Empirical Results: Weighted Average Sentiment

3.2.1 Results: TF-IDF Weighting Scheme

Historically, firms with the most positive sentiment[26], firms with the most positive YoY change in sentiment and firms with the most positive sentiment trend using the TF-IDF weighting scheme outperformed the market by at least 2.76% per year and outperformed their most extreme counterparts [27] by at least 4.62% per year with significance at the 1% level (Table 5). The long-sides of these sentiment-based strategies contributed meaningfully and significantly (i.e., >50%) to the overall long-short strategies (compare the values in column 7 to those in column 10 in Table 5), which is desirable since there are impediments against shorting stocks.

[26] Positivity-to-negative signal construction is used as a proxy to measure sentiment.
[27] A portfolio of shorts is the 20% of stocks in a quintile return spread analysis with the lowest signal values in the cross-section that an investor sells and is forecasted to underperform the most.

Table 5: Weighted Average Sentiment Signals Using TF-IDF Weight Scheme
where Sentiment is Measured using Positivity-to-Negativity
Russell 3000 May 2010 – December 2017
(Go to Signal Definitions - Exhibit 8) (Go to the Cover Page)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
		Signal Sort Order	Strategy Start Date	Average Firm Count in Each Quintile Bin	Average Monthly Spearman Correlation	Hit Rate Monthly Spearman Correlation	Annualized Average Monthly Long - Market	Annualized Monthly IR (Long - Market)	Hit Rate Average Monthly Long - Market	Annualized Average Monthly Long - Short	Annualized Monthly IR (Long - Short)	Hit Rate Average Monthly Long - Short
[1]	Positive to Negative TF-IDF Sum	D	201005	458	0.020	76.1%	2.76%	0.94	67.4%	5.18%	1.30	68.5%
[2]	p-value	NaN	NaN	NaN	0.000	0.000	0.011	NaN	0.001	0.001	NaN	0.000
[3]	Positive to Negative TF-IDF Sum YoY Change	D	201005	412	0.013	68.5%	3.14%	1.48	67.4%	4.62%	1.60	71.7%
[4]	p-value	NaN	NaN	NaN	0.000	0.000	0.000	NaN	0.001	0.000	NaN	0.000
[5]	Positive to Negative TF-IDF Sum Demeaned PreviousFour Calls	D	201005	423	0.014	69.6%	3.08%	1.45	65.2%	4.62%	1.63	71.7%
[6]	p-value	NaN	NaN	NaN	0.000	0.000	0.000	NaN	0.002	0.000	NaN	0.000

Note: Values that are shaded in green (red) are statistically significant at least at the 10% level and are consistent (inconsistent) with our ex-ante hypothesis. D = descending sort and A = ascending sort. Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 02/01/2018.

Added Benefit of TF-IDF Weighting versus Simple Weighting

Historically, relative to the simple average sentiment, the TF-IDF weighted average showed improvement on both the return and the risk side. **The annualized information ratio on the long-short quintile return spread, a proxy for economic performance, improved by 52% to 1.81 from 1.19. On the return side, the long-short quintile return spread improved by 64 basis points (bps) to 6.45% from 5.81%.** On the risk side, the maximum drawdown from monthly rebalancing improved to -3.30% from -6.14% (Table 6).

Table 6: TF-IDF Weighting Scheme versus Simple Average Weighting Scheme
Long-Short Quintile Return Spread for the Positivity-to-Negative Sentiment Signal
Russell 3000 May 2010 – Dec. 2017
(Go to the Cover Page)

	Annualized Information Ratio	Annualized Return	Annualized Volatility	Max Drawdown
Weighted Average (TF-IDF)	**1.81**	**6.45%**	**3.56%**	**-3.30%**
Simple Average	1.19	5.81%	4.47%	-6.14%

Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 10/01/2019.

3.2.2 Results: Call Similarity Weighting Scheme

Historically, sentiment-based signals using call similarity scores as the weighting scheme exhibited forecasting power in the U.S. market and had higher information ratios relative to the ones using simple sentiment averages. Firms with the most positive sentiment improvement and the largest change in language usage during calls outperformed the market and their most extreme counterparts by at least 1.72% and 3.69% per year, respectively, at the 1% statistical significance level. In Table 7, we show two such examples of sentiment improvement: change in YoY sentiment (see row [1]) and change in sentiment trend (see row [3]).

Table 7: Weighted Average Sentiment Change using Call Similarity Weighting Scheme where the Sentiment is Measured using Positivity-to-Negativity
Russell 3000 May 2010 – December 2017

(Go to Signal Definitions - Exhibit 10)

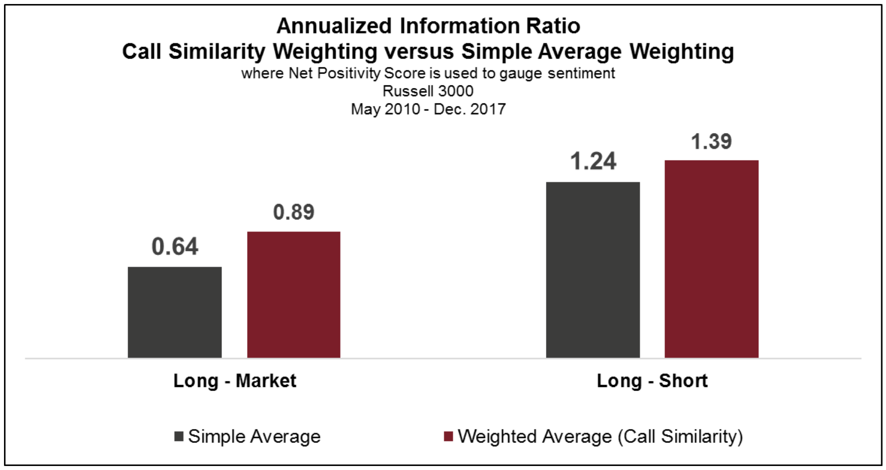
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
		Signal Sort Order	Strategy Start Date	Average Firm Count in Each Quintile Bin	Average Monthly Spearman Correlation	Hit Rate Monthly Spearman Correlation	Annualized Average Monthly Long - Market	Annualized Monthly IR (Long - Market)	Hit Rate Average Monthly Long - Market	Annualized Average Monthly Long - Short	Annualized Monthly IR (Long - Short)	Hit Rate Average Monthly Long - Short
[1]	YoY' Sentiment Improvement + Largest Language Change	D	201005	427	0.014	67.4%	1.72%	0.95	62.0%	3.69%	1.42	69.6%
[2]	p-value	NaN	NaN	NaN	0.000	0.001	0.010	NaN	0.016	0.000	NaN	0.000
[3]	Sentiment Improvement + Largest Language Change Relative to the Previous 4 Calls	D	201005	378	0.012	64.1%	2.28%	1.05	56.5%	3.71%	1.20	64.1%
[4]	p-value	NaN	NaN	NaN	0.000	0.005	0.004	NaN	0.175	0.001	NaN	0.005

Note: Values that are shaded in green (red) are statistically significant at least at the 10% level and are consistent (inconsistent) with our ex-ante hypothesis. D = descending sort and A = ascending sort. Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 02/01/2018.

Call Similarity Weighting versus Simple Weighting – Improvement in Risk-Reward Tradeoff

Sentiment-based signals using call similarity scores as their weighting scheme historically improved the information ratio of those strategies over simple average constructs. The annualized information ratio on the market-adjusted portfolio using the Net Positivity sentiment score as a proxy improved by 39% to 0.89 from 0.64. The annualized information ratio on the long-short quintile return spread improved by 12% to 1.39 from 1.24 (Exhibit 16).

Exhibit 16: Information Ratio of Sentiment Change Signals using Call Similarity Weighting Scheme versus Simple Average Weighting Scheme



Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 10/01/2019.

3.3 Results: Gauging Call Level Transparency via Executives' Behaviors and Decisions:

3.3.1 Transparency via Abundance or Scarcity of References to Guidance

Historically, firms whose executives referenced the most instances of guidance outperformed their most extreme counterparts by 5.84% per year at the 1% statistical significance level (Table 8).

Table 8: Number of References to Guidance
Russell 3000 May 2010 – December 2017

(Go to Signal Definitions – Exhibit 11) (Go to the Cover Page)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
		Signal Sort Order	Strategy Start Date	Average Firm Count in Each Quintile Bin	Average Monthly Spearman Correlation	Hit Rate Monthly Spearman Correlation	Annualized Average Monthly Long - Market	Annualized Monthly IR (Long - Market)	Hit Rate Average Monthly Long - Market	Annualized Average Monthly Long - Short	Annualized Monthly IR (Long - Short)	Hit Rate Average Monthly Long - Short
[1]	Number of References to Guidance	D	201005	custom sort	0.021	63.0%	3.08%	1.06	65.2%	5.84%	1.13	63.0%
[2]	p-value	NaN	NaN	NaN	0.001	0.009	0.004	NaN	0.002	0.002	NaN	0.009

Note: Values that are shaded in green (red) are statistically significant at least at the 10% level and are consistent (inconsistent) with our ex-ante hypothesis. D = descending sort and A = ascending sort. Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 10/01/2019.

The second noteworthy takeaway is that this signal's predictive efficacy extends robustly to the larger-cap S&P 500 universe. Since 2010, the buy portfolio has outperformed the (equal-weighted) index by 1.96% per year and has outperformed the short portfolio by 5.14% per annum with statistical significance at the 10% level or better (Table 9).

Table 9: Number of References to Guidance – Large Cap Results
S&P 500 May 2010 – December 2017 (Go to Signal Definitions – Exhibit 11)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
		Signal Sort Order	Strategy Start Date	Average Firm Count in Each Quintile Bin	Average Monthly Spearman Correlation	Hit Rate Monthly Spearman Correlation	Annualized Average Monthly Long - Market	Annualized Monthly IR (Long - Market)	Hit Rate Average Monthly Long - Market	Annualized Average Monthly Long - Short	Annualized Monthly IR (Long - Short)	Hit Rate Average Monthly Long - Short
[1]	Number of References to Guidance	D	201005	96	0.021	54.3%	1.96%	0.65	59.8%	5.14%	0.75	59.8%
[2]	p-value	NaN	NaN	NaN	0.038	0.348	0.074	NaN	0.047	0.041	NaN	0.047

Note: Values that are shaded in green (red) are statistically significant at least at the 10% level and are consistent (inconsistent) with our ex-ante hypothesis. D = descending sort and A = ascending sort. Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 10/01/2019.

3.3.2 Language Similarity in the Prepared Remarks

Firms that had the most similarity in language usage during their past five most recent earnings calls outperformed their most extreme counterparts with the least similarity by 2.85% per annum with statistical significance at the 1% level (row [1] in Table 10).

The long-short quintile return spread improved by 32% to 3.75% per year from 2.85% if the prepared remarks section is solely considered (row [3] in Table 10). One plausible explanation is that the prepared remarks are well scripted. Our hypothesis is that the path of least resistance is to have similar language in the prepared remarks from call-to-call until there is one or more material changes that compel the firm to deviate the language usage, for instance, to satisfy their legal and fiduciary responsibilities. See rows [1] and [3] in Table 10.

Table 10: Language Similarity between Calls
Russell 3000 May 2010 – December 2017
(Go to Signal Definitions – Exhibit 12) (Go to the Cover Page)

		Signal Sort Order	Strategy Start Date	Average Firm Count in Each Quintile Bin	Average Monthly Spearman Correlation	Hit Rate Monthly Spearman Correlation	Annualized Average Monthly Long - Market	Annualized Monthly IR (Long - Market)	Hit Rate Average Monthly Long - Market	Annualized Average Monthly Long - Short	Annualized Monthly IR (Long - Short)	Hit Rate Average Monthly Long - Short
[1]	Language Similarity of the Past 5 Calls - **Executives' Remarks**	D	201005	430	0.017	63.0%	1.08%	0.41	56.5%	2.85%	0.71	52.2%
[2]	p-value	NaN	NaN	NaN	0.000	0.009	0.260	NaN	0.175	0.052	NaN	0.602
[3]	Language Similarity of the Past 5 Calls - **Prepared Remarks**	D	201005	429	0.016	69.6%	1.68%	0.77	58.7%	3.75%	1.13	59.8%
[4]	p-value	NaN	NaN	NaN	0.000	0.000	0.035	NaN	0.076	0.002	NaN	0.047

Note: Values that are shaded in green (red) are statistically significant at least at the 10% level and are consistent (inconsistent) with our ex-ante hypothesis. D = descending sort and A = ascending sort. Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 02/01/2018.

3.3.3 The Decision of Timing - How Early Do Firms Introduce Numbers

Historically, firms that introduced numbers earlier on earnings calls outperformed the market and their most extreme counterparts by 1.68% and 3.97% per year, respectively, at the 1% statistical significance level (see row [1] in Table 11).[28]

3.3.4 Self-Attribution Bias - The Blame Game

Firms whose executives attributed softness in their numbers on the economy or the industry underperformed the market and their most extreme counterparts by 1.31% and 2.39% per year, respectively. See row three in Table 11.

Table 11: The Decision of Timing via How Early Do Firms Reference Numbers & Self-Attribution Bias via the Blame Game
Russell 3000 May 2010 – December 2017

(Go to Signal Definitions – Exhibit 13, Exhibit14) (Go to the Cover Page)

		Signal Sort Order	Strategy Start Date	Average Firm Count in Each Quintile Bin	Average Monthly Spearman Correlation	Hit Rate Monthly Spearman Correlation	Annualized Average Monthly Long - Market	Annualized Monthly IR (Long - Market)	Hit Rate Average Monthly Long - Market	Annualized Average Monthly Long - Short	Annualized Monthly IR (Long - Short)	Hit Rate Average Monthly Long - Short
[1]	Referencing Numbers Earlier	D	201005	494	0.008	52.2%	1.68%	0.92	57.6%	3.97%	1.38	58.7%
[2]	p-value	NaN	NaN	NaN	0.027	0.602	0.012	NaN	0.117	0.000	NaN	0.076
[3]	The Blame Game	A	201005	494	0.008	64.1%	1.08%	0.53	51.1%	2.39%	0.84	62.0%
[4]	p-value	NaN	NaN	NaN	0.007	0.005	0.143	NaN	0.755	0.023	NaN	0.016

Note: Values that are shaded in green (red) are statistically significant at least at the 10% level and are consistent (inconsistent) with our ex-ante hypothesis. D = descending sort; A = ascending sort. Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 02/01/2018.

Diversification Benefits

Table 12 has pairwise time-series correlations of long-short quintile return spreads for seven standalone strategies: the four from this section and the three in the same class (i.e., language complexity, numerical transparency and analyst forecast dispersion) that we introduced in our NLP II publication.

[28] Since the format of earnings calls may be firm-specific, we also examined this signal relative to a firm's historical past before making the cross-sectional comparison. The inference does not change and the magnitude of the economic performance is similar. The results are available upon request.

Our analysis suggests that the interactions (i.e., correlations) between the four newly introduced signals in this section are historically low (<= 0.5). The two strategies 'How Early Firms Introduce Numbers' and 'The Blame Game' offer the most benefits of diversification historically.

There is some moderate positive correlation of 0.46 between the guidance-related signal and the call similarity signal. One plausible explanation is that firms that are more likely to reference more instances of guidance are doing well financially and less likely to have disclosures that require drastically altering of their language during calls.

Table 12: Pairwise Return Correlation Matrix of Long-Short Quintile Return Spreads Russell 3000 May 2010 – Dec. 2017

	NLP III - Gauging Call Level Transparency via Executives' Behaviors and Decisions			**NLP II** - Gauging Call Level Transparency via Executives' Behaviors and Decisions		
	The Blame Game	References to Guidance	Language Similarity between Prepared Remarks	Language Complexity Gunning Fog Index	Numerical Transparency Score	Analyst Forecast Dispersion
The Decision of Timing - How Early Firms Reference Numbers	-0.08	-0.10	0.14	-0.12	0.72	0.51
The Blame Game		0.24	0.14	0.34	0.14	0.21
References to Guidance			0.46	0.22	0.22	0.17
Language Similarity between Prepared Remarks				0.37	0.39	0.35

Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 10/01/2019.

The correlations between the newly introduced signals in this publication and those from NLP II are moderately positively correlated (i.e., <= 0.50). The highest pairwise correlation of 0.72 is between the weighted average Numerical Transparency Score and the simple average construct, which is expected.

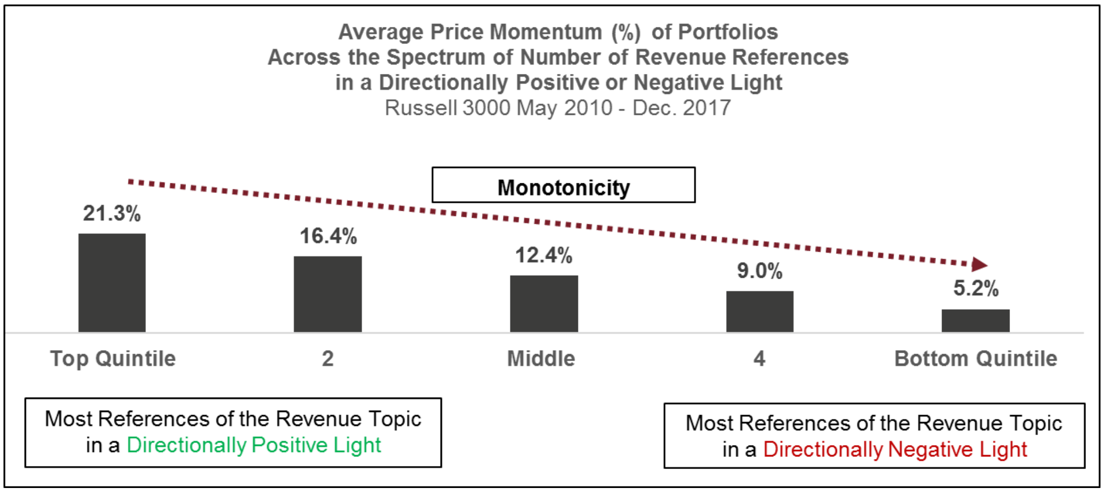
The key takeaway is that a composite signal that blends the individual signals from this set of signals should have a stronger forecasting power at a lower volatility due to the benefits of diversification.

4. Empirical Results: Controlling for Alpha & Risk Signals

Up to this point, we have showed historical results for the standalone strategies. The resulting predictability, however, may be in part or entirely be driven by other commonly used alpha or risk strategies that do not use the content from earnings call transcripts.

For instance, firms whose executives referenced the most instances of revenue-related topics in a directionally positive light should intuitively also exhibited the strong intermediate-term price momentum[29] historically as well due to the underlying financial strength of the firms. We confirm this empirically in Exhibit 17 and the analysis indicates the (strong) positive correlation between the two strategies. This begs the question whether the predictability of the signal from the tagging of the revenue topic is driven by its underlying correlation to price momentum.

Exhibit 17: Average Price Momentum of Portfolios across the Spectrum of Number of Revenue References in a Directional Positive Light



Source: S&P Global Market Intelligence Quantamental Research, as of 02/01/2018.

To validate the three sets of newly introduced signals have additive predictive power, we examine their economic performance after controlling for the following eight stock-level financial characteristics in Exhibit 18, which empirical literature has shown their historical ability to predict stock returns in the cross-section.

[29] It is a well documented phenomenon in empirical asset pricing literature in that stocks that have outperformed in the past 12-month skipping the most recent calendar month continue to outperform the next calendar month.

Exhibit 18: Eight Financial Metrics that Historically Have Stock Selection Power in the Cross-Section

Control Signals	Details	Sorting Order
Beta	CAPM beta 60 months	A
Market capitalization	Natural log of market capitalization	A
Valuation	Book-to-market	D
Price momentum	12-month return excluding the most recent month's	D
Asset growth	One-year change in asset growth	A
Gross Profitability	Gross profit scaled by total assets	D
Analyst Revision	3-month revision in analyst consensus EPS FY1 estimate	D
Earnings Surprise	Standardized unexpected earnings surprise	D

Note: D = descending sort; A = ascending sort
Source: S&P Global Market Intelligence Quantamental Research, as of 11/01/2019.

Using the signal from the tagging of the revenue topic as an example, our analysis suggests that the signal has strong positive tilts (i.e., correlations) to quality, to analyst revision and to earnings surprise in addition to price momentum. In other words, the 20% of stocks whose executives reference the revenue topic the most in a directionally positive (negative) light are also the firms that exhibit with the highest (lowest) quality, that experience the highest positive (lowest) analyst revision and that have the most positive (negative) earnings surprise. See Appendix A.4.

In light of these results, we neutralize the correlations such that the standalone results are not driven by their positive correlations to the eight stock-level characteristics. We use the Fama-Macbeth[30] framework to control for the eight financial metrics from Exhibit 18.

Our results indicate that the newly introduced signals still provide additive economic performance after the adjustments ranging from 2.02% to 6.69% per year with statistical significance at the 5% level or better (see column [1] in Table 13). The additive performance also robustly extends to the large-caps (i.e., Russell 1000) where the annual excess returns range from 1.52% to 4.04% with statistical significance at the 10% level or better (see column [3] in Table 13). **The key takeaway is that the three sets of the newly introduced signals have historically shown additive economic performance above the traditionally used signals.**

[30] Fama-Macbeth framework uses a cross-sectional approach whereas Fama-French uses a time-series one.

Table 13: Excess Long-Short Quintile Return Spreads after Controls
Russell 1000 & 3000 May 2010 – December 2017

(Go to the Cover Page)

		[1]	[2]	[3]	[4]
		Russell 3000		Russell 1000	
Signal Category	Signals	Excess Long-Short Quintile Return Spread	t-statistic	Excess Long-Short Quintile Return Spread	t-statistic
Textual Tagging & Categorization	Revenue Topic + Directionally Positive	6.69%	3.66	4.01%	2.12
	Net Income Topic + Directionally Positive	5.02%	3.72	3.21%	2.29
	Revenue Topic + Directionally Positive + Guidance	3.84%	3.43	3.36%	3.00
	Net Income Topic + Directionally Positive + Guidance	3.76%	4.72	3.08%	3.15
Weighted Avg Sentiment	TF-IDF Weighting Scheme Avg Positivity to Negativity ExecsAll	4.49%	5.53	1.95%	2.08
	Language Similarity Weighting Scheme Net Positivity Relative to the Previous Four Calls	2.46%	3.68	1.52%	1.87
Gauging Call Level Transparency via Executives' Behaviors & Decisions	Abundance or Scarcity of Guidance References	4.69%	3.39	3.75%	2.54
	Average Language Similarity in the Most Recent Five Calls	2.63%	3.67	1.15%	1.34
	The Decision of Timing - How Early Do Firms Introduce Numbers	3.24%	3.95	1.78%	2.35
	Self-Attribution Bias - The Blame Game	2.02%	2.50	2.41%	2.61

Note: Values that are shaded in green (red) are statistically significant at least at the 10% level and are consistent (inconsistent) with our ex-ante hypothesis. D = descending sort and A = ascending sort. Source: S&P Global Market Intelligence Quantamental Research. All returns and indices are unmanaged, statistical composites and their returns do not include payment of any sales charges or fees an investor would pay to purchase the securities they represent. Such costs would lower performance. It is not possible to invest directly in an index. Past performance is not a guarantee of future results. Data as of 02/01/2018.

5. Data & Methodology

Textual Data Analytics (TDA)

TDA was launched in September 2019 and is productized from the NLP II publication. It is an off-the-shelf NLP solution that outputs 40+ earnings call related predictive and descriptive analytics that could be consumed via a SQL database en masse (back to cover page).

Transcripts

The transcripts data set was added to the S&P Global Market Intelligence's Xpressfeed product in September 2017. The historical coverage in the Russell 3000 universe starts in calendar Q1 2008. Among its key features, the data set captures the different segmentations of earnings calls in the follow ways:

- Sections (e.g., prepared remarks section vs. Q&A section),
- Speaker types (e.g., managers, sell-side analysts, shareholders etc.)
- Professionals (e.g., Tim Cook) where the individual professional identifiers serve as a unique key that connects the transcripts data set with the S&P Global Market Intelligence's Professionals and Estimates data sets.

We impose a lag of three trading days to sufficiently account for the latency between an earnings call and its transcription. In the past five years, 99% of all earnings call transcripts are transcribed and hit the database within 24 hours of an earnings call.

Within the Russell 3000 universe, there is an average of 2400+ distinct firms since 2008 that have earnings call transcripts. The main reason for a missing earnings call transcript is that a firm does not hold earnings calls (e.g., Berkshire Hathaway).

6. Conclusion

S&P Global Market Intelligence's earnings call transcripts data is an unstructured data set that is of primary source that offers a differentiated source of alpha. This publication continues our efforts of introducing three new and differentiated sets of systematic stock selection signals using the content from the transcripts and the Textual Data Analytics product.

The signals from the three categories: I) textual tagging and categorization of market moving topics, II) weighted average sentiment and III) gauging call level transparency via executives' behaviors and decisions have historically demonstrated additive forecasting power above the traditionally used strategies such as value, momentum, quality, analyst revision and earnings surprise. The additive economic performance ranged from 2.02% to 6.69% per year with statistical significance at the 5% level or better in the U.S. market since 2010.

Appendix

A.1 Descriptions of Earnings Call Transcripts
Russell 3000 January 2008 – December 2017

(Return to the Main Body of the Paper)

Description	A conference call between the management of a publicly traded company, sell-side analysts and other attendees to discuss the firm's latest financial results
Major Sections	Three major sections: I) prepared remarks by executives II) questions by sell-side analysts III) unscripted responses by executives to analysts' questions
Major Participants	Executives (predominately CEOs and CFOs) and analysts; Executives say about 83% of all words (43% in the prepared remarks). Analysts say about 16% of the words in the form of questions (Exhibit A.2)
Textual	Since calendar Q1 2008, an average earnings call in the U.S. has about 300 sentences about 7000 words and lasts approximately 48 minutes

Source: S&P Global Market Intelligence Quantamental Research, as of 08/01/2019.

A.2 Descriptor Tags
S&P 100 Index January 2008 – December 2017

(Return to the Main Body of the Paper)

Topic Category	Topic Tags
Operating Income	{"ebit", "operating income", "operating profit", "operating earning"}
Cash Flow	{"cash flow", "operating cash flow", "cash flow operations", "free cash flow"}
Shareholder Return	{" buyback", "dividends", "dividend per share", "share repurchase", repurchased million shares"}

A.3 Meaning of Columns in Tables Containing Empirical Results

(Return to the Main Body of the Paper)

Signal Name, Sort Order, Start Date, Firm Count

- Column 1: the sort order of a signal where 'D' or -1 is descending and 'A' or 1 is ascending
- Column 2: the date back-tests commenced for a signal
- Column 3: the average number of firms in a quintile bucket in our sample period; In this publication, the 'custom sort' phrase denotes that a signal's positive scores are descendingly sorted in the first four quintiles. The bottom quintile has scores that are non-positive. This applies to the signals in the category of topic identification.

Signal Strength Metrics

- Column 4: the average monthly information coefficient (i.e., Spearman correlation) that is used to assess a signal's historical predictive strength
- Column 5: the monthly hit rate for column 4 - the percent of the months where the IC > 0

Active Long Metrics

- Column 6: the annualized average monthly market-adjusted return of the long portfolio
- Column 7: annualized information ratio of column 6
- Column 8: the monthly hit rate for column 6 where the market-adjusted return of the long-side > 0

Long-Short Metrics

- Column 9: the annualized long-short returns
- Column 10: annualized information ratio of column 9
- Column 11: the monthly hit rate for column 9 where the monthly long-short return > 0

A.4 Average Financial Characteristics of Quantile Portfolios across the Spectrum of Number of Revenue References in a Directional Positive Light
Russell 3000 May 2010 – December 2017

(Return to the Main Body of the Paper)

Traditional Signals	Top Quintile	2	Middle	4	Bottom Quintile
CAPM beta 60 months	1.21	1.27	1.40	1.31	1.32
Market Capitalization (MM in USD)	7930	8572	8721	7731	4794
Book-to-market	0.37	0.42	0.51	0.63	0.73
12-month return excluding the most recent month's	21.3%	16.4%	12.4%	9.0%	5.2%
One-year change in asset growth	15.0%	14.5%	13.0%	13.3%	14.7%
Gross profit scaled by total assets	43.1%	37.5%	29.8%	19.5%	16.0%
3-month analyst revision EPS FY1 estimate (demeaned in the cross-section)	0.27%	0.16%	0.00%	-0.13%	-0.35%
Standardized unexpected earnings surprise (demeaned in the cross-section)	0.64	0.36	0.00	-0.15	-0.54

References

Ali, A., Liu, M., Xu, D., and Yao, T., "Corporate Disclosure, Analyst Forecast Dispersion, and Stock Returns" (August 2012 2001). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=556704

Ang, J.S. and Ciccone, S. J., "Analyst Forecasts and Stock Returns" (June 2001). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=271713

Cohen, L. and Malloy, C. J., "Lazy Prices" (Latest Version - March 2019). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1658471

Egan, M. (2018, May 3). Analysis: Elon Musk Picked a Bad Time to Burn Bridges on Wall Street? Retrieved from https://money.cnn.com/2018/05/03/investing/elon-musk-tesla-earnings-call/index.html

Fama, E. F. and French, K. R., "Profitability, investment and average returns" (2006). Journal of Financial Economics 82, 491–518.

Fama, E. and K. French (2008). Average returns, B/M, and share issues. Journal of Finance 63(6), 2971–2995.

Fama, E. F. and French, K. R., "Dissecting Anomalies" (2008). The Journal of Finance, 63: 1653–1678. doi: 10.1111/j.1540-6261.2008.01371.

Fama, E. F. and French, K. R., "A Five-Factor Asset Pricing Model" (May 2014). Fama-Miller Working Paper. Available at SSRN: http://ssrn.com/abstract=2287202 or http://dx.doi.org/10.2139/ssrn.2287202

Fama, E. and J. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81(3), 607–636.

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.

Loughran, T., AND B. McDonald. "IPO First-day Returns, Offer Price Revisions, Volatility, and Form S-1 Language." *Journal of Financial Economics* 109 (2013): 307-326.

Loughran, T., AND B. McDonald. "Measuring Readability in Financial Disclosures." *Journal of Finance* 69 (2014): 1643-1671.

Loughran, T., AND B. McDonald. "The Use of Word Lists in Textual Analysis." *Journal of Behavioral Finance* 16 (2015): 1-11.

Zhou, D. "The Blame Game" (June 2014). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2447042

Our Recent Research

December 2019: The "Trucost" of Climate Investing: Managing Climate Risks in Equity Portfolios

Nearly 45% of global investors surveyed in a 2017 investor study cited performance as a primary concern in sustainable investing. Our latest research suggests investors' fears are misplaced – carbon-sensitive portfolios have similar returns and significantly better climate characteristics than portfolios constructed without carbon emission considerations.

- Highly profitable firms are likely to be leaders in reducing their carbon emission levels.
- There is no degradation in fundamental characteristics for the carbon-sensitive portfolios compared to the baseline portfolio, even though the difference in constituents can be as high as 20%.
- Carbon-sensitive portfolios were observed as having significant reductions in water use, air pollutants released and waste generated.

October 2019: #ChangePays: There Were More Male CEOs Named John than Female CEOs

This report examines the performance of firms that have made female appointments to their CEO and CFO positions. Our research finds that firms with female CEOs and/or CFOs:.

- Are more profitable and generated excess profits of $1.8 trillion over the study horizon.
- Have produced superior stock price performance, compared to the market average.
- Have a demonstrated culture of Diversity and Inclusion, evinced by more females on the company's board of directors.

June 2019: Looking Beyond Dividend Yield: Finding Value in Cash Distribution Strategies

Examines the relationship between yield-oriented strategies (dividend yield, buyback yield, and combined shareholder yield) and future stock return, across multiple countries/regions. Also provides insights into two additional topics:

- Which company fundamental characteristics support and enhance future shareholder payouts?
- Under which interest rate environment should investors favor yield-oriented strategies?

June 2019: The Dating Game: Decrypting the Signals in Earnings Report Dates

The first part of this report focuses on companies that deviate from a historical reporting pattern, while the second part examines a related topic – the market's reaction to companies that postpone a previously scheduled (announced) earnings release date.

- "Advancers" (companies that advance their earnings report date by at least 6 days) are likely to report improving year-year on sales, better earnings surprises, and more positive conference call sentiment readings than their industry group peers and "delayers" (companies that delay their earnings report date by at least 6 days).

- Advancers outperform delayers by over 7% on an annualized basis (Russell 3000). This return rises to 8.80% (Russell 2000) and falls to 2.21% (Russell 1000).
- The annualized return to stocks identified as buy candidates and tagged as advancers is 10.77%, compared to 6.29% for buy candidates tagged as delayers.
- Companies that postpone a previously announced earnings release date underperform the broad market by 2.44% in the 3 days surrounding the announcement. These companies are also likely to report deteriorating fundamentals.

May 2019: Bridges for Sale: Finding Value in Sell-Side Estimates, Recommendations, and Target Prices

This report looks at the informativeness of analyst recommendation revisions, target price revisions, and estimate dispersion, primarily within the post-2002 regulatory environment, and finds significant results in all three areas:

- Investors should focus on shifts in consensus recommendations, as the recommendation level by itself often reflects pro-management and high-growth biases.
- Target prices, labeled by some practitioners as "fiction" likewise provide insight into changing analyst attitudes. The six-month change in target price gap (the spread between target and market price) produces statistically significant results globally.
- Analyst estimate dispersion acts as an indicator of corporate quality – high quality companies have more stable revenue and income streams that are more amenable to forecasting

February 2019: U.S Stock Selection Model Performance Review

February 2019: International Small Cap Investing: Unlocking Alpha Opportunities in an Underutilized Asset Class

January 2019: Value and Momentum: Everywhere, But Not All the Time

November 2018: Forging Stronger Links: Using Supply Chain Data in the Investing Process

September 2018: Their Sentiment Exactly: Sentiment Signal Diversity Creates Alpha Opportunity

September 2018: Natural Language Processing – Part II: Stock Selection: Alpha Unscripted: The Message within the Message in Earnings Calls

July 2018: A Case of 'Wag the Dog'? - ETFs and Stock-Level Liquidity

June 2018: The (Gross Profitability) Trend is Your Friend

May 2018: Buying the Dip: Did Your Portfolio Holding Go on Sale?

March 2018: In the Money: What Really Motivates Executive Performance?

February 2018: The Art of the (no) Deal: Identifying the Drivers of Canceled M&A Deals

January 2018: U.S Stock Selection Model Performance Review

September 2017: Natural Language Processing - Part I: Primer

July 2017: Natural Language Processing Literature Survey

June 2017: Research Brief: Four Important Things to Know About Banks in a Rising Rate Environment

April 2017: Banking on Alpha: Uncovering Investing Signals Using SNL Bank Data

March 2017: Capital Market Implications of Spinoffs

January 2017: U.S. Stock Selection Model Performance Review 2016

November 2016: Electrify Stock Returns in U.S. Utilities

October 2016: A League of their Own: Batting for Returns in the REIT Industry - Part 2

September 2016: A League of their Own: Batting for Returns in the REIT Industry - Part 1

August 2016: Mergers & Acquisitions: The Good, the Bad and the Ugly (and how to tell them apart)

July 2016: Preparing for a Slide in Oil Prices -- History May Be Your Guide

June 2016: Social Media and Stock Returns: Is There Value in Cyberspace?

April 2016: An IQ Test for the "Smart Money" – Is the Reputation of Institutional Investors Warranted?

March 2016: Stock-Level Liquidity – Alpha or Risk? - Stocks with Rising Liquidity Outperform Globally

February 2016: U.S. Stock Selection Model Performance Review - The most effective investment strategies in 2015

January 2016: What Does Earnings Guidance Tell Us? – Listen When Management Announces Good News

December 2015: Equity Market Pulse – Quarterly Equity Market Insights Issue 6

November 2015: Late to File - The Costs of Delayed 10-Q and 10-K Company Filings

October 2015: Global Country Allocation Strategies

September 2015: Equity Market Pulse – Quarterly Equity Market Insights Issue 5

September 2015: Research Brief: Building Smart Beta Portfolios

September 2015: Research Brief – Airline Industry Factors

August 2015: Point-In-Time vs. Lagged Fundamentals – This time i(t')s different?

August 2015: Introducing S&P Capital IQ Stock Selection Model for the Japanese Market

July 2015: Research Brief – Liquidity Fragility

June 2015: Equity Market Pulse – Quarterly Equity Market Insights Issue 4

May 2015: Investing in a World with Increasing Investor Activism

April 2015: Drilling for Alpha in the Oil and Gas Industry – Insights from Industry Specific Data & Company Financials

March 2015: Equity Market Pulse – Quarterly Equity Market Insights Issue 3

February 2015: U.S. Stock Selection Model Performance Review - The most effective investment strategies in 2014

January 2015: Research Brief: Global Pension Plans - Are Fully Funded Plans a Relic of the Past?

January 2015: Profitability: Growth-Like Strategy, Value-Like Returns - Profiting from Companies with Large Economic Moats

November 2014: Equity Market Pulse – Quarterly Equity Market Insights Issue 2

October 2014: Lenders Lead, Owners Follow - The Relationship between Credit Indicators and Equity Returns

August 2014: Equity Market Pulse – Quarterly Equity Market Insights Issue 1

July 2014: Factor Insight: Reducing the Downside of a Trend Following Strategy

May 2014: Introducing S&P Capital IQ's Fundamental China A-Share Equity Risk Model

April 2014: Riding the Coattails of Activist Investors Yields Short and Long Term Outperformance

March 2014: Insights from Academic Literature: Corporate Character, Trading Insights, & New Data Sources

February 2014: Obtaining an Edge in Emerging Markets

February 2014: U.S Stock Selection Model Performance Review

January 2014: Buying Outperformance: Do share repurchase announcements lead to higher returns?

October 2013: Informative Insider Trading - The Hidden Profits in Corporate Insider Filings

September 2013: Beggar Thy Neighbor – Research Brief: Exploring Pension Plans

August 2013: Introducing S&P Capital IQ Global Stock Selection Models for Developed Markets: The Foundations of Outperformance
July 2013: Inspirational Papers on Innovative Topics: Asset Allocation, Insider Trading & Event Studies

June 2013: Supply Chain Interactions Part 2: Companies – Connected Company Returns Examined as Event Signals

June 2013: Behind the Asset Growth Anomaly – Over-promising but Under-delivering

April 2013: Complicated Firms Made Easy - Using Industry Pure-Plays to Forecast Conglomerate Returns.

March 2013: Risk Models That Work When You Need Them - Short Term Risk Model Enhancements

March 2013: Follow the Smart Money - Riding the Coattails of Activist Investors

February 2013: Stock Selection Model Performance Review: Assessing the Drivers of Performance in 2012

January 2013: Research Brief: Exploiting the January Effect Examining Variations in Trend Following Strategies

December 2012: Do CEO and CFO Departures Matter? - The Signal Content of CEO and CFO Turnover

November 2012: 11 Industries, 70 Alpha Signals -The Value of Industry-Specific Metrics
October 2012: Introducing S&P Capital IQ's Fundamental Canada Equity Risk Models

September 2012: Factor Insight: Earnings Announcement Return – Is A Return Based Surprise Superior to an Earnings Based Surprise?

August 2012: Supply Chain Interactions Part 1: Industries Profiting from Lead-Lag Industry Relationships

July 2012: Releasing S&P Capital IQ's Regional and Updated Global & US Equity Risk Models

June 2012: Riding Industry Momentum – Enhancing the Residual Reversal Factor

May 2012: The Oil & Gas Industry - Drilling for Alpha Using Global Point-in-Time Industry Data

May 2012: Case Study: S&P Capital IQ – The Platform for Investment Decisions
March 2012: Exploring Alpha from the Securities Lending Market – New Alpha Stemming from Improved Data

January 2012: S&P Capital IQ Stock Selection Model Review – Understanding the Drivers of Performance in 2011

January 2012: Intelligent Estimates – A Superior Model of Earnings Surprise

December 2011: Factor Insight – Residual Reversal

November 2011: Research Brief: Return Correlation and Dispersion – All or Nothing

October 2011: The Banking Industry

September 2011: Methods in Dynamic Weighting

September 2011: Research Brief: Return Correlation and Dispersion

July 2011: Research Brief - A Topical Digest of Investment Strategy Insights

June 2011: A Retail Industry Strategy: Does Industry Specific Data tell a different story?

May 2011: Introducing S&P Capital IQ's Global Fundamental Equity Risk Models

May 2011: Topical Papers That Caught Our Interest

April 2011: Can Dividend Policy Changes Yield Alpha?

April 2011: CQA Spring 2011 Conference Notes

March 2011: How Much Alpha is in Preliminary Data?
February 2011: Industry Insights – Biotechnology: FDA Approval Catalyst Strategy

January 2011: **US Stock Selection Models Introduction**

January 2011: **Variations on Minimum Variance**

January 2011: **Interesting and Influential Papers We Read in 2010**

November 2010: **Is your Bank Under Stress? Introducing our Dynamic Bank Model**

October 2010: **Getting the Most from Point-in-Time Data**

October 2010: **Another Brick in the Wall: The Historic Failure of Price Momentum**

July 2010: **Introducing S&P Capital IQ's Fundamental US Equity Risk Model**