Conditional Asset Pricing with Text-managed Portfolios

Jian Feng* Jiantao Huang† Shiyang Huang‡ Ran Shi§

September 2024

Abstract

We construct managed portfolios based on textual analysis of firms' earnings conference call transcripts and investigate their asset pricing implications. Loadings on the text-managed portfolios can explain a comparable amount of stock-level return variation as those on the conventional characteristics-based factors. Combining the earnings call text and firm characteristics enhances the conditional mean-variance efficiency of factor portfolios but not the predictive power for stock returns. Drawing on the insights of Kozak and Nagel (2024), our evidence suggests that the earnings call text contains information about the return covariances missing from characteristics. Text-managed portfolios are exposed to investor sentiments and forecast macroeconomic outcomes.

Keywords: factor models, mean-variance efficiency, conditioning information, text data.
JEL Classification Codes: C11, G11, G12.

*Faculty of Business and Economics, The University of Hong Kong; jianfeng21@connect.hku.hk.
†Faculty of Business and Economics, The University of Hong Kong; huangjt@hku.hk.
‡Faculty of Business and Economics, The University of Hong Kong; huangsy@hku.hk.
§University of Colorado Boulder; ran.shi@colorado.edu.

1 Introduction

The empirical finance literature has reported various firm characteristics that tend to predict their stock returns. Based on these characteristics, a common practice is building *managed portfolios* from sorts or regressions to summarize the investment opportunities of individual stocks.[1] If a few factors can achieve the same (conditional) mean-variance efficiency as all the individual stocks, they *span* the investment opportunity. Expected stock returns can then be explained by their dynamic loadings on the spanning factors, giving rise to valid conditional asset pricing models.

A natural question arises: can we always construct a small set of factors—based on given firm characteristics—that deliver the same mean-variance efficient frontier as the entire cross-section of stocks? In a recent paper by Kozak and Nagel (2024), it is shown that the answer is no for most heuristic factors constructed in practice,[2] unless the firm characteristics used to create these factors can also explain all the priced variation in conditional return covariances (in addition to expected returns). The latter condition is more likely to hold when researchers consider a large set of characteristics.

In this paper, we investigate a new category of firm characteristics—text data from the earnings conference call transcripts—and study its implications for constructing spanning factors in conditional asset pricing models. Like many firm characteristics, conference call text offers valuable insights about a firm's financial performance. More importantly, conference calls provide a platform for market participants to gather information about the management team's narrative and future outlook. They can potentially offer orthogonal information to existing characteristics that are based primarily on accounting variables and historical market prices; hence, they help satisfy the Kozak and Nagel (2024) condition for identifying spanning asset pricing factors.[3]

[1] For example, Fama and French (1993, 2015) and Hou et al. (2015) sort stocks on their characteristics to create factors. Rosenberg (1974) and Fama and French (2020) use cross-sectional return regressions on characteristics to construct least square slope factors.

[2] Kozak and Nagel (2024) show that, in theory, one can always find factors that are the managed portfolios of return-predicting characteristics, such that these factors span the investment opportunities of individual stocks. The factors are the generalized least square slope portfolios from cross-sectional return regressions. However, this construction is unpopular in practice, partly due to its reliance on the large conditional covariance matrix of individual stock returns.

[3] We recognize that other text information such as news, social media, and regulatory filings could also

The information in the conference call text has two major asset pricing implications. Conditional on the text, a latent factor model can explain a comparable amount of time-series variation in stock returns as the conventional firm characteristics. However, combining these two types of information does not significantly improve over using each type independently.

As a measure of mean-variance efficiency, or equivalently, the likelihood of factor spanning, the maximal out-of-sample Sharpe ratio (SR) of text-managed factor portfolios is close to that using firm characteristics only. Unlike predicting expected stock returns, incorporating conference call text into the conditioning information set of traditional characteristics can substantially improve mean-variance efficiency.

These findings align with the theoretical insights from Kozak and Nagel (2024). Expanding the conditioning information set with conference call text leads to higher mean-variance efficiency, even though the text does not add incremental predictive power for expected returns. This suggests that conference call text captures variation in return covariances that are overlooked by traditional characteristics, highlighting the informational advantages beyond return predictions from analyzing "big" unstructured financial text data.

Our exercise is challenging for two reasons. The first conceptual challenge is to build spanning factors *jointly* from many characteristics (including ones extracted from the earnings conference call transcripts). Our solution builds on the theoretical results in Kozak and Nagel (2024). When the linear combinations of characteristics are sufficient in explaining both the expected returns and conditional covariances (which is, again, more likely to hold when including many characteristics carrying different information), the instrumented principal component analysis (IPCA) of Kelly et al. (2019) can identify the factors (at the population level). The insights from Kelly et al. (2019) allow us to interpret the factors as the leading principal components of managed portfolios constructed from both the conventional characteristics and our text-based ones.

Another implementation challenge is to extract quantitative signals from text documents that are are specific to each firm and easy to interpret. In nature, it is a process of mapping a firm's conference call $\mathcal{D}$ to a numeric encoding vector $\boldsymbol{d}$. Modern neural-network-based

help achieve factor spanning. Our methodology can be directly used to study such text data. Other sources of information, such as granular investor holdings, specially encoded in the fashion of Gabaix et al. (2023), can also be incorporated into our framework.

methods can generate $\boldsymbol{d}$ vectors of manageable dimensions, but the elements in $\boldsymbol{d}$ can be hard to interpret. More importantly, the pre-training process to "learn" these models introduces look-ahead biases into $\boldsymbol{d}$ (Glasserman and Lin, 2023; Sarkar and Vafa, 2024).[4] The conventional approach of encoding $\mathcal{D}$ based on word frequencies will suffer from ultra-high dimensions—with each dimension representing the "loadings" of a document on a specific word—especially when working with *multi*-gram dictionaries. Our solution is to use the well-known topic modelling approach of Blei et al. (2003) for dimension reduction. Each dimension of $\boldsymbol{d}$ from this approach can be interpreted as the firm's loadings on a topic.

We estimate topics separately for the scripted presentation sessions—which include opening remarks, discussions of financial performance, and the "forward guidance"—and the more interactive Q&A sessions.[5] In our sample, the presentation and Q&A sessions each contribute 32 stable topics, yielding a total of 64 conference call topics. Most topics are highly interpretable and can be categorized into three main themes: general topics (e.g., "economic uncertainty" and "supply chain"), industries (e.g., "pharmaceutics" and "oil"), and topics related to technology vintages (e.g., "cloud service" and "renewable energy"). The presentation and Q&A sessions share common topics such as "inflation" and various industries. However, some distinctive topics emerge in the Q&A sessions, such as "forecasting and speculation" as well as "optimism," while the presentation sessions uniquely focus on topics like "profits" and "business operation."

Firms' topic loadings exhibit significant cross-sectional variation, which we use as instruments for dynamic factor loadings in IPCA estimation. These loadings provide quantitative measures of firms' exposures to common themes relevant to managers, analysts, and investors. We add these 64 topic loadings to the 35 characteristics studied in Kelly et al. (2019). We conduct expanding-window IPCA estimations to assess how well different subsets of conditioning variables explain the variation in stock-level expected returns *out-of-sample*. We further investigate whether factor portfolios constructed using conference call topic loadings can help span individual stock returns. To achieve this, we build tangency portfolios

[4]This issue can be addressed if researchers (pre-)train their own language models using time-stamped expanding text corpus instead using the off-the-shelf ones, but the resulting $\boldsymbol{d}$ vectors can be more difficult to interpret.

[5]Matsumoto et al. (2011) conduct heterogeneity analysis into the information content of the two sessions.

of the IPCA factors using recursive estimates of mean-variance efficient (MVE) portfolio weights and report the out-of-sample Sharpe ratios (SR).

In our sample over the 2003 to 2022 period, the IPCA models conditional on either the 32 presentation, 32 Q&A, or the 64 combined topic loadings can explain around 24%/0.14% variation in the realized/expected returns out-of-sample. In comparison, the numbers are 27% and 0.15% for the 35 characteristics from (Kelly et al., 2019, KPS). Adding the conference call topic loadings to the KPS characteristics does not offer much improvement in explaining time-series variation in stock returns.

Nevertheless, the textual information embedded in conference call transcripts significantly improves the conditional mean-variance efficiency. The text-managed tangency portfolios conditional on either Q&A or presentation topic loadings command annualized SRs of around 1.0–1.2 out-of-sample, and combining both sources of textual information builds an even more efficient portfolio that earns an out-of-sample SR of 1.64, almost comparable to that of the MVE portfolio conditional on 35 KPS characteristics (SR = 1.71). More importantly, conditioning on both 64 topic loadings and 35 KPS characteristics improves the maximal out-of-sample SR to 2.76.

The presentation sessions, the Q&A sessions, and the KPS characteristics tend to capture rather different cross-sectional information. We show that the three tangency portfolios conditional on these three sources of information are only weakly correlated with each other. In addition, the text-managed MVE portfolios—either based on the conference call presentation or Q&A sessions—deliver returns that are not explained by traditional factor models of Fama and French (2015) and Hou et al. (2015), as well as IPCA factors constructed from 35 KPS characteristics.

The two sessions of the conference calls yield conditional MVE portfolios that carry heterogeneous information. While the Q&A-managed MVE portfolios can be positively predicted by the sentiment measures from Baker and Wurgler (2006) and Huang et al. (2015), we do not detect similar patterns in the presentation-managed MVE portfolio. The findings are consistent with the Q&A sessions emanating specific topics such as "forecasting and speculation," "economic uncertainty," and "optimism."

We also delve deeper into the topic loadings to understand which topics are important

using a leave-one-out procedure. We repeat our IPCA estimations by dropping each topic and using the remaining ones to construct the tangency portfolios. If a topic is particularly salient for the mean-variance efficiency, we will observe a detectable drop in the Sharpe ratios. In the Q&A sessions, excluding "cloud service," "capital structure," or "human capital" leads to -0.53 to -0.17 decreases in annualized out-of-sample Sharpe ratios. In contrast, topic loadings on "business operation," as well as industries such as "steel," "semiconductor and electronics," "asset management", and "energy" emerge as crucial topics in the presentation sessions.

Concerns may be raised as our topic model estimation is performed on the whole corpus of earnings conference calls in our sample. To alleviate these concerns, we repeat our analysis using an online version of topic modeling. Effectively, our topic loadings are now also estimated using an expanding window. The asset pricing results (e.g., time-series fit and mean-variance efficiency) using these topic loadings are largely unchanged.

Our conference call text-managed portfolios contain important forward-looking information about future employment, inflation, and credit spread. At a short horizon (one month), both the Q&A- and presentation-managed tangency portfolio returns positively predict changes in the consumer price index, echoing our findings that both sessions share the common topic "inflation." The Q&A-managed portfolio returns also negatively predict credit spread over the next month, consistent with the important topic "capital structure" that we find in the Q&A session. At a longer horizon (one year), Q&A-managed tangency portfolio returns positively predict changes in the nonfarm-employment rate and negatively predict changes in the overall unemployment rate, which is consistent with a unique topic of "human capital" in the Q&A session.

Related Literature. Our paper contributes to the literature on textual analysis in financial economics, especially in asset pricing. As a specific class of "big data" in finance (Goldstein et al., 2021), text has become a more prevalent data source for empirical works, as surveyed in Gentzkow et al. (2019), Das et al. (2014) and Loughran and McDonald (2020). In the asset pricing literature, text data have been used to capture expected returns (Tetlock, 2007; García, 2013; Ke et al., 2022), create novel risk measures (Manela and Moreira, 2017;

Hassan et al., 2019; Hanley and Hoberg, 2019; Giglio et al., 2024), conduct risk management (Engle et al., 2020), and recover asset pricing factors (Bybee et al., 2023; Lopez-Lira, 2023; Aleti and Bollerslev, 2024).

Closest to our work are Lopez-Lira (2023) and Bybee et al. (2023), which also exploit information in loadings estimated from topic models to instrument for the factor exposures. Lopez-Lira (2023) studies the implications of risk disclosures in annual reports (form 10-K) on identifying priced factors. Bybee et al. (2023) construct narrative factors based on firms' return covariances with respect to loadings on news topics. Earnings conference calls are more timely than 10-K and more standardized in terms of conveying firm-specific news, making it an ideal starting point to empirically evaluate the theory of Kozak and Nagel (2024) on conditional spanning (with enriched firm characteristics such as text data).

The literature on conditional asset pricing dates back at least to the seminal theory work of Hansen and Richard (1987). Empirical studies involve factor pricing conditioning on macroeconomic variables (Ferson and Harvey, 1999) and firm-specific characteristics (Kelly et al., 2019); efficient use of conditioning information for mean-variance analysis (Ferson and Siegel, 2001); model testing and evaluation (Lewellen and Nagel, 2006; Nagel and Singleton, 2011); conditional mean-variance spanning (Kozak and Nagel, 2024). We contribute to this literature by characterizing the informational role of earnings calls in return prediction and mean-variance analysis.

Our use of text-managed portfolios to forecast aggregate outcomes echoes the works of Lamont (2001) and Kelly and Pruitt (2013), although our portfolios are not constructed to track any economic outcomes by design. The practice of building managed portfolios as traded proxies to economic "state variables" is related to the mimicking portfolio approach to estimating factor risk premiums (Huberman et al., 1987; Balduzzi and Robotti, 2008; Giglio and Xiu, 2021). Textual analysis of news data has also been proven useful for forecasting market returns and economic activities (Larsen and Thorsrud, 2019; Bybee et al., 2024; van Binsbergen et al., 2024). Different from our exercise, forecasting variables in this literature are not tradable asset returns.

The rest of the paper is organized as follows. Section 2 briefly summarizes the theoretical insights of Kozak and Nagel (2024) to clarify the impacts of adding text data to characteristics

on factor spanning. Our estimation method and its theory foundations are then discussed. Section 3 introduces our data sources and sample construction. Section 4 describes the common topics in earnings conference calls. Section 5 presents our results on using the topic loadings to measure expected returns and construct spanning factors. Section 6 documents the forecasting power of these factors for macroeconomic outcomes. Section 7 concludes.

2 Theory and Method

We review in this section the implications of enriching the conditioning information set (with, for example, firm-specific text documents in addition to the standard characteristics) for portfolio analysis. Special attention is paid to spanning the mean-variance efficient frontier with a relatively small number of factors. We then present our methodology of incorporating firm-specific text information to identify and estimate these factors.

To fix ideas, consider a panel of n realized stock returns in excess of the risk-free rate $\boldsymbol{r}_{t+1} = [r_{1,t+1}, \ldots, r_{n,t+1}]'$. Assume that stock i's characteristics and other text-based information (e.g., disclosure or news coverage) observed by the econometrician at time t is concatenated into a vector $\boldsymbol{x}_{i,t}$ of dimension m. Define the $n \times m$ time-varying matrix $\boldsymbol{X}_t = [\boldsymbol{x}_{1,t}, \cdots, \boldsymbol{x}_{n,t}]'$ and denote by $\boldsymbol{\mu}_t = \mathbb{E}[\boldsymbol{r}_{t+1} \mid \boldsymbol{X}_t]$ the conditional expected returns and by $\boldsymbol{\Sigma}_t = \text{var}[\boldsymbol{r}_{t+1} \mid \boldsymbol{X}_t]$ the conditional covariance matrix.

We are interested in the possibility of finding p ($p < n$) factors $\boldsymbol{f}_{t+1} \in \mathbb{R}^p$ that offer the same investment opportunity characterized by the n stocks. That is, by investing the p assets defined by the factors, an investor can achieve the same maximal conditional Sharpe ratio of the n individual stocks. We follow the exposition of Kozak and Nagel (2024) and assume that the conditional expected returns are linear in the vector of firm-specific information.

Assumption 1. $\boldsymbol{\mu}_t = \boldsymbol{X}_t \boldsymbol{b}_t$ *for some vector* $\boldsymbol{b}_t \in \mathbb{R}^m$.

The ideal benchmark. Under Assumption 1, one can always construct the factors that span the investment opportunities offered by the individual stocks. Define the factors as $\boldsymbol{f}_{t+1} = \boldsymbol{X}_t' \boldsymbol{\Sigma}_t^{-1} \boldsymbol{r}_{t+1}$. Since

$$\mathbb{E}[\boldsymbol{r}_{t+1} \mid \boldsymbol{X}_t] = \text{cov}[\boldsymbol{r}_{t+1},\ \boldsymbol{f}_{t+1} \mid \boldsymbol{X}_t] \times \boldsymbol{b}_t,$$

the factors $\boldsymbol{f}_t$ must achieve conditional mean-variance efficiency, and the number of factors equals the dimension of $\boldsymbol{x}_{i,t}$, i.e., $p = m$.

Managed portfolios. In practice (e.g., the Fama and French (2015) factors, and the MSCI Barra or the Axioma factors), simple managed portfolios $\boldsymbol{f}_{t+1} = \boldsymbol{X}_t'\boldsymbol{r}_{t+1}$—and their transformations based on functions of $\boldsymbol{X}_t$—are used to construct the factors. These portfolios, unlike the ideal benchmarks, can be constructed without resorting to the large covariance matrix $\boldsymbol{\Sigma}_t$, which is often difficult to estimate in empirical studies. The managed portfolios are not mean-variance efficient unless the following assumption about the covariance matrix holds (Kozak and Nagel, 2024).

Assumption 2. *$\boldsymbol{\Sigma}_t$ is such that $\boldsymbol{\Sigma}_t\boldsymbol{X}_t = \boldsymbol{X}_t\boldsymbol{B}_t$ for some nonsingular $m \times m$ matrix $\boldsymbol{B}_t$.*

A direct implication of Assumption 2 is that the $n \times m$ matrix of covariances $\text{cov}[\boldsymbol{r}_{t+1},\ \boldsymbol{f}_{t+1} \mid \boldsymbol{X}_t]$ is a linear function of $\boldsymbol{X}_t$, when the factors are managed portfolios $\boldsymbol{f}_{t+1} = \boldsymbol{X}_t'\boldsymbol{r}_{t+1}$. Assumption 2 is also equivalent to a requirement that the stock returns' loadings on the factors that are not the managed portfolios—namely the omitted factors (Giglio and Xiu, 2021)—must be in the nullspace of $\boldsymbol{X}_t$ (Kozak and Nagel, 2024).

Assumptions 1 and 2 are both more likely to hold when richer information is embedded into the vectors $\boldsymbol{x}_{i,t}$, $i = 1, \ldots, n$. As a result, the managed portfolios are more likely to span the investment opportunity of the individual stocks, when either the expected returns or the covariances (or both) are better explained by $\boldsymbol{X}_t$. Firm characteristics are commonly used as the explanatory variables in the empirical finance literature, and including more characteristics could help satisfy the two assumptions. A concern may arise in this process as the newly added firm characteristics will likely be highly correlated with the existing ones.

Text information such as firm disclosures and news coverage could be helpful by offering new information that is more orthogonal. We focus on the earnings conference calls in this paper along this venue. Before presenting our approach of incorporating earnings conference call information into $\boldsymbol{X}_t$, we discuss the issue of dimension reduction.

Dimension reduction. The managed portfolios still yield the same number of factors as the dimension of $\boldsymbol{x}_{i,t}$ ($p = m$). The following assumption can ensure that the first p ($p \leq m$) principle components of the (transformed) managed portfolios could achieve the same degree

of mean-variance efficiency as the n individual stocks.

Assumption 3. *The conditional mean* $\boldsymbol{\mu}_t$ *and covariance* $\boldsymbol{\Sigma}_t$ *are such that* $\boldsymbol{\mu}_t = \widetilde{\boldsymbol{X}}_t \boldsymbol{b}_t$ *and* $\boldsymbol{\Sigma}_t \widetilde{\boldsymbol{X}}_t = \widetilde{\boldsymbol{X}}_t \boldsymbol{B}_t$, *where* $\widetilde{\boldsymbol{X}}_t = \boldsymbol{X}_t \boldsymbol{\Gamma}$ *for some* $m \times p$ *matrix* $\boldsymbol{\Gamma}$.

If Assumption 3 holds, we can achieve dimension deduction by studying the following latent factor structure of individual stock returns:

$$\boldsymbol{r}_{t+1} = \boldsymbol{\beta}_t \boldsymbol{f}_{t+1} + \boldsymbol{\epsilon}_{t+1}, \quad \boldsymbol{\beta}_t = \boldsymbol{X}_t \boldsymbol{\Gamma}, \quad \mathrm{cov}_t(\boldsymbol{\epsilon}_{t+1}) \boldsymbol{X}_t = \boldsymbol{0}, \tag{1}$$

where $\boldsymbol{\beta}_t$ denote the dynamic factor loadings instrumented by $\boldsymbol{X}_t$, and $\boldsymbol{\epsilon}_{t+1}$ are the error terms that are orthogonal to $\boldsymbol{X}_t$ and $\boldsymbol{f}_{t+1}$ and have zero conditional means. Once again, the condition $\mathrm{cov}_t(\boldsymbol{\epsilon}_{t+1}) \boldsymbol{X}_t = \boldsymbol{0}$ is more likely to hold if more independent conditional information $\boldsymbol{X}_t$ are included in the model (1).

We can estimate the factors $\{\boldsymbol{f}_t\}_{t=1}^T$, as well as the matrix $\boldsymbol{\Gamma}$, using the instrumented principle component analysis (IPCA) estimator proposed by Kelly et al. (2019),

$$\{\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{f}}_1, \ldots, \widehat{\boldsymbol{f}}_T\} = \underset{\boldsymbol{\Gamma}, \{\boldsymbol{f}_t\}_{t=1}^T}{\arg\min} \sum_{t=0}^{T-1} (\boldsymbol{r}_{t+1} - \boldsymbol{X}_t \boldsymbol{\Gamma} \boldsymbol{f}_{t+1})' (\boldsymbol{r}_{t+1} - \boldsymbol{X}_t \boldsymbol{\Gamma} \boldsymbol{f}_{t+1}). \tag{2}$$

The estimator can be interpreted as performing a principle component analysis on the (transformed) managed portfolios $(\boldsymbol{X}_t' \boldsymbol{X}_t)^{-1} \boldsymbol{X}_t' \boldsymbol{r}_{t+1}$, or the OLS "slope portfolios" from running a cross-sectional regression of the individual stock returns $\boldsymbol{r}_{t+1}$ on the conditional information matrix $\boldsymbol{X}_t$.

Summing up, expanding the information set beyond standard characteristics (e.g., market βs, valuation ratios, and technical signals, etc.) with alternative data such as text could help capture more variation in the conditional expected returns and covariances. As a result, managed portfolios—and their transformations or principle components—are more likely to achieve conditional mean-variance efficiency in theory. Empirically, the Kelly et al. (2019) IPCA estimator can be used to identify the factors derived from the managed portfolios.

2.1 Encoding the Text Information

Enriching the firm-specific information with text means working with vectors of the form

$$\boldsymbol{x}_{i,t} = [\text{characteristic}_{i,t}, \text{text}_{i,t}, \ldots],$$

for stock i at time t. It is straightforward to put firm characteristics into these vectors, but not so for text. Ideally, for any given text document $\mathcal{D}$, we would like to have a mapping $\mathcal{D} \mapsto \mathbb{R}^K$, from the document to a vector of dimension K. This is a process called *encoding*. In this section, we lay out our text encoding method for performing conditional asset pricing.

The simplest way of encoding is perhaps counting words: for a dictionary of W words, we have a W-dimension vector $\boldsymbol{d} = [d_1, \ldots, d_W]'$, where d_j is the count of word j in document $\mathcal{D}$. The dimension of $\boldsymbol{d}$ tends to be large (the same as the dictionary size), especially when bigrams, trigrams, or more sophisticated combinations of words are included in the dictionary.

To reduce the encoding dimension, we consider the following probabilistic specification: for firm i's word document observed at time t, its word count vector $\boldsymbol{d}_{i,t}$ is generated from a multinomial distribution

$$\boldsymbol{d}_{i,t} \sim \text{Multinomial}\left(L_{i,t},\, \sum_{k=1}^{K} \phi_{i,t}^{(k)} \boldsymbol{\theta}_k\right), \tag{3}$$

where $L_{i,t} = \mathbf{1}'\boldsymbol{d}_{i,t}$ is the sum of word counts—the document length if we put only unigrams into the dictionary—in document $\mathcal{D}_{i,t}$. The vectors $\{\boldsymbol{\theta}_k\}_{k=1}^K$ denote different probability distributions over the dictionary, which are sometimes regarded as common *topics* shared among the collection of documents under study. Each topic represents a group of semantically related words that capture latent structures of the text data. For the text information of firm i observed at time t, its *loading* on the k-th topic is $\phi_{i,t}^{(k)}$. In our applications, we aim to set $K \ll W$ and encode the document using the topic loading vector $\boldsymbol{\phi}_{i,t} = [\phi_{i,t}^{(1)}, \ldots, \phi_{i,t}^{(K)}]'$.

Summing up, instead of representing a document with a large vocabulary of words, we reduce the document to a low-dimensional vector of topics. A text document $\mathcal{D}_{i,t}$ for firm i observed at time t is first mapped to a sparse word count vector $\boldsymbol{d}_{i,t}$ and then to a low-dimensional vector $\boldsymbol{\phi}_{i,t}$. The elements in $\boldsymbol{\phi}_{i,t}$ measure the likelihood that document $\mathcal{D}_{i,t}$ is related to different topics.

Estimating the loadings $\boldsymbol{\phi}_{i,t}$ and the topic vectors $\{\boldsymbol{\theta}_k\}_{k=1}^K$ under equation (3) is as if performing a factor analysis of count data, because equation (3) implies

$$\mathbb{E}[\boldsymbol{d}_{i,t}] = \sum_{k=1}^{K} \left(L_{i,t}\phi_{i,t}^{(k)}\right) \boldsymbol{\theta}_k,$$

where the topic vectors serve the role of "common" factors, and the product terms in the parenthesis are factor loadings. We adopt the well-known Bayesian approach of Blei et al. (2003) by assigning Dirichlet priors on $\boldsymbol{\phi}_{i,t}$ and performing variational inference. This implementation is also known as the latent Dirichlet allocation (LDA). LDA is commonly known as a topic modelling tool, but it can also be thought of as a form of dimensionality reduction by representing documents as mixtures of topics. The representation retains the main themes of the documents and removes noise from irrelevant or rare words unrelated to common topics.

We use the perplexity score to select the number of LDA topics. The traditional perplexity score, as described by Blei et al. (2003), is

$$\text{Perplexity} = \exp\left\{-\frac{\sum_{i,t}\log p(\boldsymbol{d}_{i,t})}{\sum_{i,t} L_{i,t}}\right\}, \tag{4}$$

where $L_{i,t}$ is the number of words in the document $\boldsymbol{d}_{i,t}$, and $p(\boldsymbol{d}_{i,t})$ is the likelihood of observing $\boldsymbol{d}_{i,t}$ given the LDA model. Overall, a lower perplexity score indicates a better fit of the LDA model.

3 Data Description

We obtain the conference call transcripts between 2003 and 2022 from Refinitiv StreetEvents, together with the exact date of the conference call and the information about the conference speaker. The conference call typically consists of presentation and question-and-answer (Q&A) sessions. In the presentation session, the executives disclose information that they want to emphasize. The presentation session will be followed by a Q&A session in which the outside investors can directly communicate with the firm managers. Figure 3 displays examples of the presentation and Q&A sessions. Before the formal LDA estimation, we need to preprocess the text data. Details about the data cleaning are presented in Appendix A.

We next collect the data of individual stock returns from the Global Factor Data provided by Jensen et al. (2023). We eliminate the stocks whose market caps are below the 1st percentile of NYSE stocks. As explained in Section 2, we estimate the IPCA model in Kelly et al. (2019) using different sets of cross-sectional information as instruments. One important source of instruments is the collection of firm characteristics studied by Kelly et al. (2019).

Since the sample used in Kelly et al. (2019) ends in May 2014, we reconstruct these firm characteristics and update the sample until December 2022, built on the Global Factor Data provided by Jensen et al. (2023). The list of 35 firm characteristics that we finally use in the empirical analyses is shown in Table A1, with the perfectly matched characteristics in Panel A and the additional close substitutes in Panel B.

After we merge the conference call data with the sample of individual stocks, we obtain a sample of 185,184 unique earnings calls between the years 2003 and 2022. Panel A of Table 1 reports the total number of firms, the number of firms holding at least one earnings conference call within the previous four quarters, and the corresponding coverage ratio. It can be clearly seen that the coverage ratios increase over time, from only 40% in December 2003 to higher than 70% in the past ten years. Panel B of Table 1 further presents the sample coverage for the large firms whose market caps are above the 50th percentile of NYSE stocks. Interestingly, large firms are more likely to hold earnings conference calls than small firms, with about 95% coverage rates in the recent subsample.

Finally, we use several aggregate economic variables. The sentiment indices in Baker and Wurgler (2006) and Huang et al. (2015) are downloaded from the authors' websites. Other macro variables are from the FRED-MD dataset constructed by McCracken and Ng (2016).

4 Topics in Conference Calls

In this section, we summarize the LDA estimation results in both the Q&A and presentation sessions of conference calls. Specifically, we conduct the estimation using the "gensim" Python package, which requires us to choose both the number of topics and the random seed that initializes the model parameters.

To determine the number of topics, we rely on the perplexity score defined in equation (4). We point out that the estimation results based on different random seeds could be different; hence, we repeat the LDA estimation ten times (that is, using ten different random seeds) and report the average log perplexity scores for each number of topics. Figure 1 plots the average log perplexity scores for different numbers of topics (20, 30, 40, 50, and 60). We can clearly see that the perplexity score is the lowest when the number of topics is 40 (or 50) in

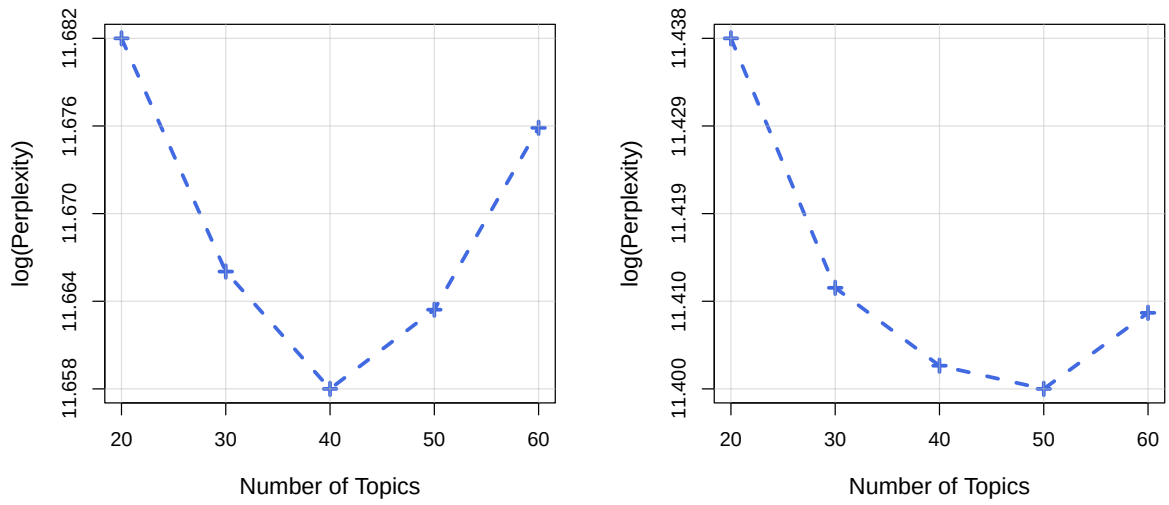
Table 1: Sample Description

	Panel A. Full sample			**Panel B.** Above 50th NYSE Firm Size		
Year	# of firms	# of firms holding conference calls	Coverage ratio	# of firms	# of firms holding conference calls	Coverage ratio
2003	3978	1573	0.40	959	590	0.62
2004	3777	1674	0.44	929	585	0.63
2005	3910	1843	0.47	929	625	0.67
2006	3807	1969	0.52	896	626	0.70
2007	3734	2041	0.55	874	641	0.73
2008	3660	2127	0.58	888	673	0.76
2009	3351	2113	0.63	853	664	0.78
2010	3216	2117	0.66	863	688	0.80
2011	3273	2225	0.68	871	727	0.83
2012	3117	2214	0.71	874	753	0.86
2013	3155	2183	0.69	911	796	0.87
2014	3221	2304	0.72	967	844	0.87
2015	3391	2450	0.72	995	887	0.89
2016	3101	2331	0.75	949	871	0.92
2017	3113	2398	0.77	952	884	0.93
2018	3189	2462	0.77	980	925	0.94
2019	3204	2474	0.77	999	952	0.95
2020	3374	2553	0.76	1090	1035	0.95
2021	3739	2720	0.73	1247	1172	0.94
2022	3674	2734	0.74	1186	1135	0.96

The table describes the sample of firms holding conference calls. We incorporate all CRSP firms listed in the US and on the NYSE, AMEX, or NASDAQ that have CRSP share codes of either 10 or 11. The firms with market caps below the 1st percentile of NYSE stocks are eliminated from our analysis. In Panel A, we report in the table the total number of firms, the number of firms holding at least one earnings conference call within the previous four quarters, and the corresponding coverage ratio in the last month of each year. We repeat a similar analysis in Panel B for the subsample of firms with market caps above the 50th percentile of NYSE stocks.

the Q&A (or the presentation) session. Therefore, we choose 40 and 50 topics in Q&A and presentation sessions in the LDA estimation, respectively.

We now exemplify how we handle the issue of initial random states in the LDA estimation. It is well known that the variational Bayesian method (which is adopted in the "gensim" package) heavily relies on the random starting values. To ensure that our empirical results are not driven by a particular random seed, we repeat the LDA estimation for 10 random seeds and figure out the stable topics that appear in most of the estimates across initial random states. We demonstrate the intuitions behind the algorithm for finding stable topics in the Q&A session, but the same logic is applied to the presentation session, with the details presented in Appendix B. Starting with the 400 topics from ten LDA estimates (40 Q&A topics per random seed), we combine the ones with high correlations into a single topic.



(a) Q&A topics **(b)** Presentation topics

Figure 1: Log Perplexity Scores of LDA Model for Different Numbers of Topics

The figure plots the log perplexity scores of the LDA model for different numbers of topics in Q&A (see Panel (a)) and presentation (see Panel (b)) sessions of conference calls.

Moreover, we keep the topics only if they have been merged at least five times and call them stable topics. For instance, if the topic "Capital Structure" is essential in the Q&A session, it should emerge as a dominant common theme even if different initial states are used in the estimation, and the estimates of its probability distributions over terms should be highly correlated across random seeds.[6]

Using the algorithm in Appendix B, we obtain 32 stable topics in both Q&A and presentation sessions. Table A2 presents the topic names, the average loadings of firms' conference call texts on these topics, and the key terms in each topic based on the full-sample LDA estimation. Since the LDA estimation is entirely unsupervised, we manually assign a label to each topic based on our understanding of the key terms. Several observations are noteworthy.

First, most of the topics have clear meanings and interpretability. For example, one of

[6]The past literature has also acknowledged that the LDA estimation is subject to the initial random states (see the Online Technical Appendix of Hansen et al. (2018)). Specifically, Hansen et al. (2018) compute the perplexity scores across five random states and select the estimates with the lowest perplexity score. However, as pointed out in their paper and confirmed by our empirical analysis, perplexity scores are rather similar across random seeds; hence, selecting one particular random seed is subject to high model uncertainty and is likely to omit important topics. Unlike their paper, we consider the estimates of topics across 10 random seeds and propose an algorithm (see Appendix B) to find stable topics.



(a) Q&A topics



(b) Presentation topics

Figure 2: Common and Specific Topics in Q&A and Presentation Sessions

The figure plots the common (red fonts) and specific (green fonts) topics in Q&A and presentation sessions. A topic's font size corresponds to the average loadings of firms' conference call texts on these topics.

the important topics in the Q&A session is "Cloud Service", in which the key terms are

> *"cloud, subscription, billings, recurring revenue, gotomarket, automation, ecosystem, subscription revenue, intelligence, workflow, hardware, hybrid, professional services."*

The above list of words is unmixed, and most of the terms are closely related to the topic "Cloud Service".

Second, the Q&A and presentation sessions contain both heterogeneous and common information. Figure 2 displays the common (highlighted in red) and specific (highlighted in green) topics in Q&A and presentation sessions, with the font sizes corresponding to the average loadings of firms' conference call texts on these topics.[7] While "Forecasting and Speculation", "Economic Uncertainty", and "Capital Structure" are important topics in the Q&A sessions, the presentation session is dominated by topics such as "Organizational Management", "Profits", and "Business Operation". These topics are labelled as Q&A-specific or presentation-specific topics. Some topics are important in both sessions, such as "Inflation", "Pharmaceutics", and "Banking and Lending", and we call them common topics. Overall, there are 13 Q&A/presentation-specific and 19 common topics.

[7] We define the common topics based on the correlation between the probability distributions over terms of two Q&A and presentation topics. In particular, we require two common topics to have a correlation coefficient of no less than 0.7. We further check whether the two topics with a high correlation share common information by reading their key terms.

Third, the firms' loadings on the LDA topics are able to capture interpretable cross-sectional information. Figure 3 gives such examples. In Panel (a), we show a small piece of the Q&A session of Broadwing Communications in 2002 Q4. Through the conversation between the analyst and the CFO, we infer that "Capital Structure" is the most important topic, and the LDA model is indeed able to capture the loadings on these common themes. Panel (b) of Figure 3 presents another example of the presentation session, in which the COO of Dun & Bradstreet Corporation heavily discusses key terms such as leadership and team members; hence, the LDA model assigns the largest weight to the "Organizational Management" topic. Overall, the LDA model successfully translates the common themes that managers and analysts are the most concerned about into simple quantitative measures—the topic loadings.

Fourth, many topics (e.g., "Defense", "Health Care", "Pharmaceutics", "Steel", etc.) represent industry-specific exposures. Unlike topics "Economic Uncertainty" or "Macro Environment" that possibly affect all the firms, most industry-specific loadings drive only a relatively small fraction of companies. Therefore, using the topic loadings as conditional instruments in the IPCA model implies that a large number of latent factors are possibly needed to span the mean-variance efficient frontier.

5 Asset Pricing Performance

In this section, we investigate whether the firms' loadings on Q&A or presentation topics that we uncover in Section 4 carry useful information for asset pricing. In particular, employing the IPCA model described in Section 2, we study whether these topic loadings can explain the time-series variations in asset returns, as well as whether they can improve the mean-variance efficiency. We also conduct empirical analyses to explore whether the Q&A and presentation sessions capture investor sentiment and understand the relative importance of Q&A and presentation topics. We finally show how to construct real-time portfolios based on the online LDA analysis of the conference call transcripts.

Broadwing Communications, 2002Q4 Earnings call	
Bardone, Analyst (Question)	Hey, guys, congratulations on securing the restructuring. Just wanted to ask three quick questions. First, I think Tom, you were talking about some deleveraging targets through 2005, which looked pretty substantial. Could you be more specific about how the deleveraging is going to take place? Are you talking about operational cash flows being used to buy back debt early in the public markets, or are you talking about potentially an equity deal here and whether you've made commitments to the banks on that front? The second question I have was on the Oak Hill paper...
Schilling, CFO (Answer)	Hey, David. Tom. I'll take the first couple questions then probably turn it over to Jack for the last one. On the senior debt deleveraging I was talking about I was specifically talking about our senior debt which is our senior secured debt which is our bank credit facility and our capital leases, as well as the 7% notes. And that -- the deleveraging there is coming from the junior capital we're raising arranged by Goldman Sachs of $350m, along with the operational cash flows during the next three years that will be generated from the business. On the convert terms of Oak Hill...
Topic Loadings: Capital Structure, 0.403 (Highest)	

(a) Q&A example

Dun & Bradstreet Corp, 2004Q3 Earnings Call	
Alesio, COO	... Our team members' focused leadership drove our risk management results in this quarter ... Let me show an example of how our team members' leadership benefited our e-business results in the quarter. In this other example, we have a customer which is a large company in the high tech industry that has used D&B Solutions for many years. In some cases they needed access to deeper and timelier information than their D&B current solution offered. And they wanted to offer access to this business insight more broadly through their organization. So, members of our sales and marketing solutions team partnered with our Hoover's team to meet our customers' needs. Through teamwork and results leadership they showed how our customer could achieve deeper, more accurate, and more timely business insight by adding Hoover's information to their current D&B Solution. As a result, our customer purchased an enterprise wide license and a full subscription to Hoover's data. This is another great example of how our team members are leading with the interest of our customers first in order to drive superb results...
Topic Loadings: Organizational Management, 0.270 (Highest)	

(b) Presentation example

Figure 3: Conference Call Examples in Q&A and Presentation Sessions

The figure displays two examples of the Q&A (Panel (a)) and presentation (Panel (b)) conference calls, respectively. In each example, we show both the texts and their loadings on the most important topics, with the key terms highlighted in red.

5.1 Time-Series Fit

We first explore the out-of-sample (OOS) time-series fit of the IPCA model using topic loadings and firm characteristics as instruments. To construct the OOS estimates, we conduct an expanding-window exercise. For each month τ, we use all the data through month τ to estimate the IPCA model and denote the estimated model parameters as $\hat{\boldsymbol{\Gamma}}_\tau$, which implies that the dynamic betas in month τ are $\hat{\boldsymbol{\beta}}_\tau = \boldsymbol{X}_\tau \hat{\boldsymbol{\Gamma}}_\tau$. Next, we construct the OOS latent factors as $\hat{\boldsymbol{f}}_{\tau+1} = \left(\hat{\boldsymbol{\Gamma}}_\tau^\top \boldsymbol{X}_\tau^\top \boldsymbol{X}_\tau \hat{\boldsymbol{\Gamma}}_\tau\right)^{-1} \hat{\boldsymbol{\Gamma}}_\tau^\top \boldsymbol{X}_\tau^\top \boldsymbol{r}_{\tau+1}$. Since the constructed latent factors are tradable, we can estimate their risk premia using the time-series averages: $\hat{\boldsymbol{\lambda}}_{f\tau} = \frac{1}{\tau} \sum_{t=0}^{\tau-1} \left(\hat{\boldsymbol{\Gamma}}_\tau^\top \boldsymbol{X}_t^\top \boldsymbol{X}_t \hat{\boldsymbol{\Gamma}}_\tau\right)^{-1} \hat{\boldsymbol{\Gamma}}_\tau^\top \boldsymbol{X}_t^\top \boldsymbol{r}_{t+1}$. Note that the estimation of $\hat{\boldsymbol{\Gamma}}_\tau$, $\hat{\boldsymbol{\lambda}}_{f\tau}$, and the individual stock return weights, $\left(\hat{\boldsymbol{\Gamma}}_\tau^\top \boldsymbol{X}_\tau^\top \boldsymbol{X}_\tau \hat{\boldsymbol{\Gamma}}_\tau\right)^{-1} \hat{\boldsymbol{\Gamma}}_\tau^\top \boldsymbol{X}_\tau^\top$, rely on only the information up to time τ; hence, they can be used to evaluate the OOS performance.

Following Kelly et al. (2019), we report both the total and predictive time-series R^2 for individual stock returns, defined as follows:

$$\text{total } R^2 = 1 - \frac{\sum_t (\boldsymbol{r}_{t+1} - \hat{\boldsymbol{\beta}}_t \hat{\boldsymbol{f}}_{t+1})^\top (\boldsymbol{r}_{t+1} - \hat{\boldsymbol{\beta}}_t \hat{\boldsymbol{f}}_{t+1})}{\sum_t \boldsymbol{r}_{t+1}^\top \boldsymbol{r}_{t+1}} \quad \text{and}$$

$$\text{predictive } R^2 = 1 - \frac{\sum_t (\boldsymbol{r}_{t+1} - \hat{\boldsymbol{\beta}}_t \hat{\boldsymbol{\lambda}}_{ft})^\top (\boldsymbol{r}_{t+1} - \hat{\boldsymbol{\beta}}_t \hat{\boldsymbol{\lambda}}_{ft})}{\sum_t \boldsymbol{r}_{t+1}^\top \boldsymbol{r}_{t+1}}.$$

Table 2 reports the OOS total and predictive R^2 of the IPCA model using as instruments the i) 32 Q&A topic loadings, ii) 32 presentation topic loadings, iii) 64 Q&A plus presentation topic loadings, iv) 35 firm characteristics, and v) both 64 topic loadings and 35 firm characteristics. The full sample ranges from January 2003 to December 2022. To ensure that the first in-sample has sufficient time-series information to estimate model parameters, we require the out-of-sample to start from January 2008. Panel A reports the time-series fit for individual stocks, whereas Panel B presents those for managed portfolios.[8] We consider

[8] Similarly, we can define the total and predictive time-series R^2 for managed portfolio returns ($\boldsymbol{z}_{t+1} = \boldsymbol{X}_t^\top \boldsymbol{r}_{t+1}$), as follows:

$$\text{total } R^2 = 1 - \frac{\sum_t (\boldsymbol{z}_{t+1} - \boldsymbol{X}_t^\top \hat{\boldsymbol{\beta}}_t \hat{\boldsymbol{f}}_{t+1})^\top (\boldsymbol{z}_{t+1} - \boldsymbol{X}_t^\top \hat{\boldsymbol{\beta}}_t \hat{\boldsymbol{f}}_{t+1})}{\sum_t \boldsymbol{z}_{t+1}^\top \boldsymbol{z}_{t+1}} \quad \text{and}$$

$$\text{pred } R^2 = 1 - \frac{\sum_t (\boldsymbol{z}_{t+1} - \boldsymbol{X}_t^\top \hat{\boldsymbol{\beta}}_t \hat{\boldsymbol{\lambda}}_{ft})^\top (\boldsymbol{z}_{t+1} - \boldsymbol{X}_t^\top \hat{\boldsymbol{\beta}}_t \hat{\boldsymbol{\lambda}}_{ft})}{\sum_t \boldsymbol{z}_{t+1}^\top \boldsymbol{z}_{t+1}}.$$

Table 2: Out-of-Sample Time-Series Fit

		$p = 5$	$p = 10$	$p = 15$	$p = 20$	$p = 25$	$p = 30$
		Panel A. Individual stocks					
32 Q&A topics	total R^2	21.28	22.15	22.50	22.78	23.03	23.27
	pred R^2	0.14	0.14	0.13	0.13	0.14	0.14
32 PRE topics	total R^2	21.69	22.58	22.98	23.26	23.53	23.75
	pred R^2	0.14	0.13	0.14	0.14	0.14	0.14
64 Q&A + PRE topics	total R^2	22.05	22.99	23.41	23.72	23.99	24.24
	pred R^2	0.13	0.12	0.12	0.13	0.13	0.13
35 Firm Characteristics	total R^2	24.00	25.13	25.79	26.30	26.71	27.04
	pred R^2	0.14	0.13	0.14	0.15	0.15	0.14
64 Q&A + PRE topics + 35 Firm Chars	total R^2	24.46	25.86	26.58	27.03	27.42	27.76
	pred R^2	0.09	0.12	0.15	0.15	0.15	0.16
		Panel B. Managed portfolios					
32 Q&A topics	total R^2	99.07	99.69	99.81	99.87	99.92	99.97
	pred R^2	0.93	0.93	0.92	0.93	0.93	0.93
32 PRE topics	total R^2	99.13	99.68	99.80	99.87	99.93	99.98
	pred R^2	0.94	0.92	0.93	0.93	0.92	0.92
64 Q&A + PRE topics	total R^2	98.23	99.25	99.45	99.55	99.62	99.69
	pred R^2	0.86	0.83	0.84	0.85	0.86	0.86
35 Firm Characteristics	total R^2	98.55	99.33	99.63	99.77	99.86	99.93
	pred R^2	0.86	0.84	0.84	0.85	0.84	0.83
64 Q&A + PRE topics + 35 Firm Chars	total R^2	96.29	98.16	98.75	98.96	99.11	99.23
	pred R^2	0.60	0.68	0.72	0.71	0.70	0.71

The table reports the out-of-sample time-series fit of the IPCA model using as instruments the i) 32 Q&A topic loadings, ii) 32 presentation (PRE) topic loadings, iii) 64 Q&A plus presentation topic loadings, iv) 35 firm characteristics, and v) both 64 topic loadings and 35 firm characteristics. The Q&A and presentation topics come from the full-sample LDA estimation. The number of latent factors in the IPCA estimation, denoted by p, equals 5, 10, 15, 20, 25, and 30. The total and predictive R^2 are calculated in the cross-section of individual stock returns in Panel A and managed portfolios in Panel B. The out-of-sample ranges from January 2008 to December 2022.

different numbers of latent factors (5, 10, 15, 20, 25, and 30). Note that the OOS fit is not mechanically inflated when more latent factors are used in the estimation.

We observe that both Q&A and presentation topics carry invaluable information in explaining time-series variations in asset returns. For instance, when the number of latent factors is 25, the total and predictive R^2 in the IPCA models using both Q&A and presentation topic loadings as instruments are 24% and 0.13%, almost comparable to the model based on the traditional firm characteristics (that delivers 27% total and 0.15% predictive R^2). However, combining the textual information embedded in the conference calls with firm characteristics does not significantly enhance the time-series fit for stock returns.

5.2 Mean-Variance Efficiency

We further examine whether the firm-specific topic loadings extracted from the conference call transcripts improve the mean-variance efficiency. To do so, we construct the OOS MVE portfolios implied by the IPCA models. As described in Section 5.1, we conduct an expanding window analysis. In each month τ, we estimate $\hat{\mathbf{\Gamma}}_\tau$ and $\hat{\boldsymbol{f}}_{\tau+1}$, as well as the sample mean and covariance matrix of $\hat{\boldsymbol{f}}$, denoted as $\hat{\boldsymbol{\lambda}}_{f\tau}$ and $\hat{\mathbf{\Sigma}}_{f\tau}$, respectively. The mean-variance efficient (MVE) portfolio is defined as

$$\hat{\boldsymbol{f}}_{\tau+1}^{MVE} = \hat{\boldsymbol{\lambda}}_{f\tau}^{\top} \hat{\mathbf{\Sigma}}_{f\tau}^{-1} \hat{\boldsymbol{f}}_{\tau+1}. \tag{5}$$

Note that the MVE portfolio in equation (5) is conditional on different cross-sectional information sets (e.g., Q&A and presentation topic loadings, 35 firm characteristics, as well as their combinations); hence, it is more precise to call them conditional MVE portfolios.

Table 3 reports the annualized OOS Sharpe ratios of the MVE portfolios implied by several IPCA models. The OOS period ranges from 2008 to 2022. When we use Q&A or presentation topic loadings as the cross-sectional conditional information (see the first two rows of Table 3), we need more than 20 latent factors to achieve the nearly optimal OOS Sharpe ratio. For example, when the number of latent factors is 25, the OOS Sharpe ratios of the MVE portfolios conditional on Q&A or presentation topic loadings are 1.20 or 1.06, respectively, compared to only 0.47 or 0.86 in the five-factor IPCA models. Using both these two sets of topic loadings can further increase the OOS Sharpe ratio to 1.37 when $p = 25$, or even 1.64 when $p = 30$ (see row 3 of Table 3). The lack of sparsity observed in Table 3 is possibly due to the fact that many topics represent industry-specific exposures, as we point out in Section 4. Hence, finding a sparse latent factor model conditional on many industry exposures is unrealistic. In the following analyses, we focus on the empirical results based on the IPCA models with 25 or 30 latent factors.

We also categorize the topics into common and specific topics (see Figure 2). To quantify the relative importance of the Q&A- and presentation-specific topics, we report in Table 4 the OOS Sharpe ratios of the MVE portfolios conditional on i) 19 Q&A common topic loadings plus 32 presentation topic loadings (that is, drop 13 Q&A-specific topics) and ii) 19 presenta-

tion common topic loadings plus 32 Q&A topic loadings (that is, drop 13 presentation-specific topics). It can be clearly seen that removing either Q&A or presentation-specific topics from the cross-sectional information set will lead to deteriorating OOS Sharpe ratios. Hence, both Q&A and presentation sessions of conference calls contain common and specific information that is useful for building MVE portfolios.

Moreover, we confirm that the IPCA model using the 35 firm characteristics as instruments has satisfactory performance, with an OOS Sharpe ratio higher than 1.7 when $p \geq 20$ (see row 4 of Table 3). More importantly, combining the textual information with the firm characteristics can further improve the mean-variance efficiency: When $p = 25$, the OOS Sharpe ratio of the MVE portfolio conditional on 64 conference call topics and 35 firm characteristics is 2.67 (see row 5 of Table 3).

Table 3: Out-of-Sample Sharpe Ratios of MVE Portfolios

Cross-Sectional Information Sets:	$p = 5$	$p = 10$	$p = 15$	$p = 20$	$p = 25$	$p = 30$
32 Q&A topics	0.47	0.75	0.51	0.77	1.20	1.10
32 PRE topics	0.86	0.99	0.93	1.14	1.06	1.16
64 Q&A + PRE topics	0.63	0.70	0.99	1.26	1.37	1.64
35 Firm Characteristics	0.71	1.24	1.59	1.74	1.79	1.71
64 Q&A + PRE topics + 35 Firm Chars	0.23	1.52	2.27	2.37	2.67	2.76

The table reports the annualized out-of-sample Sharpe ratios of the MVE portfolios implied by the IPCA models using as instruments the i) 32 Q&A topic loadings, ii) 32 presentation (PRE) topic loadings, iii) 64 Q&A plus presentation topic loadings, iv) 35 firm characteristics, and v) both 64 topic loadings and 35 firm characteristics. The Q&A and presentation topics come from the full-sample LDA estimation. The number of latent factors in the IPCA estimation, denoted by p, equals 5, 10, 15, 20, 25, and 30. The out-of-sample ranges from January 2008 to December 2022.

One may be concerned that the 2008 global financial crisis drives our results. To relieve this concern, we repeat the same analyses in Table 3 using a different OOS period — January 2013 to December 2022 — and report the results in Table A3. We affirm that the textual information embedded in conference call scripts is able to provide incremental cross-sectional conditional information beyond the traditional firm characteristics.

But how does the textual information enhance the mean-variance efficiency? Note that the conditional MVE portfolio of the entire space of individual stocks is always defined as $\boldsymbol{\Sigma}_t^{-1}\mathbb{E}_t[\boldsymbol{r}_{t+1}]$. Therefore, the topic loadings in the conference call scripts either provide the incremental ability to predict stock returns, which relates to the expected returns $\mathbb{E}_t[\boldsymbol{r}_{t+1}]$,

Table 4: Common vs Specific Topics

	$p = 5$	$p = 10$	$p = 15$	$p = 20$	$p = 25$	$p = 30$
Drop 13 Q&A specific topics	0.74	0.75	0.92	1.12	1.19	1.21
(19 Q&A common topics plus 32 presentation topics)						
Drop 13 PRE specific topics	0.42	0.55	0.56	0.84	0.90	1.32
(19 presentation common topics plus 32 Q&A topics)						
64 Q&A + PRE topics	0.63	0.70	0.99	1.26	1.37	1.64

The table reports the annualized out-of-sample Sharpe ratio of the MVE portfolios implied by the IPCA model using as instruments the i) 19 Q&A common topic loadings plus 32 presentation topic loadings (see the row "Drop 13 Q&A specific topics") and ii) 19 presentation common topic loadings plus 32 Q&A topic loadings (see the row "Drop 13 PRE specific topics"). As a comparison, we also display the estimation results based on all the 64 Q&A plus presentation topic loadings. The Q&A and presentation topics come from the full-sample LDA estimation. The number of latent factors in the IPCA estimation, denoted by p, equals 5, 10, 15, 20, 25, and 30. The out-of-sample ranges from January 2008 to December 2022.

or help capture better the conditional covariance matrix $\mathbf{\Sigma}_t$.

Nevertheless, the first channel is almost impossible since Table 2 shows that including the textual information in the IPCA model cannot increase the predictive R^2 compared to the model composed of only 35 firm characteristics. In other words, the improvement in the OOS Sharpe ratio originates mainly from the better proxy for the conditional covariance term $\mathbf{\Sigma}_t$. The second channel is more likely to be true when the information embedded in conference call transcripts is significantly different from firm characteristics. We confirm this point in Table 5: The MVE portfolios conditional on Q&A or presentation topic loadings and firm characteristics are weakly correlated at best, which implies that the conference calls indeed provide rather heterogeneous cross-sectional information. Our observation also echoes the intuition in Kozak and Nagel (2024): Additional characteristics can help even if they are unrelated to expected returns but help capture major sources of the covariance term.

Table 5: Correlation Coefficients among MVE Portfolios: 32 Q&A topics, 32 Presentation Topics, and 35 Firm Characteristics

	$p = 25$			$p = 30$		
	32 Q&A topics	32 PRE topics	35 Firm Chars	32 Q&A topics	32 PRE topics	35 Firm Chars
32 Q&A topics	1.00	0.16	0.23	1.00	0.09	0.24
32 PRE topics	0.16	1.00	-0.07	0.09	1.00	0.03
35 Firm Chars	0.23	-0.07	1.00	0.24	0.03	1.00

The table reports the correlation coefficients among the MVE portfolios implied by the IPCA model using as instruments the i) 32 Q&A topic loadings, ii) 32 presentation (PRE) topic loadings, and iii) 35 firm characteristics. The Q&A and presentation topics come from the full-sample LDA estimation. The number of latent factors in the IPCA estimation, denoted by p, equals 25 and 30.

As a final exercise in this subsection, we test whether the traditional factor models can explain our text-managed MVE portfolios. We consider five benchmark factor models: i) Fama and French (1993) three-factor model (FF3), ii) Fama and French (2015) five-factor model (FF5), iii) five Fama-French factors plus momentum (FFC6), iv) Hou et al. (2015) four-factor model (Q4), and v) the characteristics-managed MVE portfolio (p equals 25 or 30 in the IPCA model). We regress the text-managed MVE portfolio returns, which are conditional on either 32 Q&A or 32 presentation topic loadings, on the five benchmark models and report the estimates of common intercepts, as well as their t-statistics based on the Newey-West standard errors with 12 lags (see Newey and West (1987)). We show in Table 6 that none of the five benchmark models can explain the text-managed MVE portfolios, which always have significantly positive alphas of larger than 1% per month.[9] Therefore, none of these models can explain the OOS performance of either the Q&A- or presentation-managed MVE portfolios.

Table 6: Can Canonical Factor Models Explain Text-Managed MVE Portfolios?

	$p = 25$		$p = 30$	
	Q&A-managed	PRE-managed	Q&A-managed	PRE-managed
FF3	1.4%	1.3%	1.4%	1.4%
t-stat	(3.300)	(2.974)	(3.038)	(3.866)
FF5	1.2%	1.3%	1.1%	1.5%
t-stat	(2.958)	(3.184)	(2.715)	(4.552)
FFC6	1.2%	1.3%	1.1%	1.5%
t-stat	(3.000)	(3.587)	(2.740)	(4.943)
Q4	1.2%	1.2%	1.1%	1.4%
t-stat	(3.179)	(2.943)	(2.806)	(4.034)
CHAR-managed	1.1%	1.7%	1.0%	1.6%
t-stat	(2.453)	(3.236)	(2.118)	(3.487)

The table reports the estimates of alphas in the text-managed MVE portfolios conditional on either 32 Q&A or 32 presentation topic loadings. The Q&A and presentation topics come from the full-sample LDA estimation. The number of latent factors in constructing the MVE portfolios is either 25 or 30. We run the univariate time-series regression by regressing the text-managed MVE portfolio returns on several factor models: i) Fama and French (1993) three-factor model (FF3), ii) Fama and French (2015) five-factor model (FF5), iii) five Fama-French factors plus momentum (FFC6), iv) Hou et al. (2015) four-factor model (Q4), and v) the MVE portfolio conditional on 35 firm characteristics (based on the IPCA model, denoted as CHAR-based). We report the estimates of the intercepts (alphas) and the t-statistics based on Newey-West standard errors with 12 lags. The sample ranges from January 2008 to December 2022.

[9] We normalize the text-managed MVE portfolio returns to have the same volatility as the market portfolio (4.9% per month) in the sample between 2008 and 2022.

5.3 Text-Managed Portfolios and Investor Sentiment

In this subsection, we investigate whether the text-managed MVE portfolio returns are driven by investor sentiment. According to the previous literature (e.g., Baker and Wurgler (2006)), sentiment should have stronger effects on the stocks that are harder to arbitrage or more difficult to value. The underlying explanation is that arbitrageurs find these stocks more costly and risky to trade due to the potentially higher degree of noise trader risk. Therefore, following high sentiment states, the future returns on these stocks tend to experience more negative drops than others that are relatively easier to value and arbitrage.

We run the following predictive regression to study whether sentiment predicts text-managed MVE portfolio returns:

$$r_{t+1}^{\text{text}} = a + b \times \text{sentiment}_t + \text{controls}_{t+1} + \text{error}_{t+1}, \tag{6}$$

where r_{t+1}^{text} is either the Q&A- or presentation-managed MVE portfolio return, sentiment_t is the one-month lagged sentiment, and controls_{t+1} are the contemporaneous controlled factors (e.g., FF5). We consider two sentiment measures from Baker and Wurgler (2006) and Huang et al. (2015), respectively. Both two papers provide the total sentiment index, along with the measure orthogonal to several macro variables.

Table 7 shows that only the Q&A-managed MVE portfolio is positively predicted by the one-month lagged sentiment indices, and the results are robust to the additional controlled factors. Therefore, the cross-sectional information in Q&A sessions tends to tilt the MVE portfolio towards the stocks that are relatively less costly to arbitrage and relatively easier to value. The positive coefficient estimates are consistent with Stambaugh et al. (2012), in which they also find long-short strategies of many anomalies become more profitable following high sentiment. On the contrary, in the same predictive regression of the presentation-managed MVE portfolio, the coefficient estimates of sentiment are insignificant. This stark contrast provides support for the unique feature of the Q&A session—for example, the interaction between managers and investors—that enables us to capture invaluable conditional information related to investor sentiment.

Table 7: Investor Sentiment and Text-Managed MVE Portfolios

	Q&A Managed MVE Portfolio				Presentation Managed MVE Portfolio			
	BW (2006)		HJTZ (2015)		BW (2006)		HJTZ (2015)	
	Total	Orthogonal	Total	Orthogonal	Total	Orthogonal	Total	Orthogonal
	Panel A. Univariate regressions							
b	0.245	0.205	0.210	0.102	0.061	0.060	-0.027	-0.108
t-stat	(8.437)	(4.912)	(4.313)	(1.276)	(1.227)	(1.083)	(-0.489)	(-1.215)
R^2	5.49%	3.67%	3.86%	0.48%	-0.19%	-0.20%	-0.49%	0.62%
	Panel B. Control for FF5 plus Q&A- / PRE-managed MVE portfolios							
b	0.234	0.199	0.204	0.119	0.075	0.073	-0.025	-0.102
t-stat	(6.594)	(5.827)	(4.461)	(1.936)	(1.213)	(1.110)	(-0.347)	(-1.156)
R^2	18.83%	17.51%	17.70%	15.12%	3.25%	3.25%	2.82%	3.74%

The table reports the regression results in equation (6). Two sentiment measures come from Baker and Wurgler (2006) (BW (2006)) and Huang et al. (2015) (HJTZ (2015)), which provide the total sentiment index, along with the measure orthogonal to several macro variables. In Panel A, we consider the univariate predictive regressions. In Panel B, we add as controlled factors the Fama-French five factors and the Q&A-/presentation-managed MVE portfolios. For example, when using sentiment to predict Q&A-managed MVE portfolio returns, we use presentation-managed MVE portfolio return as a control. We report the point estimates of the coefficient of sentiment, b, its t-statistic based on the Newey and West (1987) standard errors with 12 lags, and the adjusted R^2. The Q&A and presentation topics come from the full-sample LDA estimation. The number of latent factors in the IPCA estimation is 25. The sample period ranges from January 2008 to December 2022.

5.4 Which Topics Are Important?

To answer this question, we adopt the leave-one-out procedure. We first study the relative importance of each topic separately in either Q&A or presentation sessions. Take the Q&A topics, for example. In the IPCA estimation, we drop one Q&A topic and use the remaining 31 Q&A topics to construct the conditional MVE portfolios. If a topic is particularly salient for the mean-variance efficiency, we will detect a significant drop in the OOS Sharpe ratio. We repeat the same process 32 times and quantify the relative topic importance in either the Q&A or presentation sessions.

Table 8 reports the results based on the IPCA models with 30 latent factors. We report a topic if eliminating this topic leads to a drop in the OOS Sharpe ratio no less than 0.10.[10] In the Q&A session, excluding "Cloud Service", "Capital Structure", or "Human Capital" leads

[10]Note that dropping some topics can occasionally lead to even higher OOS Sharpe ratios. Unlike the in-sample measure, the OOS Sharpe ratio does not mechanically increase when a new instrument is included in the model.

to the largest drop in the annualized OOS Sharpe ratio (from -0.53 to -0.17) of the Q&A-managed MVE portfolio. In contrast, the performance of the presentation-managed MVE portfolio is the most sensitive to the topics "Business Operation", "Steel", "Semiconductor and Electronics", "Asset Management", and "Energy".

Table 8: Topic Importance I

32 Q&A topics, excluding:	SR_{oos}	ΔSR_{oos}	32 PRE topics, excluding:	SR_{oos}	ΔSR_{oos}
	1.10			1.16	
Panel A. General topics					
Capital Structure	0.72	-0.38	Business Operation	1.05	-0.11
Human Capital	0.93	-0.17			
Panel B. Industry					
Cloud Service	0.57	-0.53	Steel	0.99	-0.17
			Semiconductor	1.03	-0.13
			Asset Management	1.05	-0.11
			Energy	1.05	-0.11

The table reports the relative topic importance in Q&A or presentation. We sequentially exclude one of the 32 Q&A or presentation topics in the IPCA estimation and report the corresponding annualized Sharpe ratio of the out-of-sample MVE portfolio conditional on the remaining 31 Q&A or presentation topics (see column SR_{oos}), as well as the drop in the Sharpe ratio relative to the benchmark analysis in Table 3 (see column ΔSR_{oos}). The Q&A and presentation topics come from the full-sample LDA estimation. The number of latent factors in the IPCA estimation is 30. The out-of-sample ranges from January 2008 to December 2022.

Furthermore, we explore the relative importance of all 64 Q&A and presentation topics jointly. As shown in Figure 2, there are 19 common topics in Q&A and presentation sessions whose word distributions are rather homogeneous. Instead of excluding one common topic per time, we eliminate two common topics in Q&A and presentation simultaneously (e.g., two "Cloud Service" topics) and construct the conditional MVE portfolios using the remaining 62 topics. For the 13 Q&A-specific and 13 presentation-specific topics, we still use the leave-one-out procedure as in Table 8. The results are presented in Table 9. We show that when examining the 64 topics jointly, "Capital Structure", "Cloud Service", and "Business Operation" are still the top three most important topics, consistent with Table 8. One may be concerned that "Cloud Service" mainly relates to industry-specific exposures and should not drive the entire cross-section of stock returns. However, we find that most firms that have significant loadings on the "Cloud Service" topic belong to the "Business Service" industry,

which is the largest among all 48 Fama-French industries. Moreover, the "Business Service" industry mainly consists of firms that provide professional services to all the other industries, hence closely connected to the entire cross-section of listed firms.

Table 9: Topic Importance II

64 Q&A + PRE topics, excluding:	SR_{oos}	ΔSR_{oos}
	1.64	
Panel A. General topics		
Capital Structure (Q&A-specific)	1.20	-0.44
Business Operation (PRE-specific)	1.47	-0.17
Organizational Management (PRE-specific)	1.52	-0.12
Panel B. Industry		
Cloud Service (Common)	1.40	-0.24
Semiconductor (Common)	1.54	-0.10
Steel (PRE-specific)	1.54	-0.10

The table reports the topic importance in conference calls. For the Q&A-specific and presentation-specific topics, we sequentially exclude one of them. For the common topics in both sessions, we exclude two topics (e.g., two "Cloud Service" topics in both Q&A and presentation) from the IPCA estimation. We next report in column SR_{oos} the annualized Sharpe ratio of the out-of-sample MVE portfolio conditional on the remaining 62 (excluding two common topics) or 63 topics (excluding one specific topic), as well as the drop in the Sharpe ratio relative to the benchmark analysis in Table 3 (see column ΔSR_{oos}). The Q&A and presentation topics come from the full-sample LDA estimation. The number of latent factors in the IPCA estimation is 30. The out-of-sample ranges from January 2008 to December 2022.

5.5 Real-Time Strategies with Online LDA

The full-sample LDA is likely subject to the critique that the estimates of probability distributions over terms, $\mathbf{\Theta}$, rely on the full-sample information and hence, the estimates of firm i's topic loadings in month t, $\boldsymbol{\phi}_{it}$, is also susceptible to the lookahead bias. Note that $\boldsymbol{\phi}_{it}$ is directly from the time-t conference call transcript of firm i, so the only source of forward-looking bias originates from $\mathbf{\Theta}$. In fact, the lookahead bias is a general issue in pretrained language models (e.g., Chatgpt), as pointed out by Sarkar and Vafa (2024).

We extend our analysis to construct a purely real-time trading strategy based on the online LDA, which was initially proposed to overcome the computational issue in the LDA model by Hoffman et al. (2010). In particular, in each month τ, we estimate the LDA model using only the conference call transcripts from month 1 to τ and obtain the estimates of

topic loadings $\hat{\boldsymbol{\phi}}_{i\tau}$, which will be later used as instruments in the IPCA model. The primary issue underlying the online LDA analysis is that the estimates of probability distributions over terms, $\hat{\boldsymbol{\Theta}}_\tau$, are evolving. Therefore, it is difficult to visualize the topics as in Figure 2 and Table A2 or study the topic importance as in Section 5.4. For interpretation purposes, the analysis based on the full-sample LDA is more transparent and interpretable than that based on the online LDA.

The online LDA model heavily relies on the initial states, as we indicate for the full-sample LDA in Section 4. To handle this issue, we repeat the online LDA estimation (40 and 50 topics for Q&A and presentation sessions, respectively) using 10 different random seeds, for each of which we construct both the OOS predicted stock returns ($\hat{\boldsymbol{\beta}}_\tau \hat{\boldsymbol{f}}_{\tau+1}$ and $\hat{\boldsymbol{\beta}}_\tau \hat{\boldsymbol{\lambda}}_{f\tau}$) and MVE portfolio ($\hat{\boldsymbol{\lambda}}_{f\tau}^\top \hat{\boldsymbol{\Sigma}}_{f\tau}^{-1} \hat{\boldsymbol{f}}_{\tau+1}$) following the same expanding-window procedure as before. Finally, we take the simple averages of $\hat{\boldsymbol{\beta}}_\tau \hat{\boldsymbol{f}}_{\tau+1}$, $\hat{\boldsymbol{\beta}}_\tau \hat{\boldsymbol{\lambda}}_{f\tau}$, and $\hat{\boldsymbol{\lambda}}_{f\tau}^\top \hat{\boldsymbol{\Sigma}}_{f\tau}^{-1} \hat{\boldsymbol{f}}_{\tau+1}$ across 10 random seeds. The final step can be interpreted as integrating out the uncertainty from initial random states in the LDA estimation.

Table 10 reports the OOS time-series fit for individual stock returns, as well as the Sharpe ratio of the MVE portfolio, conditional on two cross-sectional information sets: i) Q&A and presentation topic loadings and ii) topic loadings plus 35 firm characteristics. We show that the IPCA models composed of online LDA topic loadings are able to deliver satisfactory OOS time-series fit, with more than 22% total R^2 and about 0.12% predictive R^2. Similar to the results based on the full-sample LDA, the IPCA model is not sparse in the mean-variance efficient analysis. Moreover, when the number of latent factors is 25, the MVE portfolio conditional on only the online LDA topic loadings commands an OOS Sharpe ratio of 1.38 per year (compared to 1.37 in the full-sample LDA). We also confirm that the conference call transcripts contain significant incremental, real-time, cross-sectional information beyond the traditional firm characteristics: Using both online LDA topic loadings and firm characteristics, we can achieve an OOS Sharpe ratio of 2.35 when $p = 25$.

Table 10: Out-of-Sample Performance based on Online LDA

		$p = 5$	$p = 10$	$p = 15$	$p = 20$	$p = 25$	$p = 30$
		Panel A. Time-Series Fit					
Q&A + PRE topics	total R^2	22.20	23.21	23.67	23.99	24.27	24.53
	pred R^2	0.12	0.11	0.12	0.12	0.12	0.12
Q&A + PRE topics + 35 Firm Chars	total R^2	24.61	26.01	26.73	27.23	27.63	27.95
	pred R^2	0.08	0.09	0.13	0.13	0.13	0.13
		Panel B. Sharpe Ratios of MVE Portfolios					
Q&A + PRE topics		0.49	0.55	1.00	1.29	1.38	1.51
Q&A + PRE topics + 35 Firm Chars		0.25	1.08	2.14	2.23	2.35	2.47

The table reports the out-of-sample time-series fit and the Sharpe ratio of the MVE portfolio implied by the IPCA model using as instruments the i) both Q&A and presentation topic loadings and ii) both Q&A and presentation topic loadings plus 35 firm characteristics. The Q&A and presentation topics come from the online LDA estimation. We assume that the numbers of topics are 40 and 50 in the Q&A and presentation sessions, respectively, and repeat the online LDA estimation using 10 different random seeds. The model estimates, including the predicted stock returns and MVE portfolios, are simple averages of the estimates across 10 seeds. The number of latent factors in the IPCA estimation, denoted by p, equals 5, 10, 15, 20, 25, and 30. The total and predictive R^2 are calculated in the cross-section of individual stock returns. The out-of-sample period ranges from January 2008 to December 2022.

6 Macro Information in Text-Managed Portfolios

Past literature often shows that the textual information embedded in newspapers contains a forward-looking component that can forecast macro quantities (see Bybee et al. (2024), Bybee et al. (2023), and van Binsbergen et al. (2024)). Motivated by these papers, we go on to study whether the MVE portfolios conditional on Q&A and presentation sessions of conference calls can predict several important macro state variables. In particular, we study eight economic variables from FRED-MD: 1) real industrial production, 2) real personal consumption expenditures, 3) nonfarm employees, 4) total housing starts, 5) consumer price index, 6) unemployment rate, 7) BAA minus FED funds rate, and 8) 10-year treasury yield. For each variable Y_t , we regress its future one-month or one-year changes on the text-managed MVE portfolio returns. For the first five variables, we compute their log growth rates, i.e., $\log(Y_{t+1}/Y_t)$ or $\log(Y_{t+12}/Y_t)$. For the final three variables, we compute their first-order differences, i.e., $Y_{t+1} - Y_t$ or $Y_{t+12} - Y_t$.

Table 11 presents the predictive regression results using the text-managed portfolios from

Table 11: Predicting Macro Variables, Online LDA

	Industrial Production	Personal Consumption	Non-farm Employment	Housing Starts	CPI	Unemploy Rate	BAA minus Fed Rate	10-year Yield
Panel A. Predicting One-Month Changes								
Predictor = Q&A-Managed MVE Portfolio								
Coefficient	0.033	0.053	0.095	0.034	0.388	-0.107	-0.280	0.109
t-stat	(0.670)	(0.675)	(1.649)	(0.596)	(3.404)	(-1.533)	(-1.932)	(1.590)
R^2	-0.45%	-0.28%	0.35%	-0.44%	14.57%	0.58%	7.30%	0.64%
Predictor = Presentation-Managed MVE Portfolio								
Coefficient	-0.036	-0.063	0.012	0.054	0.263	0.017	-0.257	0.043
t-stat	(-0.708)	(-1.122)	(0.336)	(0.900)	(2.138)	(0.394)	(-1.539)	(0.544)
R^2	-0.43%	-0.16%	-0.55%	-0.27%	6.40%	-0.53%	6.09%	-0.38%
Panel B. Predicting 12-Month Changes								
Predictor = Q&A-Managed MVE Portfolio								
Coefficient	0.044	0.013	0.169	0.052	0.182	-0.120	-0.072	0.121
t-stat	(0.630)	(0.264)	(2.920)	(0.519)	(1.370)	(-2.052)	(-1.003)	(0.954)
R^2	-0.37%	-0.55%	2.32%	-0.29%	2.77%	0.89%	-0.03%	0.92%
Predictor = Presentation-Managed MVE Portfolio								
Coefficient	-0.115	-0.161	0.045	0.049	0.009	0.036	-0.065	0.012
t-stat	(-1.363)	(-0.950)	(0.343)	(0.372)	(0.076)	(0.249)	(-0.542)	(0.092)
R^2	0.76%	2.04%	-0.36%	-0.32%	-0.55%	-0.43%	-0.13%	-0.55%

The table reports the regression results in which the text-managed MVE portfolio returns are used to forecast the future one-month or one-year changes in eight economic variables. In the first five columns, we predict the log growth rates of the economic variables, i.e., $\log(Y_{t+1}/Y_t)$ or $\log(Y_{t+12}/Y_t)$. In columns 6–8, we forecast the first-order differences of economic variables, i.e., $Y_{t+1} - Y_t$ or $Y_{t+12} - Y_t$. We report the point estimates of the coefficient of the text-managed portfolio return, its t-statistic based on the Newey and West (1987) standard errors with 12 lags, and the adjusted R^2. The Q&A and presentation topics come from the online LDA estimation. The number of latent factors in the IPCA estimation is 25. The sample period ranges from January 2008 to December 2022.

the online LDA analysis.[11] Panel A shows that in the one-month horizon, both the Q&A- and presentation-managed MVE portfolios significantly predict CPI changes. That is, when we observe higher text-managed portfolio returns, the inflation rate tends to increase over the next month. This observation echoes our previous finding that "inflation" is a common topic in both Q&A and presentation sessions. We also find that the Q&A portfolio return has higher predictive power than the presentation one, as indicated by the time-series R^2. The Q&A-managed portfolio returns also negatively predict credit spread over the next month, consistent with the fact that the important topic "capital structure" is detected only in the Q&A session.

[11] We repeat the same analysis for the text-managed portfolios based on the full-sample LDA in Table A5 and observe similar empirical results.

Panel B further indicates that the positive Q&A-managed MVE portfolio return is associated with higher nonfarm employment and a lower unemployment rate in the future year. Conversely, we do not detect the same predictive power in the presentation-managed portfolio. This stark contrast can be potentially reconciled by the finding that "human capital" is a unique topic in the Q&A session. Overall, the text-managed portfolios contain important forward-looking information about future employment, inflation, and credit spread.

7 Conclusion

Under a conditional asset pricing framework, Kelly et al. (2019) find that firm characteristics—when used as instruments for dynamic factor loadings—can capture sizable variation in the cross-section of stock returns. We find that when the characteristics are replaced with information embedded in the earnings conference call text, the same model describes stock-level returns with similar accuracy. However, both sources of information seem to have saturated their explanatory power, as combining them does not bring improvements.

On the other hand, combining earnings call text with characteristics improves the mean-variance efficiency of factor portfolios, suggesting that text-managed portfolios play a crucial role in spanning investment opportunities. While not offering incremental predictive power for returns, conditioning on text data seems promising for factor spanning. These empirical findings are consistent with the theoretical insights from Kozak and Nagel (2024).

Finally, our work still falls into the camp of reduced-form statistical modeling for the cross-section of returns. Text-managed portfolios constructed in our analysis cannot be directly linked to economic forces. We do offer some suggestive evidence for interpreting these portfolios. For example, the expected returns on the tangency portfolio of Q&A-managed factors are explained by measures of investor sentiments. Its returns can also predict the unemployment rate dynamics at the one-year horizon. In addition, both the presentation- and Q&A-managed efficient portfolio returns are also predictive of one-month and one-year changes in inflation.

References

Aleti, Saketh, and Tim Bollerslev, 2024, News and Asset Pricing: A High-Frequency Anatomy of the SDF, *Review of Financial Studies* .

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

Balduzzi, Pierluigi, and Cesare Robotti, 2008, Mimicking portfolios, economic risk premia, and tests of multi-beta models, *Journal of Business & Economic Statistics* 26, 354–368.

Blei, David M, Andrew Y Ng, and Michael I Jordan, 2003, Latent dirichlet allocation, *Journal of Machine Learning Research* 3, 993–1022.

Bybee, Leland, Bryan Kelly, and Yinan Su, 2023, Narrative asset pricing: Interpretable systematic risk factors from news text, *Review of Financial Studies* 36, 4759–4787.

Bybee, Leland, Bryan T Kelly, Asaf Manela, and Dacheng Xiu, 2024, Business news and business cycles, *Journal of Finance, forthcoming* .

Das, Sanjiv Ranjan, et al., 2014, Text and context: Language analytics in finance, *Foundations and Trends® in Finance* 8, 145–261.

Engle, Robert F, Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebel, 2020, Hedging Climate Change News, *Review of Financial Studies* 33, 1184–1216.

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

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and Kenneth R French, 2020, Comparing Cross-Section and Time-Series Factor Models, *The Review of Financial Studies* 33, 1891–1926.

Ferson, Wayne E, and Campbell R Harvey, 1999, Conditioning variables and the cross section of stock returns, *Journal of Finance* 54, 1325–1360.

Ferson, Wayne E, and Andrew F Siegel, 2001, The efficient use of conditioning information in portfolios, *Journal of Finance* 56, 967–982.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics nonparametrically, *Review of Financial Studies* 33, 2326–2377.

Gabaix, Xavier, Ralph SJ Koijen, Robert Richmond, and Motohiro Yogo, 2023, Asset embeddings, *Available at SSRN 4507511* .

García, Diego, 2013, Sentiment during recessions, *Journal of Finance* 68, 1267–1300.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, 2019, Text as data, *Journal of Economic Literature* 57, 535–574.

Giglio, Stefano, Theresa Kuchler, Johannes Stroebel, and Xuran Zeng, 2024, Biodiversity risk, *Available at SSRN 4410107* .

Giglio, Stefano, and Dacheng Xiu, 2021, Asset pricing with omitted factors, *Journal of Political Economy* 129, 1947–1990.

Glasserman, Paul, and Caden Lin, 2023, Assessing look-ahead bias in stock return predictions generated by gpt sentiment analysis, *arXiv preprint arXiv:2309.17322* .

Goldstein, Itay, Chester S Spatt, and Mao Ye, 2021, Big data in finance, *Review of Financial Studies* 34, 3213–3225.

Hanley, Kathleen Weiss, and Gerard Hoberg, 2019, Dynamic interpretation of emerging risks in the financial sector, *Review of Financial Studies* 32, 4543–4603.

Hansen, Lars Peter, and Scott F Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 587–613.

Hansen, Stephen, Michael McMahon, and Andrea Prat, 2018, Transparency and deliberation within the fomc: A computational linguistics approach, *Quarterly Journal of Economics* 133, 801–870.

Hassan, Tarek A, Stephan Hollander, Laurence Van Lent, Markus Schwedeler, and Ahmed Tahoun, 2023, Firm-level exposure to epidemic diseases: Covid-19, sars, and h1n1, *Review of Financial Studies* 36, 4919–4964.

Hassan, Tarek A, Stephan Hollander, Laurence Van Lent, and Ahmed Tahoun, 2019, Firm-level political risk: Measurement and effects, *The Quarterly Journal of Economics* 134, 2135–2202.

Hoffman, Matthew, Francis Bach, and David Blei, 2010, Online learning for latent dirichlet allocation, *Advances in neural information processing systems* 23.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Huang, Allen H, Reuven Lehavy, Amy Y Zang, and Rong Zheng, 2018, Analyst information discovery and interpretation roles: A topic modeling approach, *Management Science* 64, 2833–2855.

Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou, 2015, Investor sentiment aligned: A powerful predictor of stock returns, *Review of Financial Studies* 28, 791–837.

Huberman, Gur, Shmuel Kandel, and Robert F Stambaugh, 1987, Mimicking portfolios and exact arbitrage pricing, *The Journal of Finance* 42, 1–9.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance?, *Journal of Finance* 78, 2465–2518.

Ke, Tracy Z, Bryan T Kelly, and Dacheng Xiu, 2022, Predicting returns with text data, *Available at SSRN 4416687* .

Kelly, Bryan, and Seth Pruitt, 2013, Market expectations in the cross-section of present values, *Journal of Finance* 68, 1721–1756.

Kelly, Bryan T, Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

Kozak, Serhiy, and Stefan Nagel, 2024, When do cross-sectional asset pricing factors span the stochastic discount factor?, *Available at SSRN 4443643* .

Lamont, Owen A, 2001, Economic tracking portfolios, *Journal of Econometrics* 105, 161–184.

Larsen, Vegard H., and Leif A. Thorsrud, 2019, The value of news for economic developments, *Journal of Econometrics* 210, 203–218.

Lewellen, Jonathan, and Stefan Nagel, 2006, The conditional capm does not explain asset-pricing anomalies, *Journal of Financial Economics* 82, 289–314.

Lopez-Lira, Alejandro, 2023, Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns, *Available at SSRN 3313663* .

Loughran, Tim, and Bill McDonald, 2020, Textual analysis in finance, *Annual Review of Financial Economics* 12, 357–375.

Manela, Asaf, and Alan Moreira, 2017, News implied volatility and disaster concerns, *Journal of Financial Economics* 123, 137–162.

Matsumoto, Dawn, Maarten Pronk, and Erik Roelofsen, 2011, What makes conference calls useful? the information content of managers' presentations and analysts' discussion sessions, *The Accounting Review* 86, 1383–1414.

McCracken, Michael W, and Serena Ng, 2016, Fred-md: A monthly database for macroeconomic research, *Journal of Business & Economic Statistics* 34, 574–589.

Nagel, Stefan, and Kenneth J Singleton, 2011, Estimation and evaluation of conditional asset pricing models, *Journal of Finance* 66, 873–909.

Newey, Whitney K, and Kenneth D West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Rosenberg, Barr, 1974, Extra-market components of covariance in security returns, *Journal of Financial and Quantitative Analysis* 9, 263–274.

Sarkar, Suproteem K, and Keyon Vafa, 2024, Lookahead bias in pretrained language models, *Available at SSRN 4754678* .

Stambaugh, Robert F, Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.

Tetlock, Paul C, 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62, 1139–1168.

van Binsbergen, Jules H, Svetlana Bryzgalova, Mayukh Mukhopadhyay, and Varun Sharma, 2024, (almost) 200 years of news-based economic sentiment, *Available at SSRN 4261249* .

Appendices

Appendix A Preprocessing Conference Call Data

We select the conference call transcripts of the US publicly listed firms from the full sample provided by Refinitiv StreetEvents. That is, we keep 371,659 conference call transcripts among the original 542,472 unique transcripts during the sample period 2003–2022. We further follow Hassan et al. (2023) to keep earnings conference calls since they typically contain both a presentation and a Q&A session, the main focus of our paper. Using this filter, we obtain 240,723 earnings conference call transcripts.

We conduct standard procedures to preprocess the conference call data following the previous literature (see, e.g., Hansen et al. (2018); Huang et al. (2018); Bybee et al. (2023, 2024)). Specifically, we remove non-alphabetic characters and stop words. This step is conducted using the Python package "NLTK". Next, we remove narrative words (e.g., would, would like to), person names, greeting words, and common narratives (such as "hello", "slide", and "please") from conference call texts. Moreover, to account for grammatical variations of English words, we conduct a standard light lemmatizing of derivative words. The rules below are applied in the order given:

(a) Replace trailing "sses" with "ss".

(b) Replace trailing "ies" with "y".

(c) Remove trailing "s", "ly", and "ed" sequentially, and replace remaining "ed" with "e".

(d) Replace trailing "ing" with "e". For remaining trailing "ing" that follow a pair of identical consonants, remove "ing" and one consonant. Remove remaining trailing "ing".

(e) Remove words with less than 3 letters.

It is well-known in textual analysis that bi-grams (a combination of adjacent single words) can capture important semantics beyond uni-grams. Therefore, we generate a set of bi-grams as all pairs of (ordered) adjacent uni-grams and include them in the bag of words. Finally, as too frequent or infrequent words generally contain no information for understanding the topic structure underlying the textual data, we remove words that appear in more than 10% (or less than 0.1%) of the entire sample of conference call transcripts.

Appendix B Algorithm for Finding Stable Topics

This section describes the algorithm for finding stable topics in the LDA estimation across different random seeds that initialize the model parameters in the variational Bayesian estimation. Throughout this section, we demonstrate the procedure using the Q&A topics, but the same algorithm is applied to the presentation topics.

As Figure 1 indicates, the data likelihood, averaged across 10 random seeds, is the highest (equivalently, the perplexity score is the lowest) when the number of Q&A topics is 40. We, therefore, collect all 400 Q&A topics (40 topics across 10 random seeds) and try to find stable topics among them. Note that if a topic is salient and stable, it should show up in most of the 10 random seeds. Conversely, the inconsequential topics, which are driven mainly by estimation noise, should appear in a small proportion of random seeds. Motivated by this logic, we design the following algorithm to find stable topics.

Step 1. We collect all the LDA topics across 10 random seeds and denote the set of topics by $\mathcal{L}$. Let $|\mathcal{L}|$ denote the total number of candidate topics (e.g., 400 in the Q&A session). In this step, we keep the estimates of probability distributions over terms, $\boldsymbol{\Theta}$, as well as the average loadings on all $|\mathcal{L}|$ topics, $\boldsymbol{\phi}$, with $\sum_l \phi_l = 1$. Note that $\boldsymbol{\Theta}$ is a $|\mathcal{L}| \times |V|$ matrix, with the l-th row denoting the probability distribution over terms in topic l.

Step 2. We search for stable topics based on the pairwise correlation between two topics' probability distributions over terms. Specifically, we compute $\text{corr}(\boldsymbol{\theta}_p, \boldsymbol{\theta}_q)$, where $\boldsymbol{\theta}_l$ is the l-th row of $\boldsymbol{\Theta}$. Note that $\boldsymbol{\theta}_l$ tends to be a sparse vector; hence, we compute the pairwise correlations using only the top 1% most frequent words implied by the LDA model.

Step 3. For each iteration, we pick the pairs $\boldsymbol{\theta}_{p^\star}$ and $\boldsymbol{\theta}_{q^\star}$ that have the highest correlation, and combine them into one new topic r, as follows:

$$\boldsymbol{\theta}_r = \frac{\boldsymbol{\theta}_{p^\star}\phi_{p^\star} + \boldsymbol{\theta}_{q^\star}\phi_{q^\star}}{\phi_{p^\star} + \phi_{q^\star}}, \quad \phi_r = \phi_{p^\star} + \phi_{q^\star}. \tag{A1}$$

Merging two topics as in equation (A1) ensures the coherent probabilistic interpretation of the model estimates; that is, $\boldsymbol{\theta}_r$ remains as a probability distribution over terms in the corpus. After the successful combination of two topics, we will remove $\boldsymbol{\theta}_{p^\star}$ and $\boldsymbol{\theta}_{q^\star}$ from $\boldsymbol{\Theta}$ but incorporate $\boldsymbol{\theta}_r$ as a new row in $\boldsymbol{\Theta}$. This step will be repeated until $\text{corr}(\boldsymbol{\theta}_p, \boldsymbol{\theta}_q)$ is less than a certain threshold κ (which is currently chosen to be 0.6).

Step 4. We define stable topics as the rows in $\boldsymbol{\Theta}$ that are merged no less than five times in Step 3. This step requires a stable topic to appear at least five times across ten random seeds.

Appendix C Additional Tables

Table A1: List of 35 Firm Characteristics

Characteristic Code	Characteristic Name in Jensen et al. (2023)	Matched Characteristic Name in Kelly et al. (2019)
Panel A. Exactly matched characteristics		
at_me	assets-to-market	assets-to-market
bidaskhl_21d	the high-low bid-ask spread	bid-ask spread
be_me	book-to-market equity	book-to-market
ni_me	earnings-to-price	earnings-to-price
ivol_ff3_21d	idiosyncratic volatility from FF3	idiosyncratic volatility with respect to FF3
at_gr1	asset growth	investment
at_be	book leverage	leverage
ret_60_12	long-term reversal	long-term reversal
beta_60m	market beta	market beta
market_equity	market equity	market capitalization
noa_at	net operating assets	net operating assets
oaccruals_at	operating accruals	operating accruals
opex_at	operating leverage	operating leverage
prc_highprc_252d	current price to high price over last year	price relative to its 52-week high
ebit_sale	profit margin	profit margin
niq_at	quarterly return on assets	return on assets
ni_be	return on equity	return on equity
niq_be	quarterly return on equity	return on equity
ebit_bev	return on net operating assets	return on net operating assets
ret_1_0	short-term reversal	short-term reversal
turnover_126d	share turnover	turnover
Panel B. Closely related characteristics		
fcf_me	free cash flow-to-price	cash flow-to-book
sti_gr1a	change in short-term investments	cash-to-short-term-investment
cash_at	cash-to-assets	cash-to-short-term-investment; total assets
gp_at	gross profits-to-assets	gross profitability
ppeinv_gr1a	change PPE and inventory	ratio of change in PPE to the change in total assets
ret_12_1	price momentum t-12 to t-1	return from 12 to 2 months before prediction
ret_6_1	price momentum t-6 to t-1	return from 6 to 2 months before prediction
ret_12_7	price momentum t-12 to t-7	return from 12 to 2 months before prediction
sale_gr1	sales growth	sales-to-assets
prc	price per share	sales-to-price
sale_me	sales-to-market	sales-to-price
dsale_dsga	change sales minus change SG&A	the ratio of SG&A costs to sales
dolvol_var_126d	coefficient of variation for dollar trading volume	unexplained volume
dolvol_126d	dollar trading volume	unexplained volume

The table reports the list of 35 firm characteristics used in empirical analyses. Built upon a large dataset provided by Jensen et al. (2023), we select 35 of them that have been used in Kelly et al. (2019) (and originally in Freyberger et al. (2020)). Panel A presents the exactly matched firm characteristics (that is, those with the same definition in both Jensen et al. (2023) and Kelly et al. (2019)). Panel B displays additional characteristics that are close substitutes for the ones used in Kelly et al. (2019).

Table A2: Key Terms in Q&A and Presentation Topics

Topic name	Loading	Key terms
		Panel A. Q&A topics
Forecasting and Speculation	0.105	ballpark, delta, time line, baseline, go forward, 3year, sure understand, exact number, missing, contemplate, bit lower, still working, pretty close, somewhere around, noise
Economic Uncertainty	0.066	recession, ball, worried, worry, sudden, throw, make money, shut, whole lot, lot different, bet, massive, amazing, lot money, know exactly, still lot, crystal
Macro Environment	0.061	backdrop, investor day, macro environment, thoughtful, pretty consistent, helpful followup, resilient, normalization, macroeconomic, delta, great followup, framework, data point
Financial Performance	0.050	midsingle, midsingle digits, automation, high single, investor day, midsingledigit, midteens, high singledigit, margin profile, teens, broadbased, yearonyear, inflationary
Litigation and Negotiation	0.043	time line, burn, language, pilot, patent, timeline, restrictions, royalty, cash burn, letter, dialogue, resolution, resolved, resolve, nearly future, finalize, negotiating
Optimism	0.042	incredible, amazing, massive, thoughtful, great opportunities, solve, tech, team done, come together, runway, lot different, pivot, phenomenal, evolution, early day, tremendous amount
Capital Structure	0.042	covenants, share buyback, return capital, stock price, leverage ratio, capital shareholders, excess cash, terms loan, share price, deleveraging, cost capital, capital market, return shareholders
Pharmaceutics	0.029	dose, therapy, disease, phase iii, efficacy, cohort, enrollment, endpoint, cell, enrolled, cancer, physicians, patients population, clinical trial, tumor
Cloud Service	0.029	cloud, subscription, gotomarket, billings, use case, hardware, logos, sales cycle, smb, new logos, motion, ecosystem, upsell, automation, recurring revenue
Supply Chain	0.025	lead time, shortages, chain issues, cancellations, order book, ordering, automation, chinese, shutdown, absorption, order come, constrained, assembly, supply constraints, meet demand
Inflation	0.023	inflationary, steel, inflationary pressure, price cost, tariffs, input cost, freight, price actions, shortages, volume growth, material cost, cost inflation, resin, metal, increase price
Banking and Lending	0.023	mortgage, originations, loan growth, lending, lenders, borrowers, underwriting, chargeoffs, nim, rate environment, delinquencies, refi, multifamily, beta, deferral, cre, ltv, hikes, deposits growth
Retail and E-commerce	0.016	ecommerce, assortment, merchandise, apparel, women, backtoschool, freight, aur, markdowns, men, shopping, holiday season, new stores, footwear, outdoor
Medical Devices	0.014	reps, surgeons, surgery, physicians, doctors, implant, sales reps, reimbursement, surgical, pump, tissue, knee, consumables, spine, lab, heart, dental, pain, new accounts, wound
Global Markets	0.014	russia, brazil, emea, france, globe, euro, chinese, emerging market, pacific, part world, asia pacific, korea, middle east, italy, rest world, constant currency, ukraine, africa, international business, spain
Renewable Energy	0.013	renewable, solar, electric, hydrogen, carbon, grid, rate case, transmission, megawatts, coal, battery, legislation, emissions, nuclear, reliability
Human Capital	0.013	attrition, tech, health care, wage, new clients, consulting, staffing, clients base, recruiting, merchant, payroll, existing clients, financial services, employers, outsourcing, labor market
Oil	0.012	rigs, drilling, basin, drill, permian, barrels, frac, acreage, lateral, pad, sand, eagle, delaware, midstream, oil price, exploration, commodity price, crew, zone, gulf
Household Goods	0.012	skus, private label, pet, shelf, consumption, elasticity, household, walmart, beauty, ecommerce, mass, grocery, consumer demand, pack, retail partners
Semiconductor	0.010	semiconductor, wafer, chip, memory, foundry, fab, node, semi, laser, silicon, dram, display, logic, inspection, sensors
Transportation: Manufacturing	0.010	dealers, truck, aircraft, boat, electric, dealerships, new car, auction, rental, aftermarket, airlines, defense, ford, aerospace, new vehicle
Health Care	0.009	payers, physicians, medicare, health care, reimbursement, nursing, pharmacy, doctors, medicaid, clinic, acute, health system, primary care, membership, outpatient, clinicians
Insurance	0.009	agents, underwriting, reinsurance, loss ratio, frequency, severity, workers, broker, cat, homeowners, accident, expenses ratio, net retention, institutional, workers comp, book business, annuity
Transportation: Service	0.009	freight, truck, miles, rail, intermodal, container, port, terminal, surcharge, rental, wage, airlines, truckload, hub, transit, trucking
Education	0.008	school, students, education, rigs, enrollment, university, offshore, vessels, campus, college, cohort, district, campuses, graduate, attrition
Electronics	0.007	chip, battery, asp, modules, sensors, cell, silicon, camera, display, laser, computing, design win, hardware, semiconductor, electric
Catering	0.007	restaurants, guests, franchisees, menu, wage, dining, staffing, samestore, samestore sales, chicken, breakfast, beef, kitchen, pizza, loyalty
Defense	0.006	defense, aircraft, aerospace, military, aftermarket, airlines, boeing, flight, aviation, max, satellite, navy, airplanes, army, pilot
Entertainment	0.006	sports, advertisers, audience, streaming, title, entertainment, political, studio, film, theaters, monetization, television, stations, subscription, programming, radio
Biotech	0.006	lab, instruments, pharma, diagnostic, sample, cell, vaccine, consumables, assay, panel, screening, sequencing, placement, academic, life science
Telecommunication	0.005	churn, subscribers, arpu, broadband, fiber, satellite, cable, sub, spectrum, wireless, streaming, tmobile, tower, puerto, rico
Hospitality	0.005	hotel, occupancy, vegas, rental, casino, club, rent, leisure, hospitality, resort, san, revpar, hawaii, vacation, travelers

Panel B. Presentation topics		
Profits	0.049	net revenue, compared quarter, stockholders, period increase, approximately compared, primarily attributable, period prior, increased approximately, expenses third, decrease primarily, adjusted net, decreased compared, revenue third, salaries, offset increase
Organizational Management	0.049	leadership team, stakeholders, pillars, prioritize, bestinclass, transform, strategic priority, streamline, profitable growth, team continue, value creation, team members, unlock, aimed, worldclass
Inflation	0.043	inflationary, price actions, material cost, freight, inflationary pressure, ebit, lower volume, lower sales, input cost, chemical, organic sales, shortages, selling price, cost inflation, volume growth
Pharmaceutics	0.034	disease, cancer, therapy, dose, cell, clinical trial, phase iii, therapeutic, cohort, tumor, efficacy, inhibitor, enrollment, preclinical, endpoint
Events	0.033	amazing, incredible, deck, remarkable, outsized, massive, extraordinary, crisis, friends, war, bright, recession, surge, alltime, letter
Business Operation	0.030	adjusted operating, growth across, driven strong, investor day, macroeconomic, broadbased, profitable growth, point yearoveryear, year driven, macro environment, return shareholders, industryleading, backdrop
Banking and Lending	0.028	mortgage, originations, loan growth, noninterest, lending, interest margin, chargeoffs, loan portfolio, linked quarter, credit loss, tangible, noninterest income, noninterest expenses, total loan, net chargeoffs
Cloud Service	0.027	cloud, subscription, billings, recurring revenue, gotomarket, automation, ecosystem, subscription revenue, intelligence, workflow, hardware, hybrid, professional services
Financial Reporting	0.024	nongaap operating, nongaap net, nongaap basis, quarter nongaap, nongaap gross, gaap net, nongaap earnings, reconciliation gaap, expect nongaap, gaap operating, nongaap results
Retail and E-commerce	0.022	merchandise, ecommerce, assortment, new stores, shopping, stores sales, fulfillment, loyalty, apparel, instore, comparable stores, comparable sales, associates, margin rate, freight
Semiconductor and Electronics	0.021	semiconductor, sensors, chip, battery, display, silicon, modules, optical, wafer, design win, data center, laser, foundry, memory, fab, nextgeneration
International Finance	0.021	constant currency, currency basis, emea, pacific, adjusted operating, asia pacific, brazil, russia, translation, reported basis, impact foreign, local currency, euro, currency translation, currency exchange
Legal and Regulatory Matters	0.016	reliance, company website, warrants, undue, undue reliance, new future, except required, place undue, update revise, per basic, historically fact, events otherwise, including limited, whether results, securities law
Renewable Energy	0.016	renewable, electric, emissions, megawatts, solar, grid, carbon, transmission, rate case, reliability, renewable energy, nuclear, regulated, electricity, storm
Oil	0.014	per day, drilling, barrels, basin, drill, boe, barrels oil, rigs, lateral, permian, eagle, commodity price, acreage, oil equivalents, pad
Fiscal Periods	0.013	fiscal quarter, half fiscal, end fiscal, year fiscal, fiscal fourth, fiscal first, fiscal compared, fiscal second, fiscal third, compared fiscal, full fiscal, fiscal expect, revenue fiscal, months fiscal, guidance fiscal
Defense	0.013	defense, aerospace, military, aircraft, satellite, aviation, booktobill, aftermarket, flight, navy, army, commercial aerospace, rocket, aerospace defense, missile
Medical Devices	0.012	surgery, surgeons, surgical, physicians, therapy, clearance, heart, implant, pain, reps, tissue, medical devices, spine, reimbursement, dental
Insurance	0.011	underwriting, agents, combined ratio, investment income, reinsurance, loss ratio, written premium, expenses ratio, catastrophe, net investment, accident, annuity, workers, investment portfolio, casualty
Household Goods	0.009	pet, consumption, beverage, ingredients, foodservice, household, snacks, coffee, flavor, fresh, wellness, grocery, consumer demand, beauty, skus
Asset Management	0.009	assets management, wealth, aum, institutional, inflows, advisory, advisers, fixed income, capital market, outflows, management fees, index, wealth management, brokerage, assets class
Biotech	0.009	diagnostic, vaccine, lab, instruments, sample, consumables, life science, animal, pharma, assay, sequencing, imaging, panel, laboratory, variant
Health Care	0.008	health care, medicare, physicians, payers, pharmacy, clinic, reimbursement, net revenue, nursing, health system, care provider, prescription, admissions, clinicians, home health
Entertainment	0.008	sports, entertainment, streaming, audience, advertisers, studio, fans, film, music, title, programming, venues, advertising revenue, broadcast, television
Education	0.008	students, school, education, enrollment, university, campus, college, career, district, learners, academic, cohort, campuses, k12, educational
Catering	0.008	restaurants, guests, franchisees, menu, samestore, dining, samestore sales, team members, restaurants sales, wage, systemwide, offpremise, guests experience, companyowned, shack
Energy	0.007	rigs, esg, vessels, offshore, drilling, marine, gulf, lng, pipe, rigs count, activity level, pumping, rental, sea
Steel	0.007	steel, metal, mill, mining, aluminum, per tons, coal, carbon, scrap, metrics tons, valueadded, selling price, imports, export, iron
Hospitality	0.006	hotel, occupancy, resort, revpar, hospitality, casino, vegas, leisure, rent, guests, venues, terms loan, residents, club, vacation
Telecommunication	0.006	subscribers, fiber, broadband, churn, arpu, wireless, cable, spectrum, satellite, tmobile, verizon, subscribers growth, postpaid, voice, rural
Freight	0.005	aircraft, airlines, freight, miles, truck, flight, delta, gallon, passenger, aviation, fuel price, pilot, airport, max, terminal
Metals and Mining	0.004	mining, gold, copper, ounces, metal, coal, silver, pound, creek, grade, carbon, ore, mill, minerals, print

The table presents the topic names, the average loadings of firms' conference call texts on these topics, and the key terms in each topic based on the full-sample LDA estimation. We obtain stable topics following the algorithm described in Appendix B.

Table A3: Robustness Check: Out-of-Sample Sharpe Ratios From 2013

	$p = 5$	$p = 10$	$p = 15$	$p = 20$	$p = 25$	$p = 30$
32 Q&A topics	0.72	0.90	0.42	0.86	1.33	1.18
32 PRE topics	1.37	1.31	1.18	1.15	1.24	1.24
64 Q&A + PRE topics	1.11	0.93	1.36	1.72	1.61	1.74
35 Firm Characteristics	1.10	1.32	1.72	1.82	1.90	1.91
64 Q&A + PRE topics + 35 Firm Chars	0.29	1.60	2.46	2.50	2.88	3.02

The table differs from Table 3 in that the out-of-sample ranges from January 2013 to December 2022.

Table A4: Topic Importance: IPCA Model with $p = 30$

32 Q&A topics, excluding:	SR_{oos}	ΔSR_{oos}	32 Presentation topics, excluding:	SR_{oos}	ΔSR_{oos}
Cloud Service	0.57	-0.53	Steel	0.99	-0.17
Capital Structure	0.73	-0.38	Semiconductor and Electronics	1.03	-0.13
Human Capital	0.94	-0.17	Energy	1.05	-0.11
Litigation and Negotiation	1.01	-0.09	Business Operation	1.05	-0.11
Semiconductor	1.05	-0.05	Asset Management	1.06	-0.11
Forecasting and Speculation	1.09	-0.02	Entertainment	1.07	-0.09
Defense	1.09	-0.02	Cloud Service	1.10	-0.06
Oil	1.09	-0.01	Education	1.10	-0.06
Banking and Lending	1.10	-0.01	Events	1.10	-0.06
Supply Chain	1.10	-0.01	Insurance	1.10	-0.06
Pharmaceutics	1.10	0.00	Metals and Mining	1.11	-0.06
Electronics	1.11	0.00	Medical Devices	1.12	-0.04
Entertainment	1.11	0.01	Freight	1.13	-0.04
Medical Devices	1.11	0.01	Defense	1.13	-0.04
Retail and Ecommerce	1.12	0.01	Fiscal Periods	1.13	-0.04
Telecommunication	1.12	0.02	Hospitality	1.13	-0.03
Education	1.14	0.03	Financial Reporting	1.14	-0.03
Household Goods	1.14	0.03	Household Goods	1.14	-0.03
Catering	1.14	0.04	Biotech	1.14	-0.02
Insurance	1.14	0.04	Pharmaceutics	1.14	-0.02
Global Markets	1.14	0.04	Health Care	1.16	-0.01
Health Care	1.15	0.04	Retail and Ecommerce	1.16	0.00
Macro Environment	1.15	0.04	Profits	1.16	0.00
Transportation: Manufacturing	1.15	0.04	Legal and Regulatory Matters	1.16	0.00
Renewable Energy	1.15	0.05	Catering	1.16	0.00
Financial Performance	1.15	0.05	International Finance	1.16	0.00
Hospitality	1.15	0.05	Inflation	1.17	0.00
Inflation	1.16	0.05	Telecommunication	1.18	0.02
Optimism	1.16	0.06	Banking and Lending	1.19	0.03
Transportation: Service	1.17	0.07	Oil	1.20	0.04
Economic Uncertainty	1.18	0.07	Organizational Management	1.21	0.04
Biotech	1.20	0.10	Renewable Energy	1.21	0.05

The table reports the relative topic importance in Q&A and presentation conference call texts, respectively. We sequentially exclude one of the 32 Q&A or presentation topics in the IPCA estimation and report the corresponding annualized Sharpe ratio of the out-of-sample MVE portfolio (see column SR_{oos}), as well as the drop in the Sharpe ratio relative to the benchmark analysis in Table 3 (see column ΔSR_{oos}). The Q&A and presentation topics come from the full-sample LDA estimation. The number of latent factors in the IPCA estimation is 30. The out-of-sample ranges from January 2008 to December 2022.

Table A5: Predicting Macro Variables, Full-Sample LDA

	Industrial Production	Personal Consumption	Non-farm Employment	Housing Starts	CPI	Unemploy Rate	BAA minus Fed Rate	10-year Yield
Panel A. Predicting One-Month Changes								
Predictor = Q&A-Managed MVE Portfolio								
Coefficient	0.091	0.086	0.163	0.055	0.417	-0.205	-0.281	0.091
t-stat	(0.941)	(0.737)	(1.536)	(0.811)	(4.057)	(-1.588)	(-2.382)	(1.314)
R^2	0.26%	0.18%	2.10%	-0.26%	16.95%	3.65%	7.36%	0.27%
Predictor = Presentation-Managed MVE Portfolio								
Coefficient	0.093	0.011	0.023	0.122	0.146	0.004	-0.187	0.027
t-stat	(1.359)	(0.219)	(0.999)	(1.561)	(1.516)	(0.099)	(-1.547)	(0.344)
R^2	0.31%	-0.55%	-0.51%	0.93%	1.57%	-0.56%	2.94%	-0.49%
Panel B. Predicting 12-Month Changes								
Predictor = Q&A-Managed MVE Portfolio								
Coefficient	0.116	0.087	0.224	0.084	0.234	-0.207	-0.109	0.175
t-stat	(1.559)	(1.530)	(2.167)	(0.839)	(1.639)	(-1.984)	(-1.503)	(1.425)
R^2	0.79%	0.19%	4.50%	0.15%	4.94%	3.73%	0.64%	2.52%
Predictor = Presentation-Managed MVE Portfolio								
Coefficient	-0.060	-0.112	0.039	0.075	-0.097	0.010	-0.009	-0.047
t-stat	(-0.958)	(-0.959)	(0.318)	(0.953)	(-0.922)	(0.089)	(-0.128)	(-0.726)
R^2	-0.21%	0.69%	-0.41%	0.01%	0.38%	-0.55%	-0.55%	-0.34%

The table reports the regression results in which the text-managed MVE portfolio returns are used to forecast the future one-month or one-year changes in eight economic variables. In the first five columns, we predict the log growth rates of the economic variables, i.e., $\log(Y_{t+1}/Y_t)$ or $\log(Y_{t+12}/Y_t)$. In columns 6–8, we forecast the first-order differences of economic variables, i.e., $Y_{t+1} - Y_t$ or $Y_{t+12} - Y_t$. We report the point estimates of the coefficient of the text-managed portfolio return, its t-statistic based on the Newey and West (1987) standard errors with 12 lags, and the adjusted R^2. The Q&A and presentation topics come from the full-sample LDA estimation. The number of latent factors in the IPCA estimation is 25. The sample period ranges from January 2008 to December 2022.