QUANTIFYING SOURCES OF FIRM-LEVEL RISK THROUGH CONFERENCE CALL TRANSCRIPTS

A BERTOPIC APPROACH

ANDREW KAPLAN

ANNA KORSUNOVA

Bachelor Thesis

Stockholm School of Economics

2024



Quantifying Sources of Firm-Level Risk Through Conference Call Transcripts: a BERTopic Approach

Abstract:

We use BERTopic to develop an alternative approach to measuring topic-specific risk for firms from quarterly earnings conference calls. This approach allows us to both discover and quantify sources of firm risk based on statements made by company executives. We show the topics identified by the model are coherent and that the aggregate risk patterns for each are intuitive based on historical context. Seven out of the thirteen risks we study are significant and positively correlated with future realised stock volatility. After controlling for firm fixed effects, only two remain significant, suggesting that our measure may primarily reflect permanent differences in firm risk. We decompose the variance structure of each risk and find that global economy, inflation, and supply chain risk are the most systematic. We also find that for cybersecurity, firms exhibit heterogeneous exposure to the aggregate risk along with large and persistent differences in firm-specific risk levels.

Keywords:

Firm risk, systematic risk, earnings conference call transcripts, topic modeling, BERTopic, natural language processing

Authors:

Andrew Kaplan (25622)
Anna Korsunova (25564)

Tutor:

Zoltan Racz, Postdoc Fellow, Department of Finance

Examiner:

Ramin Baghai, Associate Professor, Department of Finance

Bachelor Thesis
Bachelor Program in Business and Economics
Stockholm School of Economics
© Andrew Kaplan and Anna Korsunova, 2024

1 Introduction

For firms with publicly traded stock options, we can estimate the level of risk the firm is exposed to by using its implied volatility.[1] Alternatively, historical volatility in the stock price itself may provide some useful indication. However, beneath this volatility figure lie many different sources of risk: firms can be exposed to changes in interest rates, aggregate fluctuations in the economy, supply chain pressures, or an array of other topics. This study focuses on measuring precisely these underlying sources of uncertainty. We analyse the discussions of risk by firm executives during quarterly earnings conference calls using a state-of-the-art large language model called BERT. In this analysis, we are able to automatically detect topics relating to risk, and build a quantitative measure of firm-level risk on a quarterly basis for 13 different sources of risk. A primary goal is to measure topic-specific risk without the need to define exogenously what a topic is. We also want to study the extent to which these conference calls provide a useful signal for a partially new set of risks previously not researched. Finally, we decompose the variance of each risk measure to understand their behaviour in the economy such as which risks emerge systematically, and which are the most heterogeneous between firms and sectors.

An existing strand of literature looks at corporate risk disclosures from annual reports to build a similar measure (e.g. Zhu, Yang, and Moazeni, 2016; Yang, 2013). However, the information content in these disclosures is limited as they are often boilerplate statements (Schrand and Elliott, 1998). They list all possible risks regardless of their actual impact on the firm, which inevitably washes out any helpful signal. In fact, when Yang (2013) tested 30 risks using a Latent Dirichlet Allocation (LDA) model on 10-K filings, only 8 were significantly associated with realised stock volatility in the two months after the filing.

Instead, we look at earnings conference calls as they offer a promising alternative to annual reports for studying firm risk. During the calls, executives discuss not only past performance but also the future outlook of the firm. Matsumoto, Pronk, and Roelofsen (2011) argue that they should hold more material information than the aforementioned risk disclosures as they are less scripted and a large portion is prompted by outside analyst questions. They also benefit from being released at shorter intervals — quarterly rather than annually. Due to these advantages, earnings calls have recently attracted attention as a data source for measuring firm risk. Several papers have established empirically that such calls do hold material information predictive of future outcomes (e.g. Hassan, Hollander, van Lent, and Tahoun, 2019; Kim, Muhn, and Nikolaev, 2023; Wu, 2024; Nazli, Karagozoglu, and Zhou, 2021). This line of research is still nascent, however, so we offer several contributions.

Firstly, to the best of our knowledge, the only risk topics that have been covered previously using earnings conference calls from the 13 we study are supply chain (Wu, 2024 and Ersahin, Giannetti, and Huang, 2024), cybersecurity (Nazli et al., 2021), regulatory (Calomiris, Mamaysky, and Yang, 2020), interest rates (Ying, 2023), tax (Hassan et al., 2019), and inflation (Chava, Du, Shah, and Zeng, 2022). That would mean the extension of this literature to the following risks is in itself a contribution: global economy, loans, oil & gas, litigation, contracts, foreign exchange, and banking risk. Furthermore, our measure allows us to build an aggregate index of risk for each topic. Many of these risks do not have such a measure even outside the literature that studies conference calls.

Additionally, most of the existing literature relies on predetermined dictionaries of text

[1] Risk here refers to the risk the firm faces, not the risk an investor has by holding the firm's stock.

to define a topic. Whilst they proved to be robust and effective, such a requirement can be problematic if topics change over time; predefined topics can easily become outdated especially if they deal with current events. It also raises the potential for bias, favouring whatever the researcher decides is a topic, which can vary between disciplines, countries, and languages. Our approach, on the other hand, defines topics based exclusively on the contents of the conference calls we study. New topics that emerge, or those that evolve over time, can be easily included in the analysis by retraining the model on the most recent calls. Furthermore, training requires very few exogenous parameters and is easy to replicate. Two exceptions to the reliance on predefined topics are Kim et al. (2023), who use a ChatGPT approach instead, and Chava et al. (2022), who use a different variation of BERT. Our method still adds to this analysis because it is able to *discover* topics. Furthermore, the standard BERT approach requires the manual construction of a large sample of data which still requires someone to implicitly define a topic.

Thirdly, the literature is currently very fragmented with many papers covering a single risk in isolation. A notable exception is again Kim et al. (2023) who study three at a time. Still, we measure more (13 in total), all using the same methodology on the same sample. We argue that this strand of research would benefit from a consolidated approach that can be applied to all risk topics. This is motivated in part because any measure that extracts quantitative information from conference calls will have a large amount of measurement noise. We rely on conference calls being unscripted, but this also means that a lot of the variation between firms in measured risk might come from randomness in what questions are asked and how they are answered. Furthermore, any textual information that is quantified will inherently contain measurement noise, the extent of which can depend a lot on the methodology. We wish to answer questions such as which risks systematically appear in the economy, but measurement noise may cause all risks to appear idiosyncratic. What is needed then is a measure of many risks, where it is not the absolute values that matter but their relative ones. For example, we find that 1.17% of supply chain risk is explained by time fixed effects, but this result is difficult to interpret on its own. However, noting that this figure is larger than for all but two other risks suggests that exposure to the risk is relatively systematic.

Finally, to the best of our knowledge, we are the first to apply BERTopic to earnings conference calls to measure risk. A well-established but older approach for topic analysis uses the Latent Dirichlet Allocation (LDA) model. This model relies on a bag-of-words approach which means word order is ignored. BERTopic is based on Google's recent pre-trained large language model BERT (Bidirectional Encoder Representations from Transformers) which transforms text into a continuous vector representation that captures its meaning (Kenton and Toutanova, 2019). Unlike traditional approaches, the model can understand context by allowing the meaning of words to be influenced by those around them. As mentioned, Chava et al. (2022) use another variation of BERT that is specifically tasked with classifying text snippets as inflation risk, but again, this model requires new training data for every topic which can be difficult to scale. Some papers have employed BERTopic for stock market prediction using sentiment in social media posts (Zhu and Yen, 2024) and to predict corporate bond defaults using news headlines (Tang, Bu, Zuo, and Wu, 2024). We contribute here by finding a new application for the model. An interesting result then will be the extent to which it holds promise for future research.

Some discussion is warranted as to exactly what we are able to measure using conference calls. We measure the *perceived* risk of the firm management which naturally could deviate

from actual risk. Some risks are generally unforeseen by the public, for example, few noticed the impending collapse of mortgage-backed securities in the early 2000s yet it was obvious the risk was there in hindsight. There are also many reasons a firm might discuss risk and not all of them signal actual risk. For example, the fact that executives bring the subject up might indicate that the company is prepared to deal with the risk. It could be that firms discuss topics only after shocks take place, even if they do not expect the heightened risk to continue. Nevertheless, as discussed previously, there are empirical results that show at least in some cases that this measure is a good proxy for actual risk.

In order to measure topic-specific risk, we begin by identifying all discussions of any risk in a transcript. These discussions are then grouped together based on their semantic similarity, with the resulting groups forming topics. That is, a topic forms when enough discussions of risk are similar enough to each other. Forty-eight of these topics emerged based on our sample and hyper-parameters, of these the thirteen previously mentioned were selected to be evaluated. We then establish a set of requirements for a topic to be valid. Firstly, it must be coherent — the words associated with the topic should make sense. Second, the model we use should accurately classify documents into the correct topic. Third, aggregate patterns in risk for each source should be intuitive and justified by real-world events. Finally, as would be expected of any measure of risk, increased risk should correlate with increased stock volatility on a firm level. After evaluating and validating each of our topics with these four steps, we analyse their variance decomposition to help answer questions such as which risks are systematic.

We find that BERTopic is able to discover interesting and coherent topics, and the majority of documents are correctly classified according to their risk. Furthermore, when evaluated against existing indices of topic-related events, our aggregate measures of risk provide intuitive variations over time.

We further find that when controlling for time and sector fixed effects, 7 out of 13 of our measures are positive and significantly associated with realised stock volatility during a 90-day holding period. However, adding firm fixed effects leads to only 2 remaining positive and significant. This suggests that our measure of risk primarily accounts for permanent differences in risk exposures between firms. The existing literature has, at times, included or omitted firm fixed effects in the main specification,[2] but we argue that these can substantially impact the estimates. We also find that discussion of some sources of risk leads to *lower* volatility. This might indicate that for some topics, acknowledgement of a risk in an earnings call signals that the firm is prepared for it, rather than predicting future uncertainty. Nevertheless, the aggregate measures for these risks provide intuitive patterns so they are indicative of the overall risk for that topic in the economy.

When decomposing the variance of each of our measures, we find that global economy, inflation, and supply chain risk are the most systematic out of the 13 studied. We also find that for cybersecurity, firms exhibit large heterogeneity in the levels of risk *and* exposure to the aggregate risk. In fact, it is the only topic where a model of stable firm-specific exposure to aggregate risk can explain a significant portion of variance.

Despite promising findings, our approach has a limitation. Not every topic can be easily discovered. In particular, those with a diverse vocabulary are hard for the model to cluster into a single topic. For example, politics as a topic covers everything from healthcare, wars,

[2] Hassan et al. (2019) include them but some follow-up papers (e.g. Kim et al., 2023; Nazli et al., 2021; Wu, 2024) do not.

taxes, regulation, and much more. We do isolate some of these, but the model is not able to distinguish (automatically) that they all relate to politics.[3] What we can capture then is very specific topics, because, in reality, whilst these all clearly relate to politics, they also relate to many other risks. The model can discover the most basic components of a topic, in our case, we capture taxes and regulation in our risks because these are harder to subdivide.

Aside from the specific contributions, what our paper aims to offer is a robustness check to the recent literature that studies topic-specific risk in conference call transcripts. We do not define the risks ourselves, nor do we filter any that do not work.[4] We aim to see the extent to which the data in these calls can generalise to more risks, and we explore the potential issues with the interpretation of findings.

2 Data

In order to measure the perceived risk of firm executives, we use quarterly earnings conference calls. These are calls where executives present a discussion of the past quarter's results, and outside analysts can ask questions. Several databases provide transcripts of these calls. We use Seeking Alpha which is a website that publishes much more than just transcripts, including market news and investment reports. We retrieved every available transcript between the start of their data (April 2004) and when we downloaded them (October 2024). The transcripts also include metadata such as the firm name and ticker.

To validate our measure of risk, we analyze its relationship with the firm's stock volatility and returns. For this, we use daily security data from the Center for Research in Security Prices (CRSP). We use the adjusted returns to investors (ret) which account for stock splits and dividends. Forward-looking volatility is calculated based on a 90-day holding period beginning the day when the conference call was held. For lagged volatility, we use a 90-day holding period that ends the day before the call. The same applies to average returns.

We also need data on each firm's sector and the value of assets as controls in our regressions. Different firms have different fiscal quarters, so Q1 for one firm might happen at the same time as Q3 for another. To align the conference calls with the correct (calendar) time period, we need this information. For all of this, we use Compustat's quarterly fundamentals data. Namely, for sector we use the SIC code and for assets we use the variable atq. The dataset also provides the calendar quarter that corresponds to each fiscal quarter.

Transcript coverage by quarter is summarised in Figure A.1. Near the beginning and end of our data, there are very few transcripts, so these quarters are excluded for plotting purposes as they can make the results very volatile. The original sample contained 238,420 transcripts. 14,939 were filtered out because they were less than 100 words, not a transcript, a duplicate, or were missing information in the title. A further 107,109 were lost because we could not reliably match them to Compustat or they were not in the intersection with CRSP.[5] Given that Seeking Alpha covers non-American firms and CRSP does not, this is to be expected. 15,361 of these are also because the CRSP data ends in 2023 so we lose all

[3] Hassan et al. (2019) themselves note that topic models likely lack the power to discern politics as its own topic. Topics can be merged manually using the model, but this partially defeats the purpose of endogenously defining them.

[4] Although we do select topics that make sense in the context of risk after obtaining them from the BERT model, this selection occurs before conducting any further validation exercises.

[5] We rely on the stnd_compname package in Stata for matching. This package standardises and cleans

transcripts from 2024. In any remaining cases, the attrition should be random so we have no reason to suspect this will cause any issues other than loss of statistical power in our tests.

This leaves 122,710 transcripts with an average of 6,949 words each covering 4,808 U.S. listed firms.[6] The 107,109 transcripts lost to merging were still included in training BERTopic as more documents can make the detection of topics easier.

In our measure, we look for discussions of risk. The terms "risk officer" and "risk credit officer" can confound this as they are commonly listed at the beginning of transcripts along with the call participants' names. Following the methodology of Hassan et al. (2019), we remove these.

3 Empirical Approach

3.1 Classifying Topics

The first part of creating our risk measure requires us to identify and define which topics exist in our conference calls. This is done in four steps: 1) extract and limit attention to only the few lines in a transcript that are dedicated to discussing risk, 2) convert these discussions into a vector of continuous values that captures the meaning of the text, 3) identify clusters of similar discussions, and 4) label the clusters and set key terms that define them as a topic.

To identify snippets of text pertaining to risk specifically, we follow the same approach as Hassan et al. (2019). We take the 10 words on either side of synonyms for "risk" or "uncertainty" found in the transcript as defined by the Oxford Dictionary. Each resulting 21-word snippet is what we consider one document.[7] Note, we use "snippet" and "document" interchangeably throughout this paper. We also filter out the so-called harbour statements — boilerplate statements about how the conference call is forward-looking and that actual results may differ materially. These are generic and do not hold any topic-specific information, nor do they convey any information on the overall risk of the firm. So, if a risk snippet comes from the first half of the transcript and contains two or more of the following words — safe, harbour, forward, looking, actual, statements, statement, future, events, SEC, or results, — it is filtered out. The result is that a transcript will have just over 5 risk snippets on average.

In order to be able to cluster similar documents together, we must define what "similar" means. A simple approach would be to group topics that share similar words, however, this ignores the possibility that context influences the meaning of words (for example, "interest" might refer to either the cost of borrowing money or someone's curiosity about a certain subject). We use what is known as an embedding model based on BERT (Bidirectional Encoder Representations from Transformer) (Kenton and Toutanova, 2019). This allows text to be converted into a high-dimensional vector of continuous values. The position in

names so for example, "North America" becomes "NA".

[6] Some additional small number of transcripts are filtered out in regressions depending on the variables being used and their availability. Furthermore, 1,400 transcripts that do not contain the exact conference call date are filtered out of regressions. The total number of transcripts in each regression is available at the bottom of every table.

[7] "Document" is the typical term used to describe a given piece of text in the natural language processing literature.

the vector space captures the text's meaning, and two sets of text close to each other in said space are thought to mean similar things. The embedding model has been pre-trained on large quantities of text and shown empirically to capture meaning very well. Every document is first passed through this model and the steps that proceed use the vector representations produced.

There are many approaches to clustering the vectors. We use a model known as BERTopic[8] (Grootendorst, 2022). It includes a series of algorithms that reduce the dimensions of the vectors produced by the embedding model and then group them in a way similar to K-means clustering.[9] For a topic to form, there need to be enough documents that are similar enough to each other. Not all documents are assigned a topic — if they are not close enough to a cluster, they are considered outliers.

Once clusters are formed, that is, topics have been identified, they need to be labelled so humans can actually interpret what they represent. Two steps are involved in this. Firstly, to determine which words and terms constitute a topic, a term-frequency inverse-document-frequency algorithm is used. Formally, this calculates the following for each term i in each topic t: $tf_{it} \times log(N/df_i)$[10] where tf_{it} is the count of that term in the topic, N is the number of topics, and df is the number of topics that contain the term i. This places higher weight on words that appear more often within a topic, but do not appear in others. So words like "quarter" or "question" appear a lot in each topic but also are present in every topic, so they have low information content. Conversely, "supply chain" is unlikely to appear in any topic other than supply chain, so it has high information content. Then another model called KeyBERT selects the top 10 key terms from all those with high information content which are collectively known as the topic's representation.

Importantly, the algorithm uses not just words, but also *bi-grams* and *tri-grams* — terms composed of 2 or 3 words. These *n-grams* only form from combinations of words that appear together very often, so using the same example, "supply chain" is a term that commonly is together. When the model is evaluating a snippet with this term in it, it evaluates the two words together rather than looking at "supply" and "chain" separately. This approach is common in the natural language processing literature since words in combination may have a different meaning than individually. Furthermore, to avoid noise in the topics, we remove stop-words[11] and the actual synonyms for risk since these appear in all documents.

One key assumption that sets this model apart from, for example, the more typical Latent Dirichlet Allocation (LDA) approach is that each document pertains only to a single topic. Given how short our documents are, this is a close approximation to the truth, however, it does not always hold and so this is a limitation of the model.

3.2 Quantifying Risk

At this point, every risk snippet from every transcript has been used to train the BERTopic model and is now assigned a topic (or is an outlier). The original goal, however, was to get a firm-quarter-level measure of risk for every topic. Since a transcript can have many of these

[8]For documentation on the library used visit: https://maartengr.github.io/BERTopic/api/bertopic.html

[9]In fact, K-means clusters can be used as a substitute for the default algorithm.

[10]There are many variants that achieve a similar goal. This is also not the standard version as we use topic frequencies rather than document frequencies.

[11]Stop-words are words like "a", "the", and "are" which are common and hold very little meaning.

risk snippets, we need a way to aggregate them together so the transcript as a whole has a score. For this, we have two approaches:

The discrete measure. The first measure classifies each snippet as a single topic and then counts the number of occurrences of each topic in a given transcript. So, if two of the risk snippets in a transcript mention regulatory risk, the score for that quarter would be two. The score can range from zero to the number of risk snippets that are in a document. This approach has a very simple interpretation — more mentions of a risk equals more of that risk. The drawback of this approach is that it has limited power to detect topics, and a snippet is allowed to only contain a single topic. So, a firm discussing inflation as a result of the pandemic may be classified as pandemic risk rather than inflation, even though it is a clear signal of both. Also, more ambiguous discussions that signal some amount of a given risk may go unclassified. We evaluate later to what extent this might be an issue.

The continuous measure. Even though BERTopic is a classification algorithm, it also outputs the probability that a given document pertains to each topic (for an example of how the model compares topics for a particular document, see Figure A.3). In theory, a higher probability indicates that the document is more similar to that topic. Instead of classifying each of the snippets directly, the continuous measure uses the probability distribution of all topics for a single document; each risk snippet has an *n-dimensional* vector, where n is the number of topics and each element is the probability that that snippet pertains to that topic. To obtain a measure for each transcript, we compute the average of the probability vectors for all its snippets, with each value in the resulting vector representing the average probability that each snippet pertains to a given topic. This inherently controls for the overall risk level because we are taking an average rather than a sum (dividing by the number of mentions of risk). In cases when there are no risk snippets, each risk is given its average probability. We justify this because a low probability directly signals a lack of a given topic. Contrarily, if a topic scores an average probability it indicates ambiguity. In other words, a low score will not arise simply because the topic in question is not in the snippet, it is also a requirement that another topic is clearly there. Without any mentions of risk, we have no such signal. Of course, the (reasonable) assumption here is that a firm still has risk even if it does not mention it, and that the composition of risk is only revealed if management chooses to discuss it. We further justify this assumption and our approach of averaging vectors by testing an alternative where we add together the probability vectors instead, or impute zero when there are no snippets. In both cases, we obtain very unintuitive aggregate patterns of topic-specific risks, which are highly correlated with each other and/or the aggregate patterns in total risk. This means we are not actually measuring each particular topic. The fundamental issue with creating a continuous measure like this is the model is not designed to allow for more than one topic per document, so it is very difficult to evaluate topics in isolation. The resulting interpretation of this measure is that a higher score for a given topic does not equate to more risk overall, but simply more of the firm's risk is related to that topic. The relationship is likely not perfect or proportional, however, as otherwise simply interacting the percentage of each risk with the total risk should have worked but it too suffered from the same issues.[12] It is instead just an indication of the tilt a firm faces towards a particular risk. Due to the

[12]Words that have high information content for one topic tend to do so in several. This means that the raw *n-gram* scores produced by the model are highly correlated between topics in a document. We also tried controlling for this by demeaning the *n-gram* score sums but this results in something very similar to simply averaging the probabilities. Indeed one is still inherently controlling for the overall level of risk this way. We use the simpler and easier-to-interpret approach of averaging them. For more details on how these *n-gram* scores are calculated see Figure A.3.

difficulties in interpreting this measure, our focus throughout the analysis is on the discrete version. However, we include this here still as an interesting exercise and because in some cases it does provide insightful results.

Measurement noise must be addressed as well for any measure that quantifies textual data, especially this one. There are many sources of this noise, first, the transcripts themselves are a noisy signal of firm risks. A lot is dependent on the choices of management to mention or not to mention a particular risk, or for an analyst to ask a certain question. Much cross-sectional variation originates from this or subtle differences in wording that may affect the measure. Furthermore, whilst the number of mentions of a given risk seems to be a fair proxy for the level of that risk, it certainly is not a perfect correlation. Finally, there could be classification errors, and two documents both classified as a given topic may not indicate equivalent levels of that risk.

What we measure is also not exactly risk, it is the risk *perceived* by management. This does not mean it is wrong per se but it may be biased and it does not measure unforeseen risks. It also will not be as precise as other measures such as implied volatility that relies on the extensive research of many hundreds or thousands of market participants.

4 Risk Measures & Validation

4.1 Topics

We now present the topics that the model has discovered in our data. The first step is to evaluate whether they are coherent, and then we will test to what extent the model is able to sort documents into the correct topics. Naturally, these two are minimum requirements for a measure of topic-specific risk.

Based on our training sample and model hyperparameters, 48 topics emerged, all of which can be viewed in Table A.1. We chose the 13 which can most clearly be related to a source of risk and broad enough to be interesting. For example of those we excluded, topic 32 seems to discuss business management and strategy, which is a reasonable topic, but there is no clear connection to what such a risk would represent. Topic 38 exclusively discusses pipelines which is too specific. All topics can be viewed in Table 1, along with their 10-term representations. The representations are not the only words used to define a topic, in fact, most words have some value associated with every topic, but the representations capture the meaning of the topic. We argue that they do this very well — with few exceptions, all terms in the table have a clear connection to the topic, and in turn, the topic is well-defined based on the terms in the table.

Table 1: Topic Representations
This table shows the top 10 terms associated with each topic, known as the topic representations. Topic representations are the terms most associated with each topic, endogenously generated by a combination of the term-frequency inverse-document-frequency algorithm and the model KeyBERT after the topic clusters have been identified. A term may consist of up to three words.

Topic Name	Representation
Cyber Security	cybersecurity, cyber security, cyber, cyber attacks, security industry, security systems industry, security systems, cloud based, systems industry, data security
Global Economy	global economies, global economy, global economic, economic growth, global economic conditions, global growth, global macroeconomic, gdp growth, global economic environment, world economy
Loans	loans loans, total loans, loans, higher loans, performing loans, loans loan, loans higher, nonperforming loans, consumer loans, loan portfolios
Oil & Gas	oil gas industry, oil demand, crude oil prices, business oil, gas industry, gas markets, gas service industry, oil prices, oil companies, oil market
Supply Chain	demand inventory, inventory supply chain, inventory supply, supply chains, inventory levels, supply chain challenges, supply chain constraints, higher inventory, inventory management, supply chain disruptions
Interest Rates	rising rate environment, rising rates, rates economic, market rates, rates going, lower rates, rising rate, higher rates, term rates, increase rates
Litigation	court decision, courts, court case, future litigation proceedings, litigation subject discuss, court, litigation proceedings, future litigation, future litigation pertaining, settlement agreement
Regulation	regulatory requirements, conditions future regulatory, obtain necessary regulatory, future regulatory, include company failure, regulatory process, regulatory authorities, company failure, necessary regulatory, operations include company
Contracts	contracts think, contracts believe, contracts contracts, contract negotiations, long term contracts, limitation contracts, limitation contracts terminated, think contracts, large contracts, longer term contracts
Taxes	effective tax rate, tax rate quarter, tax rates, tax rate, tax rate going, effective tax, tax rate approximately, tax positions quarter, reduction tax, tax provision

Continued on next page

Table 1: Topic Representations

Topic Name	Representation
Inflation	inflationary pressures, economy inflation, continued inflation, economic inflation, inflation recession, inflation economic, regarding inflation, rising inflation, inflation going, increase inflation
Foreign Exch.	foreign currencies, impact foreign currency, currencies detailed sec, foreign currency, foreign currency exchange, foreign countries currencies, foreign currency fluctuations, conducting business foreign, business foreign countries, foreign currency translation
Banking	banking industry, banking business, banking environment, corporate banking, commercial banking, banking management, banking sector, banking lines business, community banking, banking relationships

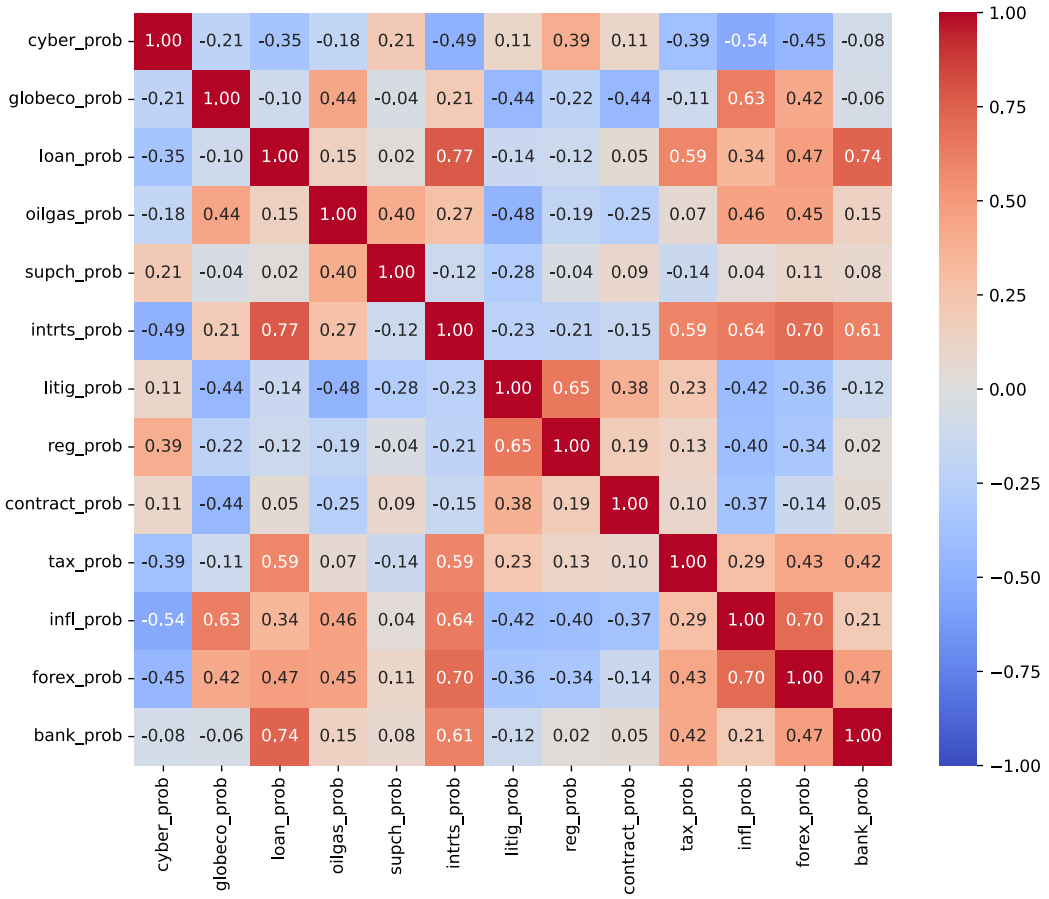
Topic representations are just one aspect, and it very well could be that the top 10 terms look promising but the topic as a whole is not coherent. As an additional validation step, we evaluate to what extent each pair of topics are similar. If indeed we have a good measure of meaning, then we should be able to see intuitive relationships between the topics. That is, two topics we know are similar should also be similar in the eye of the model.

There are different ways to measure the similarity of topics, a simple one is based on the continuous measure. Each document that the model analyses gets assigned a vector of probabilities for each topic. So, every document has *some* probability of being every topic, even if it is very low. What we do is correlate the probabilities for each topic within each document. A particular document might score high on the banking topic, the question we are exploring here is what other topics are scoring highly in that same document? More generally, which topics tend to have a high probability when banking risk does? Extend this to every pair of topics and the results are in Figure 4. Again, we should note that an assumption of the model is one topic per document. In so far as this is violated, these correlations could also capture correlations in the risks themselves. In other words, if a single risk snippet contains more than one risk, then these probabilities may also be correlated because the two risks appear at the same time. However, we are assuming for our analysis that the one-topic assumption is a reasonable approximation, given documents are only 21 words long.

Topics that have a high correlation are similar in the sense that they have a shared language. We previously asked which risks' probabilities co-move with banking, and we find reassuringly that the top three are loans, interest rates, and foreign exchange. Some more examples that are positively correlated are regulation and litigation; the global economy and inflation; and banking and loans. There are more examples, but these findings indicate that the similarity between topics is intuitive.

This exercise in itself has an interesting finding. Namely, cybersecurity is most related to regulation. This would seem to indicate that the cybersecurity topic has a lot to do with the actual regulation surrounding it (e.g. GDPR in the EU) and is not exclusive to actual risks of hacking and data breaches.

Figure 1: Topic Similarities
This figure shows the similarities between topics. Every risk snippet has a probability vector with an element for every topic. We correlate these probabilities within each risk snippet. A higher probability indicates shared language between topics.



The next step is we must establish that the model is actually able to classify documents into the correct topic. To do this, we manually validated 320 risk snippets. 260 of these pertain to potential false positives, i.e. they are classified as one of our 13 topics, and 60 are those that were classified to topics other than the 13 we selected (possible false negatives). Table 2 shows the results.

Let us assume for this analysis that the error rates are on the model level and do not change from topic to topic. Then, we can estimate what the actual model error rate is using the Wilson score to calculate a confidence interval. In our sample, 21/260 were false positives, so our true error rate lies within 5.3% and 12.0% to a 95% confidence level. We found 7/60 false negatives, since this is a much smaller sample the interval is larger at 5.8% to 22.2%.

These error rates are not trivial. In particular, we must be concerned about the selection of which documents are incorrectly classified. If documents from one topic are systematically being classified in another topic, then the latter will be biased with whatever effects the former risk has. Even if it is random which topic is misclassified, average effect sizes of other risks are unlikely to be the same as that for the topic the documents are being wrongly assigned to. This may introduce some bias in the coefficients towards the average of all risks as the measure is not a pure signal of one topic. For the false negatives, we worry about

which types of documents go unclassified. It is highly likely not to be random, but instead, it is the topics that offer a slightly weaker signal and/or mention other topics as well. This could mean we only lose power, or if subtle mentions of risk signal something different then we could be missing out on part of the measure of risk.

Table 2: Manual Quality Sample Results

This table shows the results from 320 manually validated randomly generated samples. One sample here corresponds to a single risk snippet. 260 samples are from documents the model classified as one of the risk topics we study (20 per topic), 60 correspond to cases when the model classified them as a topic other than the thirteen being studied ("None"). # are the number that were incorrectly classified. % is this number divided by total sampled in that category.

	Cyber	**Globeco**	**Loan**	**Oilgas**	**Supch**	**Intrts**	**Litig**	**Reg**	**Contr.**	**Tax**	**Infl**	**Forex**	**Bank**	**None**
#	6	0	0	2	0	0	1	3	1	6	0	1	1	7
%	30%	0%	0%	10%	0%	0%	5%	15%	5%	30%	0%	5%	5%	12%

We also decompose the percentage of transcripts that registered as having some exposure to each of the risks in Table 3. In line with the previous concerns over false negatives, most risks are in less than 5% of transcripts. Just under a third of the transcripts have mentioned at least one source of (our thirteen) risks. There are two possible explanations. The first is that there are simply not that many discussions that are a clear signal of a certain topic. Many might be ambiguous or only loosely related which would (and should) not be classified as a certain topic. This could be an issue if these still offer *some* information about the topic. Another source of this error is when the one topic per document assumption is violated. In this case, it might be a clear signal of two topics but of course, only one can be chosen.

Nevertheless, the vast majority of documents are correctly classified. Hence, we feel comfortable moving on with the analysis but need to keep in mind how these factors might affect the interpretation of our results.

Table 3: Percentage of Transcripts with Discrete Risk

This table shows how many transcripts have at least one mention of each risk using the discrete measure. Total represents how many transcripts have at least one of *any* risk, since transcripts can mention more than one risk, this total is not the sum of topic specific percentages.

Risk	Percentage > 0
cyber_dis	4.33 %
globeco_dis	5.91 %
loan_dis	3.43 %
oilgas_dis	5.38 %
supch_dis	4.07 %
intrts_dis	2.27 %
litig_dis	2.51 %
reg_dis	1.97 %
contract_dis	2.02 %
tax_dis	2.92 %
infl_dis	1.55 %
forex_dis	1.50 %
bank_dis	1.01 %
Total	30.73 %

4.2 Aggregate Risk

We now aggregate the risk across transcripts for each quarter. These aggregate patterns are interesting contributions on their own given many topics do not have existing measures of risk. We will also analyse the co-movements between risks on the aggregate level. More importantly, however, this serves as another validation check. Aggregate patterns in risk should be intuitive, as an obvious example we should see loan and banking risk increase around 2008. To facilitate this sort of analysis, we selected outside validation indices for most topics that establish some historical context. To aggregate them, we averaged together the topic measures across all transcripts in a given quarter. Averaging rather than adding them controls for the amount of data we have in each period. Both the discrete and continuous aggregate measures of risk can be found in Figure A.4.

We find that, in general, both the discrete and continuous variations of the measure qualitatively complement each other. We will reiterate here that the two are measuring slightly different things: the discrete measure describes the overall *level* of risk, whereas its continuous counterpart measures the *relative* level. This is why, for example, supply chain and many other risks that spiked in 2020 in the discrete measure, dipped in the continuous one; there was a much stronger signal of other risks such as the pandemic or the global economy (sub-figure (d)) which mechanically brought down all other risks.

Without being a subject matter expert, it is difficult to judge whether the aggregate patterns are in fact intuitive. The validation measures in Table 4 were plotted alongside our discrete aggregate indices in Figure 2. Contract and litigation risk do not have a measure because, to the best of our abilities, we could not find one that was representative of the risk and covered any significant period of time. The remaining measures are intended to reflect some aspect of the topic that we deem informative of the aggregate risk. This means that most are not (and certainly should not be) directly correlated with our measures of risk.

Table 4: Validation Indices

This table outlines the validation measures we use at the aggregate level. Figure 2 plots these alongside our discrete measure of risk. Not all measures are intended to be perfectly correlated with our risk measures, but they reflect *some* aspect of the topic that is useful to compare to. Two are missing a measure due to lack of available data.

Risk	Validation Measure	Definition	Reference
Cybersecurity	Nasdaq CTA Cybersecurity Index	Tracks the performance of cybersecurity companies.	Nasdaq (2024)
Global economy	Cboe Volatility Index (VIX Index)	Implied volatility of the S&P 500 index.	Cboe Global Markets (2024)
Loans	Delinquency rate on all loans, all commercial banks	Delinquency rate for all loans and all commercial banks in the U.S.	Federal Reserve Bank of St. Louis (2024a)

Continued on next page

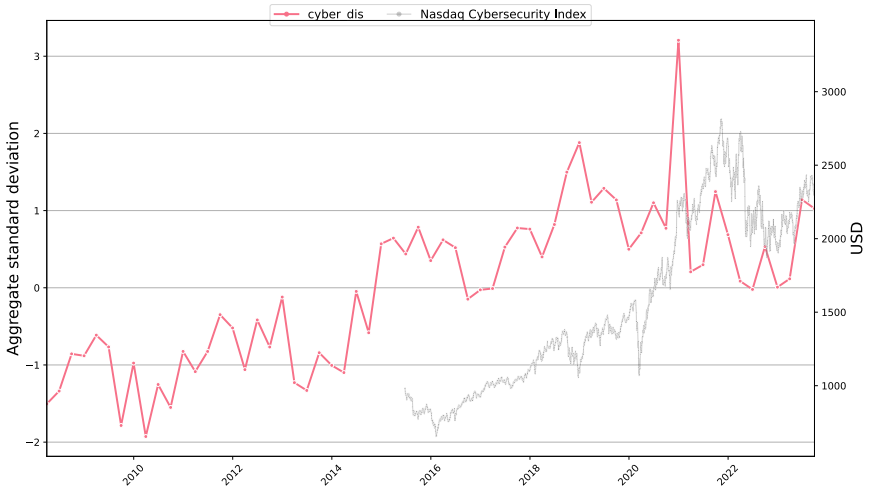
Table 4: Validation Indices

Risk	Validation Measure	Definition	Reference
Oil & gas	Global price of WTI crude, U.S. dollars per barrel	Crude oil prices.	Federal Reserve Bank of St. Louis (2024c)
Supply chain	Global Supply Chain Pressure Index (GSCPI)	Combines data on transportation costs and manufacturing to measure the state of supply chains worldwide.	Federal Reserve Bank of New York (2024)
Interest rates	Federal Funds effective rate	Federal Funds rate in the U.S.	Federal Reserve Bank of St. Louis (2024b)
Regulatory	U.S. Economic Policy Uncertainty Index	Measures policy-related economic uncertainty.	Economic Policy Uncertainty (2024)
Tax	U.S. federal corporate tax rate	Historical U.S. federal corporate tax rates.	Trading Economics (2024)
Inflation	Consumer Price Index (CPI)	Tracks inflation by measuring changes in the price level of a market basket of goods.	U.S. Bureau of Labor Statistics (2024)
Forex	U.S. Dollar Index (DXY)	Measures the value of the U.S. dollar relative to a basket of currencies.	The Wall Street Journal (2024b)
Banking	Dow Jones U.S. Banks Index	Tracks the performance of companies in the banking sector.	The Wall Street Journal (2024a)

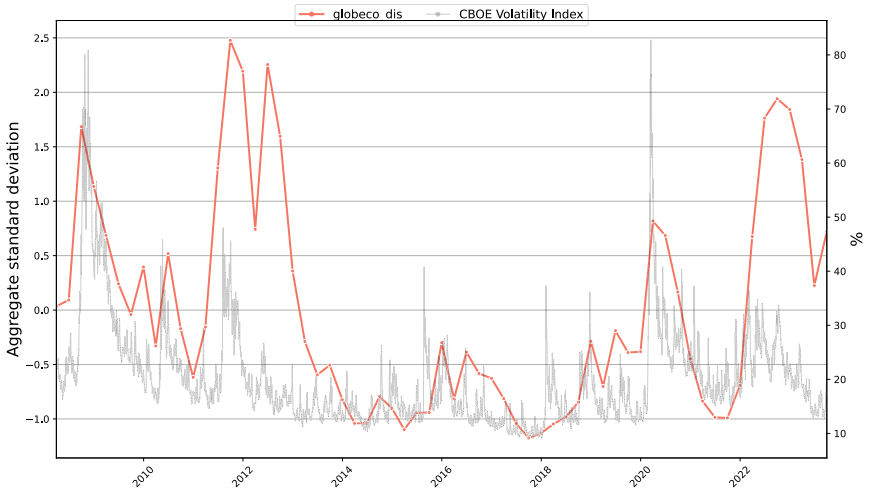
Figure 2: Aggregate Measures with Validation Indices

This figure shows the aggregate discrete measures of risk broken down by individual risk types including the validation measures outlined in Table 4. Our measure always uses the left axis, the validation index on the right one. Aggregate measures of risk are calculated by averaging the level of risk across transcripts. They are then standardised on the aggregate level.

(a) Cybersecurity



(b) Global economy



(c) Loans

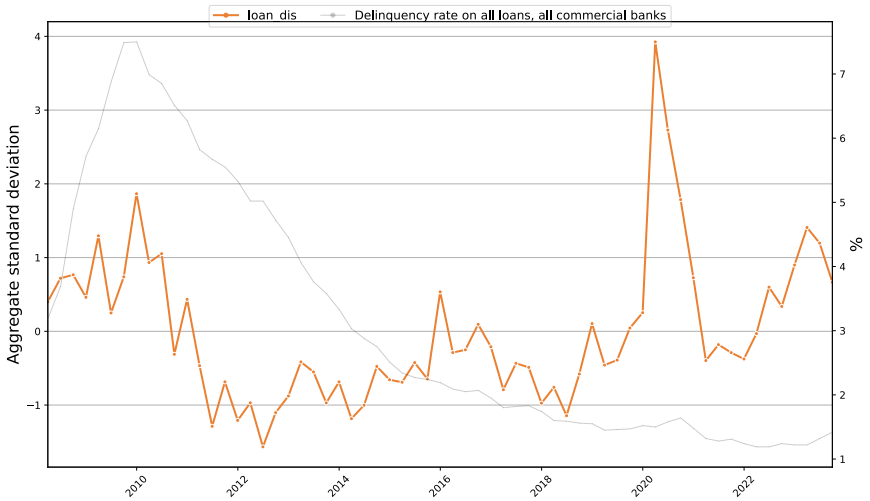
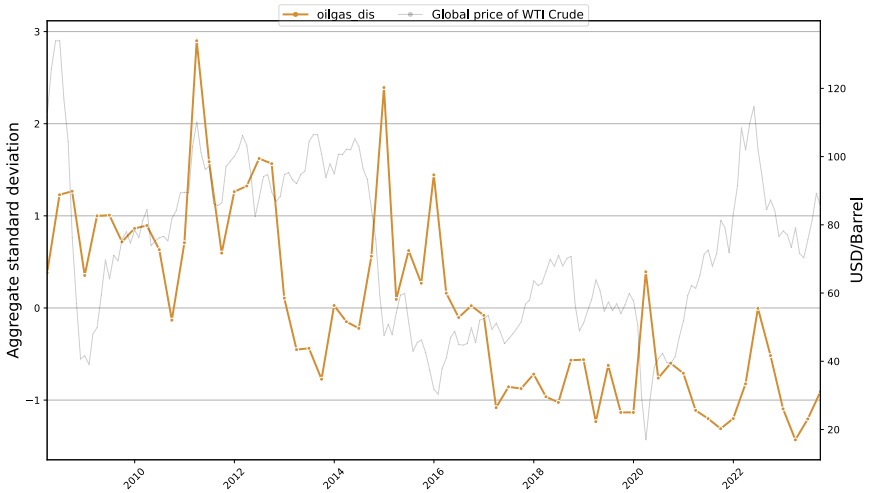
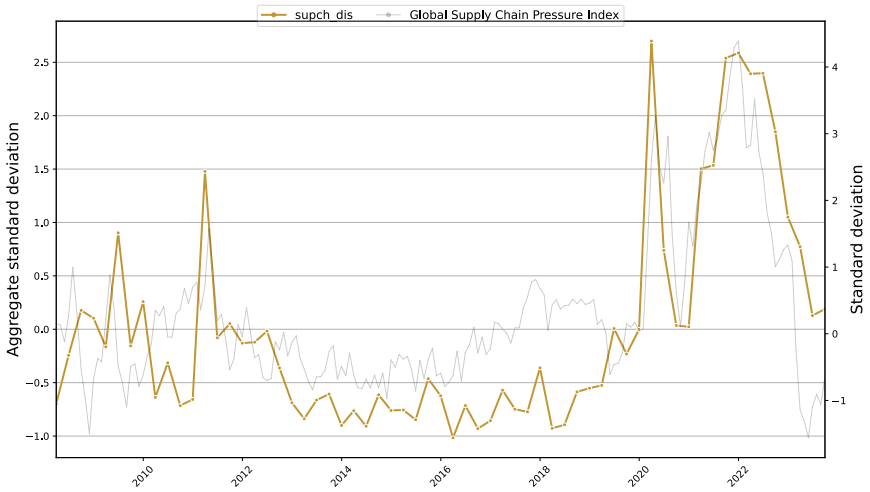


Figure 2
Continued

(d) Oil, gas, & energy



(e) Supply chain



(f) Interest rates

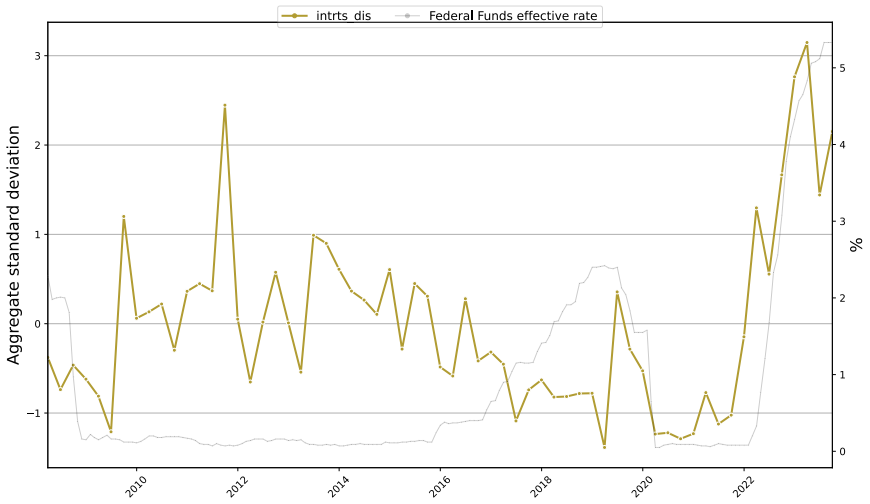
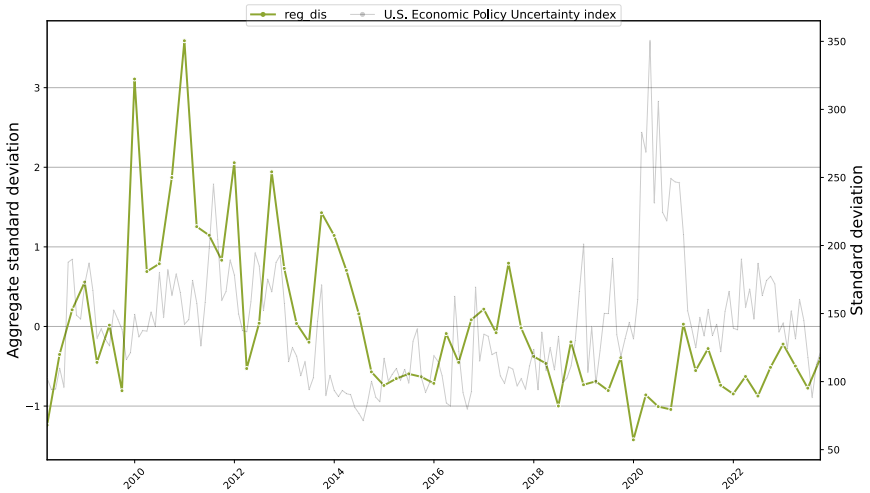
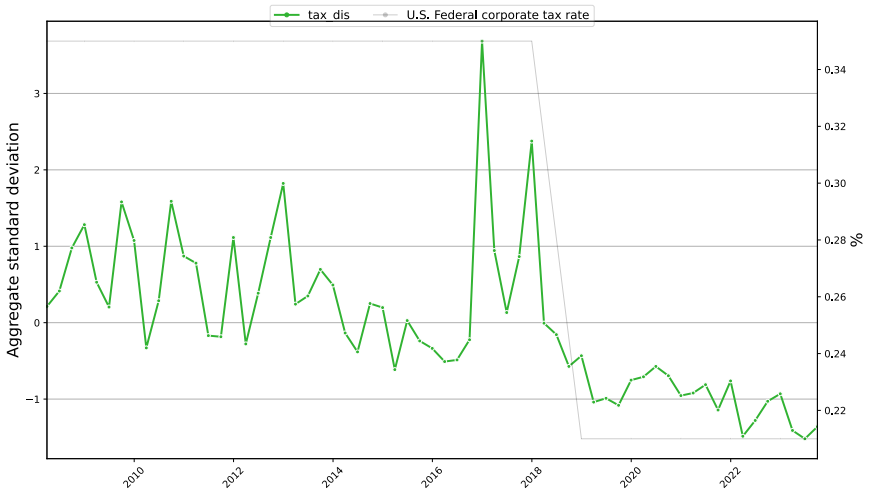


Figure 2
Continued

(g) Regulation



(h) Tax



(i) Inflation

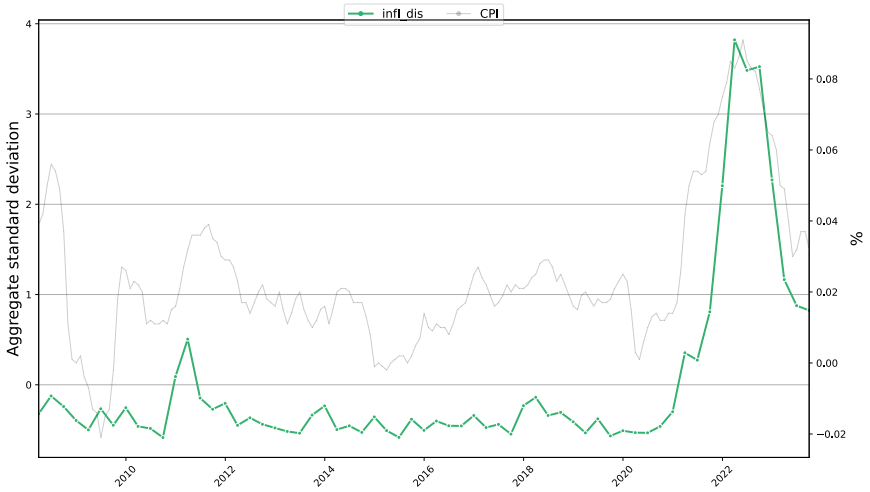
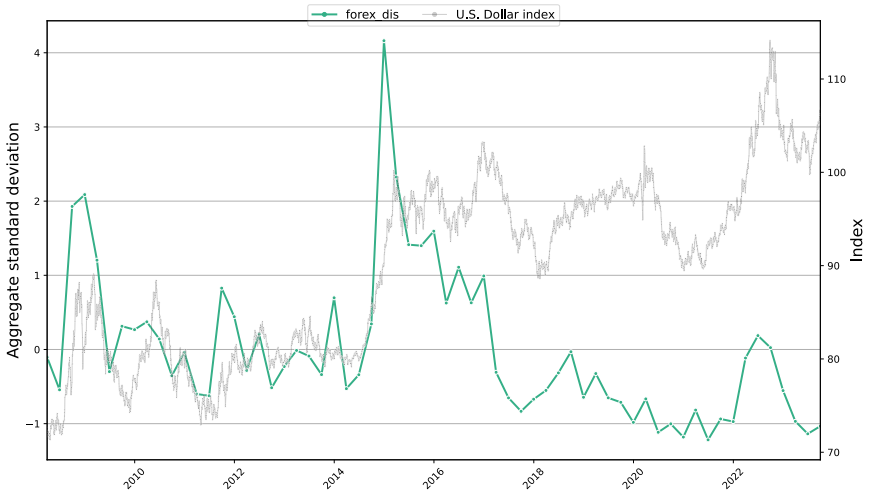
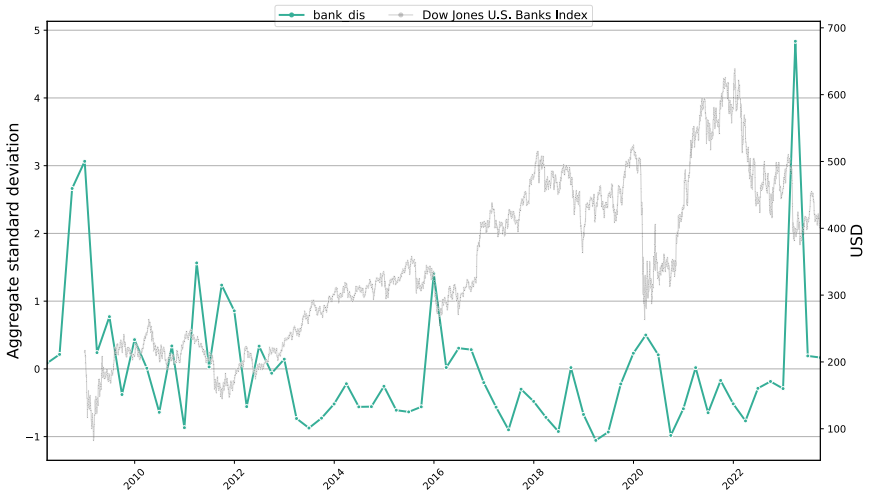


Figure 2
Continued

(j) Foreign exchange rates



(k) Banking



We now dedicate a discussion to each of the aggregate measures in Figure 2 to provide more historical context when needed and observe when the risk measures do and do not line up with their validation counterparts.

Cybersecurity is of recent interest which is why data is more limited. Nazli et al. (2021) apply the methodology in Hassan et al. (2019) to create a cybersecurity risk measure. They note that ransomware attacks have become more frequent in recent years. Governments have also started enacting new policies as concerns grow. This is reflected in the trend of perceived cybersecurity risk, up until recently. We also notice that the shares of cybersecurity firms rallied amongst the spike in risk in late 2020 and early 2021. Afterwards, as risk began to trend downwards again, we see these firms have a drop in share price, presumably because the risk of cybersecurity attacks diminished and the market for their services was expected to have less demand.

The global economy needs less introduction; there was a major spike in 2008 and in the subsequent Eurozone crisis. Another spike is at the onset of the pandemic, and the next spike is during 2022 and 2023 when there were fears of a recession (Washington Post, 2023). In general, most spikes in the VIX correlate with the risk measure, as expected, since in this case the measures are very similar. In fact, using the quarterly measures of the VIX the two have a correlation coefficient of about 0.7.

Loan shows a clear spike in 2010, as is expected since in the aftermath of the 2008 crisis delinquency rates reached their highs for this period. The spike in 2020 is a good example of when perceived risk may not necessarily be realised.[13] Whilst there was a small uptick in delinquency rates after the pandemic, it was much smaller than might have been expected.

Oil & gas risk at first glance appears not to correlate well with oil prices. One should observe, however, that spikes in the risk index are almost always correlated with large *movements* in price, not necessarily in any one direction. The spikes in 2011, and 2022 correlate with large upward movements in oil, whereas 2008, 2014, 2016, and 2020 come with large downward movements. Indeed then this is a very pure indication of uncertainty.

Supply chain risk has been a major issue thrice in recent years. Once around 2010, 2020 with the pandemic, and another time during the global supply chain crises of 2021-2022. The risk measure fits this narrative perfectly, even the relative levels of the measure line up well with the validation index. That is, the size of the spikes is correct, which is not always true in our other measures.

Interest rates in the latter half of the period correlate well with our measure of risk. One might note that the large spike ahead of 2012 is not proceeded by any changes in interest rates, which, as mentioned previously, is not incompatible with our measure. We are measuring the *perceived* risk of changes in the interest rate, but the very fact that we measure uncertainty means that the risk may never be realised. It is very plausible that firms expected rate increases after the dramatic drop in 2008, but over the years, as the Federal Reserve kept rates low, such perception began to dissipate. Confirming this is a study by Qadan, Shuval, and David (2023) where they created "a VIX-style measure of forward-looking volatility about Treasury yields to capture uncertainty about interest rates". Indeed their measure spikes in 2009 and then again in 2010 and 2011, decreasing after that and increasing again only in 2020.[14] Conversely, Ying (2023) in their study of earnings calls find

[13] This is true for any type of risk, but especially when what we measure can be an incorrect perception of risk.

[14] Ideally we would use this measure to compare to ours instead, but the data was not available for our

that perceived interest rate risk has steadily decreased between 2003 and 2021 until it spikes again, so we might have a reason to be less confident in this result.

Regulation until 2018 exhibits a similar trend to the U.S. Economic Policy Uncertainty Index. Amidst new banking regulations imposed after the 2008 crises, the rise and then gradual decline of this risk makes sense. In 2020, however, the EPU index spikes whereas our measure remains flat. One obvious explanation is that the EPU is not regulatory risk and that the uncertainty here pertained to other political issues (such as the election). What this confirms is that this risk indeed is a narrow measure of specifically regulatory risk, and does not reflect broader political uncertainty per se like that of Hassan et al. (2019). Calomiris et al. (2020) built their own measure of aggregate regulatory risk using conference calls and found a similar trend, although the actual fluctuations are different. However, they also compare their measure to others in the literature and find that they all differ dramatically.

Tax risk spikes just before the only major related event that occurred in our sample, that is the *Tax Cuts and Jobs Act* of 2018. This cut the corporate from as high as 39% to a flat 21%. Clearly, there was a lot of uncertainty preceding the act which is reflected in the aggregate plot, and all the uncertainty was resolved after the policy was enacted. Of course, one might be concerned over when the law was *passed* not when it was enacted. To this end, it was not even introduced until November 2nd, 2017, around the first spike.

Inflation figures seem to correlate closely with the aggregate risk measure, with the exception being in 2008. Chava et al. (2022) built their own measure of inflation risk using conference call transcripts, and our aggregate measure lines up very closely with theirs with the exception again in 2008. This may be because firms exclusively discussed the issue in conjunction with another topic. In this case, we might have some documents that were not classified as inflation even though they mentioned it. Indeed in Figure A.4, we do see that the continuous measure rose around that time which would indicate that there were discussions of inflation even if they were not classified as such by the discrete approach.

Foreign exchange rate risk seems to spike when there are sudden or large movements in the value of the US dollar. Notably, 2008, 2010, 2012, 2015, and 2023. We will note though that whilst there seems to be a relationship here, it is not perfect. Interestingly, the risk is trending downward, one explanation is that despite there still being large movements in the dollar after 2015, there are also extended periods of relative calm. Before 2015, when risk on average was higher, the dollar appeared to be more volatile. If we split the period into before and after 2015, we find that the standard deviation of the per cent daily index changes in earlier years is 20-25% higher than in later years. This is robust regardless of whether we set the split to 2014, 2015, or 2016, so it is not explained by the sudden break in 2015.

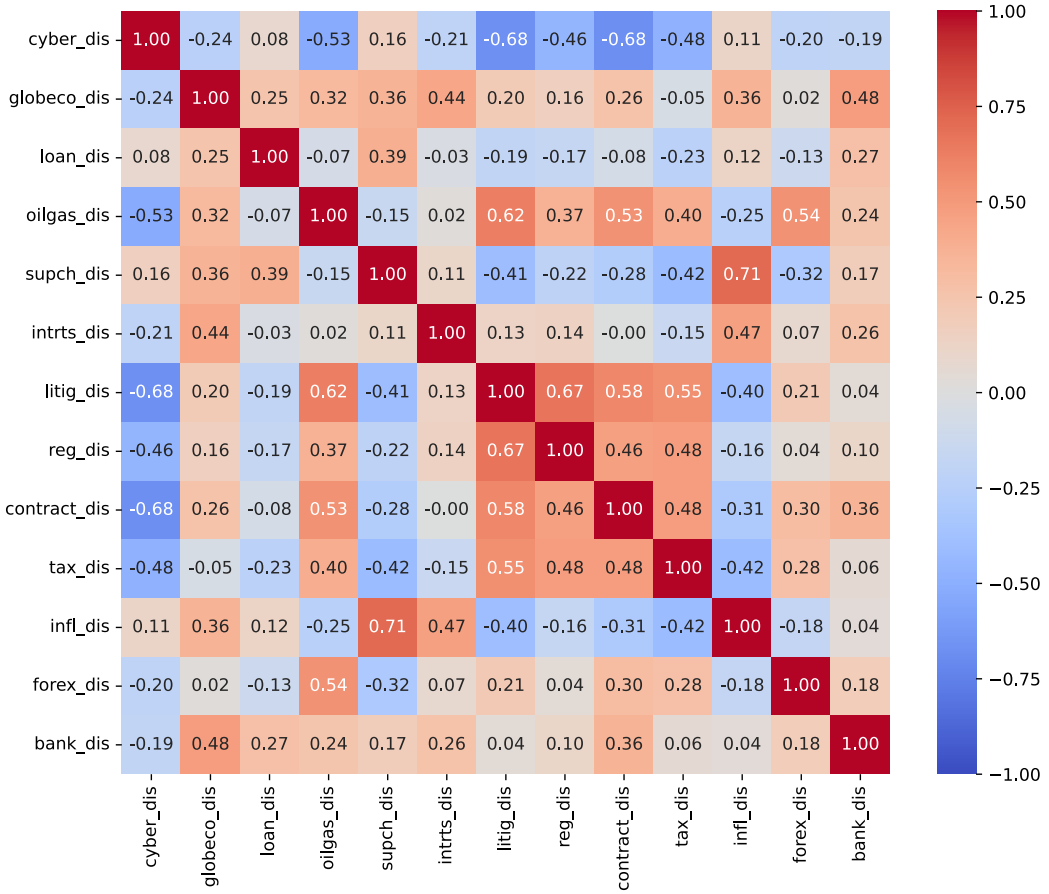
Banking risk spikes appear to be highly correlated with drops in the banking stock index. Furthermore, the risk spikes as expected in 2008, and again in 2023 around the collapse of Silicon Valley Bank and First Republic Bank, and rising fears of further contagion.

Figure 4 also notes the correlation between the aggregate (discrete) risks. We find here that banking risk is positively correlated with the global economy, loan risk, interest rates, and contract risk. Contract risk is likely to be in large part counter-party risk, and indeed we saw it spike around 2008 when many firms were failing (Figure A.4). So this result seems justifiable. Other notable cases that are positively correlated are oil & gas and foreign

own plotting.

exchange; inflation and supply chain; inflation and interest rates; regulation, litigation, tax, and contracts. Given we have a relatively short time series, and this is on the aggregate level, these correlations should only be an indication. For example, litigation and cybersecurity risk are negatively correlated. We do not interpret this as there being any real relationship here, it just so happens that one is trending up whilst the other down.

Figure 3: Aggregate Topic Correlations
This figure shows the correlations between aggregate (discrete) risk measures shown in Figure A.4. Risk was aggregated by averaging the measure for each topic across transcripts in a given quarter.



What we have shown in this section is that the aggregate risk patterns generally reflect real events in a way that would be expected. Furthermore, the aggregate correlations between risks are intuitive. Combined with the topic representations and correlation exercises of the previous sub-section, we thus argue that indeed we are able to capture events for the actual topics which the model has produced.

4.3 Volatility & Returns

The next question that logically follows is whether we are actually measuring risk. High levels of risk would indicate that a firm is exposed to large potential shocks that can significantly change the value of the company. That is, we would expect riskier firms on average to have higher volatility in their publicly traded securities. To test this, we focus our analysis on the relationship between each firm's measures of risk, and the realised stock volatility over

a 90-day holding period. If what we measure is risk, there should be a positive relationship.

Of course, this will only apply to our discrete measure of risk, since its continuous counterpart inherently controls for the overall risk. Our analysis is thus focused on the former, but the results from the latter may still be interpreted as the relative importance of each risk. Our main specification for either measure is as follows:

$$Vol_{it} = \delta_t + \delta_i + \beta Risk_{it} + \lambda ln(A_{it}) + \epsilon_{it} \tag{1}$$

Where Vol represents a 90-day holding period starting on the day the conference call was held, δ_t the set of time fixed effects, δ_i firm fixed effects, $Risk$ represents the vector of either the discrete or continuous risks variables (β is the corresponding vector of coefficients), and $ln(A)$ is the log of the firm's assets. All errors are clustered on the firm level as motivated in Hassan et al. (2019).[15] We also tested regressing each risk individually, but this did not change the conclusions of the analysis. We include log of assets as a control for the size of a firm since this is a strong predictor of its volatility and could easily be correlated with aspects of our measure. Our set of controls is motivated by the specification in Hassan et al. (2019), however, we will also explore other alternatives.

We should note that several later authors (e.g. Kim et al. 2023; Wu 2024; Nazli et al. 2021) omit the firm fixed effects in their main specifications. In some cases, this does not significantly change their analysis, but for Kim et al. (2023) it turned their coefficient for cybersecurity risk insignificant. There is a reason this might be appropriate as it is possible that firms have different permanent exposures to risk which are revealed based on whether they discuss the risk or not. Such an effect would be omitted if we were to control for firm fixed effects. For this reason, we test this alternative specification as well where instead of firm fixed effects we use sector.

Before analysing the topics, we should first establish that the discussions of risk themselves do hold information about risk and future volatility. That is, rather than introducing any topic-specific risks, we run our main specification on just the total risk. Total risk is defined as the number of times a firm discusses risk in a transcript. The results are in Table 5. We see that indeed discussion of risk is generally followed by periods of higher volatility, even after controlling for time and firm fixed effects. This indicates that there is variance within this measure that is informative about risk.

[15] They run a series of falsification exercises and show that in simulations the distribution of the resulting t-statistic is correct.

Table 5: 90-Day Volatility Regressed on Total Risk

The following table regresses realised stock volatility over a 90-day holding period on total risk. Total risk (*context_count*) is defined as the number of synonyms for risk or uncertainty in a transcript, after filtering for safe-harbour statements. Realised stock volatility is measured as the standard deviation of investor returns over a 90-day holding period beginning on the day the conference call was held. We control for the log of firm assets (*log_atq*). Volatility is winsorised at the 1st and 99th percentiles and then standardised. Sector is defined as the two-digit SIC code. All errors are clustered at the firm level.

	(1)	(2)
VARIABLES	90 Vol	90 Vol
context_count	0.00742***	0.00267***
	(0.00104)	(0.000600)
log_atq	-0.196***	-0.116***
	(0.00397)	(0.0111)
Constant	1.427***	0.851***
	(0.0311)	(0.0832)
Observations	120,869	120,894
R-squared	0.504	0.657
Time FE	Yes	Yes
Sector FE	Yes	n/a
Sector x Time FE	Yes	No
Firm FE	No	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

We also find that firm fixed effects do make a large difference to the effect size. In fact, the coefficient becomes just 35% of the size when only sector fixed effects are used. This would seem to indicate a large portion of the information signalled about risk in conference calls pertains to permanent differences between firms rather than the variation over time. Of course, this only holds true to the extent that there are no other unobserved differences that can explain the change in coefficients. Certainly, these may exist, such as a CEO who tends to not share much information when risks are high. Hassan et al. (2019) tested this particular possibility with CEO fixed effects but found no notable difference in their results.

With the conviction that there is risk to be measured in the conference calls, we now move on to breaking this down by topic. Figure 4 runs our specification in Equation 1 as well as the alternative that replaces firm with sector fixed effects. Table A.2 and Table A.3 show the full regression coefficients and additional specifications as well. For the discrete measure, we find that with the sector fixed effects, 7 out of 13 of the risks are positive and significantly associated with realised volatility to a 5% significance level or more: loan, oil & gas, interest, litigation, regulatory, contract, and banking risk. After controlling for firm fixed effects, only 2 remain significant: loan and banking risk. We argued though that such a control would remove any permanent differences in risk that we measure. The fact that the majority of risks then become insignificant would suggest that what our measure picks up is precisely these differences. Of course, we cannot rule out that there is a confounding factor as well, but much of the literature has argued for the validity of omitting firm fixed effects (e.g. Kim et al. 2023; Wu 2024; Nazli et al. 2021).

The discrete measure is the number of mentions of a particular risk in a transcript, all other units are standardised. After considering measurement noise, the coefficients are economically significant. One mention of banking risk, before controlling for firm fixed effects, is associated with around 0.1 standard deviation increase in the firm's volatility over the next 90 days. Given also that banking risk is a *small* aspect of the overall risk profile of a firm, this is large. Additionally, this is over 13 times larger than the coefficient for our overall risk in Table 5[16]. This means that either banking risk is very important, or our overall measure has a lot of measurement noise and the topic-specific measure provides a cleaner signal. The continuous measure tests the relative importance of a given topic, and since banking risk is not significant in either direction, it suggests the latter result is more likely.

There are also some negative coefficients in both specifications, namely, global economy, interest rates, and inflation (depending on the controls). One explanation is that for certain risks, discussing them may signal increased risk on the aggregate level since indeed the aggregate indices made sense, but on the firm level, they indicate that the firm is more prepared. That is, firms will discuss a given risk more when it is more important, but the fact that they discuss it also signals that they are more prepared and hence we would expect volatility to be lower. The relationship between stock volatility and the risks where we find negative coefficients has not been studied previously using conference calls. Hence, this is a novel finding and might indicate that not all risks can be measured on the firm level using these transcripts.

We can also further interpret the continuous measures (Figure 4. We find that after controlling for firm fixed effects, loan and litigation risk are more important than the "average" topic to a 1% significance level. The topics it is being compared to are the 48 in Table A.1. This interpretation may not hold if these risks are correlated with the total risk, however, because then they do reflect not just the composition but the level of risk. We checked this possibility, however, and found the litigation continuous measure is actually negatively correlated (-0.06 R^2) which would make our original conclusion slightly stronger. Loan risk on the other hand is positive but with a very small R^2 of 0.05. Litigation risk being of above average importance raises the question of selection in our measure. If a lawsuit has come to the attention of company executives and is important enough to be discussed on an earnings call with analysts, it is likely to be a large case. We then might presume here what we measure is not the relationships between litigation in general, but only the tail end of the largest cases. In this case, we would certainly expect it to be a very important risk.

In both specifications, we found supply chain and cybersecurity risk to not have significant coefficients, contrary to the findings of Wu (2024) and Nazli et al. (2021).[17] Wu (2024) actually use concurrent quarter volatility which would include on average 30 days of stock volatility *prior* to the earnings call, and Nazli et al. (2021) use only a thirty-day window, so it is possible that the results are highly sensitive to this assumption. Of course, this could also indicate that classification algorithms (like ours) are not well suited for these risks. It is possible that they violate an assumption of the model such as being commonly mentioned in conjunction with another topic. An alternative is that the relationship for these topics does not hold in our time period.

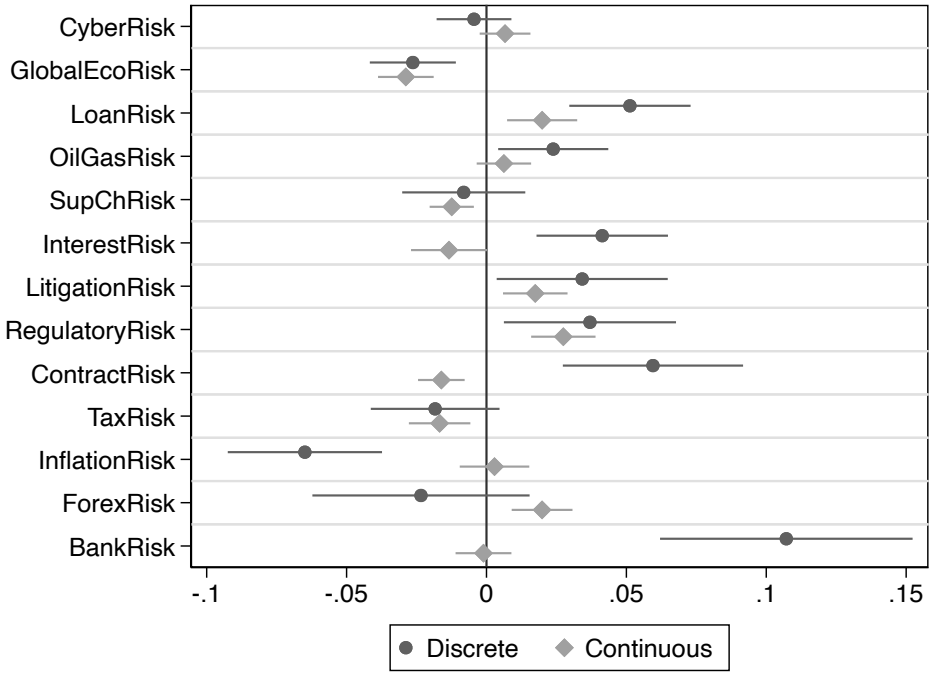
[16]These can be directly compared because total risk is the number of mentions of risk, and the discrete measure is the number of mentions of a particular risk, hence we have the same units.

[17]Note that for the other risks that have been covered by previous papers such a comparison is not available.

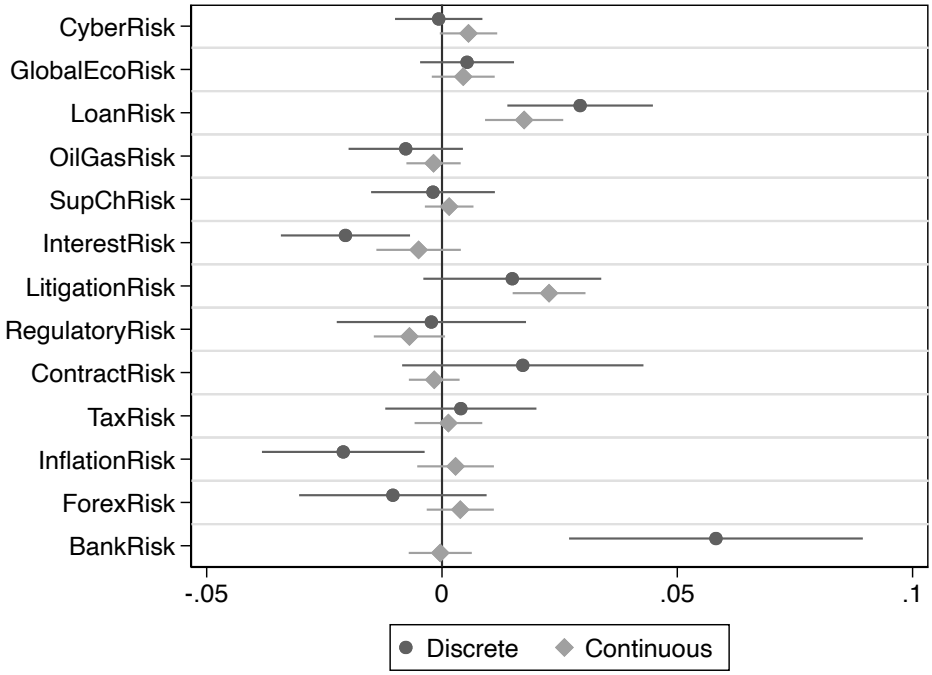
Figure 4: Realised Volatility Regression Coefficients

This figure summarises the 95% confidence intervals for the regression of a 90-day holding period volatility on each of the risk factors. Exact coefficients and standard errors can be found in Figure A.2 (discrete) and Figure A.3 (continuous). Sub-figure (a) includes time, and sector fixed effects as well as their interaction. Sector is defined by the two-digit SIC code. Sub-figure (b) is the specification in Equation 1 and includes firm and time fixed effects. Both specifications include the logarithm of firm assets as a control. Volatility is calculated as the standard deviation of daily returns over a 90-day holding period beginning on the day of the conference call. It is winsorised at the 1st and 99th percentiles and then standardised. The continuous measure is standardised, the discrete measure is the number of mentions of that topic in a transcript. All errors are clustered on the firm level.

(a) Sector and time fixed effects



(b) Firm and time fixed effects



We need to investigate further why the firm fixed effects made such a large difference. If it the case that firms have large permanent differences in risk exposure, and this is what we are measuring, it would be a mistake to add this control. Conversely, if we are missing some confound that is not accounted for by firm size or sector differences, then our results are spurious.

One possibility is that our measure is a lagging indicator of risk. That is, firms discuss a risk only after it has already had a chance to affect the firm in some way. This would mean that a recent previous period could see higher volatility, and only then our measure would increase. In this case, we might still see a positive relationship in future volatility because it is auto-correlated (Mandelbrot, 1967). It is also possible this relationship is only one-sided in that firms discuss risk after a negative shock. To test these possibilities we ran two regressions in Table 6. The first repeats our main specification in Equation 1 but substitutes future volatility for past volatility. The second does the same but with previous 90-day average returns.

Table 6: Regression Test for Lagging Risk

The following regressions test lagged volatility and returns on our risk measures. Regression (1) regresses the previous 90-day volatility on each risk measure, including firm and time fixed effects. Regression (2) is the same but with previous 90-day returns. Lagged returns and volatility are calculated using a 90-day holding period ending the day prior to the earnings call. All errors are clustered on the firm level. Volatility is winsorised at the 1st and 99th percentiles and then standardised. Returns is also standardised. "*volatilityL*" is the previous 90-day volatility. "*log_atq*" is the log of assets.

	(1)	(2)
VARIABLES	Previous 90-day vol	Previous 90-day ret
cyber_dis	-0.00150	0.00551
	(0.00440)	(0.00434)
globeco_dis	0.00711	-0.0141**
	(0.00483)	(0.00701)
loan_dis	0.0298***	-0.0275***
	(0.00828)	(0.00820)
oilgas_dis	-0.0108*	-0.0305***
	(0.00643)	(0.00930)
supch_dis	-0.00495	-0.0291**
	(0.00674)	(0.0123)
intrts_dis	-0.0189***	0.00577
	(0.00713)	(0.00927)
litig_dis	0.0287***	-0.0245**
	(0.00977)	(0.0124)
reg_dis	-0.00458	0.00709
	(0.0107)	(0.0149)
contract_dis	0.0112	-0.0338**
	(0.0140)	(0.0159)
tax_dis	0.00450	0.0158
	(0.00836)	(0.0117)
infl_dis	-0.00771	-0.00865
	(0.00774)	(0.0132)
forex_dis	-0.0274***	-0.00857
	(0.00993)	(0.0142)
bank_dis	0.0493***	-0.0360*
	(0.0171)	(0.0211)
Observations	120,892	120,892
R-squared	0.668	0.189
Controls	Yes	Yes
Time FE	Yes	Yes
Firm FE	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

We will discuss two potential interpretations of the results in Table 6. First, as mentioned, conference calls may represent a lagging indicator of risk or shocks. The second is that markets might incorporate the information about increased risk before the conference call.

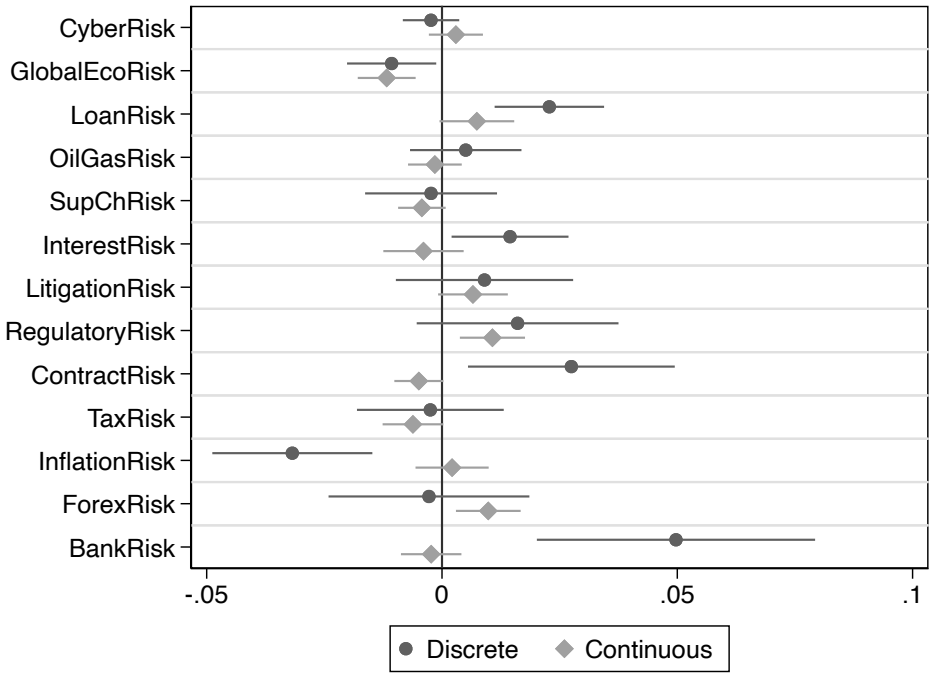
If we had a lagging indicator of risk, there would be a much clearer relationship between risk and previous 90-day volatility. However, only 3 are significant and positive to a 1% level, and 3 are also negative and significant (with an additional one to a 10% level). On the other hand, 6 of the risks are significant and negatively related to previous 90-day returns. Since those that are positive are not significant, it seems most measures are associated with negative returns in the previous quarter. It might be then that firms only discuss risks after difficult times when returns have been low or a negative shock has occurred. Still, it is not clear whether this would necessarily confound our estimate of future volatility.

The alternative interpretation is that markets recognise the increased risk the firm is exposed to before the conference call. This could be the case if the risks we measure correlate with negative changes in the firm's expected return. For example, a firm discussing a large litigation case against them now faces a chance of a large loss with no potential upside — even in the state of the world where they win, they would have had to pay large legal fees and possibly lost customers as the company's reputation worsened. It would thus be the case that any amount of this risk decreases the expected return of the firm. If investors realise this, the firm's valuation should (and does) fall even before the conference call has a chance to discuss it.

In any case, as a robustness check, we repeat our specification in (a) from Figure 4 with sector and time fixed effects, but additionally, control for previous 90-day returns and volatility to see if this has the same effect as controlling for the firm. The results are plotted in Figure 5 and the full regressions are in Table A.2 and Table A.3. Naturally, we do find attenuated coefficients, however, 4 remain positive and significant. This would seem to indicate two things. Firstly, part of the firm fixed effect difference may come from partially controlling for some lagged effect. However, we have also ruled out the possibility that this explains *all* the differences. After controlling for previous returns and volatility, we still find that 4 of our coefficients are significant and positive. It might be that the remaining differences are due to permanent exposures between firms, in which case we are in fact measuring risk.

We should argue that the reason this is not our primary specification is that by controlling for previous period risk, we are inherently restricting ourselves to only the initial variance of a stock during a period of elevated risk. If a rise in risk affects the volatility of more than one subsequent period, we can only exploit the effect of the first. This is because, after that initial rise in volatility, future periods will control for past volatility, and attenuate the effect of the risk itself. We also would not want to control for previous negative returns if the higher risk was being priced in by the market since this would again confound the effect of the risk itself.

Figure 5: Realised Volatility Regression Coefficients With Additional Controls
This figure summarises the 95% confidence intervals for the regression of a 90-day holding period volatility on each of the risk factors. Exact coefficients and standard errors can be found in Figure A.4 (discrete) and Figure A.5 (continuous). The regression includes time, sector, and their interaction fixed effects. Sector is defined by the two-digit SIC code. We additionally control for the size of firm assets, previous 90-day average returns, and previous 90-day average volatility. Previous 90-day implies a 90-day holding period ending on the day prior to when the conference call was held. Volatility is winsorised at the 1st and 99th percentiles and then standardised. The continuous measure is standardised, the discrete measure is the number of mentions of that topic in a transcript. All errors are clustered on the firm level.



5 Variance Decomposition

In so far as one is confident in our measure of risk, the next question we ask is where do fluctuations in each risk come from. Namely, how much of the variance comes from differences across time or sectors, and how much remains to be differences between firms. To answer this, we decompose the variance of each topic into time, sector, and their interaction fixed effects. A primary motivation is to evaluate the extent to which each risk is systematic. This is the same analysis that has been performed in Hassan et al. (2019) and Kim et al. (2023) and so we also wish to compare findings. Our results will lead to the conclusion that independently performing this variance decomposition on a single risk can be difficult to interpret, and we suggest that the results only be compared *between* risks.

The approach is simple, begin by controlling for only time fixed effects and record the R^2 of the regression. Add sector fixed effects to the same regression and record the incremental R^2. We then do the same for their interaction and for firm fixed effects. Each of these increments represents the share of variance that the fixed effect can explain. Residual variance is referred to as "firm level". This consists of a permanent component (firm FE), and one that varies over time. Table 7 shows the results of this for our 13 measures of risk.

We find remarkably low levels of variation explained by time fixed effects. However, this

Table 7: Variance Decomposition of Risk

This table shows the variance decomposition of each risk using its discrete measure. In the standard model, the percentages represent the incremental R^2 for each fixed effect (time, sector, and their interaction), and firm level is the residual variance after including all fixed effects. Breakdown decomposes the firm level variation into a permanent component (firm fixed effect) and the remaining residual is the variation of a firm across time, after controlling for all fixed effects. The beta model is constructed by regressing for each firm its risk on the quarterly mean, and then including the interaction of the estimated beta for each firm-topic pair with its quarterly mean in the standard model. The idiosyncratic component corresponds to the remaining firm level variance in the standard model that still cannot be explained. In calculating the beta model, around 6% of firms with only one observation are dropped. This can account for why idiosyncratic variation is marginally higher in some cases than in the standard model. Sector is defined as the two-digit SIC code.

	Standard				Breakdown		Beta model
Risk	Time	Sector	S × T	Firm Level	Firm FE	Resid	Idiosyncratic
Cyber.	0.11%	3.97%	1.17%	94.75%	45.38%	44.92%	84.13%
Global econ.	2.73%	1.05%	5.15%	91.07%	9.30%	73.44%	91.06%
Loan	0.22%	15.64%	2.34%	81.80%	19.77%	44.07%	81.82%
Oil & gas	0.36%	11.18%	4.00%	84.45%	10.93%	58.81%	84.43%
Supply chain	1.17%	2.70%	5.95%	90.18%	8.11%	73.09%	90.17%
Interest rates	0.28%	6.43%	2.54%	90.76%	12.80%	68.95%	90.77%
Litigation	0.25%	2.20%	3.51%	94.05%	18.93%	70.32%	94.04%
Regulation	0.14%	1.23%	3.74%	94.89%	8.23%	82.00%	94.91%
Contract	0.13%	1.14%	5.00%	93.73%	10.41%	77.65%	93.72%
Tax	0.57%	0.53%	3.81%	95.09%	8.44%	82.02%	95.08%
Inflation	3.03%	0.84%	5.59%	90.54%	5.75%	76.55%	90.56%
Foreign Exch.	0.55%	0.26%	3.15%	96.04%	10.74%	81.63%	96.03%
Bank	0.33%	3.78%	2.52%	93.37%	7.81%	79.20%	93.35%

is in line with the previous literature who repeat the same analyses on their set of risks (Hassan et al. (2019) for political risk, Calomiris et al. (2020) for regulatory risk, and Kim et al. (2012) for AI, climate, and again political risk). We reassuringly still find that global economy and inflation have by far the largest time effects. After these two, it is supply chain risk; this is not surprising in light of the recent (global) supply chain crisis. It is possible, however, that the results are primarily driven by just this event. To probe this theory we split the sample before and after (including) 2020, and found the latter to have nearly twice as much variance explained by time fixed effects despite it being a much shorter time period. Of course, more research would be needed to establish this as fact, however, it is worth noting that the systematic nature of a risk may change over time.

On the other hand, cybersecurity and contract risk comparatively have small systematic components. The former also has the lowest sector and time interaction variance. This would indicate that overall exposure to cybersecurity risk is fairly constant over time, relative to how much it varies permanently between firms and sectors. Given that our aggregate trend in Figure 2 is relatively steady with few sudden spikes, this would make sense. For contract risk we see a very high interaction effect, this means that the risk is still somewhat systematic but disproportionally affects key sectors.

Cybersecurity, interest rate, loan, and oil & gas risk have large sector effects. This would indicate that certain sectors for these risks are permanently exposed to different levels of this

risk. Conversely, for systematic issues like the global economy or inflation the sector fixed effects are small, suggesting few permanent differences across sectors. However, their time and sector interacted effects are the largest so certain sectors are disproportionately affected across time.

We also find that cybersecurity risk has by far the largest permanent difference in risk exposure between firms. In fact, nearly half of all topic-related risk can be explained this way. In line with our previous finding that there is little variation across time, this makes sense. The next highest is loan risk, since loan holdings are concentrated among few firms, this seems intuitive. On the other end, inflation has the lowest difference between firms. This is because the risk is very time-dependent, a firm in the middle of 2016, when inflation was low and there was little reason to suspect this would change, would have no reason to be talking about this. Hence, the variance is mostly across time.

Our time fixed effects are intended to capture systematic risk. However, if firms have heterogeneous exposure to the aggregate level of risk, this might fail to capture the full extent of how systematic a risk is. In order to investigate this possibility, we also test an alternative model of systematic risk exposure as described in Hassan et al. (2019). Namely, we create a firm-level beta for each risk and firm pair by regressing firm-level risk on its quarterly mean, and then including the interaction of the beta for each firm with the quarterly mean of each risk into the standard specification with full time and sector fixed effects. We find, as did they, that such a model does not explain well the variation in risk, with one exception, cybersecurity risk. In fact, for all other risks the firm level variation is almost identical to the "idiosyncratic" one. For cybersecurity, the model explains 10% of the variance which in comparison is a very large effect.[18] Combined with the fact that a large part of cybersecurity risk can be explained by permanent firm differences, it would seem to indicate there is much larger heterogeneity between firms with regard to this risk than the others. Specifically, firms have very different but relatively constant exposures to the aggregate risk, on top of the fact that each firm has a different level of the risk itself at all times. However, we cannot conclude that such heterogeneity in exposures does not also exist for the other risks as if the true betas were to vary over time, the model would also fail to explain very much variance.

As a final point, we argue that because all studies thus far have found similarly low levels of time variation across risks, interpreting the results in absolute terms is difficult. Measurement noise will mechanically present itself in the firm-level variation. However, we might assume that the amount of measurement noise is constant across different risk types, in which case the relevant values of variance still are informative. For example, Hassan et al. (2019) found that time fixed effects explain only 0.81% of the variation in political risk. They concluded that this was lower than what the theoretical literature would predict, however, this is higher than what most of our risks have, indicating it is a more systematic risk. Of course, our measure could have more measurement noise, but this is the very reason why a single methodology is needed across risk. Indeed, Kim et al. (2023) found for their measure(s) of political risk found time fixed effects to explain 1-5% of the variation. Even within the same paper, the results can be very different between similar measures. Therefore, the absolute values are not very informative, so a repeated study that included explicit political risk alongside other measures could then present more compelling evidence.

[18] As outlined in the table, idiosyncratic risk for some topics exceeds standard risk by a small margin. This can be explained by around 6% of firms with only one observation being dropped in the calculation of the betas. We tested the standard model with these dropped as well and this increases the explanatory power of all the beta models by around 1%, but makes no difference to the conclusions of the analysis.

The closest risk we have is regulation, which is a subset of political risk, but here we do confirm their findings as we find the time fixed effect portion to be one of the lowest, on par with contract and cybersecurity risk.

6 Discussion

We have established in this paper that BERTopic can effectively extract topics from conference calls. We also find that in building our aggregate measures of risk they show intuitive patterns over time. Our regression analysis provided support for 7 of the measures of risk, contingent on there being no significant differences between firms that are not controlled for by sector or size effects. Even after controlling for firm fixed effects, we find that loan and banking risk are still significant and positively correlated with future volatility.

The fact that only 2 risks remain significant and positive when adding firm fixed effects is suggestive of a few things. First, our approach might only measure permanent differences between firms. We discussed in the construction of our topics that the model lacks power for subtle signals of topics. It might then be the case that clear indications of a certain topic primarily reveal these permanent differences, whereas all time variation originates from more subtle statements. Second, future research should carefully consider controlling for firm fixed effects only when necessary. We argue that it is not always appropriate as this also controls for differences in permanent levels of risk between firms. However, it is difficult to assess what confounding factors may also play a role, so they should always be used as a robustness check.

Furthermore, depending on the specification, 3 of our topics (global economy, inflation, and interest rates) are negatively correlated with future volatility.[19] This indicates that in some cases, discussion of risk in a conference call is not a good measure of that firm's risk because it instead signals that it is well prepared. Since the aggregate indices still appear convincing, however, we argue it is still a good proxy for the aggregate level of risk. One might wonder what makes these risks different. It would seem that all three of these topics are macroeconomic, whereas the majority of others deal in micro issues. An exception might be foreign exchange, but if one refers to the regressions in Figure 4, forex risk is also negative, and we could lack the power to detect a significant effect. Might it be then, that these types of risks are not well suited to the analysis of conference calls? Future research is needed to evaluate this question more.

Due to the availability of data, the time period we and other authors are able to study is very short. Nevertheless, we cover some years that others do not. It is possible that the relationship observed between topic risk discussions in earnings calls is time-varying. The COVID-19 pandemic was quite an unprecedented time, with many unforeseen events impacting firms all at once. It could have diverted management's attention during the calls to the pandemic itself. It is not inconceivable that this could affect our measures. We tested the main regression specification in Equation 1 with firm fixed effects using only data prior to 2020Q1, the results show that 7 coefficients are positive but were not significant enough to confirm the hypothesis; we may lack the statistical power given the large truncation of the data.

[19] We cannot compare findings with Chava et al. (2022) and Ying (2023) because they do not repeat a similar regression of future volatility.

In our variance decomposition, we also found evidence that an analysis of risk in the context of conference calls would benefit from being applied to many risks at once. Measures of risk in conference call transcripts mechanically have noise which makes interpretations of the absolute value of results difficult. We found interesting results, however, by comparing risks to each other. Included in this, we found that supply chain risk tends towards being systematic, and cybersecurity risk presents itself very heterogeneously in terms of exposure to aggregate risk and permanent differences between firms.

Our analysis is not without its limitations. In the validations of risk snippets, we noticed some do discuss more than one risk, which violates a central assumption of the model. Furthermore, the model has a non-trivial error rate in classifying risks. We conclude that whilst BERTopic is an interesting method for discovering topics and it does seem to effectively measure some sources of firm risk, further research should look to bridge the gap between our model and the existing literature. Namely, we need a convincing continuous measure of risk that can both account for subtleties in the level of risk a snippet signals and still allow endogenous definitions of topics.

References

Calomiris, C. W., Mamaysky, H., Yang, R., 2020. Measuring the cost of regulation: A text-based approach. Tech. rep., National Bureau of Economic Research.

Cboe Global Markets, 2024. Historical Data for Cboe VIX Index and Other Volatility Indices. Available at: https://www.cboe.com/tradable_products/vix/vix_historical_data/.

Chava, S., Du, W., Shah, A., Zeng, L., 2022. Measuring firm-level inflation exposure: A deep learning approach. SSRN.

Economic Policy Uncertainty, 2024. U.S. EPU. Available at: https://www.policyuncertainty.com/us_monthly.html.

Ersahin, N., Giannetti, M., Huang, R., 2024. Supply chain risk: Changes in supplier composition and vertical integration. Journal of International Economics 147, 103854.

Federal Reserve Bank of New York, 2024. Global Supply Chain Pressure Index (GSCPI). Available at: https://www.newyorkfed.org/research/policy/gscpi#/interactive.

Federal Reserve Bank of St. Louis, 2024a. Delinquency rate on all loans, all commercial banks. Available at: https://fred.stlouisfed.org/series/DRALACBN.

Federal Reserve Bank of St. Louis, 2024b. Federal Funds Effective Rate. Available at: https://fred.stlouisfed.org/series/FEDFUNDS.

Federal Reserve Bank of St. Louis, 2024c. Global price of WTI Crude. Available at: https://fred.stlouisfed.org/series/POILWTIUSDM.

Grootendorst, M., 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. arXiv preprint arXiv:2203.05794 .

Hassan, T. A., Hollander, S., van Lent, L., Tahoun, A., 2019. Firm-level political risk: Measurement and effects. The Quarterly Journal of Economics 134, 2135–2202.

Kenton, J. D. M.-W. C., Toutanova, L. K., 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of naacL-HLT*, Minneapolis, Minnesota, vol. 1, p. 2.

Kim, A., Muhn, M., Nikolaev, V., 2023. From transcripts to insights: Uncovering corporate risks using generative ai.

Kim, C. F., Pantzalis, C., Park, J. C., 2012. Political geography and stock returns: The value and risk implications of proximity to political power. Journal of Financial Economics 106, 196–228.

Mandelbrot, B., 1967. The variation of some other speculative prices. The Journal of Business 40, 393–413.

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

Nasdaq, 2024. Nasdaq cybersecurity index (nqcybrt) - historical data. Available at: https://www.nasdaq.com/market-activity/index/nqcybrt/historical?page=1&rows_per_page=10&timeline=y10.

Nazli, S. A., Karagozoglu, A. K., Zhou, T., 2021. Firm-level cybersecurity risk and idiosyncratic volatility. Journal of Portfolio Management 47, 110–140.

Qadan, M., Shuval, K., David, O., 2023. Uncertainty about interest rates and the real economy. The North American Journal of Economics and Finance 68, 101978.

Schrand, C. M., Elliott, J. A., 1998. Risk and financial reporting: A summary of the discussion at the 1997 aaa/fasb conference. Accounting Horizons 12, 271.

Tang, W., Bu, H., Zuo, Y., Wu, J., 2024. Unlocking the power of the topic content in news headlines: Bertopic for predicting chinese corporate bond defaults. Finance Research Letters 62, 105062.

The Wall Street Journal, 2024a. Dow jones u.s. banks index (djusbk). Available at: https://www.wsj.com/market-data/quotes/index/DJUSBK/historical-prices.

The Wall Street Journal, 2024b. U.S. Dollar Index (DXY). Available at: https://www.wsj.com/market-data/quotes/index/DXY/historical-prices.

Trading Economics, 2024. United states corporate tax rate. Available at: https://tradingeconomics.com/united-states/corporate-tax-rate.

U.S. Bureau of Labor Statistics, 2024. 12-month percentage change, Consumer Price Index, selected categories. Available at: https://www.bls.gov/charts/consumer-price-index/consumer-price-index-by-category-line-chart.htm.

Washington Post, 2023. Sharp economic slowdown triggers new recession fears. Available at: https://www.washingtonpost.com/business/2023/04/27/recession-gdp-slow-economy/".

Wu, D., 2024. Text-based measure of supply chain risk exposure. Management Science 70, 4781–4801.

Yang, B., 2013. Extending topic models for text analysis of corporate risk disclosures.

Ying, S., 2023. Measuring macroeconomic uncertainty. Ph.D. thesis, Macquarie University.

Zhu, E., Yen, J., 2024. Bertopic-driven stock market predictions: Unraveling sentiment insights. arXiv preprint arXiv:2404.02053 .

Zhu, X., Yang, S. Y., Moazeni, S., 2016. Firm risk identification through topic analysis of textual financial disclosures. In: *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, pp. 1–8.

A Appendix

A.1 AI Appendix

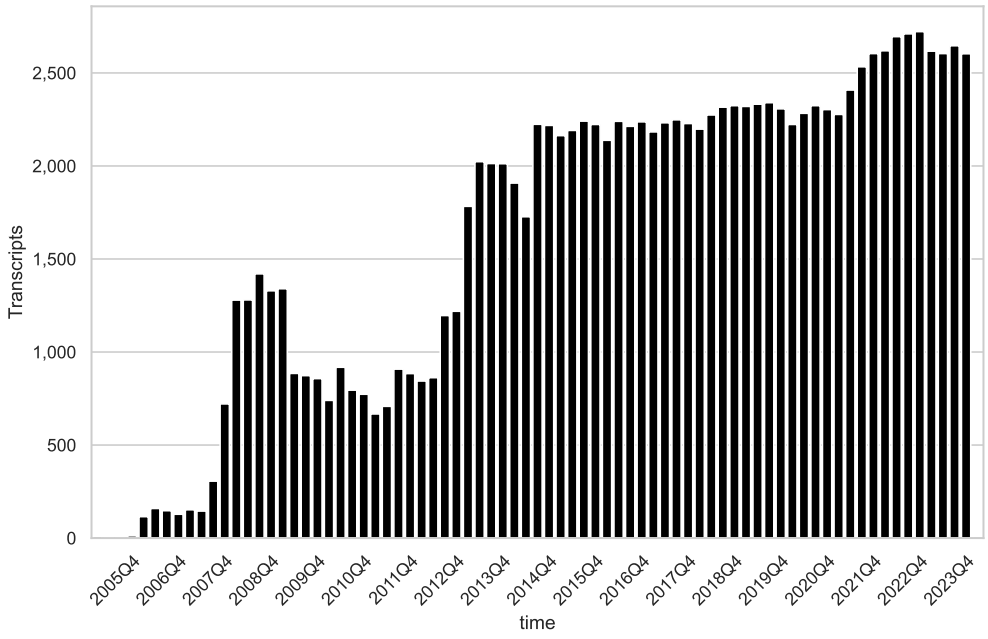
We used ChatGPT to help with bibtex citations, some coding tasks, and formatting tables in LaTeX.

A.2 Figures

Figure A.1: Filtered Data Coverage Quarterly

This figure shows total transcripts and risk mentions in our data each quarter. Results are filtered according to the steps in section 2. Sub-figure (a) shows the total number of earnings conference call transcripts, sub-figure (b) the total number of risk mentions, and sub-figure (c) takes the total number of risk mentions and divides it by the number of transcripts in that quarter. A risk mention is defined as the occurrence of one of the filtered Oxford Dictionary synonyms of "risk" or "uncertainty", not including those in safe-harbour snippets.

(a) Number of transcripts



(b) Number of risk mentions

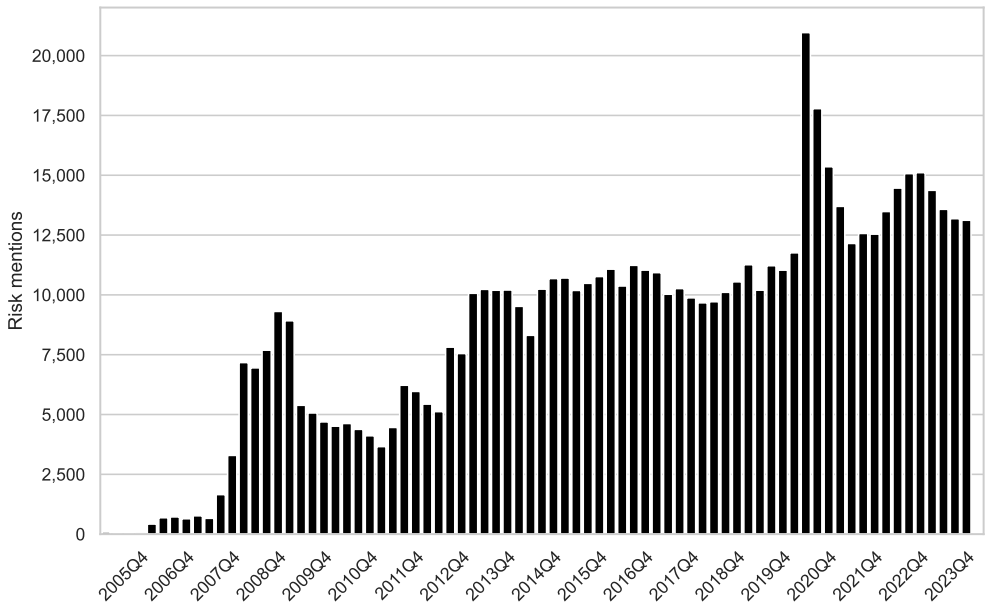


Figure A.1
Continued

(c) Risk mentions normalised

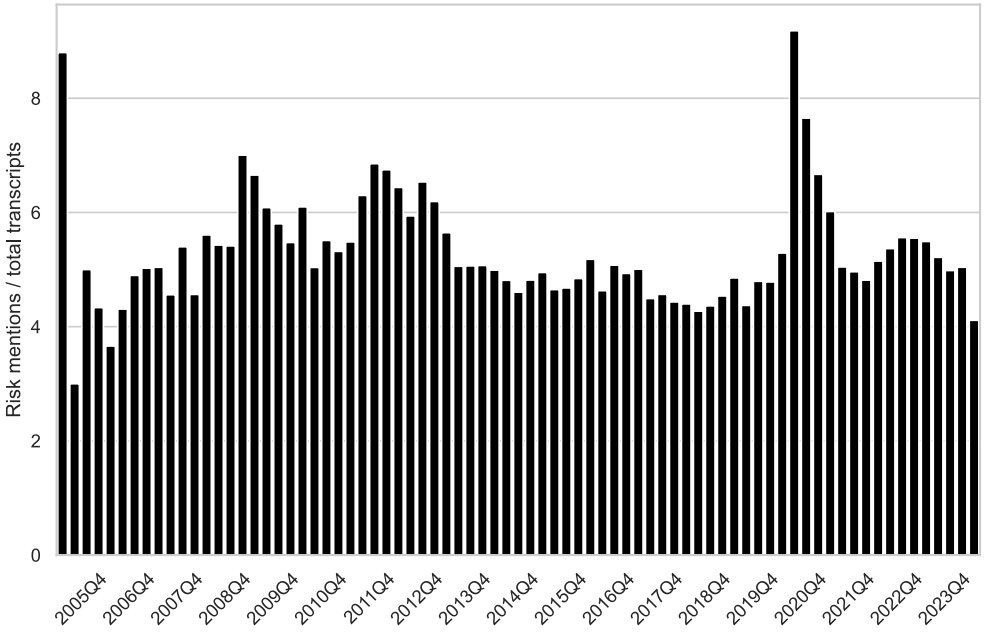


Figure A.2: Distribution of Continuous Risk

The following histograms show the distribution of our continuous measures of risk on the transcript level. The continuous measure is constructed as the average across all risk snippet probability vectors in a transcript, and it has been standardized to a mean of 0 and standard deviation of 1. For transcripts without any risk snippets, the mean of each measure is used instead.

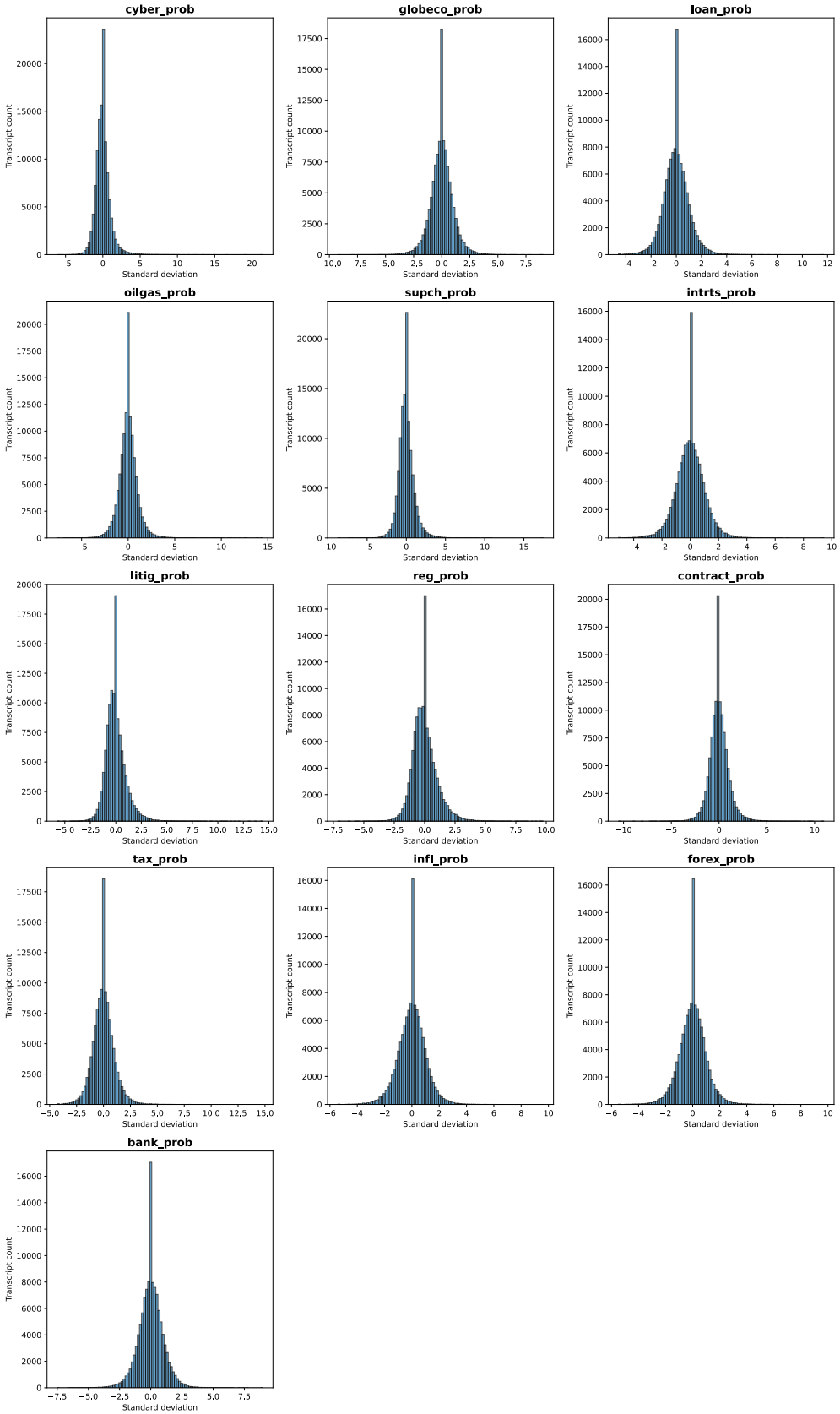


Figure A.3: Example Output From BERTopic

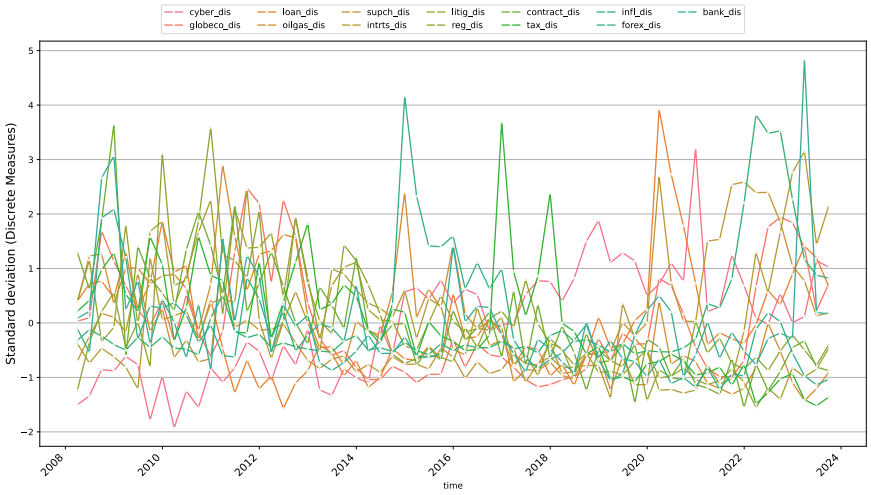
The following figure shows an example of what BERTopic sees when evaluating a topic. The left column shows a subset of the topics, filtered for conciseness. Each column represents one of the words in the snippet, the top row shows what this word is. Each cell is the value associated with the word-topic pair. Higher values indicate a stronger association. Context affects the value of a word, so high information words such as "supply" raise the score of the surrounding words. When calculating probabilities, the model evaluates the scores across all terms for all topics. The snippet selected comes from an actual conference call transcript.

	of	what	it	does	to	global	supply	chains	and	the	uncertainty	around	shipping	mean	all	of	the	things
0_the_we_and_of	0.219	0.391	0.656	1.127	1.326	1.573	1.735	1.495	1.440	1.334	1.165	1.168	0.989	0.858	0.749	0.746	0.719	0.674
1_the_of_patients_and	0.247	0.499	0.685	0.954	0.890	0.826	0.855	0.586	0.623	0.576	0.475	0.475	0.256	0.114		0.115	0.115	0.115
2_factors_and_the_filings	0.181	0.299	0.436	0.669	0.671	0.761	0.902	0.805	0.875	0.855	0.725	0.705	0.601	0.538	0.562	0.695	0.657	0.688
3_you_operator_thank_question	0.177	0.374	0.374	0.498	0.448	0.251	0.456	0.552	0.719	0.873	0.668	0.448	0.270	0.371	0.618	0.857	1.024	0.942
4_covid_pandemic_the_of	0.154	0.308	0.524	0.915	1.054	1.205	1.266	1.015	0.983	0.899	0.806	0.799	0.537	0.447	0.262	0.278	0.378	0.249
5_supply_inventory_chain_the	0.179	0.319	0.556	1.186	1.719	2.291	2.797	2.485	2.035	1.811	1.458	1.517	1.630	1.253	1.015	0.843	0.679	0.769

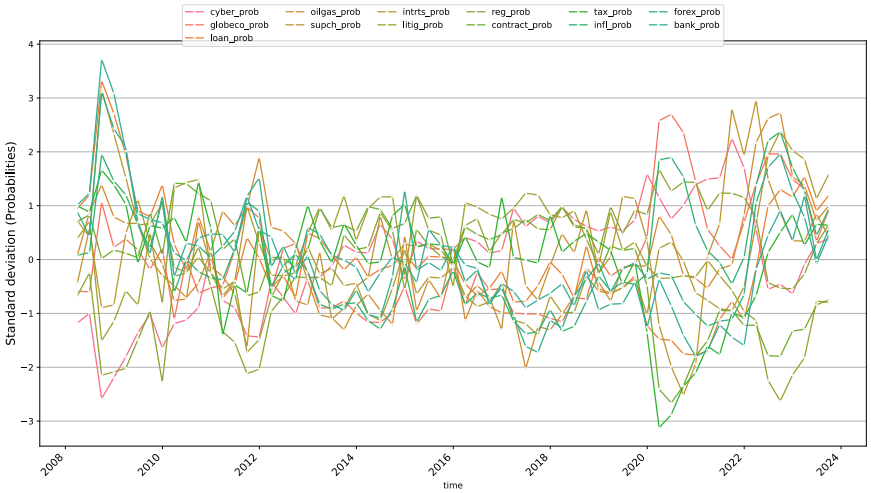
Figure A.4: Aggregate Measures of Firm Risk

This figure shows the aggregate measures of risk. Sub-figure (a) shows all the risks in a single plot. Sub-figures (b)-(h) break this down by individual risk types. The discrete measure is the total number of risk snippets classified as a particular topic divided by the number of transcripts. The continuous measure on the transcript level is the average of the probabilities across snippets in a transcript for a particular topic, on the aggregate level it is the average across all transcripts. The measures are standardised on the aggregate level.

(a) All risks discrete measure



(b) All risks probabilistic measure



(c) Cyber security

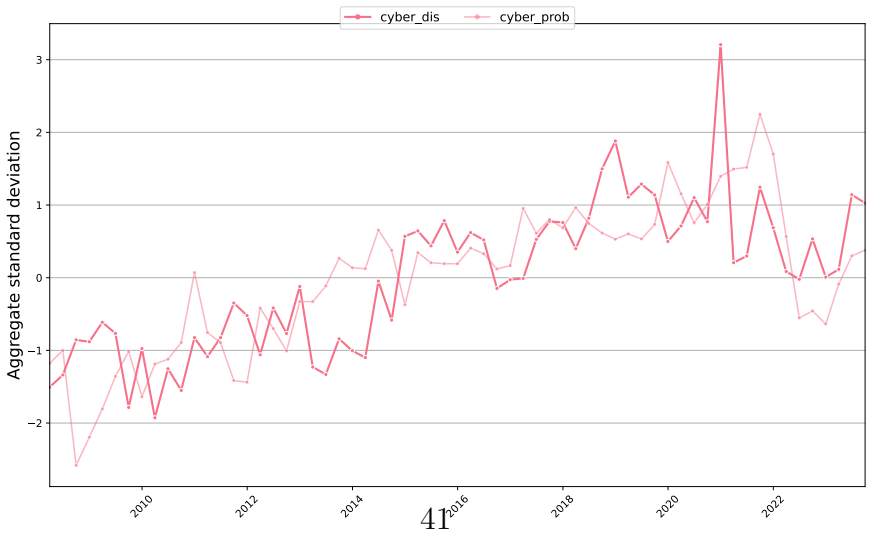
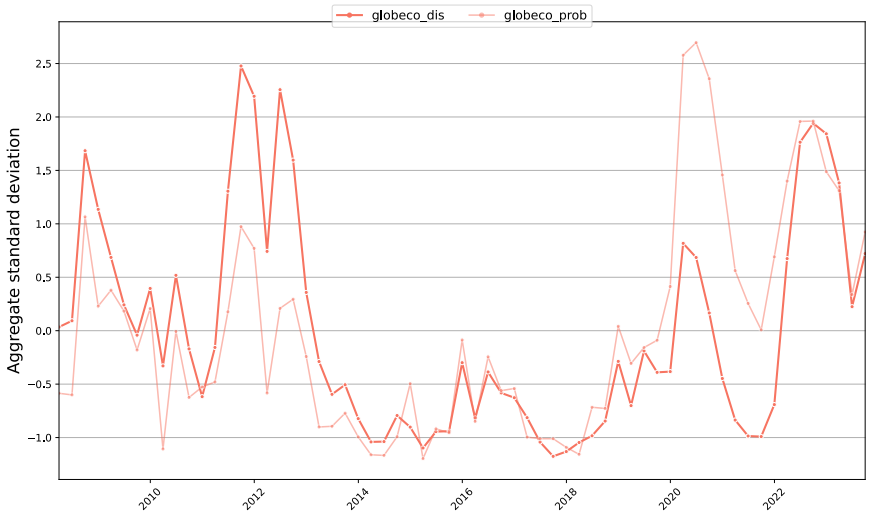
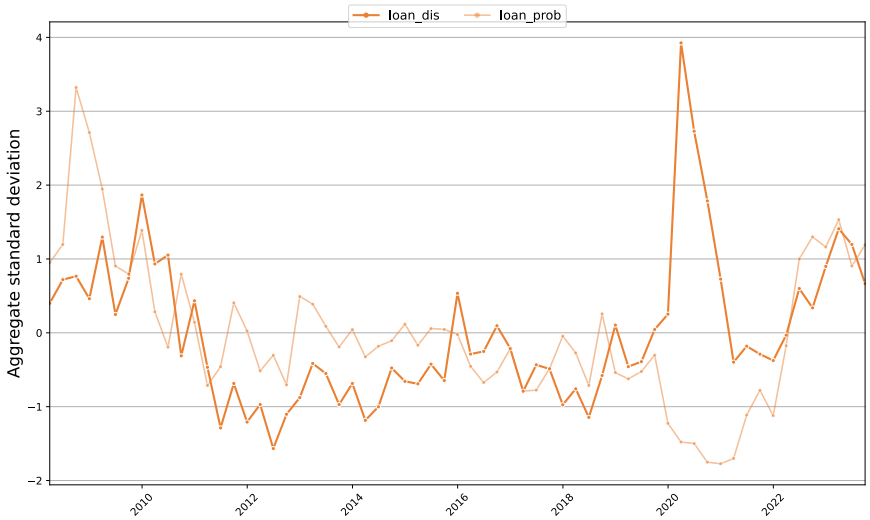


Figure A.4
Continued

(d) Global economy



(e) Loans



(f) Oil & gas

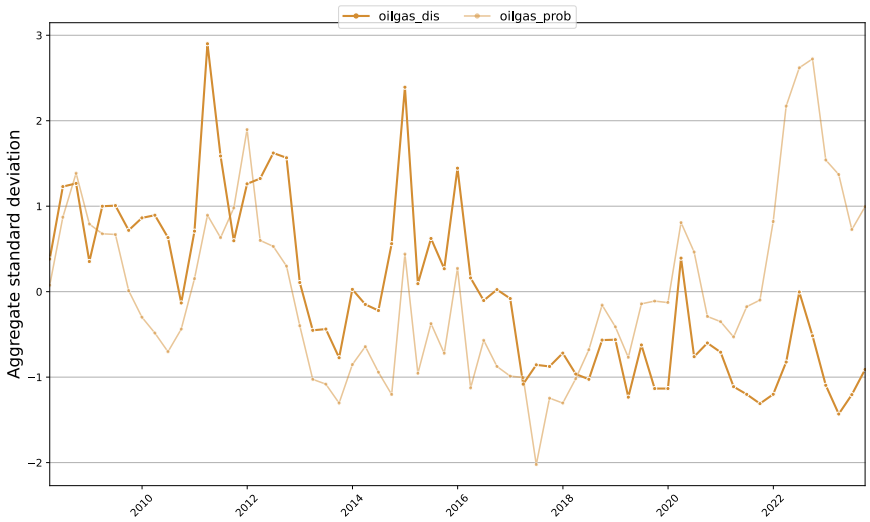
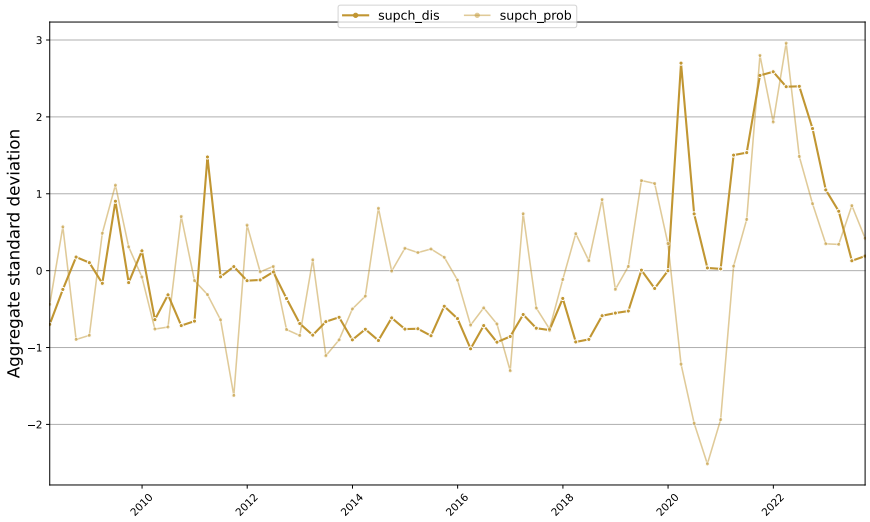
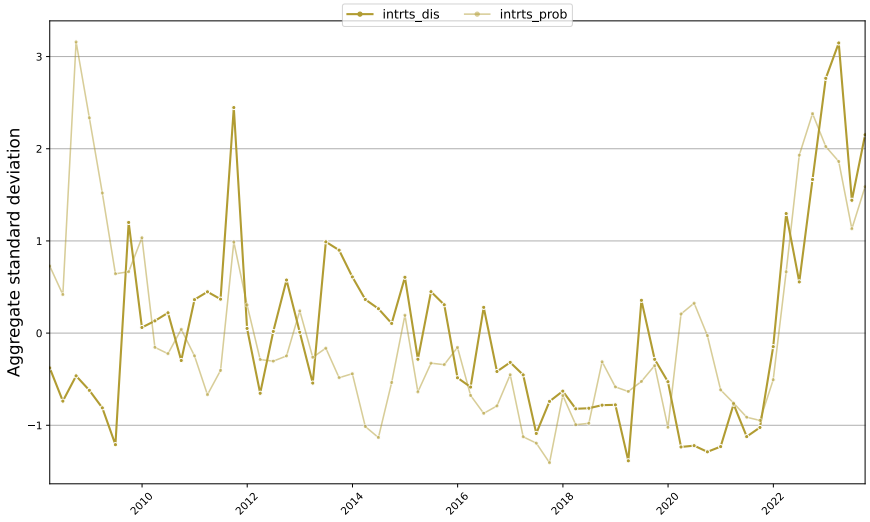


Figure A.4
Continued

(g) Supply chain



(h) Interest rates



(i) Litigation

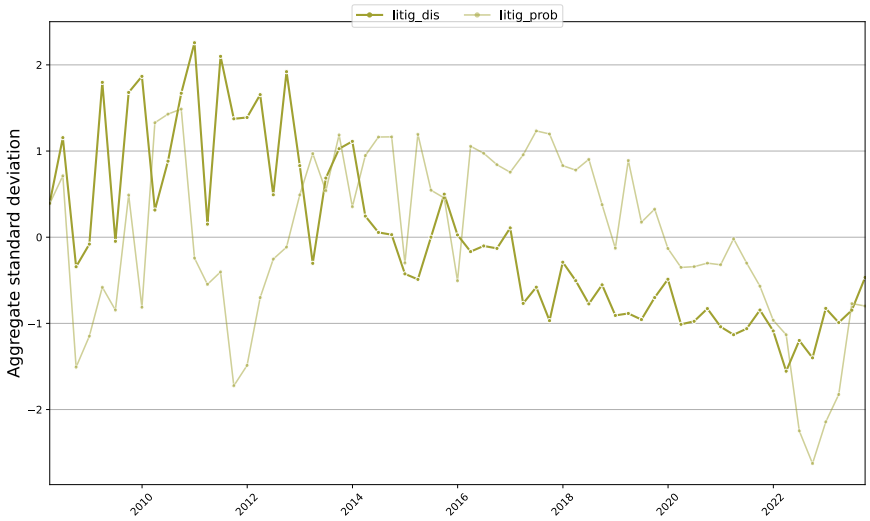
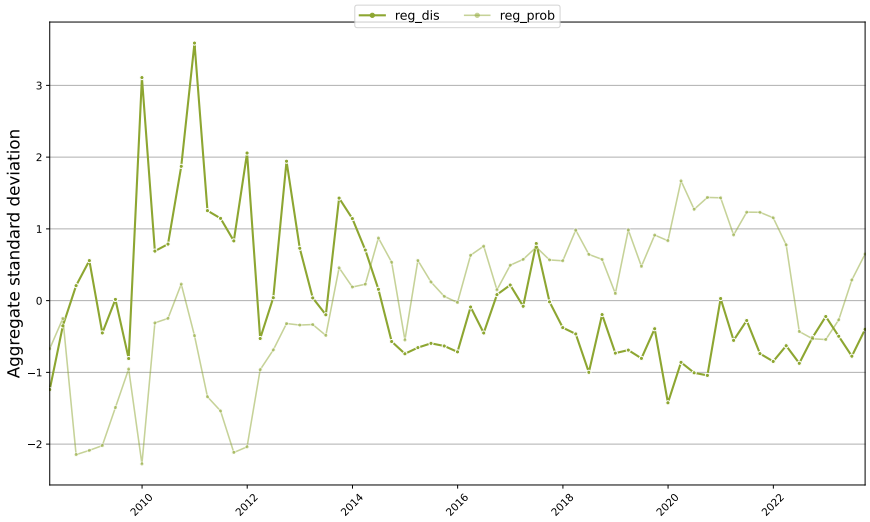
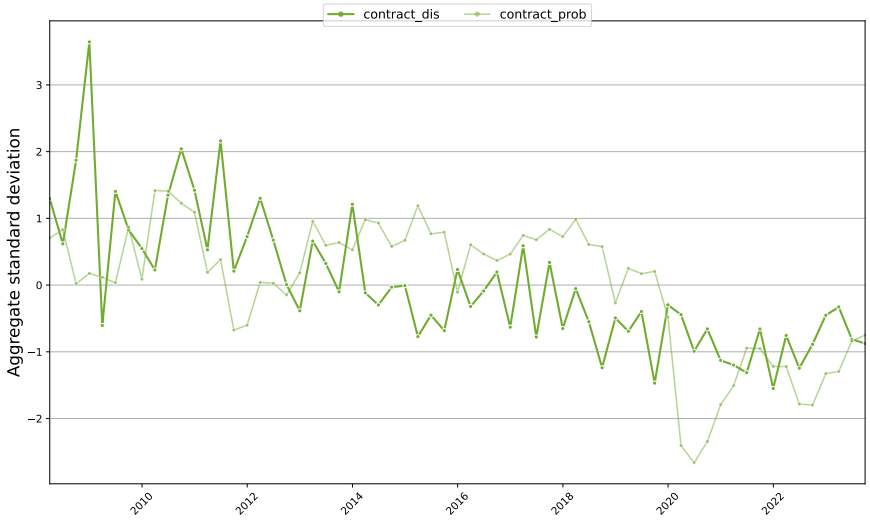


Figure A.4
Continued

(j) Regulation



(k) Contracts



(l) Tax

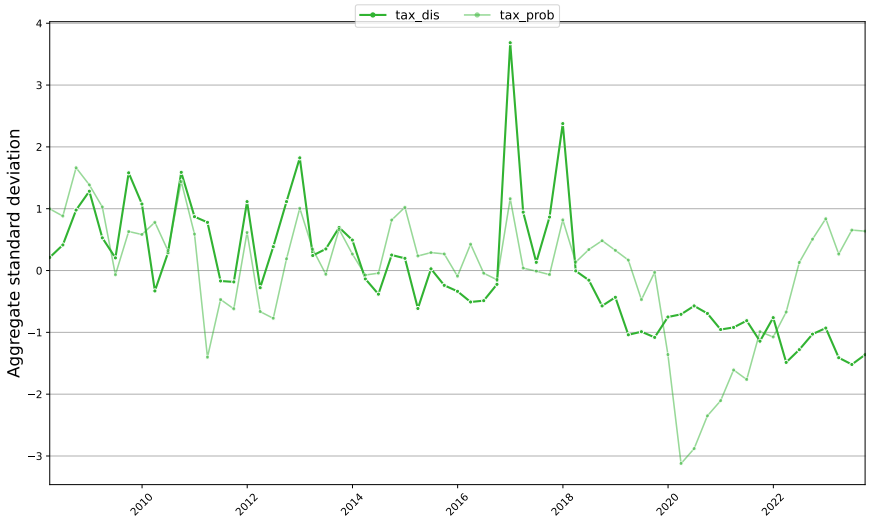
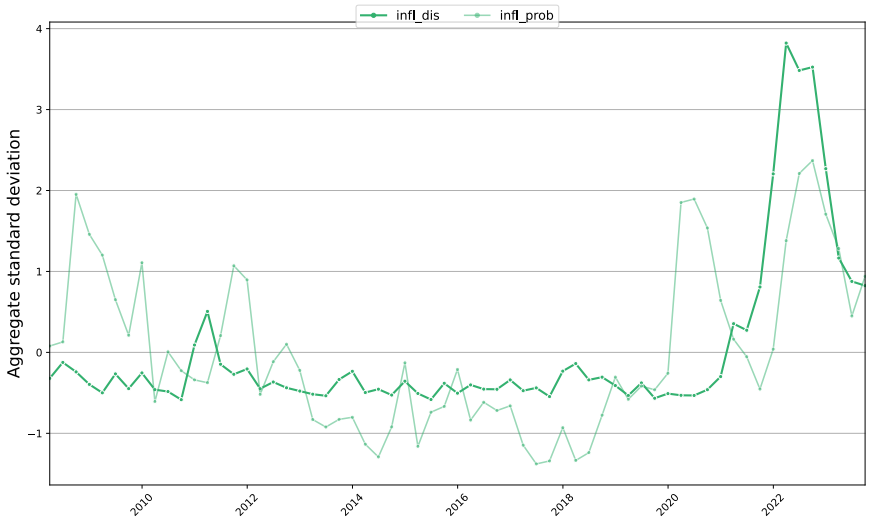
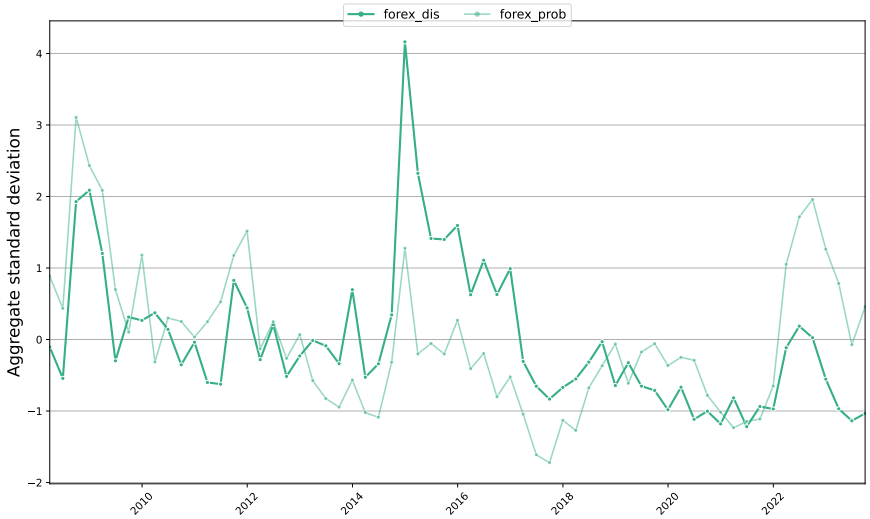


Figure A.4
Continued

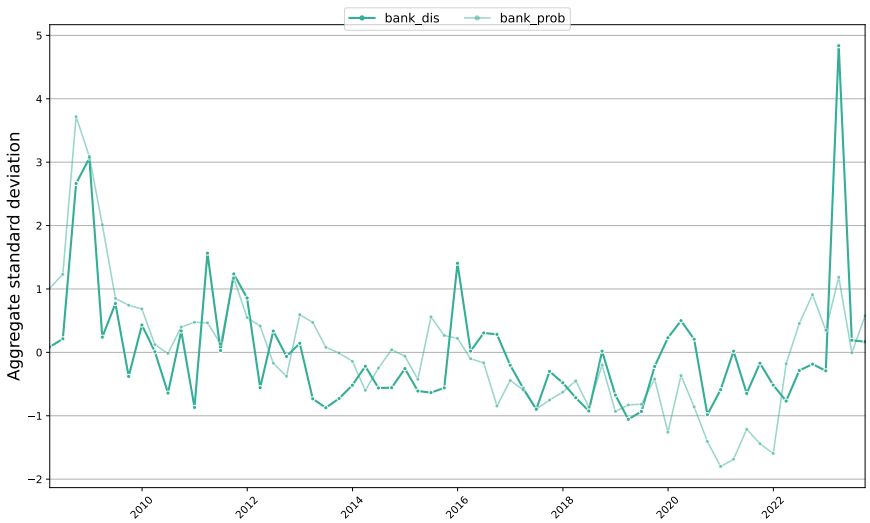
(m) Inflation



(n) Foreign exchange rates



(o) Banking



A.3 Tables

Table A.1: All Topics Generated
This table shows the top 10 terms associated with each topic, known as the topic representations. These terms were endogenously generated by the BERTopic model. The name of each topic is the first four terms in the topic representation. A term may consist of up to three words. -1 is the "topic" for outliers, that is documents that do not fit well into any of the other topics.

Name	Representation
-1-factors including-involve known factors-annual report form-factors affect	factors including, involve known factors, annual report form, factors affect, factors cause, company filings, management current expectations, report form, annual report, factors company
0-patients-patients high-patient population-high patients	patients, patients high, patient population, high patients, indications, clinical, patient, clinical trials, medical, clinical trial
1-factors annual report-factors section annual-annual report form-discussed factors section	factors annual report, factors section annual, annual report form, discussed factors section, section entitled factors, company sec filings, sec filings including, factors section company, quarterly report form, sec filings
2-associated covid pandemic-related covid pandemic-covid pandemic factors-ongoing covid pandemic	associated covid pandemic, related covid pandemic, covid pandemic factors, ongoing covid pandemic, covid pandemic, surrounding covid pandemic, pandemic covid, impact covid pandemic, effects covid pandemic, caused covid pandemic
3-thank operator question-thanks operator-operator thank question-thank operator instructions	thank operator question, thanks operator, operator thank question, thank operator instructions, thank operator thank, operator thank operator, question operator, operator question come, operator instructions question, operator question line
4-potential customers-customer base-customer concentration-customers business	potential customers, customer base, customer concentration, customers business, customers customers, customers need, customers want, business customers, support customers, lot customers
5-point think-say think-really think-people think	point think, say think, really think, people think, think point, obviously think, think say, think mean, think focus, view think
6-cybersecurity-cyber security-cyber-cyber attacks	cybersecurity, cyber security, cyber, cyber attacks, security industry, security systems industry, security systems, cloud based, systems industry, data security

Continued on next page

Table A.1: All Topics Generated

Name	Representation
7-like adjusted returns-attractive adjusted returns-adjusted returns market-good adjusted returns	like adjusted returns, attractive adjusted returns, adjusted returns market, good adjusted returns, returns market, think adjusted returns, best adjusted returns, better adjusted returns, strong adjusted returns, high adjusted returns
8-global economies-global economy-global economic-economic growth	global economies, global economy, global economic, economic growth, global economic conditions, global growth, global macroeconomic, gdp growth, global economic environment, world economy
9-loans loans-total loans-loans-higher loans	loans loans, total loans, loans, higher loans, performing loans, loans loan, loans higher, nonperforming loans, consumer loans, loan portfolios
10-leases-long term leases-leasing-lease terms	leases, long term leases, leasing, lease terms, lease space, commercial real estate, leases signed, term leases, lease maturities, new leasing
11-strong credit quality-credit quality metrics-credit quality-quality credit	strong credit quality, credit quality metrics, credit quality, quality credit, credit metrics, credit performance, credit environment, credit perspective, credit ratings, maintaining credit
12-oil gas industry-oil demand-crude oil prices-business oil	oil gas industry, oil demand, crude oil prices, business oil, gas industry, gas markets, gas service industry, oil prices, oil companies, oil market
13-based capital ratio-capital ratio total-based capital ratios-capital total based	based capital ratio, capital ratio total, based capital ratios, capital total based, capital ratios, total based capital, capital ratio, tier capital ratio, capital weighted assets, capital total
14-price increases-price increase-pricing increases-higher pricing	price increases, price increase, pricing increases, higher pricing, cost increases, increase cost, higher price, prices going, lower prices, say pricing
15-demand inventory-inventory supply chain-inventory supply-supply chains	demand inventory, inventory supply chain, inventory supply, supply chains, inventory levels, supply chain challenges, supply chain constraints, higher inventory, inventory management, supply chain disruptions
16-portfolio overall-portfolios-portfolio opportunities-portfolio say	portfolio overall, portfolios, portfolio opportunities, portfolio say, portfolio positioned, portfolio companies, portfolio provides, portfolio diversification, portfolio perspective, overall portfolio

Continued on next page

Table A.1: All Topics Generated

Name	Representation
17-drilling wells-wells drilled-drill wells-currently drilling	drilling wells, wells drilled, drill wells, currently drilling, exploration wells, drilling activity, additional wells, additional drilling, start drilling, exploration drilling
18-managing capital-increase capital-capital generation-approach capital	managing capital, increase capital, capital generation, approach capital, capital management, use capital, capital business, capital work, capital needs, capital flexibility
19-say market-understand market-market market-market like	say market, understand market, market market, market like, markets think, market people, markets market, markets want, like market, different market
20-build projects-development projects-projects construction-project development	build projects, development projects, projects construction, project development, project projects, projects projects, projects built, different projects, construction projects, types projects
21-chief financial officer-financial officer executive-executive vice-officer executive vice	chief financial officer, financial officer executive, executive vice, officer executive vice, chief executive officer, chief financial, goldman sachs, president chief financial, officer executive, executive vice president
22-gross margin-margin gross-gross margin rate-gross margins	gross margin, margin gross, gross margin rate, gross margins, gross margins depending, quarter gross margins, expect gross margin, gross margin expansion, margin year, profit margin
23-reinsurance contracts objective-reinsurance business-reinsurance market-reinsurance	reinsurance contracts objective, reinsurance business, reinsurance market, reinsurance, insurance reinsurance, reinsurance segment, investment reinsurance, profile reinsurance, reinsurance program, property reinsurance
24-second quarter revenue-year revenue growth-quarter quarter revenue-quarter revenue	second quarter revenue, year revenue growth, quarter quarter revenue, quarter revenue, quarterly revenue, revenue growth, revenue quarter, quarter revenues, year revenue, near term revenue
25-quarter fourth quarter-fourth quarter quarter-think fourth quarter-fourth quarter think	quarter fourth quarter, fourth quarter quarter, think fourth quarter, fourth quarter think, quarter second quarter, going quarter quarter, fourth quarter little, quarter quarter just, quarter quarter think, just quarter quarter

Continued on next page

Table A.1: All Topics Generated

Name	Representation
26-strong balance sheet-balance sheet strong-strengthen balance sheet-balance sheet good	strong balance sheet, balance sheet strong, strengthen balance sheet, balance sheet good, strengthened balance sheet, balance sheet think, balance sheets, balance sheet robust, robust balance sheet, improving balance sheet
27-acquisitions dispositions competition-acquisitions competition-acquisition market-acquisition market view	acquisitions dispositions competition, acquisitions competition, acquisition market, acquisition market view, acquisitions dispositions, associated acquisitions dispositions, expansion acquisitions competition, acquisitions including, acquisitions acquisition, future acquisitions
28-executive chinese-chinese economy-chinese angie chang-executive chinese angie	executive chinese, chinese economy, chinese angie chang, executive chinese angie, change china, china growth, china economy, growth china, china business, business china
29-rising rate environment-rising rates-rates economic-market rates	rising rate environment, rising rates, rates economic, market rates, rates going, lower rates, rising rate, higher rates, term rates, increase rates
30-court decision-courts-court case-future litigation proceedings	court decision, courts, court case, future litigation proceedings, litigation subject discuss, court, litigation proceedings, future litigation, future litigation pertaining, settlement agreement
31-regulatory requirements-conditions future regulatory-obtain necessary regulatory-future regulatory	regulatory requirements, conditions future regulatory, obtain necessary regulatory, future regulatory, include company failure, regulatory process, regulatory authorities, company failure, necessary regulatory, operations include company
32-business continue-manage business-business strategy-managing business	business continue, manage business, business strategy, managing business, grow business, growing business, business manage, future business, business business, impact business
33-contracts think-contracts believe-contracts contracts-contract negotiations	contracts think, contracts believe, contracts contracts, contract negotiations, long term contracts, limitation contracts, limitation contracts terminated, think contracts, large contracts, longer term contracts
34-effective tax rate-tax rate quarter-tax rates-tax rate	effective tax rate, tax rate quarter, tax rates, tax rate, tax rate going, effective tax, tax rate approximately, tax positions quarter, reduction tax, tax provision

Continued on next page

Table A.1: All Topics Generated

Name	Representation
35-range million million-million increase-million million million-million million range	range million million, million increase, million million million, million million range, million million, million million billion, say million, billion million million, million compared million, million quarter million
36-inflationary pressures-economy inflation-continued inflation-economic inflation	inflationary pressures, economy inflation, continued inflation, economic inflation, inflation recession, inflation economic, regarding inflation, rising inflation, inflation going, increase inflation
37-hedges hedge-hedge rate-hedge mitigate-hedge position	hedges hedge, hedge rate, hedge mitigate, hedge position, hedge hedge, continue hedge, hedge portfolio, hedge positions, management hedge, hedge
38-plains american pipeline-american pipeline-pipeline projects-pipeline opportunities	plains american pipeline, american pipeline, pipeline projects, pipeline opportunities, pipelines, pipeline say, pipeline pipeline, pipeline including, pipeline going, higher pipeline
39-unidentified analyst okay-unidentified analyst just-unidentified analyst got-unidentified analyst right	unidentified analyst okay, unidentified analyst just, unidentified analyst got, unidentified analyst right, unidentified analyst yes, yes unidentified analyst, unidentified analyst thanks, unidentified analyst, question unidentified analyst, unidentified analyst helpful
40-economic volatility-volatility economic-market volatility economic-market volatility production	economic volatility, volatility economic, market volatility economic, market volatility production, economy volatility, markets high volatility, volatility particularly, markets volatility, volatility currently, volatility business
41-beginning year-year beginning year-think end year-year end year	beginning year, year beginning year, think end year, year end year, year early, year going year, year going, early year, just year, start year
42-foreign currencies-impact foreign currency-currencies detailed sec-foreign currency	foreign currencies, impact foreign currency, currencies detailed sec, foreign currency, foreign currency exchange, foreign countries currencies, foreign currency fluctuations, conducting business foreign, business foreign countries, foreign currency translation
43-free cash flow-free cash flows-cash flow think-cash flow results	free cash flow, free cash flows, cash flow think, cash flow results, cash flows, cash flow perspective, cash flow year, cash flow generation, positive cash flow, strong cash flow

Continued on next page

Table A.1: All Topics Generated

Name	Representation
44-brazil market-brazilian market-brazilian economy-brazil cost	brazil market, brazilian market, brazilian economy, brazil cost, business brazil, brazil situation, situation brazil, brazil terms, revenues brazil, brazilian real
45-program cars-new vehicles-cars-automotive	program cars, new vehicles, cars, automotive, vehicles, automakers, automobile, automotive markets, car companies, automotive market
46-banking industry-banking business-banking environment-corporate banking	banking industry, banking business, banking environment, corporate banking, commercial banking, banking management, banking sector, banking lines business, community banking, banking relationships
47-subject certain assumptions-assumptions subject certain-based assumptions subject-certain assumptions subject	subject certain assumptions, assumptions subject certain, based assumptions subject, certain assumptions subject, underlying assumptions, involve assumptions, subject assumptions known, based assumptions, assumptions subject inherent, certain assumptions described

Table A.2: Regression of Volatility on Discrete Risk

The following table shows the regressions of realised stock volatility on all discrete risk measures. Controls include just the logarithm of firm assets. Volatility is defined as the standard deviation in investor returns for a 90-day holding period beginning on the day the conference call was held. Sector is defined as the two-digit SIC code. Volatility is winsorised at the 1st and 99th percentiles and then standardised. All errors are clustered on the firm level.

	(1)	(2)	(3)	(4)
VARIABLES	90 Vol	90 Vol	90 Vol	90 Vol
cyber_dis	-0.0171**	-0.0181***	-0.00446	-0.000696
	(0.00810)	(0.00595)	(0.00682)	(0.00473)
globeco_dis	-0.0447***	0.0566***	-0.0263***	0.00533
	(0.0109)	(0.00971)	(0.00784)	(0.00509)
loan_dis	-0.0471***	0.0835***	0.0513***	0.0294***
	(0.0138)	(0.0153)	(0.0111)	(0.00789)
oilgas_dis	-0.0161	0.0922***	0.0239**	-0.00770
	(0.0157)	(0.0119)	(0.0100)	(0.00620)
supch_dis	0.0457***	0.0755***	-0.00813	-0.00191
	(0.0141)	(0.0129)	(0.0112)	(0.00671)
intrts_dis	-0.194***	-0.00698	0.0414***	-0.0205***
	(0.0135)	(0.0153)	(0.0120)	(0.00700)
litig_dis	-0.0660***	-0.0123	0.0342**	0.0149
	(0.0219)	(0.0175)	(0.0156)	(0.00964)
reg_dis	-0.0667***	0.00233	0.0370**	-0.00225
	(0.0204)	(0.0179)	(0.0157)	(0.0103)
contract_dis	0.0183	0.0695***	0.0595***	0.0172
	(0.0229)	(0.0233)	(0.0164)	(0.0131)
tax_dis	-0.158***	-0.0838***	-0.0183	0.00401
	(0.0166)	(0.0141)	(0.0117)	(0.00819)
infl_dis	-0.143***	0.00222	-0.0649***	-0.0210**
	(0.0162)	(0.0150)	(0.0141)	(0.00882)
forex_dis	-0.170***	-0.0852***	-0.0234	-0.0104
	(0.0218)	(0.0216)	(0.0198)	(0.0102)
bank_dis	-0.0202	0.186***	0.107***	0.0582***
	(0.0317)	(0.0295)	(0.0230)	(0.0159)
Observations	121,309	121,161	120,869	120,894
R-squared	0.006	0.170	0.504	0.657
Controls	No	Yes	Yes	Yes
Time FE	No	No	Yes	Yes
Sector FE	No	No	Yes	n/a
Sector x Time FE	No	No	Yes	No
Firm FE	No	No	No	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A.3: Regression of Volatility on Continuous Risk

The following table shows the regressions of realised stock volatility on all continuous risk measures. Controls include just the logarithm of firm assets. Volatility is defined as the standard deviation in investor returns for a 90-day holding period beginning on the day the conference call was held. Sector is defined as the two-digit SIC code. Volatility is winsorised at the 1st and 99th percentiles and then standardised. All errors are clustered on the firm level.

	(1)	(2)	(3)	(4)
VARIABLES	90 Vol	90 Vol	90 Vol	90 Vol
cyber_prob	0.0187**	-0.00603	0.00664	0.00563*
	(0.00746)	(0.00580)	(0.00463)	(0.00311)
globeco_prob	-0.0321***	-0.0137**	-0.0288***	0.00452
	(0.00740)	(0.00613)	(0.00507)	(0.00341)
loan_prob	0.0382***	0.0340***	0.0199***	0.0175***
	(0.00915)	(0.00782)	(0.00639)	(0.00424)
oilgas_prob	-0.00683	0.0175***	0.00623	-0.00179
	(0.00782)	(0.00608)	(0.00496)	(0.00294)
supch_prob	0.0305***	-0.00674	-0.0124***	0.00153
	(0.00573)	(0.00483)	(0.00403)	(0.00263)
intrts_prob	-0.0147	-0.0114	-0.0134*	-0.00497
	(0.0104)	(0.00870)	(0.00693)	(0.00458)
litig_prob	0.0624***	0.0364***	0.0174***	0.0228***
	(0.00837)	(0.00696)	(0.00588)	(0.00395)
reg_prob	0.125***	0.0503***	0.0275***	-0.00692*
	(0.00858)	(0.00679)	(0.00587)	(0.00386)
contract_prob	-0.0762***	-0.0399***	-0.0161***	-0.00165
	(0.00606)	(0.00490)	(0.00425)	(0.00275)
tax_prob	-0.0424***	-0.0568***	-0.0167***	0.00137
	(0.00759)	(0.00639)	(0.00560)	(0.00367)
infl_prob	0.0223**	0.0262***	0.00287	0.00289
	(0.00878)	(0.00744)	(0.00634)	(0.00416)
forex_prob	0.0207***	0.0110	0.0199***	0.00389
	(0.00793)	(0.00668)	(0.00554)	(0.00364)
bank_prob	-0.0320***	-0.00218	-0.00107	-0.000369
	(0.00753)	(0.00618)	(0.00509)	(0.00342)
Observations	121,309	121,161	120,869	120,894
R-squared	0.027	0.171	0.505	0.657
Controls	No	Yes	Yes	Yes
Time FE	No	No	Yes	Yes
Sector FE	No	No	Yes	n/a
Sector x Time FE	No	No	Yes	No
Firm FE	No	No	No	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A.4: Regression of Volatility on Discrete Risk With Additional Controls
The following table shows the regression of 90-day volatility on each discrete measure of risk. 90-day volatility is the standard deviation of investor returns for a 90-day holding period that begins on the day the conference call was held. We control for the size of firm assets, previous 90-day average returns, and previous 90-day average volatility. Sector is defined by the two-digit SIC code. Volatility is winsorised at the 1st and 99th percentiles and then standardised. The continuous measure is standardised, the discrete measure is the number of mentions of that topic in a transcript. All errors are clustered on the firm level.

	(1)
VARIABLES	90 Vol
cyber_dis	-0.00233
	(0.00306)
globeco_dis	-0.0107**
	(0.00483)
loan_dis	0.0228***
	(0.00593)
oilgas_dis	0.00505
	(0.00604)
supch_dis	-0.00232
	(0.00715)
intrts_dis	0.0145**
	(0.00634)
litig_dis	0.00903
	(0.00961)
reg_dis	0.0161
	(0.0109)
contract_dis	0.0275**
	(0.0112)
tax_dis	-0.00247
	(0.00795)
infl_dis	-0.0318***
	(0.00867)
forex_dis	-0.00277
	(0.0109)
bank_dis	0.0497***
	(0.0151)
Observations	120,867
R-squared	0.669
Controls+	Yes
Time FE	Yes
Sector FE	Yes
Sector x Time FE	Yes
Firm FE	No

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.5: Regression of Volatility on Continuous Risk With Additional Controls

The following table shows the regression of 90-day volatility on each continuous measure of risk. 90-day volatility is the standard deviation of investor returns for a 90-day holding period that begins on the day the conference call was held. We control for the size of firm assets, previous 90-day average returns, and previous 90-day average volatility. Sector is defined by the two-digit SIC code. Volatility is winsorised at the 1st and 99th percentiles and then standardised. The continuous measure is standardised, the discrete measure is the number of mentions of that topic in a transcript. All errors are clustered on the firm level.

	(1)
VARIABLES	90 Vol
cyber_prob	0.00295
	(0.00293)
globeco_prob	-0.0117***
	(0.00314)
loan_prob	0.00741*
	(0.00405)
oilgas_prob	-0.00151
	(0.00291)
supch_prob	-0.00427*
	(0.00258)
intrts_prob	-0.00393
	(0.00435)
litig_prob	0.00658*
	(0.00378)
reg_prob	0.0107***
	(0.00353)
contract_prob	-0.00491*
	(0.00266)
tax_prob	-0.00617*
	(0.00329)
infl_prob	0.00214
	(0.00396)
forex_prob	0.00984***
	(0.00351)
bank_prob	-0.00230
	(0.00328)
returnL	-0.103***
	(0.00396)
Observations	120,867
R-squared	0.670
Controls+	Yes
Time FE	Yes
Sector FE	Yes
Sector x Time FE	Yes
Firm FE	No

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1