Transformers, Central Banker Speeches, Tail Stock Returns

Krisztian Gado*

September 24, 2024

Abstract

I use transformers to construct central bank speech sentiment indices using financial text trained (expert) and not financial text trained (layman) models. I find that cross-sentence information is important for forecasting, and aggregating across positive and negative sentiment loses information. Empirically, sentiment measured by expert and aggregate indices affects stock returns. The share of tail stock returns increases (decreases) in response to negative Fed (Bank of Canada) speech sentiment. Negative Fed speech sentiment has a spillover effect with the same sign on the Canadian stock market, and leads changes in negative Canadian central banker speech sentiment. The relationship is driven by Fed Governor speeches. On trading days following Fed speeches, a long-only trading strategy based on return response to negative Fed speech sentiment could have outperformed the S&P 500 by 47% and 0.61 in terms of cumulative returns and Sharpe ratio respectively.

Keywords: Central Bank Communication, Sentiment Analysis, Transformers, BERT, Asset Pricing

JEL Codes: E58, G11, G12, G32

*Gado: Brandeis University, Email: krisgado@brandeis.edu. I thank Raphael Schoenle, Catherine L. Mann, Davide Pettenuzzo, Eric T. Swanson, Michael McMahon, Ahmad Namini, Daniel Bergstresser, as well as seminar participants at Brandeis University, and the Bank of England for useful comments and discussions.

"Monetary policy is 98% talk and 2% action." Bernanke (2015)

1 Introduction

Expectations are a key channel through which central banks achieve their goals, and verbal communications are a core component of this. Whether and to what degree messages central bankers send through their speeches can be quantified, and if different types of market participants interpret these uniformly, are important and unresolved questions. While there have been many attempts to quantify verbal central bank signaling using the latest developments in BERT models[1] including but not limited to Fisher et al. (2023) and Filippou et al. (2024), these attempts have largely treated these models as black boxes. As such, they employed them in the traditions of older approaches that were not able to take advantage of the most significant innovation of BERT models: being able to numerically represent contextual information. BERT captures context by numerically representing how words' meaning is influenced by their position within a sentence, and by their connections to other sentences. This ability to understand and quantify context is crucial for accurately interpreting central bank speeches. Fisher et al. (2023) and Filippou et al. (2024) construct sentiment indices using BERT models on a sentence-by-sentence basis rather than processing larger blocks of text in one shot. Consequently, their sentiment indices do not incorporate cross-sentence information.

I propose an alternative approach to using BERT models by applying BERT to larger blocks of text when constructing sentiment indices. Additionally, I address the open question of how different market participants interpret central bank signaling by creating multiple sentiment indices, each derived from BERT models trained on different types of text. Specifically, financial text trained (expert) models, and non-financial text (layman) models. Depending on the training data, different BERT models classify different information as neg-

[1]BERT (Bidirectional Encoder Representations from Transformers) is an open source machine learning framework used for natural language processing.

ative and positive sentiment (bad and good news). I analyze the effectiveness of these indices in explaining daily tail stock returns in U.S., Canadian, and international spillover contexts, as well as in stock return forecasting. My findings indicate that aggregating positive and negative sentiment can obscure critical information, and that cross-sentence information is crucial for stock return forecasting.

Empirically, sentiment measured by expert indices and averages of expert and layman indices affects stock returns. The share of tail stock returns increases (decreases) in response to negative Fed (Bank of Canada) speech sentiment. I detect international spillovers: Negative Fed speech sentiment signaling has a spillover effect with the same sign on the Canadian stock market, and leads changes in negative Canadian central banker speech sentiment. In the U.S., the relationship is driven by Fed Governor speeches. Comparing performance only on trading days following Fed speeches, a long-only trading strategy based on return response to negative Fed speech sentiment could have outperformed the S&P 500 by 47% in cumulative returns and 0.61 in Sharpe ratio over the sample period.

I start with an intuitive example of how BERT translates contextual information into numbers, followed by how these numbers can be classified as positive (*good news*), or negative (*bad news*). I also give examples of how expert and layman models differ in terms of their classifications of (*good*) versus (*bad*) news, as well as how cross-sentence information is important.

To construct the sentiment indices, I scraped Fed speech manuscripts from the Board of Governors website, and collected Bank of Canada central banker speech manuscripts from the Bank for International Settlements (BIS). I used ten of the most popular BERT models for text classification from *Hugging Face*, a popular open-source platform aimed at democratizing machine learning. The models were chosen optimizing for comparable outputs and diverse training sets. I classified each speech using every BERT model. Post-cleaning, the dataset has 1688 Fed speeches which were given between June 1996 and November 2023, and 438 Canadian central banker speeches which were given between March 1997 and

February 2024.

I construct a novel measure to capture daily tail returns: the cross-sectional share of returns outside the standard deviation range around the mean. This is a metric that is highly correlated with kurtosis by definition, but unit changes in it are easier to interpret. Figure (1) motivates this measure graphically, as well as demonstrates how a speech charged with negative sentiment by Chairman Powell can increase the share of extreme tail returns.

I analyze whether Fed speech sentiment indices are associated with tail stock returns in the U.S. and Canada, and following the work of Iacoviello and Navarro (2019), test for international spillover effects. I find that negative Fed speech sentiment is positively associated with an increase in tail returns in both countries. The relationship is driven by Governor speeches. Conversely, negative Bank of Canada speech sentiment is linked to a decrease in Canadian tail returns, with no evidence of spillover to the U.S. Furthermore, I find that Canadian speech sentiment can be predicted by preceding Fed speech sentiment, but not vice versa, reinforcing the unidirectional spillover effect from the U.S. to Canada.

Lastly, I present a long-only trading strategy based on forecasts from regressing stock returns on preceding negative Fed speech sentiment. On trading days following Fed speeches, the strategy outperformed the S&P 500 by 47% and 0.61 in terms of cumulative returns and Sharpe ratio respectively. Using a context-inclusive sentiment index is key, as it outperforms an analogous non-context approach by 68% and 0.76 in terms of cumulative returns and Sharpe ratio respectively.

Related Literature Fed official speech sentiment has been studied by a large literature in economics and finance. Most related to this paper is the sentiment analysis Fisher et al. (2023) have done on FOMC meeting materials. In particular, Fisher et al. (2023) construct a sentiment index using a version of FinBERT, and show that if FOMC meeting materials with a five-year lagged release date were made available in real time, market expectation errors of future policy rates could be reduced. The impact of the Fed speeches, the impact of Fed speech sentiment on financial assets, and the viability of using transformer models

on financial text has also been demonstrated by Gardner et al. (2022), Gorodnichenko et al. (2023), Swanson (2023), Granziera et al. (2023), Alexopoulos et al. (2024), Schmanski et al. (2023), Mishev et al. (2020), Barbaglia et al. (2022), Chen et al. (2022), and Filippou et al. (2024).

Contribution I propose applying BERT models to larger blocks of text. This is methodologically important, since the current literature ignores contextual information.[2][3] Second, by decomposing the sentiment indices into expert and layman models, I show that models substantially disagree about positive sentiment, and they largely agree about negative sentiment. This finding speaks against taking a weighted average of positive and negative sentiments – the standard practice in the literature. It also speaks to how different types of economic agents can interpret Fed signals differently. I show that, supporting the previous literature, daily stock returns do react to an increase in negative sentiment signalled through central bank speeches both in the U.S. and in Canada. The contribution here is the finding that positive sentiment does not drive tail returns on its own, and that tail returns are driven by Governor speech sentiment specifically. I also establish these findings in an international spillovers context, showing that the Fed sentiment spills over into Canada, but find no evidence of the reverse. Lastly, the S&P 500 outperforming trading strategy shows that markets were inefficient in the past with respect to central banker speech sentiment, and that contextual information is important for trading and forecasting purposes.

The paper proceeds as follows: Section (2) describes the speeches used, other data sources, and the thought process behind the measure of tail risk used. Section (3) explains the intuition for how transformer and BERT models work, with an audience of both experts and those without a computer science degree in mind. Section (4) describes the process used to construct the sentiment indices, documents trends in these indices themselves, and contrasts

[2]The current literature tends to apply BERT sentence by sentence, or uses topic- and modifier-keyword pairs. For example, see Fisher et al. (2023), Schmanski et al. (2023), Gardner et al. (2022), and Filippou et al. (2024).

[3]As an example of why context matters, consider the sentence *unemployment has decreased*. Evaluating this sentence from a policymakers perspective, in a low-inflation environment this would generally be good news. However, in a high-inflation environment it would generally be bad news.

them. Section (5) introduces the empirical framework and the results in a U.S., Canadian, international spillovers, and Chair vs. Governor context. Section (6) presents the trading strategy. Section (7) concludes.

2 Data

I use five data sources to construct the datasets: the Fed speech data, daily stock return data, daily four Fama-French factor data, VIX data, and Bank of Canada official speech data. The sample period is from June 1996 to November 2023.

2.1 Fed Speech Data

I scrape publicly available Fed speech manuscripts from the Board of Governors, and the Board of Governors archive websites. The former was used for speeches given between 2006 and 2023, and the latter for speeches that were given between 1996 and 2005. The websites contain the specific dates when the speeches were given, and the names of the officials who gave them. The officials in the sample are chairs, vice chairs, governors, and in one instance a director. The speech data exclude testimonies before governmental agencies, and are restricted to those speeches given during events such as forums, symposiums, conferences, and commencements. I drop speeches that have fewer than 200 words. The mean number of speeches per month over the sample period is 5.1 with a standard deviation of 3.2, and the mean word count per speech is 2808.4 with a standard deviation of 970.2. The first speech in the dataset was on the 13^{th} of June, 1996, and the last was on the 16^{th} of November, 2023. Post-cleaning 1688 speeches remain, all of which are used to create the sentiment indices. Of these speeches, 1608 were given on 1354 distinct trading days, which I use to test whether speech sentiment drives tails of daily stocks returns. Excluding the Chairs, there are 1235 distinct speeches in the dataset.

2.2 Daily Stock Data

The daily stock return data are collected from CRSP (through WRDS), and pertain to common stocks with a regular security status traded on the NYSE, the Nasdaq, and the AMEX. Since my focus is on the impact of speech sentiment on the given trading days excluding dividends, I use returns calculated as the percentage difference between stocks' open and close prices. Given the research question, dividends are not considered. I use these realized returns to construct time series for the first four moments (mean, standard deviation, skew, and kurtosis), as well as my preferred measure of tail returns (the share of returns a standard deviation or more away from the mean).

To clean the data, I first drop observations where open or close prices are not available. Then, I discard microcap stocks by dropping tickers below the 30^{th} percentile of market caps on the given trading day, based on the last trading day's close prices. Then, I drop returns above and below the 99.99^{th} and the 0.01^{th} percentile of all returns across all trading days. Shares with open or close prices below one dollar are also dropped to discard penny stocks. Lastly, I drop the first trading day of all tickers to make sure there is no particular IPO day effect affecting the returns.

Post-cleaning and across the whole sample period, the equal weighted-mean, standard deviation, minimum, and maximum daily returns are: 0%, 3%, -30.5%, and 38.1% respectively. While the minimums and maximums might seem like outliers relative to the means and standard deviations, the censoring I employ is already strict given this paper's focus on tail returns. To give an example of a tail return that this censoring drops but is a reasonable daily return, consider Silicon Valley Bank's 39% decrease on the 9^{th} of March, 2023. Post-cleaning, the mean number of tickers considered per trading day is 3835.

2.3 Other Data Sources

Both the daily four Fama-French factors (smb, hml, mktrf, umd), and the CBOE VIX data were accessed through WRDS. The CBOE VIX data used pertain to the S&P500 Volatility

Index.

Canadian central banker speech manuscripts were obtained from the Bank for International Settlements (BIS), and pertain to speeches made by past and present governors, deputy governors, and senior deputy governors. There are 438 Canadian central banker speeches in the dataset. Canadian stock returns are from Compustat, and consist of stocks traded on the Toronto Stock Exchange.

2.4 Densities Over Time & Motivating the Tail Share Measure

The dependent variable of interest in the baseline specifications is the daily amount of tail risk in the U.S. stock market. Given standardized returns,[4] I construct a novel measure of daily tail risk by the share of returns a standard deviation or more away from zero. This measure is highly correlated with kurtosis by construction. However, unlike in the case of kurtosis, a unit increase in the share of returns a standard deviation or more away from the mean can be easily understood in terms of its impact on tail risk. I include both positive and negative ends of the distribution as risk, since both of these represent large deviations in pricing activity. Figure (2) shows the 30-day moving average of tail risk over time in the U.S. stock market.

3 Transformer, BERT, and Sentiment Intuition

This section provides intuition as to how and why *Bidirectional Encoder Representations from Transformers (BERT)* models are so novel and valuable,[5] with an audience of both experts and those without a computer science degree in mind.[6]

[4]The standardizing is done by demeaning and dividing returns at time t by the means and standard deviation of all returns at time t specifically.

[5]The popularity of BERT and transformer models is represented by the citation count of the research papers associated with developing them. According to arXiv as of March 2024, the paper on BERT by Devlin et al. (2019) had over 94,000 citations. The paper by Vaswani et al. (2017), which describes transformers, had over 111,000 citations.

[6]Further (technical) details beyond the intuition laid out in this paper about the mechanisms underlying transformer and BERT models can be found in Vaswani et al. (2017) and Devlin et al. (2019).

The power of BERT models comes from being able to numerically represent contextual information. Context refers to relationships words have depending on their position, as well as across sentences. BERT models are an application of transformer models. This paper uses BERT models to analyze the sentiment communicated through Fed speeches. Given a speech, the sentiment analysis through BERT has two major steps: first, get a numerical representation of the speech, preserving as much information as possible; second, classify the numerical representation obtained in step one. The classification process is what yields the sentiment indices.

3.1 Numerical Representation of a Text

BERT is able to process and numerically represent information across multiple sentences simultaneously, enabling it to capture information arising from relationships between sentences and from word positions. That is, it is able to capture context. To demonstrate the intuition behind how numerically representing context works, consider the following representative problem, where one has to fill in the missing word.

Fill in the missing word: Bob put all his life savings into Shell stock yesterday. Today the stock must have gone _____ because he cannot stop smiling.

The key to filling in the correct word *"up"* with a high degree of certainty is reading the text both before and after the missing word. BERT models are trained to do precisely this, mimicking how a human would solve this problem. The underlying mechanism is that the transformer architecture allows BERT to process text both forward and backward simultaneously. Prior to BERT and transformer models, other models were sequentially (unidirectionally) processing text. Thus, they had limited success with filling in missing words with a high degree of certainty.

The training process of BERT involved fine-tuning functions that represent textual information with numbers, doing essentially the previous fill-in-the-missing-word exercise at large scale. Hence, BERT was parameterized maximizing an objective function increasing

in correct guesses of missing words for the texts BERT was trained on. BERT achieved its performance by simultaneously processing text both forward and backward to come up with the best numerical representation for the text, and which allowed it to make its best prediction. The base training texts for BERT were the Toronto Book Corpus and all of English Wikipedia, which have about 3.3 billion words worth of text in total.

3.2 Classification

BERT models used for classification have another training step (fine-tuning process) involved. A softmax layer, which is most commonly a logit function, is added to the base BERT model. Using the ability to convert textual information along with its context to numbers, the logit function is then parameterized on a categorized (labelled) dataset. The final outputs from these models are generally probabilities and their associated categories (labels), given the underlying logistic function in the final step.

There are many labelled datasets with which this final step can be done. There is considerable heterogeneity across models in terms of how far and on what labelled datasets they are fine-tuned on. I use ten different BERT-based models where in this final step, the labels for the labelled datasets were positive, negative, or neutral given financial and non-financial texts. Section (3.3) gives examples of how sentences are classified by these models, clarifying the differences between expert (financial text trained) and layman (not financial text trained) models. Section (3.4) gives examples of how cross-sentence information affects classification.

3.3 Positive, Negative, and Neutral Sentiment

Expert models classify sentiment by *reading* the text from the point of view of a financial analyst, whereas layman models take the perspective of a person not versed in finance. Positive sentiment is to be interpreted as good news, negative sentiment as bad news, and neutral sentiment as pure information without any good or bad implications. Tables (1) and

(2) provide examples demonstrating how expert and layman models differ, classifying the sentences *our margins are growing thinner*, and *dogs are our friends* respectively.

For the first example, Table (1) shows how the representative expert model captures the financial bad news by classifying the sentence as negative with a 96% probability. Conversely, the representative layman model classifies the sentence as positive with a 65.3% probability, and neutral with a 34.3% probability by focusing on the words *growing* and *thinner*, words which are considered good news in most non-financial contexts.

For *dogs are our friends*, Table (2) shows how the expert model associates the statement as pure information without having good or bad implications with a 90.3% neutral probability. Conversely, the layman model classifies the sentence as positive with a 97.2% probability, having the value judgment of a representative person who likely has and is fond of their dog.

3.4 Cross-Sentence Information

The power of BERT, and transformer models comes from being able to capture cross-sentence information. Tables (3) and (4) provide examples of why this is crucial. Consider the sentence *"Unemployment decreased"*. While this would generally be considered good news from a financial perspective, and as indicated by the FinBERT model in Table (3), its interpretation depends on context and would be prone to error if we analyzed this sentence in isolation.

To show why context matters, consider the sentence pair: *Inflation was high. Unemployment decreased.* Traditional methods such as topic- and modifier-keyword pairs would interpret this as neutral news, since each expression would be evaluated in isolation. From an asset pricing standpoint this should be interpreted as bad news, however, since high inflation with decreasing unemployment implies that interest rates will rise. Unlike traditional methods, FinBERT is able to classify this as bad news as shown in Table (4).

4 Sentiment Indices

This section describes the construction of the sentiment indices using the intuition for sentiment built in Subsection (3.3). It describes the U.S. aggregate positive and negative sentiment indices, and compares their expert and layman decompositions. I find that the aggregate indices contain sensible patterns. The decompositions are highly correlated for the negative indices, and weakly correlated for the positive indices. This implies that negative sentiment is easier to capture than positive sentiment.

4.1 Index Construction

I use ten different BERT-based models to classify the central banker speeches. Five of these are expert, and five are layman models. I take the equal-weighted mean across all ten models to get the aggregate indices. Similarly, I take equal-weighted averages across the five expert and five layman models to get their respective indices. The specific models I use are detailed in Table (5).

As inputs, I use the first n words of a speech a model can take without batching.[7] Doing more would require batching the inputs, which entails some questionable assumptions.

Not batching requires the assumption that the first 440 words on average contain enough information to capture the sentiment of the whole speech. This assumption is a lot more reasonable than the potential ones required for batching, given that, by definition, the introductory part of a speech should clearly indicate the direction the whole speech will take.

In contrast, batching requires making assumptions about how sentiment and information are distributed across the entire text. Batching would require breaking the texts into subtexts, the size of which depend on how many words the tensor can process. Each subtext would be classified, and then one would average across. The classification can be sensitive to where the texts are cut into subtexts, however, and cross-batch information is ignored.

[7]For nine out of the ten models, n is on average the first 440 words of a speech, whereas for the model by Pérez et al. (2021) I use the first 110 words on average.

As described in Subsection (3.3), each of these models outputs the probabilities that a given speech is positive, negative, and neutral. These probabilities sum to one by construction. I interpret the average across the positive (negative) probabilities as the point estimate for the positive (negative) speech sentiment index at a given point in time.

The pseudocode for the speech classification process can be found in algorithm (1).

4.2 Aggregate Indices

The aggregate U.S. indices contain patterns aligning with significant macroeconomic events over the sample period. The patterns, shown in Figure (3)[8] are in line with the work of Fisher et al. (2023): Fed speech sentiment contains macroeconomic information.

Figure (3a) shows that the aggregate positive index has local maxima and minima at key macroeconomic events: during the end days of the Dot-com Bubble, during the Great Financial Crisis, and the Coronavirus Pandemic.

The aggregate negative index in Figure (3b) shows similar patterns as well, with local maxima at and after the Great Financial Crisis, and from the onset of the Coronavirus Pandemic. The lack of a local maximum during the end of the dot-com bubble aligns with the narrative that the event did not cause significant macroeconomic risk.

4.3 Expert and Layman Indices

Positive and negative indices should not be aggregated. The decomposed expert and layman U.S. indices in Figure (4)[9] imply that there is an advantage to handling the positive and negative series separately – in contrast to the current practice in the literature.

The positive expert and layman indices are weakly positively correlated as shown in Figure (4a). While they co-move during the Dot-com Bubble, the Great Financial Crisis, and the initial Covid period, they mostly diverge in other periods. The correlation coefficient

[8]The Figure shows one-month moving averages for representation purposes.
[9]The Figure shows six-month moving averages for representation purposes.

between the two indices is 0.33 at the daily level. The higher average level of layman models is consistent with the indices being constructed from the first parts of speeches, since these usually contain pleasantries such as *it is a pleasure to be here today.*

The negative expert and layman U.S. indices are strongly positively correlated as shown in Figure (4a). While there are occasional divergences, such as during the end of the Dot-com Bubble, and the first interest-rate-raising cycle starting in late 2015, the periods without co-movement are short lived. The correlation coefficient between the two indices at the daily level is 0.63. The strong correlation between the expert and layman models is consistent with negative sentiment being easier to capture because it involves words common across financial and non-financial circles, such as *no, disaster*, and *decrease.* An interesting caveat is that divergences between negative speech indices are relatively more likely to happen during periods of economic stability, and lack of real macroeconomic risk.

5 Results & Empirical Setup

This section presents the main empirical frameworks and the results. It starts with the analysis of the relationship between tail returns and the different sentiment indices in the U.S., Canada, and in between central banks respectively. It concludes with analyzing whether sentiment impact differs by speaker type.

5.1 U.S. Results

When I consider the impact of Fed speech sentiment on U.S. stock market tail returns, I find that negative sentiment is associated with a statistically significantly higher share of tail returns using both aggregate and expert indices. There is no evidence that negative sentiment as measured by layman models, or that positive sentiment impacts tail returns. The takeaways are: (1) As argued in Subsection (4.3), positive and negative indices capture information of different quality, and so should not be aggregated – against the standard

practice in the literature. (2) A *non-expert* layman person's interpretation of Fed signals does not impact financial markets. (3) The findings support the results of Alexopoulos et al. (2024), Gardner et al. (2022), and Swanson (2023), in that Fed speeches impact stock returns.

Empirical Setup The following regression framework was used to capture the impact of Fed speech sentiment on the share of tail returns:

$$O_t = c + \beta_1 * S_{pos,t,m} + \beta_2 * S_{neg,t,m} + \alpha + \Omega_t + u_t \tag{1}$$

In equation (1), $S_{pos,t,m}$ and $S_{neg,t,m}$ denote the positive and negative speech indices at time t, where m stands for model type (expert, layman, or both). O_t denotes the share of tail returns at t, α are Fed Chair fixed effects, and Ω_t are controls.

Each sentiment model type m is considered in a separate regression. The paper considers six specifications per model type, each adding new controls and keeping the previous ones. Specification (1) includes no controls. Specification (2) adds the four Fama-French factors. Specification (3) adds the first four moments of returns[10] at time t. Specification (4) adds the percent change in the VIX at time t.[11] Specification (5) adds the share of tail returns at $t-1$, thus O_{t-1}, to capture any persistence in the level of tail returns. Specification (6) includes the mean returns at time $t-1$, controlling for the leverage effect.

Estimation Based on aggregate and expert indices, for each unit of extra negative speech sentiment, there is an additional 0.004% to 0.017% higher share[12] of returns a standard deviation or more away from the mean, depending on the specification considered in Tables (6) and (7). While the coefficients are small in terms of their economic significance, their statistical significance is remarkable given: (1) the challenges of tail share estimation (Joanes and Gill (1998); Gilli et al. (2019)), and (2) speaker fixed effects are not controlled for

[10]Mean, standard deviation, skew, kurtosis.

[11]The percent change in the VIX is used since levels in the VIX and the standard deviation of daily returns (which were already added in specification (3)) are highly correlated by definition.

[12]In absolute terms.

due to data limitations. Table (8) shows that the layman negative sentiment index is not statistically significant for explaining tail returns under any specification. Positive sentiment is not associated with a higher share of tail returns once the return moments are controlled for in Tables (6), (7), and (8).

5.2 International Evidence: Canadian Results

Negative Bank of Canada speech sentiment, as measured by expert and aggregate indices, is negatively associated with the share of tail stock returns in the Canadian market. There is no evidence that negative sentiment, as measured by layman models, or that positive sentiment impacts tail returns. These results are important, since they provide international evidence of the relationship.

Empirical Setup I analyze the impact of Bank of Canada official speeches on Canadian stock returns in the same framework as specified in equation (1). Specification (1) includes no controls. Specification (2) adds the four Fama-French factors from the U.S., since no daily Canadian version is available. Specification (3) adds the first four moments of returns[13] at time t. Specification (4) adds the percent change in the U.S. S&P 500 VIX at time t, since I could only access the Canadian TSX VIX post-2020.[14] Specification (5) adds the share of tail returns at $t-1$, thus O_{t-1}, to capture any persistence in the level of tail returns. Specification (6) includes the mean returns at time $t-1$, controlling for the leverage effect. Contrary to the U.S. stock and Fed sentiment version in section (5.1), the sample period here is from January 2011 to February 2024. The reasons for the narrower sample period are twofold: (1) The BIS data contain multiyear gaps without speeches in the years prior to 2008. (2) The Compustat returns data do not have enough stock observations with both open and close prices per day to estimate a high moment like the tail measure.[15]

Estimation The impact of Bank of Canada speech sentiment on tail risk in the Canadian

[13]Mean, standard deviation, skew, kurtosis.

[14]The percent change in the VIX is used since levels in the VIX and the standard deviation of daily returns (which were already added in specification (3)) are highly correlated by definition.

[15]The rule of thumb I used is to have at least 170 return observations per trading day.

stock market is the opposite relative to the U.S., as presented in Tables (9), (10), and (11). Aggregate and expert negative sentiment is negatively associated with a higher share of tail returns in the Canadian stock market. For each unit of extra negative speech sentiment, there is an additional 0.017% to 0.039% lower share of returns a standard deviation or more away from the mean, depending on the specification considered. As in the U.S., there is no evidence that layman negative sentiment affects tail returns. As in the U.S. markets, once return moments are controlled for, positive sentiment is not associated with a different share of tail returns.

5.3 Spillovers: Fed Sentiment, Canadian Stocks

The impact of Fed speech sentiment on tail risk in the Canadian stock market is the same as in its U.S. counterpart. Aggregate and expert negative Fed sentiment is positively associated with a higher share of tail returns in the Canadian stock market. There is no evidence that negative sentiment, as measured by layman models, or that positive sentiment impacts tail returns. The takeaways are threefold: (1) The findings support the literature on Fed signalling spilling over to foreign markets (Iacoviello and Navarro (2019)). (2) The findings support the results in Section (5.1). (3) The implication is that the impact of negative Fed signalling is different from negative Bank of Canada signalling.

Empirical Setup I analyze the impact of Fed speeches on Canadian stock returns in the same framework as specified in equation (1). Specification (1) includes no controls. Specification (2) adds the four Fama-French factors from the U.S., since no daily Canadian version is available. Specification (3) adds the first four moments of returns[16] at time t. Specification (4) adds the percent change in the U.S. S&P 500 VIX at time t, since I could only access the Canadian TSX VIX post-2020.[17] Specification (5) adds the share of tail returns at $t-1$, thus O_{t-1}, to capture any persistence in the level of tail returns. Specification

[16]Mean, standard deviation, skew, kurtosis.

[17]The percent change in the VIX is used since levels in the VIX and the standard deviation of daily returns (which were already added in specification (3)) are highly correlated by definition.

(6) includes the mean returns at time $t-1$, controlling for the leverage effect. Contrary to the U.S. stock and Fed sentiment version in Section (5.1, the sample period here is from January 2011 to November 2023. The reasons for the narrower sample period are twofold: (1) The BIS data contain multiyear gaps without speeches in the years prior to 2008. (2) The Compustat returns data do not have enough stock observations with both open and close prices per day to estimate a high moment like the tail measure.[18]

Estimation Tables (12), (13), and (14) present the results. Aggregate and expert negative Fed sentiment is positively associated with a higher share of tail returns in the Canadian stock market in all specifications. For each unit of extra negative speech sentiment, there is an additional 0.012% to 0.04% higher share of returns a standard deviation or more away from the mean, depending on the specification considered. As in the U.S., there is no evidence that layman negative sentiment affects tail returns. As in the U.S. markets, once return moments are controlled for, positive sentiment is not associated with a different share of tail returns.

5.4 Spillovers: Bank of Canada Sentiment, U.S. Stocks

I find no evidence that U.S. tail returns are affected by Canadian speeches.

Empirical Setup I analyze the impact of Bank of Canada official speeches on U.S. stock returns in the same framework as specified in equation (1). Specification (1) includes no controls. Specification (2) adds the four Fama-French factors. Specification (3) adds the first four moments of returns[19] at time t. Specification (4) adds the percent change in the U.S. S&P 500 VIX at time t.[20] Specification (5) adds the share of tail returns at $t-1$, thus O_{t-1}, to capture any persistence in the level of tail returns. Specification (6) includes the mean returns at time $t-1$, controlling for the leverage effect. The sample period is from January 2011 to December 2023.

[18]The rule of thumb I used is to have at least 170 return observations per trading day.

[19]Mean, standard deviation, skew, kurtosis.

[20]The percent change in the VIX is used since levels in the VIX and the standard deviation of daily returns (which were already added in specification (3)) are highly correlated by definition.

Estimation Tables (15), (16), and (17) show no evidence of a spillover from Canadian speeches to the U.S. stock market under any specification which controls for return moments and beyond.

5.5 Spillovers between the Fed and the Bank of Canada

Negative Fed speech sentiment is positively associated with subsequent negative Bank of Canada speech sentiment, and is negatively associated with subsequent positive Bank of Canada speech sentiment. Negative Canadian speech sentiment is associated with a decrease in subsequent positive Fed speech sentiment. The finding supports the evidence of international spillovers established in Sections (5.3) and (5.4). That is, negative Fed speech sentiment is important, and has a persistent effect internationally. Layman indices not being significant imply that non-financially trained interpretations of central bank speeches do not spill over. The negative relationship between negative Canadian sentiment and following positive U.S. sentiment is consistent with Fed officials communicating with more caution when observing *bad news* in Canada.

Empirical Setup The following regression framework was used to capture the impact of lagged central banker speeches internationally:

$$S_{j,s,t,m} = c + \beta_1 * S_{k,pos,t-1,m} + \beta_2 * S_{k,neg,t-1,m} + u_t \tag{2}$$

In equation (2), where $s \in \{pos,\ neg\}$, $j \in \{Bank\ of\ Canada, Fed\}$, and k denotes the other bank – if $k = Fed$, $j = Bank\ of\ Canada$. $S_{j,pos,t,m}$ and $S_{j,neg,t,m}$ represent speech sentiment for positive and negative speech indices at time t, respectively, and m denotes the model type (expert, layman, or both). The variable t denotes the daily mean sentiment across all speeches given on that day, while $t-1$ refers to the mean speech sentiment from the previous time period when a speech was delivered.

Estimation For a unit increase in negative Fed speech sentiment, the following Canadian

negative sentiment will increase by 0.183 to 0.2 based on the regressions using the aggregate and expert indices, as shown in Table (18). The expert index also shows evidence of a negative relationship between negative signaling by the Fed and positive signaling by Canadian officials. There is no evidence of a lead-lag relationship using layman indices.

Table (19) shows that using expert indices, negative Bank of Canada speech sentiment leads positive Fed speech sentiment with a negative sign.

5.6 Non-Chair vs. Chair Sentiment

The impact of Fed sentiment on U.S. tail returns, and its spillover to the Canadian stock market are driven by Fed Governors. When using Chair speech sentiment, I find no evidence of these relationships once important variables are controlled for. The potential reasons are twofold: (1) Most of the speeches are given by Governors (1176 out of 1608 in the U.S., 523 out of 668 in Canada). Since the analyses focus on the tails, which are inherently noisy, there may not be enough Chair speeches to detect the relationship. (2) It is also possible that Governors' speeches are the ones revealing new information for markets.

Empirical Setup I analyze the impact of Bank of Canada official speeches on Canadian stock returns in the same framework as specified in equation (1). Details about the U.S. and Canadian stock market specifications can be found in Sections (5.1) and (5.3) respectively.

Estimation Negative Governor sentiment increases the share of tail returns in the U.S. and Canada from 0.005% to 0.01% and 0.018% to 0.027% respectively. There is some evidence that positive Fed speech sentiment impacts U.S. tail returns; however, the evidence is not robust to controlling for the daily persistence of tail risk. There is no evidence that the Chair speeches impact tail returns in either the U.S. or Canada. Tables (20), (21), (22), and (23) present the findings.

6 Trading Strategy

On trading days following Fed speeches, a long-only trading strategy based on aggregate negative sentiment outperforms the S&P 500 by 47% in cumulative returns, and 0.61 in terms of Sharpe ratio over the sample period. An identical trading strategy based on a negative sentiment index constructed without cross-sentence information (no context index), would underperform the S&P 500 by 21% in cumulative returns, and 0.15 in terms of Sharpe ratio. The tail risk return results in Section (5.1) are, however, robust to using sentiment indices constructed without including contextual information. The implications are that negative sentiment indices are a viable way to construct profitable trading strategies, that contextual information is important when using sentiment indices for stock return forecasting and trading strategies, and that markets have been inefficient with respect to contextual sentiment information.

No Context Sentiment Index A trading strategy is better suited to comparing context-inclusive vs. exclusive sentiment indices than tail return analysis, since tail returns estimation requires an order of magnitude more degrees of freedom (Joanes and Gill (1998); Gilli et al. (2019)). I construct sentiment indices where I exclude cross-sentence information akin to Fisher et al. (2023). That is, for a given speech, I apply BERT sentence by sentence, and then I take equal-weighted averages by sentiment to create the positive and negative indices.[21] I refer to these indices as **no context sentiment indices**.

To better understand the relationship between context and no context sentiment indices, I run OLS regressions of the aggregate context-based indices on the aggregate no context indices of the form:

$$S_{i,j}^{context} = c + \beta_1 S_{i,j}^{no\ context} + u_i, \tag{3}$$

where i identifies the particular speech, and $j \in \{positive, negative\}$

The no context positive (negative) index explains 77.4% (76.7%) of the variation in the

[21] I repeat this using all ten BERT models and average across to create the aggregate positive and negative indices, as in the case of the context-based approach in Section (4.1).

baseline context-inclusive positive (negative) sentiment index, as shown in Table (24).

The tail risk return results in Section (5.1) are robust to using sentiment indices constructed without including contextual information. I re-estimate the models involving Fed speeches in Section (5.1) using the no context indices created in Section (6) to see if the results hold using the no context indices. Tables (25), (26), and (27) show that the coefficient signs and estimates are consistent with those in Section (5.1). Does this mean that the additional information contained in context-inclusive indices is unimportant? Not at all, as the trading strategy below shows.

Trading Strategy The trading strategy is based on the aggregate negative sentiment index.[22][23] For simplicity, the strategy considered employs only long positions, and the main benchmark used is the S&P 500 returns on days following Fed speeches – when the trades are made. As a secondary benchmark, I use forecasts based on the aggregate no context negative sentiment index.

To forecast returns, I estimate models of the form in rolling windows:

$$ret_{i,t} = c_i + \beta_{1,i} * S_{Fed,neg,t-1,aggregate} + u_{i,t} \tag{4}$$

In equation (4) $S_{Fed,neg,t-1,aggregate}$ denotes the negative Fed speech sentiment index based on the aggregate negative index. Note that t corresponds to the first trading day after a speech on a preceding day $t-1$. $\beta_{1,i}$, and c_i are the stock specific slope and intercept coefficients. The rolling window size is the last 60 days on which speeches were made.

Algorithm (2) provides a pseudocode on how the forecasts are made, and how the trades are executed. In principle, after estimating the model and given the next speech sentiment, I make forecasts for the returns on the next trading day for the stocks whose $\beta_{1,i}$ coefficient is significant. If a stock returns forecast is above 0, and at the time its sentiment coefficient's p-value is in the lowest quartile of p-values, I take a long position in the stock at market open

[22]Recall that the aggregate negative index is the mean of the expert and layman indices.

[23]The aggregate index was chosen in the spirit of using an *ensemble-model-based* index for forecasting purposes.

of the next trading day and sell at market close. Then, I take an equal-weighted average of the returns. On these days, the benchmarks I use for comparison are the S&P 500, and the performance of a similar portfolio based on sentence-by-sentence classification-based negative sentiment.

Results The context-inclusive trading strategy outperforms the S&P 500 by 47% over the sample period, as shown in Figure (5). The strategy's Sharpe ratio is 0.61 higher than the S&P 500's on these days. Note that most of the out-performance comes from the strategy not losing as much as the S&P 500 in the 1997-2008 period, but the strategy earns consistent profits past 2008 as well. This implies that (1) between 1997 and 2008 the S&P 500 tended to perform negatively following Fed speeches, and (2) markets have been inefficient with respect to pricing in negative Fed speech sentiment.

The context-inclusive trading strategy also outperforms the trading strategy based on no context negative sentiment by 68% in terms of cumulative returns and 0.76 in terms of Sharpe ratio. It does this consistently through the whole sample period. The implication is that context provides important information for forecasting and trading purposes. An explanation as to why context matters for trading purposes, but yields comparable results in terms of explaining tail returns, is that tail returns require many more excess degrees of freedom to estimate accurately (Joanes and Gill (1998); Gilli et al. (2019)).

7 Conclusion

I addressed whether and to what degree central banker speech sentiment can be quantified and to what degree different types of market participants interpret these uniformly. Using transformers, I simultaneously constructed central bank speech sentiment indices using financial text trained (expert) and not financial text trained (layman) models. I found that cross-sentence information is important for forecasting, and aggregating across positive and negative sentiment loses information. Empirically, sentiment measured by expert indices and

averages of expert and layman indices affects stock returns. The share of tail stock returns increases (decreases) in response to negative Fed (Bank of Canada) speech sentiment. I detected international spillovers: Negative Fed speech sentiment signaling has a spillover effect with the same sign on the Canadian stock market, and leads changes in negative Canadian central banker speech sentiment. The relationship is driven by Fed Governor speeches. Comparing performance only on trading days following Fed speeches, a long-only trading strategy based on returns' response to negative Fed speech sentiment could have outperformed the S&P 500 by 47% and 0.61 in terms of cumulative returns and Sharpe ratio respectively.

Accurately measuring and understanding sentiment is important for monetary policy transmission. Going forward, there are several promising avenues for further research. Extensions could explore whether the standard deviation across the classifications of different models could be used to proxy for the clarity of central banker communication. Alternatively, in the spirit of Miller (1977) and Bali et al. (2023), can these standard deviations be used to explain returns? Is there a Fed sentiment spillover effect to other countries? How much better would more recent transformer-based models capture sentiment?

References

Alexopoulos, Michelle, Xinfen Han, Oleksiy Kryvtsov, and Xu Zhang (2024) "More than words: Fed Chairs' communication during congressional testimonies," *Journal of Monetary Economics*, 142, 103515, https://doi.org/10.1016/j.jmoneco.2023.09.002.

Araci, Dogu (2019) "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models."

Bali, Turan G, Bryan T Kelly, Mathis Mörke, and Jamil Rahman (2023) "Machine Forecast Disagreement," Working Paper 31583, National Bureau of Economic Research, 10.3386/w31583.

Barbaglia, Luca, Sergio Consoli, Sebastiano Manzan, Luca Tiozzo Pezzoli, and Elisa Tosetti (2022) "Sentiment Analysis of Economic Text: A Lexicon-based Approach," *Social Science Research Network* (4106936), `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4106936`.

Bernanke, S. Ben (2015) "Inaugurating a new blog," `https://www.brookings.edu/articles/inaugurating-a-new-blog/`.

Camacho, Jose Collados, Kiamehr Rezaee, Talayeh Riahi et al. (2022) "TimeLMs: Diachronic Language Models from Twitter," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 251–260, Dublin, Ireland: Association for Computational Linguistics, May, 10.18653/v1/2022.acl-demo.25.

Chen, Yifei, Bryan T. Kelly, and Dacheng Xiu (2022) "Expected Returns and Large Language Models," (4416687), `https://ssrn.com/abstract=4416687`.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019) "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *North*

American Chapter of the Association for Computational Linguistics, https://api.semanticscholar.org/CorpusID:52967399.

Filippou, Ilias, Christian Garciga, James Mitchell, and M Nguyn (2024) "Regional Economic Sentiment: Constructing Quantitative Estimates from the Beige Book and Testing Their Ability to Forecast Recessions," *Economic Commentary (Federal Reserve Bank of Cleveland)*, 10.26509/frbc-ec-202408.

Fisher, Eric, Rebecca McCaughrin, Saketh Prazad, and Mark Vandergon (2023) "Fed Transparency and Policy Expectation Errors: A Text Analysis Approach," *Federal Reserve Bank of New York Staff Reports* (1081), 10.59576/sr.1081.

Gardner, Ben, Chiara Scotti, and Clara Vega (2022) "Words speak as loudly as actions: Central bank communication and the response of equity prices to macroeconomic announcements," *Journal of Econometrics*, 231 (2), 387–409, https://doi.org/10.1016/j.jeconom.2021.07.014, Special Issue: The Econometrics of Macroeconomic and Financial Data.

Gilli, Manfred, Dietmar Maringer, and Enrico Schumann (2019) "Chapter 9 - Financial simulation at work: some case studies," in Gilli, Manfred, Dietmar Maringer, and Enrico Schumann eds. *Numerical Methods and Optimization in Finance (Second Edition)*, second edition edition, 189–216: Academic Press, https://doi.org/10.1016/B978-0-12-815065-8.00020-0.

Gorodnichenko, Yuriy, Tho Pham, and Oleksandr Talavera (2023) "The Voice of Monetary Policy," *American Economic Review*, 113 (2), 548–84, 10.1257/aer.20220129.

Granziera, Eleonora, Vegard H. Larsen, and Greta Meggiorini (2023) "Fed Sentiment and Expectations: Evidence from Speeches by FOMC Members," *Working Paper*, https://congresscreator.org/system/files/papers/GranzieraLarsenMeggiorini2022.pdf.

Hartmann, Jochen, Mark Heitmann, Christina Schamp, and Oded Netzer (2021) "The Power of Brand Selfies," *Journal of Marketing Research*.

Hazourli, Ahmed (2022) "FinancialBERT - A Pretrained Language Model for Financial Text Mining," 02, 10.13140/RG.2.2.34032.12803.

Huang, Allen H., Hui Wang, and Yi Yang (2023) "FinBERT: A Large Language Model for Extracting Information from Financial Text,"Technical report, 10.1111/1911-3846.12832.

Iacoviello, Matteo and Gaston Navarro (2019) "Foreign effects of higher U.S. interest rates," *Journal of International Money and Finance*, 95, 232–250.

Joanes, D. N. and C. A. Gill (1998) "Comparing measures of sample skewness and kurtosis," *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47 (1), 183–189, https://doi.org/10.1111/1467-9884.00122.

Miller, Edward M. (1977) "Risk, Uncertainty, and Divergence of Opinion," *The Journal of Finance*, 32 (4), 1151–1168, https://doi.org/10.1111/j.1540-6261.1977.tb03317.x.

Mishev, Kostadin, Ana Gjorgjevikj, Irena Vodenska, Lubomir T. Chitkushev, and Dimitar Trajanov (2020) "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers," *IEEE Access*, 8, 131662–131682, `https://api.semanticscholar.org/CorpusID:220836326`.

Pérez, Juan Manuel, Juan Carlos Giudici, and Franco Luque (2021) "pysentimiento: A Python Toolkit for Sentiment Analysis and SocialNLP tasks."

Schmanski, Bennett, Chiara Scotti, and Clara Vega (2023) "Fed Communication, News, Twitter, and Echo Chambers," Finance and Economics Discussion Series 2023-036, Board of Governors of the Federal Reserve System (U.S.), 10.17016/FEDS.2023.036.

Swanson, Eric T. (2023) "The Importance of Fed Chair Speeches as a Monetary Policy Tool," *AEA Papers and Proceedings*, 113, 394–400, 10.1257/pandp.20231073.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin (2017) "Attention is All you Need," in Guyon, I., U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett eds. *Advances in Neural Information Processing Systems*, 30: Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

8 Appendix

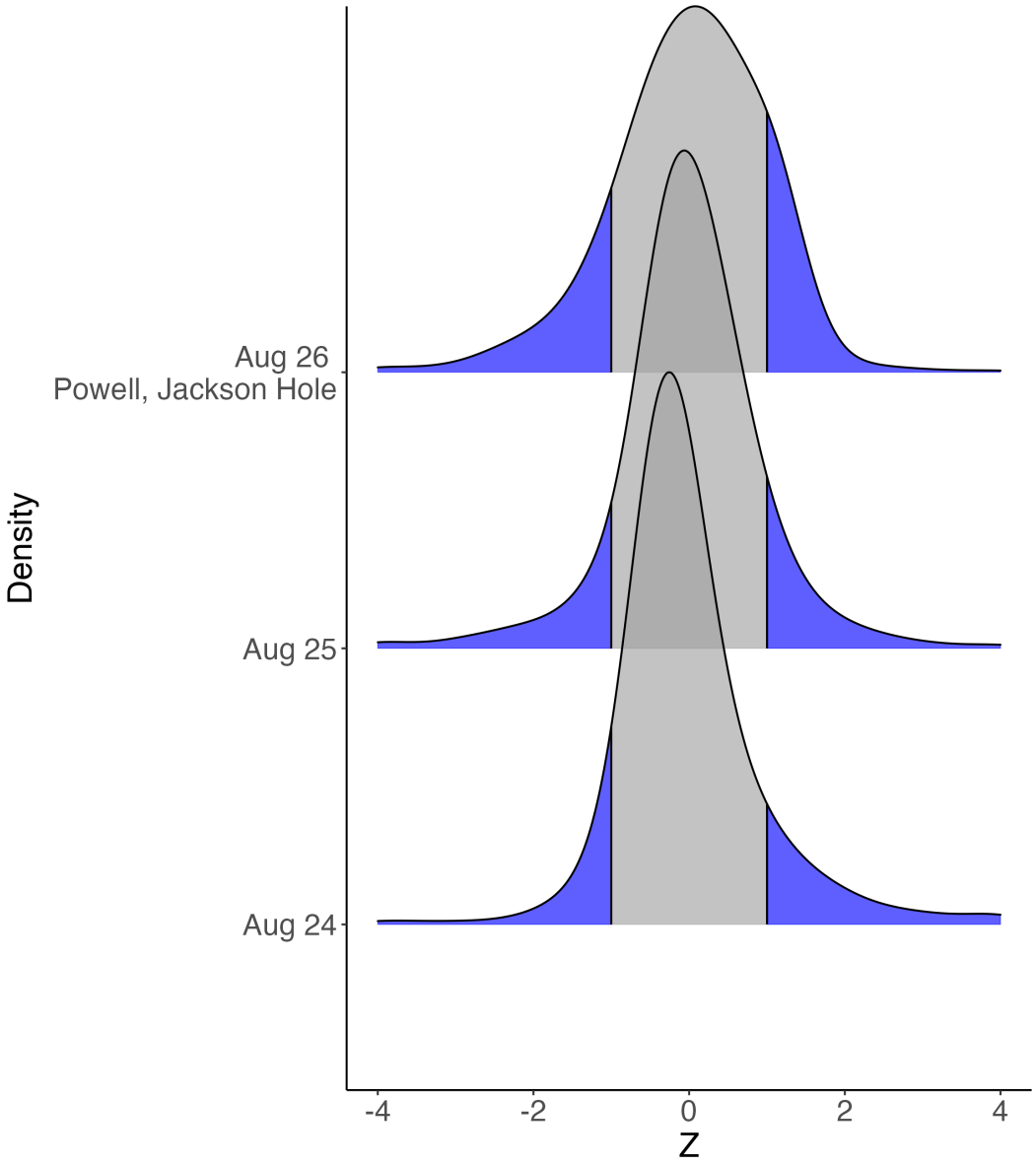


Figure 1: The evolution of standardized daily stock return densities over time between the 24^{th} of August 2022 and 26^{th} of August 2022. The 26^{th} of August marked Chairman Powell's Jackson Hole speech in which he strongly signaled that the Fed is determined to bring inflation back down to 2%. Each day was standardized using the respective means and standard deviations on the given days. The blue shaded area corresponds to Z score values one standard deviation above and below the mean. The areas of the shaded regions are 16%, 18%, and 30% respectively, showing how on the day of Chairman Powell's speech relatively *extreme* returns were much more likely to happen. Note that the densities are winsorized at -4 and 4.

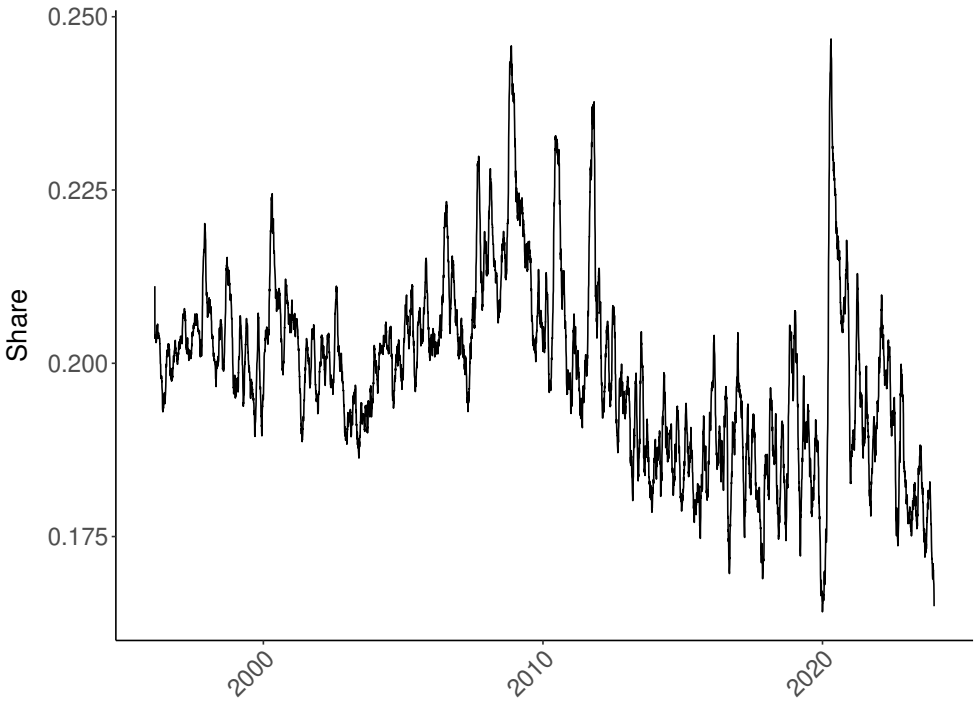


Figure 2: Right aligned 30 trading day moving average of the U.S. share of returns a standard deviation or more away from the mean, given the realized means and standard deviation on the respective trading days. The stock return data are censored as prescribed in Section (2). The stricter the censoring of extreme returns, the closer the shares over time would get to the normal distribution. With the current censoring, the lowest and highest returns are -30.5% and 38.1% respectively.

Table 1: Text sentiment analysis example by an expert (financial text trained) and a layman (not financial text trained) model type for the sentence: "Our margins are growing thinner."

Model Type	Model	Pr(Positive)	Pr(Neutral)	Pr(Negative)
Expert	FinBERT	0.015	0.024	0.960
Layman	BERTweet-SemEval2017	0.653	0.343	0.003

Table 2: Text sentiment analysis example by an expert (financial text trained) and a layman (not financial text trained) model type for the sentence: "Dogs are our friends."

Model Type	Model	Pr(Positive)	Pr(Neutral)	Pr(Negative)
Expert	FinBERT	0.051	0.903	0.046
Layman	BERTweet-SemEval2017	0.972	0.026	0.002

Table 3: Text sentiment analysis example by an expert (financial text trained) and a layman (not financial text trained) model type for the sentence: "Unemployment decreased."

Model Type	Model	Pr(Positive)	Pr(Neutral)	Pr(Negative)
Expert	FinBERT	0.607	0.040	0.353
Layman	BERTweet-SemEval2017	0.191	0.769	0.040

Table 4: Text sentiment analysis example by an expert (financial text trained) and a layman (not financial text trained) model type for the sentence: "Inflation was high. Unemployment decreased."

Model Type	Model	Pr(Positive)	Pr(Neutral)	Pr(Negative)
Expert	FinBERT	0.302	0.087	0.611
Layman	BERTweet-SemEval2017	0.041	0.872	0.087

Algorithm 1: Pseudocode for speech classifying a representative speech.

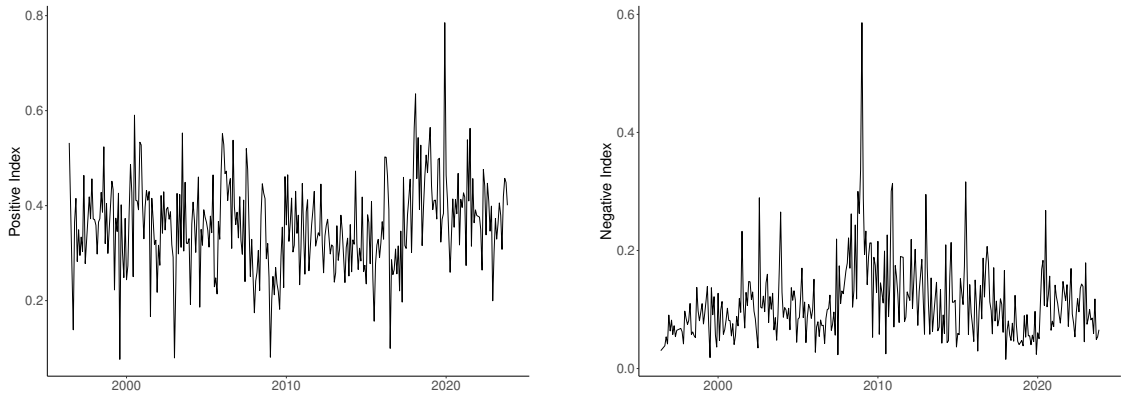
Input: speech, models, tokenizer

Output: pr(positive), pr(negative), pr(neutral)

```
take a speech
clean the speech from html artifacts
for each i in models:
    load transformer model[i] and its associated tokenizer
    tokenize the speech
    classify using the model
    extract and save the classification results
take summary statistics across the extracted results
```

Table 5: Base models used for text classification, their training % fine-tuning datasets, and the associated paper describing the process (when available). All the models are variations on the BERT model. That is, the models are fine-tuned versions of *Bidirectional Encoder Representations from Transformers*, developed by Devlin et al. (2019) at Google. The maximum number of tokens the models accept is 512, with the exception of model 9, which only takes up to 128.

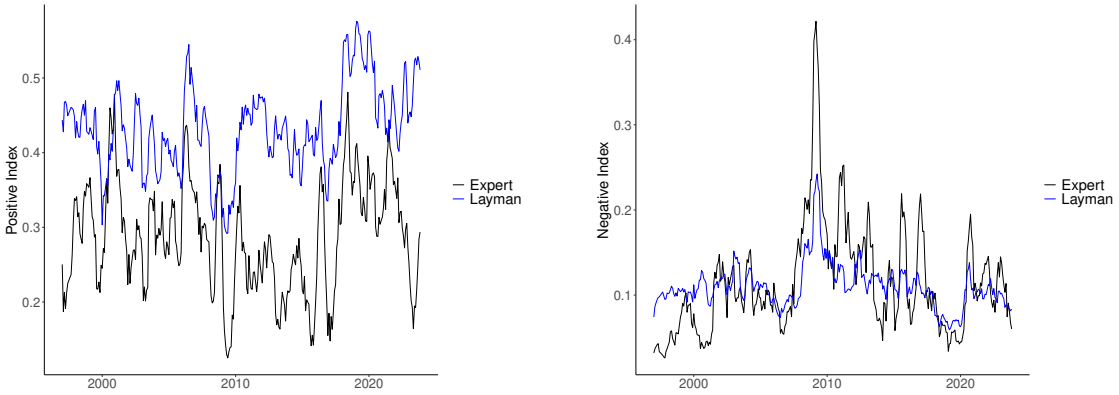
	Base Model	Training/Fine Tuned On	Paper
1	BERT	Financial PhraseBank	Araci (2019)
2	DistilRoBERTa	Financial PhraseBank	-
3	BERT	Financial PhraseBank	Hazourli (2022)
4	FinBERT	demo-org/auditor_review	-
5	BERT	10-K, 10-Q, Earnings Calls, Analyst Reports	Huang et al. (2023)
6	RoBERTa	Twitter, TweetEval	Camacho et al. (2022)
7	DistilBERT	Multilingual Sentiment Dataset	-
8	BERTweet	StockTwits-crypto	-
9	BERTweet	SemEval2017	Pérez et al. (2021)
10	RoBERTa	Twitter and Instagram Dataset	Hartmann et al. (2021)



(a) Positive sentiment index.

(b) Negative sentiment index.

Figure 3: Monthly U.S. sentiment indices from averaging the classifications of all BERT models. The monthly averaging was done for representation purposes. Note that the scales across panels (3a) and (3b) are not aligned.



(a) Positive sentiment indices.

(b) Negative sentiment indices.

Figure 4: Right aligned six-month moving averages of the U.S. sentiment indices from averaging across the expert (financial text trained) and layman (not financial text trained) BERT models separately. The six-month moving averages are presented for representation purposes. Note that the scales across panels (4a) and (4b) are not aligned.

Table 6: Fed Speeches, U.S. Stock Returns, Aggregate Models: Regression results using the indices constructed with both expert (financial text trained) and layman (not financial text trained) BERT models. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) add a control for the previous trading day's mean realized returns. Section (5.1) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.007**	0.007**	0.002	0.002	0.0003	0.0003
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Pr(neg)	0.017***	0.017***	0.007**	0.007**	0.007**	0.007**
	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.16	0.19	0.49	0.49	0.55	0.55
n	1608	1608	1608	1608	1608	1608

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7: Fed Speeches, U.S. Stock Returns, Expert Models: Regression results using the indices constructed with only expert (financial text trained) BERT models. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. Section (5.1) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.007***	0.007***	0.002*	0.002	0.001	0.001
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Pr(neg)	0.010***	0.010***	0.005**	0.005**	0.004**	0.004**
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Controls	None	FF	+ Moments	+ Δ %$VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.16	0.19	0.49	0.49	0.55	0.55
n	1608	1608	1608	1608	1608	1608

Note: *p<0.1; **p<0.05; ***p<0.01

Table 8: Fed Speeches, U.S. Stock Returns, Layman Models: Regression results using the indices constructed with only layman (not financial text trained) BERT models. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. Section (5.1) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	−0.004	−0.004	−0.005	−0.005	−0.004	−0.004
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Pr(neg)	0.008	0.007	0.0001	−0.00003	0.002	0.002
	(0.009)	(0.008)	(0.007)	(0.007)	(0.006)	(0.006)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.15	0.19	0.49	0.49	0.55	0.55
n	1608	1608	1608	1608	1608	1608

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9: Bank of Canada Speeches, Canadian Stock Returns, Aggregate Models: Regression results using the indices constructed with both expert (financial text trained) and layman (not financial text trained) BERT models, Bank of Canada official speeches, and Canadian stock returns. The regression model is as specified in equation (1). Each specification includes Bank of Canada Governor fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four U.S. Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the S&P 500 VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. The sample period here is from January 2011 to February 2024. Section (5.2) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	-0.030^{*}	-0.028^{*}	-0.011	-0.011	-0.012	-0.012
	(0.017)	(0.017)	(0.012)	(0.012)	(0.012)	(0.012)
Pr(neg)	-0.039^{*}	-0.036^{*}	-0.026	-0.026	-0.030^{*}	-0.031^{*}
	(0.021)	(0.021)	(0.016)	(0.016)	(0.016)	(0.016)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.03	0.02	0.48	0.47	0.48	0.48
n	307	307	307	307	307	307

Note: *p<0.1; **p<0.05; ***p<0.01

Table 10: Bank of Canada Speeches, Canadian Stock Returns, Expert Models: Regression results using the indices constructed with only expert (financial text trained) BERT models, Bank of Canada official speeches, and Canadian stock returns. The regression model is as specified in equation (1). Each specification includes Bank of Canada Governor fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four U.S. Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the S&P 500 VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. The sample period here is from January 2011 to February 2024. Section (5.2) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	-0.018^{*}	-0.018^{*}	-0.007	-0.007	-0.008	-0.008
	(0.010)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
Pr(neg)	-0.022^{*}	-0.021^{*}	-0.017^{*}	-0.017^{*}	-0.019^{**}	-0.019^{**}
	(0.011)	(0.011)	(0.009)	(0.009)	(0.008)	(0.008)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.03	0.03	0.48	0.48	0.49	0.48
n	307	307	307	307	307	307
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 11: Bank of Canada Speeches, Canadian Stock Returns, Layman Models: Regression results using the indices constructed with only layman (not financial text trained) BERT models, Bank of Canada official speeches, and Canadian stock returns. The regression model is as specified in equation (1). Each specification includes Bank of Canada Governor fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four U.S. Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the S&P 500 VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. The sample period here is from January 2011 to February 2024. Section (5.2) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.001	0.004	0.011	0.012	0.008	0.008
	(0.021)	(0.021)	(0.016)	(0.016)	(0.016)	(0.016)
Pr(neg)	0.002	0.009	0.014	0.016	0.007	0.007
	(0.042)	(0.042)	(0.032)	(0.034)	(0.034)	(0.034)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.01	0.01	0.47	0.47	0.48	0.48
n	307	307	307	307	307	307

Note: *p<0.1; **p<0.05; ***p<0.01

Table 12: Fed Speeches, Canadian Stock Returns, Aggregate Models: Regression results using the indices constructed with both expert (financial text trained) and layman (not financial text trained) BERT models, Fed official speeches, and Canadian stock returns. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four U.S. Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the S&P 500 VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. The sample period here is from January 2011 to November 2023. Section (5.3) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.013	0.014	0.005	0.004	0.004	0.004
	(0.010)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
Pr(neg)	0.039**	0.040**	0.021*	0.021*	0.021*	0.022*
	(0.016)	(0.016)	(0.012)	(0.012)	(0.012)	(0.012)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.02	0.02	0.49	0.49	0.49	0.49
n	668	668	668	668	668	668
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 13: Fed Speeches, Canadian Stock Returns, Expert Models: Regression results using the indices constructed with only expert (financial text trained) BERT models, Fed official speeches, and Canadian stock returns. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four U.S. Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the S&P 500 VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. The sample period here is from January 2011 to November 2023. Section (5.3) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.003	0.003	−0.003	−0.003	−0.003	−0.003
	(0.006)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)
Pr(neg)	0.020**	0.021**	0.012*	0.013*	0.013*	0.013*
	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.02	0.02	0.49	0.49	0.49	0.49
n	668	668	668	668	668	668
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 14: Fed Speeches, Canadian Stock Returns, Layman Models: Regression results using the indices constructed with only layman (not financial text trained) BERT models, Fed official speeches, and Canadian stock returns. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four U.S. Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the S&P 500 VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. The sample period here is from January 2011 to November 2023. Section (5.3) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.027*	0.026*	0.019**	0.019*	0.019*	0.019*
	(0.014)	(0.014)	(0.010)	(0.010)	(0.010)	(0.010)
Pr(neg)	0.075**	0.074**	0.037	0.038	0.038	0.038
	(0.033)	(0.033)	(0.026)	(0.026)	(0.026)	(0.026)
Controls	None	FF	+ Moments	+ $\Delta\%\ VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.02	0.02	0.49	0.49	0.49	0.49
n	668	668	668	668	668	668
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 15: Bank of Canada Speeches, U.S. Stock Returns, Aggregate Models: Regression results using the indices constructed with both expert (financial text trained) and layman (not financial text trained) BERT models, Bank of Canada official speeches, and U.S. stock returns. The regression model is as specified in equation (1). Each specification includes Bank of Canada Governor fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four U.S. Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the S&P 500 VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. The sample period here is from January 2011 to December 2023. Section (5.4) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	−0.009	−0.009	−0.003	−0.003	−0.003	−0.003
	(0.007)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)
Pr(neg)	0.005	0.002	0.008	0.008	0.007	0.008
	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.15	0.18	0.53	0.53	0.56	0.56
n	426	426	426	426	426	426

Note: *p<0.1; **p<0.05; ***p<0.01

Table 16: Bank of Canada Speeches, U.S. Stock Returns, Expert Models: Regression results using the indices constructed with only expert (financial text trained) BERT models, Bank of Canada official speeches, and U.S. stock returns. The regression model is as specified in equation (1). Each specification includes Bank of Canada Governor fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four U.S. Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the S&P 500 VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. The sample period here is from January 2011 to December 2023. Section (5.4) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	-0.007^{*}	-0.007^{*}	-0.002	-0.002	-0.002	-0.002
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Pr(neg)	0.003	0.001	0.004	0.005	0.004	0.004
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Controls	None	FF	+ Moments	+ $\Delta\%\ VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.15	0.18	0.53	0.53	0.56	0.56
n	426	426	426	426	426	426
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 17: Bank of Canada Speeches, U.S. Returns, Layman Models: Regression results using the indices constructed with only layman (not financial text trained) BERT models, Bank of Canada official speeches, and U.S. stock returns. The regression model is as specified in equation (1). Each specification includes Bank of Canada Governor fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four U.S. Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the S&P 500 VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. The sample period here is from January 2011 to December 2023. Section (5.4) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.005	0.006	0.0002	−0.001	−0.001	−0.001
	(0.009)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
Pr(neg)	0.021	0.020	0.017	0.015	0.011	0.014
	(0.022)	(0.021)	(0.014)	(0.013)	(0.013)	(0.013)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.14	0.18	0.53	0.53	0.56	0.56
n	426	426	426	426	426	426
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 18: Speech Sentiment Spillover Effect from the Fed to the Bank of Canada, as specified in equation (2). A corresponds to aggregate layman and expert sentiment indices, E corresponds to expert sentiment indices, and L corresponds to layman indices.

	Dependent variable:					
	A $Pr(pos)_t$	A $Pr(neg)_t$	E $Pr(pos)_t$	E $Pr(neg)_t$	L $Pr(pos)_t$	L $Pr(neg)_t$
	(1)	(2)	(3)	(4)	(5)	(6)
A $Pr(pos)_{t-1}$	−0.032	0.036				
	(0.051)	(0.041)				
A $Pr(neg)_{t-1}$	−0.108	0.200***				
	(0.069)	(0.063)				
E $Pr(pos)_{t-1}$			−0.048	0.004		
			(0.052)	(0.041)		
E $Pr(neg)_{t-1}$			−0.119**	0.183***		
			(0.060)	(0.060)		
L $Pr(pos)_{t-1}$					−0.033	0.007
					(0.058)	(0.028)
L $Pr(neg)_{t-1}$					−0.086	0.079
					(0.148)	(0.062)
Constant	0.397***	0.101***	0.335***	0.128***	0.463***	0.112***
	(0.024)	(0.019)	(0.022)	(0.016)	(0.039)	(0.018)
Controls	N	N	N	N	N	N
FE	N	N	N	N	N	N
R Sq.	0.005	0.032	0.008	0.029	0.001	0.005
n	426	426	426	426	426	426
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 19: Speech Sentiment Spillover Effect from the Bank of Canada to the Fed, as specified in equation (2). A corresponds to aggregate layman and expert sentiment indices, E corresponds to expert sentiment indices, and L corresponds to layman indices.

	Dependent variable:					
	A $Pr(pos)_t$	A $Pr(neg)_t$	E $Pr(pos)_t$	E $Pr(neg)_t$	L $Pr(pos)_t$	L $Pr(neg)_t$
	(1)	(2)	(3)	(4)	(5)	(6)
A $Pr(pos)_{t-1}$	0.012	0.010				
	(0.035)	(0.024)				
A $Pr(neg)_{t-1}$	−0.046	0.034				
	(0.042)	(0.030)				
E $Pr(pos)_{t-1}$			0.012	−0.001		
			(0.028)	(0.020)		
E $Pr(neg)_{t-1}$			−0.097***	0.012		
			(0.036)	(0.028)		
L $Pr(pos)_{t-1}$					0.003	0.014
					(0.042)	(0.017)
L $Pr(neg)_{t-1}$					0.049	0.059
					(0.083)	(0.038)
Constant	0.364***	0.103***	0.295***	0.111***	0.431***	0.096***
	(0.016)	(0.011)	(0.011)	(0.008)	(0.026)	(0.011)
Controls	N	N	N	N	N	N
FE	N	N	N	N	N	N
R Sq.	0.001	0.001	0.006	0	0	0.002
n	426	426	426	426	426	426
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 20: Fed Speeches excluding speeches by the Chair & Chairman, U.S. Stock Returns, Expert Models: Regression results using the indices constructed with only expert (financial text trained) BERT models. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. Further details in Sections (5.1) and (5.6).

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.007***	0.007***	0.004**	0.004**	0.003	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Pr(neg)	0.010***	0.010***	0.006**	0.005**	0.005**	0.005**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Controls	None	FF	+ Moments	+ Δ %$VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.17	0.19	0.48	0.48	0.55	0.55
n	1176	1176	1176	1176	1176	1176

Note: *p<0.1; **p<0.05; ***p<0.01

Table 21: Fed Speeches by the Chair or Chairman, U.S. Stock Returns, Expert Models: Regression results using the indices constructed with only expert (financial text trained) BERT models. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. Further details in Sections (5.1) and (5.6).

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.007*	0.006*	−0.003	−0.003	−0.003	−0.003
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Pr(neg)	0.009	0.010*	0.002	0.002	0.001	0.002
	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Controls	None	FF	+ Moments	+ Δ %$VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.13	0.18	0.5	0.5	0.55	0.55
n	432	432	432	432	432	432

Note: *p<0.1; **p<0.05; ***p<0.01

Table 22: Fed Speeches excluding speeches by the Chair & Chairman, Canadian Stock Returns, Expert Models: Regression results using the indices constructed with only expert (financial text trained) BERT models, Fed official speeches, and Canadian stock returns. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four U.S. Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the S&P 500 VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. The sample period here is from January 2011 to November 2023. Further details in Sections (5.3) and (5.6).

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.004	0.006	−0.002	−0.003	−0.003	−0.003
	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Pr(neg)	0.027***	0.027***	0.017**	0.018**	0.018**	0.018**
	(0.010)	(0.010)	(0.008)	(0.008)	(0.008)	(0.008)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.03	0.03	0.46	0.47	0.47	0.47
n	523	523	523	523	523	523
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 23: Fed Speeches by the Chair or Chairman, Canadian Stock Returns, Expert Models: Regression results using the indices constructed with only expert (financial text trained) BERT models, Fed official speeches, and Canadian stock returns. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four U.S. Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the S&P 500 VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. The sample period here is from January 2011 to November 2023. Further details in sections (5.3) and (5.6).

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	−0.001	−0.002	0.002	0.003	0.003	0.003
	(0.014)	(0.014)	(0.009)	(0.009)	(0.010)	(0.010)
Pr(neg)	−0.001	−0.003	−0.005	−0.006	−0.007	−0.007
	(0.023)	(0.023)	(0.015)	(0.016)	(0.016)	(0.016)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	-0.02	-0.04	0.58	0.58	0.57	0.57
n	145	145	145	145	145	145
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 24: Regressing the baseline context-inclusive indices on the no context mean indices as specified in equation (3).

	Dependent Variables	
	Context Pr(Pos)	Context Pr(Neg)
	(1)	(2)
Sentence by Sentence Pr(Pos)	1.549***	
	(0.020)	
Sentence by Sentence Pr(Neg)		1.029***
		(0.014)
Intercept	−0.067***	−0.052***
	(0.006)	(0.003)
Adj. R Sq.	0.774	0.767
n	1688	1688
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 25: Robustness: Fed Speeches, U.S. Stock Returns, Aggregate Models: Regression results using the indices constructed using both expert (financial text trained) and layman (not financial text trained) BERT models **applied sentence by sentence**. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. Section (5.1) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.013**	0.013***	0.004	0.003	0.002	0.002
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Pr(neg)	0.022***	0.022***	0.011***	0.011***	0.011***	0.011***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.16	0.19	0.49	0.49	0.55	0.55
n	1608	1608	1608	1608	1608	1608
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 26: Robustness: Fed Speeches, U.S. Stock Returns, Expert Models: Regression results using the indices constructed with only expert (financial text trained) BERT models **applied sentence by sentence**. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. Section (5.1) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.008**	0.008**	0.002	0.002	0.001	0.001
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)
Pr(neg)	0.013***	0.013***	0.008***	0.008***	0.008***	0.008***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Controls	None	FF	+ Moments	+ Δ %$VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.16	0.19	0.49	0.49	0.55	0.55
n	1608	1608	1608	1608	1608	1608

Note: *p<0.1; **p<0.05; ***p<0.01

Table 27: Robustness: Fed Speeches, U.S. Stock Returns, Layman Models: Regression results using the indices constructed with only layman (not financial text trained) BERT models **applied sentence by sentence**. The regression model is as specified in equation (1). Each specification includes Fed Chair fixed effects. Specification (1) includes no controls. Specification (2) adds controls for the four Fama-French Factors (hml, smb, umd, mktrf). Specification (3) adds controls for the first four moments of the realized returns on the given trading days (mean, standard deviation, skew, and kurtosis). Specification (4) adds controls for the daily percent change in the VIX. Specification (5) adds a control for the previous trading day's share of returns a standard deviation or more away from the previous trading day's mean returns. Specification (6) adds a control for the previous trading day's mean realized returns. Section (5.1) details the justification for each specification.

	O_t					
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(pos)	0.008	0.007	0.002	0.001	−0.001	−0.0004
	(0.007)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)
Pr(neg)	0.027***	0.026***	0.012	0.011	0.011	0.011
	(0.010)	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)
Controls	None	FF	+ Moments	+ $\Delta\%$ $VIXt$	+ O_{t-1}	$\bar{ret}_{t-1}$
FE	Y	Y	Y	Y	Y	Y
Adj. R Sq.	0.16	0.19	0.49	0.49	0.55	0.55
n	1608	1608	1608	1608	1608	1608
Note:						*p<0.1; **p<0.05; ***p<0.01

Algorithm 2: Pseudocode for the algorithm used to make the forecasts and trades.

Input: stock returns aligned with preceding negative speech sentiment

Output: positions to take on the next trading day

```
rolling windows = specify the rolling windows
for each i in rolling windows:
    load data in rolling window[i]
    estimate the regression model by ticker
    using the estimates, make forecasts given the next speech sentiment
drop forecast if: speech sentiment coefficient is not significant at 5%
keep forecast if: forecast > 0
by date, define: threshold = quantile(sentiment coefficient p-value, 0.25)
keep forecasts based on coefficients the p-value of which ≤ threshold
take positions in the remaining tickers on the next trading day at market open
sell at market close
```

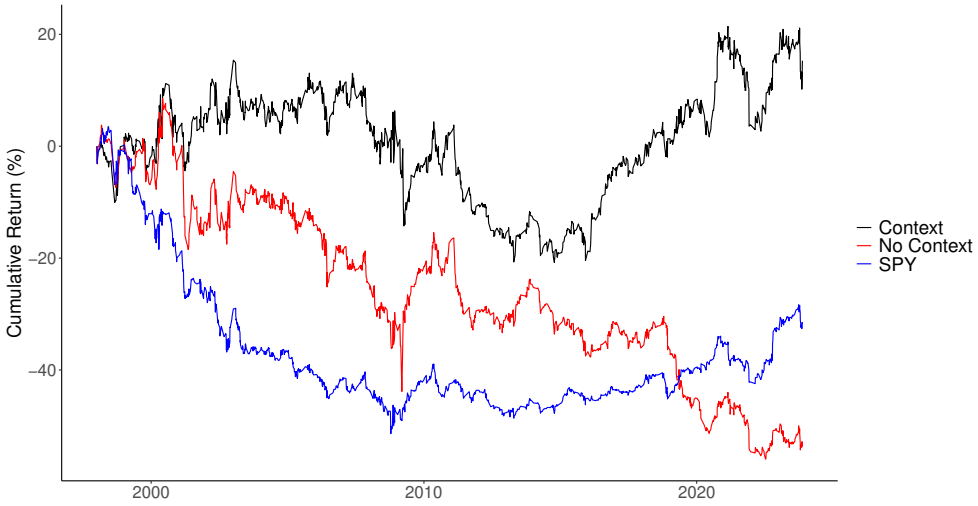


Figure 5: Cumulative returns to an equal-weighted portfolio based on the negative Fed speech sentiment forecast (Context, black) described in Sections (4.1) and (4.2), versus the S&P 500 (SPY, blue), and the sentence-by-sentence (No Context, red) classification-based negative sentiment index described in Section (6). Performance is evaluated only on trading days after the Fed speeches in the sample. In terms of cumulative returns on trading days following Fed speeches, the forecast based on context-inclusive negative Fed speech sentiment beats the S&P 500 by 47% and has a 0.61 higher Sharpe ratio over the sample period. The main strategy beats the strategy based on the sentence-by-sentence classification index by 68% and has a 0.76 higher Sharpe ratio.