Using Chat GPT to Predict Canada's Interest Rate

Arkadi Dobriakhine
300322777

2 April 2024

Major Research Paper
submitted to the Faculty of Social Sciences

University of Ottawa

Abstract

This essay presents the development and the resulting predictive power of Chat GPT based sentiment index of the Bank of Canada's (BoC) communications. The essay's objective is to detect and quantify any signals of Forward Guidance provided in the BoC policy announcements. This essay validates the resulting index by comparison with market expectations, and internal BoC staff forecasts. The findings include significant predictive power of index to forecast future policy rate changes. The essay underscores the challenges of accurately interpreting central bank communications and the potential of AI tools in enhancing this interpretation.

Keywords— Artificial Intelligence, Bank of Canada, Chat GPT, Forward Guidance

Contents

1 Introduction

Can we know the future path of monetary policy? The year 2024 has just begun and the Canadian media is once again full of conjecture on inflation and the housing mortgage rates. For the last two decades, Canadians experienced very low mortgage rates with the lowest recorded rate in 2021 at just 1.44%. However the recent interest rate increases have taken many households by surprise, with mortgage rates jumping as high as 7.5% by the end of 2023. Much of the media has been focused on housing unaffordability and inflation and the public's attention seems focused on the high interest rates. Yet, at first glance, the media and public speculation should be unnecessary—Canada provides regular updates on the future path of its policy through regular Bank of Canada (BoC) policy announcements and should, in theory, anchor any public discussions.

In fact, the BoC has for years practiced transparency in its communications. As far back as 1997, the Bank began to publish its press releases documenting any changes in its Target for Overnight Rate—the main tool it uses for influencing Canadian interest rates. The updates also include BoC statements about the present and future state of the Canadian economy and occasionally forecast the Bank's future actions directly. The Bank refers to signals present in its communications as Forward Guidance (FG) [Carney, 2012]. Yet, despite the availability of this information, there is evidence that the public continues to be unaware of the communication provided by the central bank [Coibion et al., 2019].

From an econometric perspective, one of the challenges with analyzing the communications provided by central banks is the qualitative nature of the text contained in the announcements. As we will discuss in Section 2, a number of attempts have been made to quantify and incorporate central bank communications into monetary economic models. One of several types of approaches involves using algorithms to create a "Sentiment Index". This is a numeric representation of the tone or the general sense of each announcement text. The strength of the resulting index in representing the actual sentiment contained in the text often depends on the type of algorithm selected for this task.

1.1 Use of AI: GPT 3.5, GPT 4.0

One of the recent tools or algorithms that has become available to researchers is Chat GPT as well as its Excel versions based on the GPT 3.5 and GPT 4.0 by Open AI. This is a series of software that allows easy interaction between users and a sophisticated Generative Pretrained Transformer (GPT), which is an advanced form of AI. It is part of a larger class of software called Large Language Models (LLM) that specialise in text recognition and analysis [Ali, 2023]. Anecdotal evidence points to the potential for sophisticated capabilities of this new tool including several advantages:

- Low cost of analysis: Chat GPT 3.5 is available free online and Chat GPT 3.5 and 4.0 are available to use for modest subscription costs through Microsoft Office Excel [OpenAI, 2023].
- Scalability: Because LLM are trained on hundreds of thousands of documents available online, the resulting software is able to analyze prompts in many different languages,

allowing for consistent comparison of central bank communications across countries [Open AI, 2023].

- Low programming requirements: Since Chat GPT possess great text analysis capability, users are able to interact with it using vernacular English, without use of stylised programming code[OpenAI, 2023].

1.2 Essay Objectives

This essay will attempt to follow two main objectives:

- Use Chat GPT (or similar GPT-based Open AI products*) to generate an index of BoC rate change sentiment
- Validate the Index through the use of proxy variable for market expectations and BoC internal staff forecasts.

Reaching these objectives would allow BoC communications and their impact to be evaluated directly. To achieve this, the essay will follow the following outline: we will discuss the currently available literature on central bank communications and forward guidance and discuss the definitions of forward guidance currently in use; we will formally identify the index in the context of the BoC policy rate announcements, develop our models and provide the methodology for validation of the index; we will discuss the data sources and transformations used in this paper; finally we will discuss our results, including any puzzles or challenges encountered, and in provide the conclusion including potential paths of further research.

2 Literature Review

2.1 Quantifying Forward Guidance

The use of deliberate communications, including the use of FG by central banks has been documented by monetary economists for at least two decades. There are numerous publications that analyse FG, however relatively few that attempt to find novel ways of quantifying FG directly.

[Gürkaynak et al., 2005], investigates the market's response to both the actions and public statements of the central bank of the United States the The Federal Reserve (FED). They used a high-frequency event-study approach. They focused on two factors: the "current federal funds rate target" and the "future path of policy". In their paper the former refers to the direct policy action the FED is taking while the latter is associated with central bank communications and FG. Their findings indicate a significant market reaction to central bank communication. However the presence of FG is largely inferred indirectly through the reaction of the financial markets, with no direct sentiment index being constructed.

[Campbell et al., 2012] explore the macroeconomic effects of Federal Reserve's forward guidance, specifically in the period leading up to and during the first zero lower bound episode 2008–2015. They classified FG into types and were the first to coin the terms "Delphic" (or forecasts-based) and "Odyssean" (or commitment-based) types of FG. They find that different FG types have a significant impact on markets and forecasters. However, they do not quantify FG directly. They use changes in market expectations and revisions to private forecasts as evidence of the presence and effects of FG.

[Sutherland, 2023], focuses on the impact of forward guidance on interest rate expectations. The panel data they used is very broad and includes three decades of monthly observations across two dozen western economies, including Canada. Their main goal was to measure and quantify not just the effect of FG but also disentangle different types of forward guidance. For example they discuss at length the differences between "commitment based" forward guidance "state dependent" forward guidance. The author also makes a distinction between surprise and expected FG. Many insights provided are original and unique to their paper such as categorisation of FG into 5 sub-classes and the use of innovative FG change transmission matrix to record FG. However, the data used by Sutherland for the FG index is derived from the author's personal interpretation of each central bank statement. Because the index largely relies on researcher input, the potential for researcher bias in the resulting estimates is great, so too is the time-cost of deriving the data.

[Bholat et al., 2015] provide an overview of text analysis methods currently used for studies of central bank communications. They note a relative dearth in applications compared to the use of text mining techniques in other disciplines. They list several notable papers and their methods from arithmetic word counting techniques to sophisticated neural network (supervised and unsupervised) learning algorithms. Notably, not all the techniques listed were from economic studies and none focused on the communications by the BoC. Furthermore, the techniques presented were complicated by complex programming requirements or access to extensive informational databases (as in the case of neural network algorithms). Chat GPT was not available in 2015 and so was not assessed in their overview.

The available publications reveal several gaps in academic understanding of FG. One is that few studies explicitly identify FG and many instead sidestep the issue by measuring the effects of FG. Second is that quantifying FG, in studies that attempt it, remains very labour-intensive and prone to researcher bias. Thirdly there are still relatively few publications that deal with Canada's data, while most are focused on the FED in the US. Lastly very few recent publications use LLM to analyse central bank communications and none were found to use Chat GPT. This Essay will attempt to fill the gap left by other researchers.

2.2 Definitions of Forward Guidance

Forward Guidance has become a common feature of central bank communications. However, its definition remains nuanced. Our attempt to create an index of Forward Guidance must be clear about its definition. Based on the existing literature as well as the use of the term 'guidance' by the Bank when referring to Forward Guidance, we adapt three related definitions of Forward Guidance. Note that the definitions presented below are not necessarily mutually exclusive. Often a single text can contain all three types.

2.2.1 Extraordinary Forward Guidance

This is a more deliberate form of forward guidance where the central bank commits to a specific future monetary policy path, often conditional on certain indicators and limited in time. It involves direct statements about future policy rates or actions. While this is not a form of legal obligation, any commitment-based guidance can be seen as merely a sub-set of ordinary guidance: the Bank in effect forecasts its own policy rate actions [Blinder et al., 2017]. The Bank is typically assumed to stake its reputation on following through on this type of Forward Guidance [Stein, 1989]. This definition is often referred to as "Odyssian" Forward Guidance [Campbell et al., 2012]. For the purposes of this paper, we will adopt the term used by the Bank itself-—Extraordinary Forward Guidance [Carney, 2012]. An example of Extraordinary FG can be observed in the 21 April 2009 Policy Rate Announcement [Bank of Canada, 2023a]: "Conditional on the outlook for inflation, the target overnight rate can be expected to remain at its current level until the end of the second quarter of 2010 in order to achieve the inflation target."

2.2.2 Ordinary Forward Guidance

This definition focuses on central bank communications that projects future economic states or conditions without committing to specific monetary policy action. It leaves the Bank with room to maneuver in terms of the policy rate decisions it will take in the future. This type of FG seeks to inform the public and market participants about future economic conditions. It also allows the public and the markets to infer potential future policy actions ([Bholat et al., 2015], [Svensson, 2015], [Sutherland, 2023]). This definition is often referred to as "Delphic" Forward Guidance [Campbell et al., 2012]. In order to be consistent, we will slightly transform the term used by the Bank and call this type of guidance—Ordinary Forward Guidance.[Carney, 2012]. An example of Ordinary Forward Guidance can be observed in the 25 October 2023 Policy Rate Announcement [Bank of Canada, 2023a]: "In the Bank's October projection, CPI inflation is expected to average about 3.5% through the middle of next year before gradually easing to 2% in 2025."

2.2.3 Broad Forward Guidance

A broader definition encapsulates all statements made by the Bank of Canada as part of its policy rate announcements as a type of guidance. For our purposes we call this Broad Forward Guidance. It includes general views, analyses, and intentions regarding monetary policy, without necessarily including any statements about future conditions or policy commitments by the Bank. This definition is broader than used in most other academic work. It allows to include information that may be present in communications but that does not formally fit into the previous two definitions. In 2012, the Bank labeled these statements as designed to help the market and the public understand how the Bank will respond to future developments over time in other words to "think along with us" [Carney, 2012]. This suggests a level of awareness by the Bank that at least some of the text contained in the announcements can contain information useful in terms of predicting future path of policy, even if the text does not provide explicit forward guidance (ordinary or extraordinary). In addition to its potential for including more information, the Broad Forward Guidance may be more effective

under certain conditions from a game-theory perspective. This is described in more detail by [Bassetto, 2019]. Broad Forward Guidance is the definition we will adopt for further use in the index construction.

3 Identification, Method, and Model

In this section we seek to formally identify the causal relationships between FG present in the BoC communications, market reactions, internal staff forecasts, and Governing Council decisions. We also clarify the method we are using to measure and validate these variables. Finally we establish the mathematical model that will be estimated and tested.

3.1 Identification

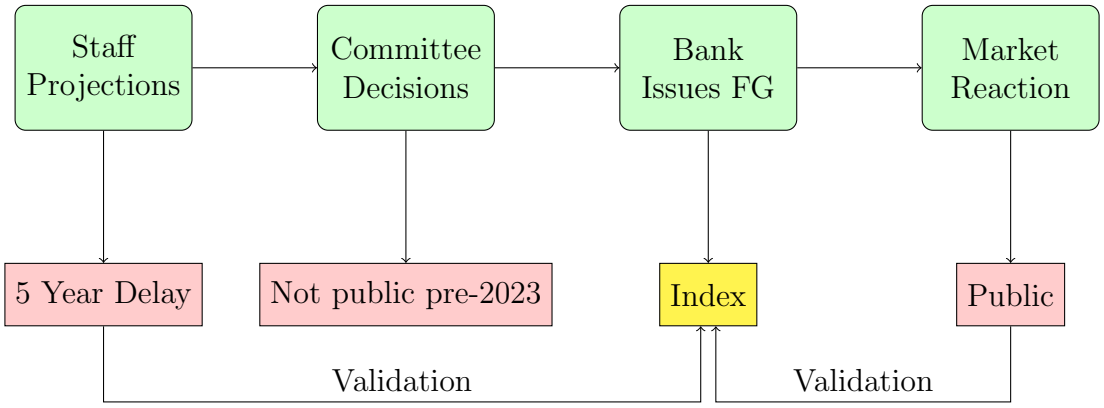

Figure 1: Forward Guidance Decision Flow Diagram

In this study, we are interested in detecting Broad Forward Guidance that may be issued by the BoC by creating a data set of index values of FG sentiment. We also want to validate the resulting index against known data. A good starting point is to visualise the decision diagram of BoC when issuing FG and the relationship of our various data sets. Figure 1 provides a visual representation. Roughly two to three weeks before each announcement, staff create or update staff projections of key economic variables, including projected future policy rate. Then, the BoC Governing Council, which includes the Bank Governor, Deputy Governor and others, makes a consensus-based decision regarding changing or keeping the current policy rate. At the same time, the Governing Council drafts and approves the accompanying policy announcement document and press release [Murray, 2013]. The next step is to then release the announcement with its policy rate decision and its accompanying comments. The wording of the announcement can reflect both the reasons behind the current decision as well as any Broad, Ordinary, or Extraordinary Forward Guidance the committee may choose to include. Finally, the markets observe the policy rate change as well as the FG and react.

Each one of these decisions is accompanied by data that can be used to represent information available at the decision node. Staff projections are released publicly with a five year delay. Committee decisions are recorded primarily in the announcements themselves, especially

prior to 2023 [Bank of Canada, 2022]. Market expectations are public in the sense that they can be deduced from price changes in financial instruments following the method used by [Gürkaynak et al., 2005] and earlier researchers. These data sets are represented by smaller red rectangles in Figure 1, with the Index as the main focus of this essay highlighted in yellow.

While different agents participate in this decision chain, it is conducted with strict timing and measures of information confidentiality[Bank of Canada, 2022]. Staff projections are not made public until five years after internal formulation. This means that the markets can not react to any changes in these projections. At the same time, the Committee is the only decision making body at the Bank that is able to make policy rate changes and compose the announcement text. Their decisions are likewise kept confidential until after the announcement [Bank of Canada, 2022]. Market Reactions is the only node that is immediately publicly available to all actors, but because of the timing of the chain of decisions market reactions occur at the end of the process. The decision chain thus implies a one-way causality link between each node. This means that it is possible to validate the index that measures Bank issued FG by comparing it separately: to the information contained in market reactions and to the information contained in staff projections. The one-way causality assumptions can also be tested mechanically by using a Granger-Causality test [Granger, 1969]. Figure 1 represents this intuition by showing "Validation" feedback loops at the bottom of the diagram.

3.2 Chat GPT

The main conclusions of this paper revolve around the index generated by Open AI's application called Chat GPT. This application comes in different versions. The two used in this paper are GPT3.5-turbo-0125 and GPT4.0-0125-preview.

Because of the number of announcement texts that make up the data set, the GPT that were used in this essay were accessed through Excel using the "=AI.ASK" set of formula functions. These are available to subscribed users using the Chat GPT for Excel add on. The use of Excel allowed for manipulation of large sets of data as well as control over some mechanical functions of the prompt such as its inherent "randomness" or "creativity".

In terms of the mechanics of interfacing with the software, it has three important components: The Text, The Prompt, and The Response. Due to the way this interaction was programmed in Excel each prompt-text-output combination is performed independently of any other output request in neighbouring Excel cells.

3.2.1 The Text

This is the text of the BoC policy announcement discussed in section 4.2. In addition to the baseline announcement text series, a second series was created with a major modification in which any references to the rate change or to the current or past rates as well as the date of the announcement were removed from the text. This second text version was modified specifically to test if the current rate changes had an impact on Chat GPT assessment of sentiment of the announcement. More on this in the Results section.

3.2.2 The Prompt

Early in the research for this paper a prompt framework was developed and refined. It consists of several parts:

- Role—This part of the prompt anchors the GPT to emulate a specific point of view or perspective.
- Content—These are additional instructions used during specific prompt testing.
- Instructions—These are general instructions provided as part of every prompt. They tend not to change prompt to prompt.
- Word Association Bins (Categories)—These are specific word or number bins that GPT was asked to match to the sentiment expressed in the announcement text. Both 11 and 7 sized bins were tested.
- Restrictions—Additional instructions restricting some methods of response by GPT. Specifically, precluding the use of older word counting algorithms and in some cases explicitly precluding the use of the knowledge of past policy rate decisions.
- Output—This is focused on the technical aspects of the output defining the format so it can be aggregated later in Excel.
- Context—This provides a general context of how the response will be used.

The prompts were designed to provide as much information as possible in order to allow Chat GPT to associate the sentiment and the FG found in the text to the word categories or bins. Breaking the prompt down into parts allowed for easier substitution and experimentation by the researcher.

Because the two models have different costs of use per word token, GPT 3.5 was used more extensively in the earlier stages of research. Prompt experimentation and comparison were done with results produced with this model. Once successful prompts were identified, GPT 4.0 was used to further test the most promising prompt candidates. During the early stages of prompt experimentation the prompts were selected based on their correlation with the policy rate changes. Some of these correlation coefficients will be discussed in the results section. The list of prompts and their text are available in Annex A for further review.

3.3 Method of Testing

One of the central challenges in Monetary Economics remains the identification of the effects of change in policy rates on economic variables of interest. This is because while central banks use change in policy rate to influence key economic variables, they do so endogenously, often in response to changes in the same variables. One of the ways to address this is to use a high frequency event study approach. This essay will use the even study method by constructing the data set around each individual announcement.

Since historic daily market data is available, High Frequency Identification will be used for validation of the FG index against changes in the market. Daily frequency allows us to keep the data window to be very narrow. This can eliminate any two-way influence from the endogenous variable. Also, the narrow window limits the influence of external factors. This can only work if the data we are studying is itself highly sensitive to changes in the policy rate. Luckily the financial markets provide such an avenue with many assets. In this Essay, we use daily changes in the Canada Government Bond yields for this purpose and assume that bond markets are at least Weak Form Efficient – that is to say, we assume that bond yields already incorporate all publicly available information. This methodology is similar to those used by [Cook and Hahn, 1989], [Kuttner, 2001], and [Gürkaynak et al., 2005]. The method is further reinforced by the observations of [Gürkaynak et al., 2005] that daily bond data are almost as good as 30 min or 1 hour window high frequency measurements in terms of the estimated effect of FG and policy rate changes on market expectations.

3.3.1 Initial Model

What we set out to test is the hypothesis that the changes in the BoC Overnight Rate[1] are explained by the FG sentiment contained in the previously released announcements. The interpretation of this relationship is that sentiment included in the overnight policy change announcements signal future changes in the overnight rate, which is the definition of Broad Forward Guidance discussed in section 2. We can thus regress:

$$\Delta OR_t = B_1 + B_2 I_{t-n} + e_t \tag{1}$$

Where ΔOR_t Stands for Overnight Rate Change at time t; I_{t-n} is the index value of previous (n periods ago) announcement; B_1, B_2 are the regression coefficients, and e_t is the error term that captures exogenous influence on the policy rate changes not captured by the sentiment index. We are making the following classical OLS assumptions about the error term:

Expectation of Zero Mean:	$E(e_t) = 0,$
Homoscedasticity:	$Var(e_t) = \sigma^2,$
No Autocorrelation:	$Cov(e_t, e_j) = 0 \quad \forall t \neq j,$
Independence:	$Cov(e_t, I_{t-1}) = 0,$
Normality:	$e_t \sim N(0, \sigma^2),$

Of particular note is the independence assumption of error term. Our main reason for being able to do so is our data construction—because we are using high frequency data it is unlikely that residuals would capture additional information that is related to the dependent variable [Gürkaynak et al., 2005]. In equation 1, the index value is the independent variable. The same assumption is repeated for other specifications using the same high frequency logic.

[1]The policy rate in Canada is also known as Overnight Rate.

3.3.2 Model of Market Expectations

Equation (1) can be estimated as is, however we can go one step further. We can use the information contained in financial market reactions to isolate a market expected change in overnight rate and a market surprise change in overnight rate change. We can then examine how the index relates to each measure of ΔOR_t. The formula we start with is the assumption that changes in the Overnight Policy rate OR_t will cause the markets to react and result in changes in the Bond Yields Y_t:

$$\Delta Y_t = \beta_1 + \beta_2 \Delta OR_t + \epsilon_t \tag{2}$$

Where ΔY_t Stands for changes in Bond Yields at time t; ΔOR_t is the change in Overnight Rate at time t; β_1, β_2 are the regression coefficients, and ϵ_t is the error term that captures exogenous influence on the yield changes not captured by the changes in the Overnight Rate. We are making the classical OLS assumptions A1-A5 about the error term, including the assumption about the independence of the error term.

Because changes to the overnight rate are announced at predetermined dates, some changes are going to be anticipated by the market while other changes will surprise the market. In our notation, the subscripts s means surprise and e means expected, thus:

$$\Delta OR_t = \Delta OR_{s,t} + \Delta OR_{e,t} \tag{3}$$

We can rewrite (2) to incorporate information contained in (3):

$$\Delta Y_t = \beta_1 + \beta_2(\Delta OR_{s,t} + \Delta OR_{e,t}) + \epsilon_t \tag{4}$$

Since we are using High Frequency Identification with a narrow data window and assume markets to be at least Weak Form Efficient [Fama, 1970], we conclude that any expected changes to the overnight rate must already be incorporated into the bond yields before the bond yield changes are recorded in our data. The narrow data window allows us to also assume that the error term is unlikely to be contaminated by exogenous influence. In other words, we can take the expectation on both sides of equation (4):

$$\mathbb{E}[\Delta Y_t] = \mathbb{E}[\beta_1 + \beta_2(\Delta OR_{s,t} + \Delta OR_{e,t}) + \epsilon_t] \tag{5}$$

Because $\mathbb{E}[\epsilon_t]$ and $\mathbb{E}[\Delta OR_{e,t}]$ are expected to be zero under our assumptions, we simplify to get the following:

$$\Delta Y_t = \beta_1 + \beta_2 \Delta OR_{s,t} \tag{6}$$

Rewriting (6), we get:

$$\Delta OR_{s,t} = -\frac{\beta_1}{\beta_2} + \frac{1}{\beta_2}\Delta Y_t \tag{7}$$

Ideally, we would like to measure $\Delta OR_{s,t}$. In other studies, this is done using futures data [Gürkaynak et al., 2005] or rate expectations surveys [Sutherland, 2023]. However, the best data available for this study is for ΔOR_t. That means we will estimate:

$$\Delta OR_t = \gamma_1 + \gamma_2 \Delta Y_t + \eta_t \tag{8}$$

Where $\Delta OR_t = \Delta OR_{s,t} + \Delta OR_{e,t}$, the γ_1, γ_2 are estimated coefficients, and η_t is the residual with standard OLS assumptions A1-A5, including the assumption about the independence of the error term.

Subtracting equation (7) from (8), yields

$$\Delta OR_{e,t} = \eta_t \tag{9}$$

In other words, when estimating equation (8), under our assumptions of OLS and Low Market Efficiency Hypothesis, fitted values are the effects of unexpected changes to the Overnight Rate, these are denoted $\Delta \hat{O}R_t$ and residuals capture effects of expected changes to Overnight Rate, these are denoted $\hat{\eta}_t$.

3.3.3 Using Market Expectations to Validate the Index

Since the above analysis allows us to deconstruct overnight rate changes into market-expected and market-unexpected components, we can use these to further validate an index we use to measure FG. Specifically, we expect under the Low Market Efficiency Hypothesis, markets will have access to overnight change announcements and will be able to react to the sentiment contained in the announcements by pricing these into bond yields. Under this assumption, if our index correctly captures the underlying sentiment, we should expect the lagged index of FG to have a statistically significant relationship with market-expected overnight rate changes $\Delta OR_{e,t}$. Similarly, since the markets do not price in the information they do not have, any market-unexpected changes in the overnight rate $\Delta OR_{s,t}$ should not be priced in and should not have a statistically significant relationship with the FG index. This means that we can rewrite the initial equation (1) to estimate both parts of ΔOR_t separately as $\Delta OR_{e,t}$ and $\Delta OR_{s,t}$. These correspond to the values obtained when estimating equations 8 $\hat{\eta}_t$ and $\Delta \hat{O}R_t$ respectively. The results will provide us with additional index validation, which will be further discussed in the results section.

$$\hat{\eta}_t = \Delta OR_{e,t} = B_{1,e} + B_{2,e} I_{t-n} + e_{e,t} \tag{10}$$

$$\Delta \hat{O}R_t = \Delta OR_{s,t} = B_{1,s} + B_{2,s} I_{t-n} + e_{s,t} \tag{11}$$

WhereB_1, B_2 are the regression coefficients and e_t is the error term from Equation 1 and $B_1 = B_{1,e} + B_{1,s}$, $B_2 = B_{2,e} + B_{2,s}$, $e_t = e_{e,t} + e_{e,t}$

When estimating this equation our hypothesis is that we will find significant relationship between the coefficient $B_{2,e}$ and $\Delta OR_{e,t}$ and no such relationship between $B_{2,s}$ and $\Delta OR_{s,t}$. Additionally when equations 10 and 11 will be used with a Vector Auto-regression (VAR) model we expect to find one way Granger-causality between I_t and $\Delta OR_{e,t}$ such that the former Granger-causes the latter. We expect to find no such relationship in the case of $\Delta OR_{s,t}$ and I_t.

3.3.4 Model of Staff Forecasts

In 2018, BoC released its internal staff forecasts together with a working paper analysing their predictive power [Champagne et al., 2018]. The staff projections are organized by quarter. They include key economic indicators, such as GDP, Inflation, and importantly for us—the overnight rate projections. These are organized by observation date and include projections eight to twelve quarters ahead for each observed quarter. These staff forecasts, together with staff comments and other unobserved information, are presented to the Governing Council prior to any decision on rate changes or policy announcement text ([Murray, 2013]). In section 3.1, we identified that through a causal chain, staff forecasts influence the Governing Council decisions and, consequently, the sentiment in the announcement text. The assumption we are making is that the Governing Council bases their decisions, at least in part, on the staff projections they receive. This assumption can be tested by estimating the effect of staff projections on policy changes:

$$OR_{t-1} = \alpha_1 + \alpha_2 SPOR_t + \nu_t \tag{12}$$

Where OR_t Stands for Overnight Rate at time t or the rate decided at the current Governing Council decision meeting; $SPOR_t$ is the Staff Projected Overnight Rate presented to the council; α_1, α_2 are estimated coefficients; and ν_t is the error term that captures exogenous influence on the policy rate changes not captured by the staff projections with the usual A1-A5 OLS assumptions, including the assumption about the independence of the error term.

3.3.5 Using Staff Forecasts to Validate the Index

If the staff projections influence the decisions this also means they must influence the sentiment and the FG in the announcement text. Since we defined any FG as largely affecting the future rate decisions, we want to find any signals the staff projections imply about future policy rate changes. We formulate this internal signal arithmetically as follows:

$$\Delta SFOR_t = SFOR_t - OR_t \tag{13}$$

Where $\Delta SFOR_t$ stands for staff forecasted change in Overnight Rate at time t; $SFOR_t$ is the staff forecast made at time t for the Overnight Rate in the next quarter; and OR_t is the current overnight rate decided by the Governing Council. If the staff forecasted rate is higher than the current overnight rate, staff are implying a rate increase in their predictions.

Any signals foreshadowing future rate changes in the announcement text is FG according to our Broad FG definition. If the Governing Council include FG in the announcement text, they do so with the knowledge of the rate change they just decided on [Murray, 2013]. The FG in the announcement text is thus the difference between what the Governing Council believe the rates will be in the future and what they are now:

$$\Delta GCOR_t = GCOR_t - OR_t \tag{14}$$

Where $\Delta GCOR_t$ stands for Governing Council forecasted change in Overnight Rate; and $GCOR_{t+n}$ is the Governing Council projected Overnight Rate at time t. However since we do not have data on the details of deliberations of the committee, the implied staff forecasted rate change $\Delta SFOR_{t+1}$ is the best proxy available for Governing Council's beliefs about the future path of policy rate. The assumption we are left to make is:

$$\Delta GCOR_t = \Delta SFOR_t \tag{15}$$

This assumption allows us to estimate the relationship between $\Delta SFOR_t$ and the sentiment or FG index I_t.This relationship will provide a second validation of the FG Index and will further be discussed in the results section:

$$I_t = \theta_1 + \theta_2 \Delta SFOR_t + \upsilon_t \tag{16}$$

Where θ_1 and θ_2 are coefficients and υ_t is the error terms with the usual A1-A5 OLS assumptions, including the assumption about the independence of the error term.

When estimating equation 16 our hypothesis is that we will find significant relationship between coefficient θ_2 and I_t. Additionally when equation 16 will be used with a Vector Autoregression (VAR) model we expect to find one way Granger-causality between $SFOR_t$ and ΔI_t such that the former Granger-causes the latter.

4 Data Source and Description

The data used in this essay is compiled from two primary sources. First, we use the BoC for the data in sub-sections 4.2, 4.3, 4.4, and 4.5. Second, we use the data generated by Open AI's Chat GPT for Excel using both GPT 3.5 and GPT 4.0 versions in sub-section 3.2. Annex A contains some highlights of the prompts used in this essay. All collected data and calculations are available from the author upon request.

4.1 The Chat GPT Output

As discussed in Section 3.2 The GPT has been asked to provide output to prompts in terms of the word category or bins. Each of these categories was per-assigned a numeric value. The response coefficients were then extracted using Excel formulas to create a list of discreet numeric responses. For each prompt-announcement combination the response was requested and received at least 5 times. In all cases an average response was calculated and retained for use to run any regressions. The variance of individual responses across the time for all prompts was between 1.55 at the low end and 6.67 at the high end with the average variance of all individual response vectors is 3.9. Variance was also calculated for each individual prompt-announcement combination. Within these, the variance was typically much smaller with only an occasional pair producing inconsistent results: the average variance was only 0.21 with many pairs having zero variance due to producing the same answer. The max variance within a single prompt-announcement pair was 12.8. In other words, a small number of prompt-announcement combinations produced widely divergent responses. Due to the low occurrence the influence of these observations was not judged to be very great. These responses were retained without modification. For all prompt-announcement combination an average response was used.

While GPT was requested to identify only the bin, in many cases it also provided a word explanation as to why that bin was selected. It is important to understand that such explanations are simply a probabilistic response to the prompt and do not constitute a true analysis of the response's significance [University of Ottawa, 2023]. Any accompanying word explanations were discarded and only the numeric response was kept.

In several cases the output returned errors. The majority of these were due to bandwidth limits Open AI imposes on users of Chat GPT. While a Visual Basic algorithm was used to throttle response requests in order to avoid reaching the bandwidth limit, in instances where this was not sufficient the requests that returned a bandwidth error were simply re-run again.

In some cases the output did not include a category or a bin selection, but did include some other text or number not related to the category or bin. These were recorded as "." in the data and were not included in to the subsequent calculations of average response calculations. The occurrences of this type of error were not numerous and since each prompt-text combination was tested at least five times, this allowed for no loss of observations.

4.2 Announcement Texts

The announcement texts are available on the Bank's website to as far back as 26 June 1997 [Bank of Canada, 2023a]. Although the most recent announcement was made on 6 March 2024, for this essay the upper limit of the announcements studied is 25 Oct 2023. The announcements have been provided at semi-regular intervals, and after 2001 there were approximately two per quarter. The announcements are typically made in the morning. BoC follows strict protocols in order to prevent leaks of any changes to the rate prematurely [Bank of Canada, 2024].

Over the course of the announcement data range, there have been some out of sequence

policy rate change announcements. These occurred in several instances. In March 2020 there were two more announcements that were focused on COVID-19 response in Canada. Another instance was in 2001 an extra announcement was made discussing the Canada's monetary response to the shock of 11 Sep 21 terrorist acts in US. Because these announcements contained unique text, they had the possibility of containing FG. These extra announcements have been incorporated into the data set without modification. Similarly, before 2001, the total number of announcement observations in the data set is 200. Before the 2001, there are fewer announcements made per year. These were included without modification.

Not all announcements provide a rate change, but all carry some text that has the potential to provide forward guidance. The text was copied manually from the Bank's website into Excel. For ease of use by "Chat GPT for Excel" software, the text was concatenated into a single cell. Some front text was removed (for example "share on Facebook" button text). Likewise, the date of the announcement was copied and stored in an adjacent cells. Initially the text of the announcement included the subject line, however, this was later removed for two of the prompts. This modification and its implications will be further discussed in the results section.

4.3 Overnight Rates

In Canada, the policy rate is more formally referred to as the Overnight Rate. These were copied from the BoC website together with the announcement texts. The values of the Overnight Rate were matched with each announcement date. Changes in the overnight rate were calculated by subtracting the current rate from the previously announced overnight rate. The number of observations is 200, and the dates covered are the same as the announcements 26 June 1997 to 25 October 2023.

4.4 Bond Yields

The 2-year Bond Yields were chosen as the variable to measure market reaction to announcements Y_t. These were chosen primarily due to availability of daily data on the BoC website [Bank of Canada, 2023b]. Since bonds with different maturity lengths are available to choose from, a series of preliminary regressions were run to find the bond maturity length most sensitive to changes in the OR_t. This turned out to be the 2-year bond yield. The use of bond yields daily data is broadly in line with existing literature, including the methods employed by [Gürkaynak et al., 2005] and [Kuttner, 2001] for the US. The dataset begins on 2 January 2001 and runs past the last included announcement 25 October 2023. For each daily observation the bond yield recorded is at the close of the day. To measure the change in bond yields the bond yield recorded on the day before the announcements date was subtracted from the bond yield on the close of the announcement date. This provides 179 observations and each one is matched to a specific announcement event. A fair criticism is that by relying on daily data we capture not just the effect of the announcement, but also other exogenous factors that may have influences the market on the same day. If higher frequency data that allowed for construction of a tighter change window around the announcements was used, this would have avoided the issue of data contamination. In earlier studies where both daily and hourly data were available, the differences in estimated results were considered small, given a large

enough data set [Gürkaynak et al., 2005]. The size of our data set, as well as the results obtained in other studies, allows us to assume that on the whole, the daily yield changes are capturing the effects of announcements alone.

4.5 Staff Forecasts

Staff Projections were first made publicly available by the BoC in November 2018 as part of a working paper [Champagne et al., 2018]. In that dataset, projections are provided quarterly, with the first observations recorded in Quarter 2, 1982. The last date currently available is Quarter 4, 2018. Each observation date includes projections made 8–12 quarters ahead.

Since projections are made eight quarters forward and our Broad FG definition does not include explicit time period, we simplify the analysis by creating an average staff projection for quarters 1–4 and 5–8. Although some observation dates include staff projections beyond eighth quarters, these are not taken into account in order to keep the data set consistent. We thus have three measures of staff forecast: one quarter ahead, four quarters ahead average, and five to eight quarters ahead average. Since each forecast is made on a specific date, in our dataset these are matched to an announcement observation made at time t. Therefor a one quarter ahead staff forecast $\Delta SFOR_t$ is denoted as being made at time t even if it refers to one or several quarters ahead.

5 Results

5.1 Visualising the data

Prior to presenting the results of regression analysis, it is best to describe the data in a visual format. What follows are graphical representations of the relationship between versions of index (different prompt-text combinations) and the changes in the overnight rates that was described by Equation 1 above.

The first index series presented in Figure 2 below focus on experimenting with asking GPT to assess sentiment and FG present in announcement text, focusing specifically on signals that may be short term in nature such as next quarter, next announcement, or over-the-next-quarter horizons. The intuition tested here is that some FG may be provided within a short term context. If so we would like to find it.

Figure 2 shows a graph of the average value returned by three different prompts. The y-axis shows the average value for a single observation and the x-axis shows the observation date. Each observation corresponds to a specific announcement. The indexes are marked in green, blue, and purple. The black line represents the changes in overnight rate and was plotted using a secondary y-axis to represent the rate changes in percent. This allows for a more direct visual comparison between two sets of data. What becomes apparent is how closely the sentiment index follows the actual path of the policy rate changes. There are several examples where a steep rise in interest rates is occurring simultaneously with hawkish FG sentiment.

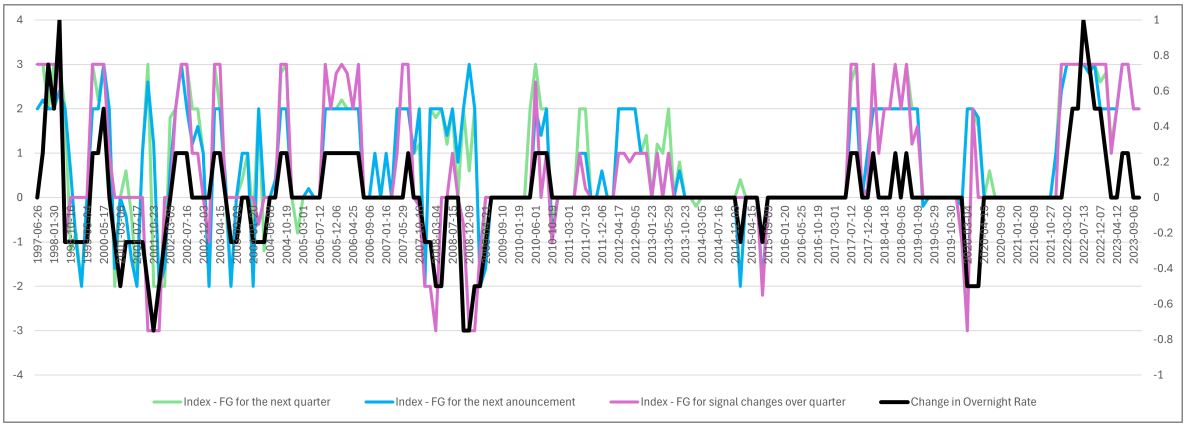

Figure 2: Comparison of several prompt results with changes in the Overnight Rate

This is especially true in the second half of the data range: examples include April 2010 to October 2010, May 2017 to December 2018, January 2022 to October 2023. There are also long periods of no change in overnight rate that are accompanied by a neutral sentiment index reading such as from September 2015 to May 2017. We note that the early half of the data range is more noisy compared to the second half. In this version of the prompts, the category bins included 7 different categories and map to discreet values in the individual responses of zero as well as between +3 and -3. Correlation coefficients between these index and change in overnight rate are likewise high. The Table 1 provides the correlation coefficients of index and overnight rate change from Figure 2. Note that the top and bottom prompts are designed to measure similar time horizons but have very different correlation coefficients. This suggests that the index results are to an extent influenced by unintuitive changes in the wording contained in the prompt. In this case the latter prompt asks GPT to search for signals as opposed to focus on likelihood of future rate changes. This approach tended to provide results with higher correlation coefficient in other prompt versions as well. More research will have to be conducted to determine the optimal prompt wording that should be used. The text of the prompts used in this essay can be reviewed in Annex A.

Table 1: Correlations with Rate Changes in Figure 2

Category	**Correlation w Rate Changes**
Index—FG for the next quarter	0.528
Index—FG for the next announcement	0.417
Index—FG for signal changes over quarter	0.756

The next index series presented in Figure 3 focus on experimenting with the role aspect of the prompt framework. Specifically, three prompts are devised to observe if changing the role from the baseline role of "an economist" to more specific roles of a "seasoned financial analyst" or a "mortgage consumer" would change the output results.

The graph is constructed in a similar manner to Figure 2 above. The left hand y-axis represents the sentiment values and the right hand y-axis represents the overnight rate changes in percent. The x-axis represents the date of each observations, which is the announcement date. The sentiment generally follows the overnight rate changes. There is more noise in the first half of the data set and less noise in the second. The role component does not seem to make a lot of difference, this could be in part because all three roles describe a financial agent type and

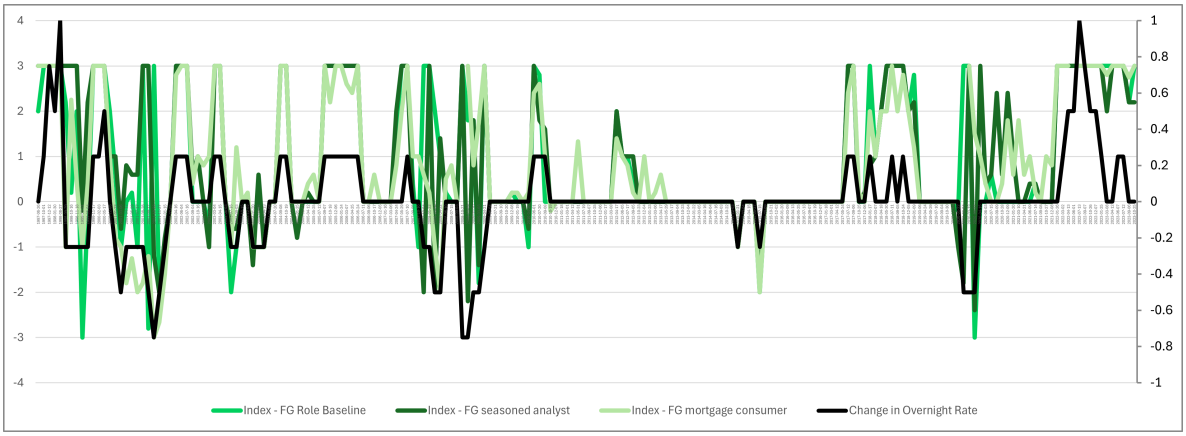

Figure 3: Comparison of several prompt results with changes in Overnight Rate

therefore likely provide similar probabilistic associations for Chat GPT. Table 2 reports the correlations with the overnight rate changes:

Table 2: Correlations with Rate Changes in Figure 3

Category	**Correlation w Rate Changes**
Index—FG Role Baseline	0.450
Index—FG Seasoned Analyst	0.508
Index—FG Mortgage Consumer	0.617

5.1.1 Collinearity Puzzle

In all cases considered above, the index appears to follow the rate changes very closely, without an apparent lead and with relatively little noise. While this does not violate the explicit assumptions made in Equation 1, this is nonetheless something we did not expect. Earlier attempts by other authors likewise point to predictive power with significant lead time [Lucca and Trebbi, 2011]. There are two ways to interpret this divergence. First, this could be an indication that the Bank of Canada, unlike it's US counterpart, does not typically provide a lot of FG in the sense of signalling their future decisions. Instead the Bank may be more focused on providing explanations to their current rate decisions. This could be especially true in the second half of the data set where we observe even less noise and less divergence between the index and the change in overnight rates. This would meant that our results are challenging one of the fundamental assumptions we made about the use of FG by BoC in sections 2 and 3—namely that it regularly provides FG. Another, more mechanical, hypothesis could be made: Chat GPT could be picking up the actual rate change referred to in the announcement text and detecting this as sentiment. If true, it is possible that the influence of these quantitative signals is so great that they dominate any qualitative signals the prompt is designed to detect. This is plausible because both the wording of the text and the LLM training Chat GPT received could be causing Chat GPT to associate announced numerical increases in interest rates with hawkish signals and the inverse for dovish signals.

We graph the index results from a new prompt-text combination to test this hypothesis. Here extensive modifications were done to the text in order to remove any mention of dates or

rate changes from the text. With this it is hoped we can avoid Chat GPT interpreting the quantitative data present in the announcements as part of the assessment of the sentiment.

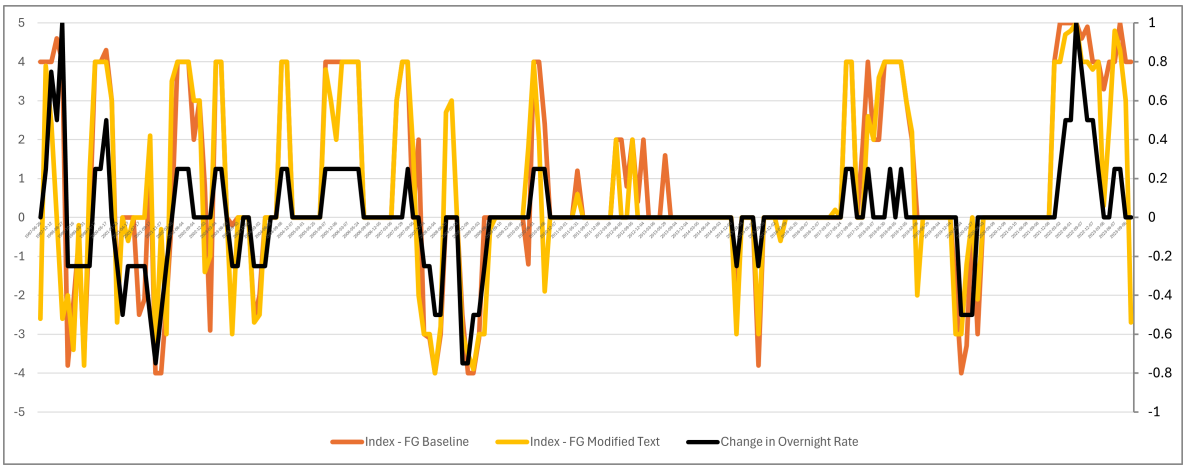

Figure 4: Comparison of modification to text of the prompt with changes in Overnight Rate

Figure 4 compares two prompt types: one a baseline prompt similar in composition to prompts shown in Figure 2 and Figure 3; and the same prompt applied to a modified version of the announcement text. We can visually see little difference in the graphs. The correlations table likewise points to similarities.

Table 3: Correlations with Rate Changes in Figure 4

Category	**Correlation w Rate Changes**
Index—FG Baseline	0.777
Index—FG Modified Text	0.648

This leaves one more possibility. Because Chat GPT was trained on millions of documents that were available online at the time of its training [Open AI, 2023], it is very likely that it was exposed to BoC announcement texts. It is possible that even with the rates removed from the announcement text used in our data, the algorithm matches the provided text to the unmodified version contained in its "memory". Without additional analytic tools released by Open AI it would not be possible to ensure that the actual rate change that occurred on a specific date does not influence the subsequent sentiment analysis performed by Chat GPT on a version of the same document. However, there is one-way we can re-examine this proposition. The results discussed so far were obtained by Chat GPT model 3.5. This model's training comprises documents available to Open AI up to September 2021. This means that any announcements that were made after this date are not part of Chat GPT "memory". We can examine the data range of September 2021 to 25 October 2023 to see an example of real world BoC announcements being indexed without Chat GPT having prior knowledge of these announcements. We can see in Figure 4 that both the index containing the text with overnight rate changes and the index with the modified text behave similarly after September 2021 and both follow the Overnight Rate Changes closely. Figure 5 provides a more "zoomed-in" version of the last 18 announcements of our data set: 8 September 2021 to 25 October 2023, with table 4 providing the correlations coefficients for this short data range.

We can see that both indexes appear to follow the path of policy rate change. This provides

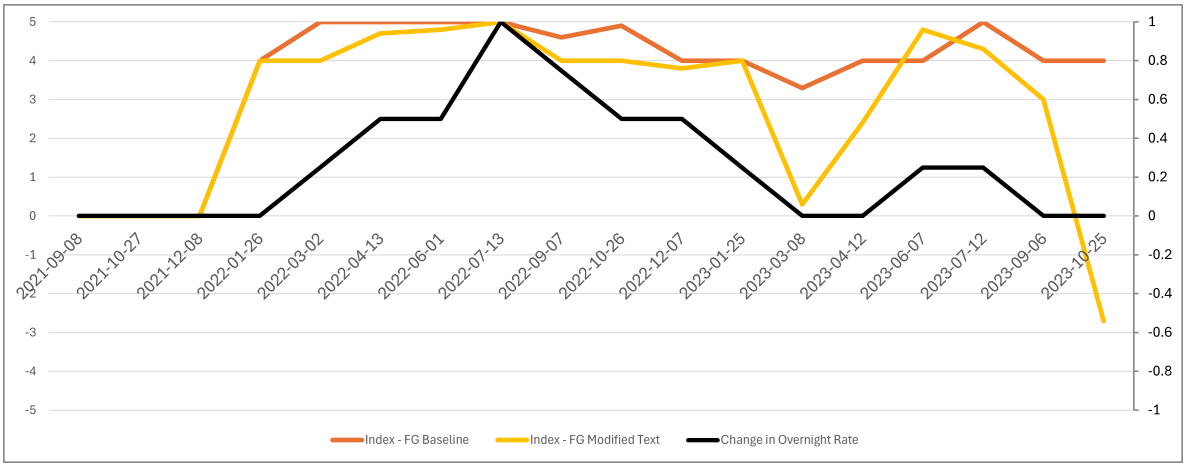

Figure 5: Comparison of modifications to text of the prompt with changes in the Overnight Rate

Table 4: Correlations with Rate Changes in Figure 5

Category	**Correlation w Rate Changes**
Index—short FG Baseline	0.557
Index—short FG Modified Text	0.654

validity to the hypothesis that the indexes built with Chat GPT do in fact capture the sentiment in the text and do not simply measure the historic policy rate changes.

One final prompt variation is to use GPT 4.0 as the underlying chat GPT model. This version completed its LLM training in December 2023 [Open AI, 2023], so we should assume its training included all announcements texts in our dataset.

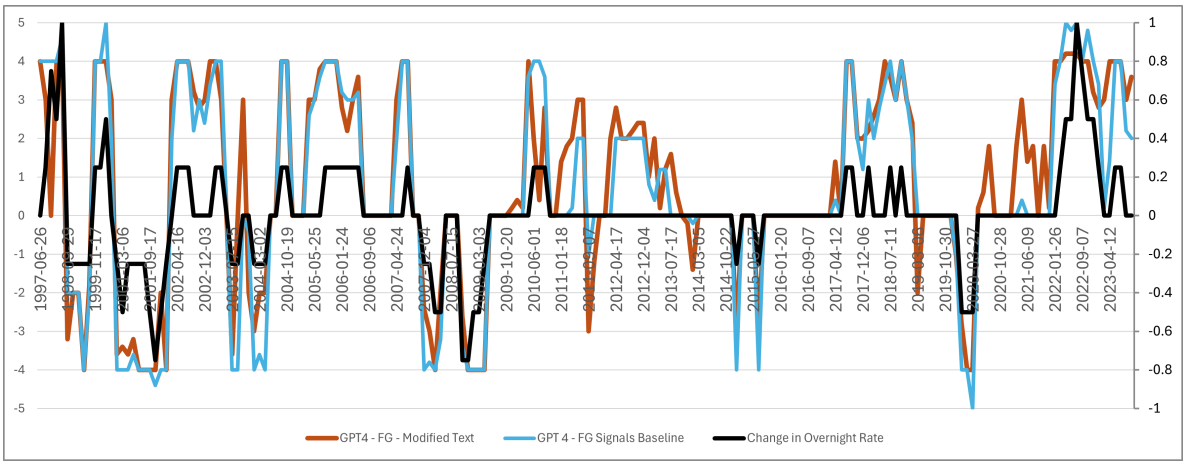

Figure 6: Comparison of Prompts using GPT 4.0

In Figure 6 and Table 5 we compare prompts that use both versions of the announcement text. Both visually and in terms of correlations, the prompts that use GPT 4 seem to follow the same overall pattern we saw earlier: more noise earlier in the data, overall closely following the changes in the overnight rate. 3.3.1. Since the prompt "GPT 4 FG Signals Baseline"

Table 5: Correlations with Rate Changes in Figure 6

Category	**Correlation w Rate Changes**
Index—GPT 4 FG Modified Text	0.745
Index—GPT 4 FG Signals Baseline	0.841

seems to have the highest correlation with the changes in overnight rate, it will be used for the validation regressions below.

5.2 Validation against Market Reactions

We know from earlier works that markets react to information provided by the central banks and this reaction can be observed using Bond Yields. In section 3.3.2, we have established equations (10) and (11) as candidates to validate the index. We are looking for a strong positive relationship between lagged values of the index and expected changes in the overnight rate. We are also looking for an absence of this relationship in the case of market surprises. Table 6 below summarises the results of the two regressions:

Table 6: Regression Results

VARIABLES	(1) Market Expectations	(2) Market Surprise
L1.Market Expectations	0.533***	
	(0.0820)	
L1.GPT 4 FG Signals Baseline	0.0360***	0.00394
	(0.00809)	(0.00243)
L2.GPT 4 FG Signals Baseline	-0.00322	-0.00641**
	(0.00893)	(0.00322)
L3.GPT 4 FG Signals Baseline	-0.0195**	0.00283
	(0.00880)	(0.00323)
L4.GPT 4 FG Signals Baseline	0.0133**	-0.00181
	(0.00663)	(0.00244)
Constant	-0.00796	-0.00778**
	(0.0108)	(0.00382)
Observations	178	179
R^2	0.635	0.031

Notes: Standard errors in parentheses. The notation "Lj" denotes the lag j.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In specification (1) in Table 6, market expectations of the changes in overnight rates were regressed on the lagged index values. Lags up to 4 were included. Due to the auto-correlation in Market Expectations, its lagged value was also included on the right hand side of the equations.

In specification (2), market surprise overnight changes were regressed on the lagged index values. Up to 4 lags were included. Since no auto-correlation was present, Market Surprise

specification did not include its lagged value.

The coefficients of interest are those of the Index estimates and are presented in Table 6. In the first specification we find the coefficients of the lagged values of the index to be significantly different from zero for three of the four estimates. However, the significance is reduced to just one of four in the second specification. Since there is considerable multicollinearity between the lagged index values in both specifications, the best tool to tell us if the lagged values have predictive power is to conduct a joint significance F-test. Testing the coefficient jointly yields an F value of 6.73. In technical terms, this causes us to reject the null hypothesis that the coefficients are jointly equal to zero. This provides evidence that they are jointly significant in the first specification. However for the second specification, this test returns a F value of only 1.38. This means that we fail to reject the null hypothesis that the coefficients are jointly equal to zero.

To test this concept further we run two Granger-causality tests: first between Market Expected changes and the index and second between Market Surprise changes and the index. The technical results are provided in tables 7 and 8. The number of lags were selected mechanically, based on the result of the AIC test with two lags used in both cases. The interpretation is that we find one-way causality between the index and the market expectations. Specifically, changes in the index Granger-cause market expected changes in overnight rate. However, no Granger-causality link is found for the market surprise overnight rate changes.

Table 7: Granger-causality Wald tests

Equation	Excluded	χ^2	Prob > χ^2
Market Expectations	GPT 4 FG Signals Baseline	18.781	0.000
GPT 4 FG Signals Baseline	Market Expectations	0.08621	0.958

Table 8: Granger-causality Wald tests

Equation	Excluded	χ^2	Prob > χ^2
Market Surprise	GPT 4 FG Signals Baseline	5.8452	0.054
GPT 4 FG Signals Baseline	Market Surprise	0.81677	0.665

Both the results of the coefficient estimation as well as Granger-causality test seem to fit the predictions made in Section 3.3 of this essay—we find that the index is able to anticipate market predicted movements in the overnight rate. This result is broadly in line with the assumptions we made earlier—markets observe BoC announcement texts and react to them. Sentiment present in the text is picked up by the markets and we detect this as a strong statistical and Granger-causal relationship between the index and the market expectations. At the same time, very little statistical relationship exists between the sentiment index and market surprise changes in overnight rate. One reason for this may be that both the index and the market are picking up the same FG signals present in announcements. This would mean that any changes that are a surprise to the market are also undetected by the index. This would certainly explain our results. We now turn towards the staff forecasts for additional insights.

5.3 Validation against Staff Forecasts

In section 3.3.5 we established that we expect a relationship to exist between staff forecasts and the index and that it should take the form of equation (16). Since we have three possible measures of staff forecasts $\Delta SFOR_t$, we include all three as separate specifications with one modification—we include one lag of the index value on the right-hand side in order to deal with autocorrelation in the index variable I_t.

Table 9: Comparison of results Staff Forecasts

	Index—GPT 4 FG Signals Baseline		
VARIABLES	(1)	(2)	(3)
L1.Index—GPT 4 FG Signals Baseline	0.690***	0.704***	0.750***
	(0.0493)	(0.0498)	(0.0516)
ΔSFOR Q1	1.507***		
	(0.264)		
ΔSFOR Q1-4		0.936***	
		(0.182)	
ΔSFOR Q5-8			0.387***
			(0.125)
Constant	-0.213	-0.253*	-0.311
	(0.139)	(0.147)	(0.192)
Observations	158	158	158
R^2	0.636	0.623	0.585

The results presented in Table 9 note the high significance of the staff forecast for determining the index value. This is true for forecasts made for the immediate next quarter as well as the average of forecasts Quarter 1–Quarter 4 and Quarter 5–Quarter 8. Note, however, that the coefficient estimates are quite different. In specification (1), forecasts made for the next quarter, the estimate is 1.5. This means that for every 1 percent point increase in staff forecasted interest rate, the average GPT index value is likely to increase by 1.5 points (on a 11 point scale between -5 and +5). This effect drops to 0.9 and just 0.3 points for a longer-term forecast change in specifications (2) and (3).

Note also the relatively large R^2 value in these specifications. This is due to the presence of lagged y value on the right hand side. Regressed individually the staff forecast coefficients provide an R^2 that is much lower as detailed in Table 10.

Table 10: Comparison of R^2 without autoregressive coefficient

	Index—GPT 4 FG Signals Baseline		
VARIABLES	(1)	(2)	(3)
R^2 with L1.Index...	0.636	0.623	0.585
R^2 without L1.Index...	0.170	0.134	0.017

We can see that the staff forecasts, as presented to the Governing Council, have some explanatory power when it comes to predicting the sentiment (as captured by the index) that

will be used in the draft announcement. However the staff forecasts do not account for all the variation in the index data. This likely means that the Governing Council does not rely exclusively on information presented by the staff when making their rate decisions, however without additional insights into the inner decision making process or to know the exact data presented to the council, we are unable to provide further insights.

We now turn to examining the one way causality assumption we made earlier. In our model the staff forecasts influence the Governing Council who then compose the sentiment in the the rate announcement. We are therefor able to conduct a series of Granger-causality tests to test this assumption.

Table 11: Granger-causality Wald tests—next Q1

Equation	Excluded	chi2	df	Prob > chi2
Index—GPT 4 FG ...	ΔSFOR Q1	58.56	4	0.000
ΔSFOR Q1	Index—GPT 4 FG ...	4.3237	4	0.364

Table 12: Granger-causality Wald tests Q1–Q4

Equation	Excluded	chi2	df	Prob > chi2
Index—GPT 4 FG ...	ΔSFOR Q1	53.619	4	0.000
ΔSFOR Q1	Index—GPT 4 FG ...	7.0443	4	0.134

Table 13: Granger-causality Wald tests Q5–Q8

Equation	Excluded	chi2	df	Prob > chi2
Index—GPT 4 FG ...	ΔSFOR Q1	15.359	4	0.004
ΔSFOR Q1	Index—GPT 4 FG ...	15.792	4	0.003

After running Granger-causality tests for all three specifications we present the results in tables 11, 12, 13. For both one quarter forecast and Quarter 1–Quarter 4 average forecasts we find that the staff forecasts one-way Granger-cause changes in the index. For the last specification for Quarter 5–Quarter 8 average forecasts we find two-way Granger-causality. The results in table 11 and 12 support our assumptions about the relationship of the index to internal forecast signals that the Governing Council receives—namely, that the staff forecast influences the FG sentiment placed in the announcements but not the reverse. However, these results depend heavily on the choice of the number of included lags. In this test, the reported results involved 4 lags which is approximately two quarters in our data. However technical considerations can also suggest the lag choice to be 1. In this case, the test reports a two-way Granger-causality for all three specifications. Two way causality puzzle notwithstanding, in all cases we observe that the staff forecasts statistically influence the sentiment index.

By establishing a statistical link both the coefficient estimation and the Granger-causality tests provide further evidence in support of the index as a measure of sentiment.

6 Conclusion

6.1 Summary of Findings

This essay has set out to investigate the capabilities of Chat GPT—a new Large Language Model developed by Open AI, to detect the Forward Guidance of the Bank of Canada. The focus was on the sentiment inherent in policy announcements. We wanted to find a way to build an index that would have some predictive power vis à vis the policy rate changes. We have reviewed the currently available literature, economic theory, developed our model, and clarified the definition of Forward Guidance we are using. We developed several prompts that allowed us to capture the sentiment index and found that it was highly correlated with the historic changes in the overnight rate. We discussed the collinearity puzzle and considered that our measurements may be contaminated by endogeneity. We found evidence against this by analysing a variety of outputs. We further validated our index by comparing it with market expectations and internal BoC staff forecasts. The coefficient estimates and some additional tests have indicated that the sentiment index matches the theory and our anticipations when it comes to its interactions with the market expectations and internal forecasts used by the Bank. Therefore, we conclude that we seem to have achieved the goals we set in section 1.2.

6.2 AI's Role in Economic Analysis

The application of Chat GPT models 3.5 and 4.0 has demonstrated significant potential of transforming qualitative communication statements into quantitative and usable data. This allows us to detect and analyse Forward Guidance directly, in a way than was not possible previously. An important consideration is also a relatively low cost and fast processing time offered by this new method.

6.3 Challenges and Implications

Despite the encouraging results, the study encountered the puzzle of collinearity between the indexes and actual policy rate changes. The potential influence of BoC announcements embedded in the AI's training data could have implications for the index's perceived validity. This emphasized the continued need to be able to rely on older models in order to be able to compare communications from before and after the model's training cut off. By extending the available data to include the more recent announcements as well as by incorporating data from other sources, it is hoped that in time this puzzle will be resolved definitively.

6.4 Future Research Directions

This research can be extended in multiple directions:

First, statistical analysis presented here can be improved. Out of the many prompts presented

here and developed over the course of the research for this essay only a few most promising ones were taken all the way to Stata regression results stage. As well, most were not tested using GPT 4.0 the newer model. It is possible that within the data already generated as part of this essay, additional insights are concealed, waiting to be teased out.

Second, many improvements can be made to the prompts that were used to generate this index. It is possible that with additional experimentation the results can become even more precise and we will be able to reveal more insights about the way Forward Guidance is used by the Bank and the markets.

Third, the body of communications studied using these prompts can be expanded to include additional communications released by BoC such as the quarterly monetary reports. As well, this method can be exported to communications made by other central banks such as in the UK and the US.

Finally, the theoretical framework of the way the FG is used by the central banks can be revised to include for example: different governor appointments and the length of time they are in office, the makeup of the governing council, political leadership at the parliamentary and federal level, concurrent macroeconomic considerations, etc.

6.5 Outlook

As central banks continue to evolve to become more transparent, we are likely to see FG communications employed even more extensively. Given the recent rise in AI-enabled tools, we are also more likely to see more communications analysis using these tools. We hope that as these tools are validated and become more accessible, that they can be used by members of the public to better inform themselves of the state of the Canadian economy. Perhaps this will go some way towards abating the typical conjecture and speculation that abounds in this media field.

7 Acknowledgments

I want to thank my wife Uliana and my daughter Anna for giving me the opportunity to study this subject at length. Also I want to thank Professor Lilia Karnizova for her countless hours trying to guide me through this research. A special thanks goes to fellow MA student David Valenta for demonstrating the feasibility and ease of use of GPT enabled research

References

[Ali, 2023] Ali, F. (2023). Gpt models explained and compared. https://www.makeuseof.com/gpt-models-explained-and-compared/. Accessed: 2024-03-26.

[Bank of Canada, 2022] Bank of Canada (2022). Imf report following transparency review. Accessed: 2024-03-27.

[Bank of Canada, 2023a] Bank of Canada (2023a). Policy rate announcements. https://www.bankofcanada.ca/press/press-releases/. Accessed: Dec 2023.

[Bank of Canada, 2023b] Bank of Canada (2023b). Selected bond yields, bd.cdn.2yr.dq.yld. https://www.bankofcanada.ca/?p=39890. Accessed: Dec 2023.

[Bank of Canada, 2024] Bank of Canada (2024). Media advisory: Bank of canada interest rate announcement - march 6, 2024. https://www.bankofcanada.ca/2024/03/bank-canada-interest-rate-announcement-march-6-2024/. Accessed: March 30, 2024.

[Bassetto, 2019] Bassetto, M. (2019). Forward guidance: Communication, commitment, or both? *Journal of Monetary Economics*, 108:69–86.

[Bholat et al., 2015] Bholat, D., Hansen, S., Santos, P., and Schonhardt-Bailey, C. (2015). Text mining for central banks. Technical report, Centre for Central Banking Studies, Bank of England.

[Blinder et al., 2017] Blinder, A., Ehrmann, M., de Haan, J., and Jansen, D.-J. (2017). Necessity as the mother of invention: monetary policy after the crisis. *Economic Policy*, 32(92):707–755.

[Campbell et al., 2012] Campbell, J. R. et al. (2012). Macroeconomic effects of federal reserve forward guidance. *Brookings Papers on Economic Activity*, 2012(1):1–80.

[Carney, 2012] Carney, M. (2012). Speech to cfa society toronto: Guidance. Bank of Canada, Speeches and appearances. Accessed: 2024-03-03.

[Champagne et al., 2018] Champagne, J., Poulin-Bellisle, G., and Sekkel, R. (2018). Evaluating the bank of canada staff economic projections using a new database of real-time data and forecasts. Staff Working Paper 2018-52, Bank of Canada.

[Coibion et al., 2019] Coibion, O., Gorodnichenko, Y., and Weber, M. (2019). Monetary policy communications and their effects on household inflation expectations. NBER Working Paper 25482, National Bureau of Economic Research. Revised November 2021.

[Cook and Hahn, 1989] Cook, T. and Hahn, T. (1989). The effect of changes in the federal funds rate target on market interest rates in the 1970s. *Journal of Monetary Economics*, 24(3):331–351.

[Fama, 1970] Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

[Granger, 1969] Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.

[Gürkaynak et al., 2005] Gürkaynak, R. S., Sack, B. S., and Swanson, E. T. (2005). Do actions speak louder than words? the response of asset prices to monetary policy actions and statements. *International Journal of Central Banking*, 1(1).

[Kuttner, 2001] Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of Monetary Economics*, 47(3):523–544.

[Lucca and Trebbi, 2011] Lucca, D. O. and Trebbi, F. (2011). Measuring central bank communication: An automated approach with application to fomc. *Federal Reserve Bank of New York*.

[Murray, 2013] Murray, J. (2013). Monetary policy decision making at the bank of canada. *Bank of Canada Review*, Autumn 2013. Accessed: 2024-03-27.

[Open AI, 2023] Open AI (2023). Text generation guide. https://platform.openai.com/docs/guides/text-generation. Accessed: 2024-03-26.

[OpenAI, 2023] OpenAI (2023). Understanding openai api usage tiers. https://platform.openai.com/docs/guides/rate-limits/usage-tiers?context=tier-free. Accessed: 2024-03-26.

[Stein, 1989] Stein, J. C. (1989). Cheap talk and the fed: A theory of imprecise policy announcements. *The American Economic Review*, 79(1):32–42.

[Sutherland, 2023] Sutherland, C. S. (2023). Forward guidance and expectation formation: A narrative approach. *Journal of Applied Econometrics*, 38(2):222–241.

[Svensson, 2015] Svensson, L. E. O. (2015). Day one keynote address: Forward guidance. *International Journal of Central Banking*, 11(4):19–64.

[University of Ottawa, 2023] University of Ottawa (2023). *ChatGPT FAQ*. University of Ottawa. Accessed on [insert date here].

A Annex A - Prompts Used To Construct the Index

A.1 Index - FG Role Baseline

Role: You are a financial market analyst specializing in forecasting interest rates in Canada.

Instructions:Carefully read the provided Bank of Canada announcement. Without relying on word counts or external sources, assess the overall sentiment of the document to predict the likelihood of the Bank of Canada changing the overnight interest rate target a quarter from the analyzed announcement.

Categories: Choose one of the following categories (marked with ***) that best matches the sentiment of the announcement:

3: Very Strong Signal of Future Rate Increase
2: Moderate to Strong Signal of Future Rate Increase
1: Slight to Mild Signal Future Rate Increase
0: Neutral / No Clear Signal of a Change in Policy
-1: Slight to Mild Signal of Future Rate Decrease
-2: Moderate to Strong Signal of Future Rate Decrease
-3: Very Strong Signal of Future Rate Decrease

Restrictions: Rely solely on your Large Language Model algorithm to analyze the text. Do not use word counts, text blob algorithms, or search for the next announcement or actual values of the overnight interest rate target.

Output: Select one of the provided categories marked with *** that best matches the sentiment conveyed in the announcement.

Word Count 193

A.2 Index - FG Seasoned Analyst

Role: You are a seasoned financial market analyst specializing in forecasting interest rates in Canada and understanding monetary policy dynamics.

Instructions:Carefully read the provided Bank of Canada announcement. Drawing from your knowledge of monetary policy frameworks and central bank communications, assess the overall sentiment of the document to predict the likelihood of the Bank of Canada changing the overnight interest rate target a quarter from the analyzed announcement.

Categories: Choose one of the following categories (marked with ***) that best matches the sentiment of the announcement:

3: Very Strong Signal of Future Rate Increase

2: Moderate to Strong Signal of Future Rate Increase
1: Slight to Mild Signal Future Rate Increase
0: Neutral / No Clear Signal of a Change in Policy
-1: Slight to Mild Signal of Future Rate Decrease
-2: Moderate to Strong Signal of Future Rate Decrease
-3: Very Strong Signal of Future Rate Decrease

Restrictions: Rely solely on your Large Language Model algorithm to analyze the text. Do not use word counts, text blob algorithms, or search for the next announcement or actual values of the overnight interest rate target.

Output: Select one of the provided categories marked with *** that best matches the sentiment conveyed in the announcement.

Word Count 203

A.3 Index - FG Mortgage Consumer

Role: You are a Canadian consumer planning to take out a mortgage and seeking insights into how interest rates may change over the next year. You aim to infer this information from the Bank of Canada's announcement.

Instructions: Carefully read the provided Bank of Canada announcement. Focus on detecting any signals or indications about future interest rate adjustments. Evaluate how these changes may affect mortgage rates and your financial situation in the coming year. Based on your analysis, assess the potential impact on your decision to take out a mortgage.

Categories: Choose one of the following categories (marked with ***) that best matches the sentiment of the announcement:

3: Very Strong Signal of Future Rate Increase
2: Moderate to Strong Signal of Future Rate Increase
1: Slight to Mild Signal Future Rate Increase
0: Neutral / No Clear Signal of a Change in Policy
-1: Slight to Mild Signal of Future Rate Decrease
-2: Moderate to Strong Signal of Future Rate Decrease
-3: Very Strong Signal of Future Rate Decrease

Restrictions: Rely solely on your interpretation of the Bank of Canada's announcement. Do not seek external financial advice or consult mortgage professionals. Use your Large Language Model algorithm to analyze the text. Do not use word counts, text blob algorithms, or search for the next announcement or actual values of the overnight interest rate target.

Output: Select one of the provided categories marked with *** that best matches the sentiment conveyed in the announcement.

Word Count 244

A.4 Index - FG for the next quarter

Role: You have the role of an economist studying monetary policy. Your task is to carefully analyse Bank of Canada announcements and note any signals of future monetary policy changes.

Instructions: I am providing you with the body of the text of a Bank of Canada announcement. I want you to read it and match the overall sentiment of the text to one of several categories. The goal is not to match the categories to the current policy rate or even to the policy change currently being announced, but rather to predict the direction of change of the policy rate over the next quarter.

Categories: I want you to match the announcement to one of the following categories that capture the likelihood of policy rate changes during the next announcement. The categories are delimited using triple *** symbols.

3: Very Likely Future Rate Increase
2: Likely Future Rate Increase
1: Low Likelihood of Future Rate Increase
0: Unlikely to see any change in policy
-1 Low Likelihood of Future Rate Decrease
-2 Likely Future Rate Decrease
-3 Very Likely Future Rate Decrease

Restrictions: I want you to rely on your Large Language Model algorithm to analyze this text. I don't want you to use word counts or text blob algorithm. As well, I do not want you to use your knowledge of actual Bank of Canada policy rate decisions to come up withe the answer. I want you to use only text of the announcement to come up with the answer.

For the output I want you to select on of the categories provided above (marked with ***) that you feel best matches the text of the announcement.

Context: Your output will be used to quantify multiple Bank of Canada announcements (over 200) and used to build a database of observations. These will then be used for regression analysis that looks into effectiveness of BoC communications.

Word Count 316

A.5 Index - FG for the next announcement

Role: You have the role of an economist studying monetary policy. Your task is to carefully analyse Bank of Canada announcements and note any signals of future monetary policy changes.

Instructions: I am providing you with the body of the text of a Bank of Canada announcement. I want you to read it and match the overall sentiment of the text to one of several categories. The goal is not to match the categories to the current policy rate or even to the change currently being announced. The goal is to predict the direction of change of the policy rate on the next announcement date.

Categories: I want you to match the announcement to one of the following categories that capture the likelihood of policy rate changes during the next announcement. The categories are delimited using triple *** symbols.

3: Very Likely Future Rate Increase
2: Likely Future Rate Increase
1: Low Likelihood of Future Rate Increase
0: Unlikely to see any change in policy
-1 Low Likelihood of Future Rate Decrease
-2 Likely Future Rate Decrease
-3 Very Likely Future Rate Decrease

Restrictions: I want you to rely on your Large Language Model algorithm to analyze this text. I don't want you to use word counts or text blob algorithm. As well, I do not want you to use your knowledge of actual Bank of Canada policy rate decisions to come up withe the answer. I want you to use only text of the announcement to come up with the answer.

For the output I want you to select on of the categories provided above (marked with ***) that you feel best matches the text of the announcement.

Context: Your output will be used to quantify multiple Bank of Canada announcements (over 200) and used to build a database of observations. These will then be used for regression analysis that looks into effectiveness of BoC communications.

Word Count 317

A.6 Index - FG for signal changes over quarter

Role: You have the role of an economist studying monetary policy. Your task is to carefully analyse Bank of Canada announcements and note any signals or indications of future monetary policy changes contained in the text.

Instructions: I am providing you with the body of the text of a Bank of Canada announcement. I want you to read it and match the overall sentiment of the text to one of several categories. The goal is not to match the categories to the current policy rate or even to the change currently being announced. The goal is to predict the direction of change of the policy rate over the next quarter.

Categories: I want you to match each announcement to one of the following categories that capture the signals or indications of future policy rate changes over the next quarter that the Bank of Canada is conveying. The categories are delimited using triple *** symbols.

3: Strong Signal of Future Rate Increase
2: Moderate Signal of Future Rate Increase
1: Mild Signal Future Rate Increase
0: Neutral / No Clear Signal of a Change in Policy
-1: Mild Signal of Future Rate Decrease
-2 Moderate Signal of Future Rate Decrease
-3: Strong Signal of Future Rate Decrease

Restrictions: I want you to rely on your Large Language Model algorithm to analyze this text. I don't want you to use word counts or text blob algorithm. As well, I do not want you to use your knowledge of actual Bank of Canada policy rate decisions to come up withe the answer. I am interested only in the signals contained in the text of the announcement.

Output: For the output I want you to select on of the categories provided above (marked with ***) that you feel best matches the text of the announcement.

Context: Your output will be used to quantify multiple Bank of Canada announcements (over 200) and used to build a database of observations. These will then be used for regression analysis that looks into effectiveness of BoC communications.

Word Count 339

A.7 GPT4 - FG - Signals

Role: You have the role of an economist studying monetary policy. Your task is to carefully analyse Bank of Canada announcements and note any signals or indications of future monetary policy changes contained in the text.

Instructions: I am providing you with the body of the text of a Bank of Canada announcement. I want you to read it and match the overall sentiment of the text to one of several categories.

Categories: I want you to match each announcement to one of the following categories that capture the signals or indications of future policy rate changes that the Bank of Canada is conveying. The categories are delimited using triple *** symbols.

5: Very Strong Signals of Future Rate Increase
4: Strong Signals of Future Rate Increase
3: Moderate Signals of Future Rate Increase
2: Mild Signals Future Rate Increase
1: Slight Signals of Future Rate Increase
0: Neutral / No Clear Signals of a Change in Policy
-1: Slight Signals of Future Rate Decrease
-2: Mild Signals of Future Rate Decrease
-3: Moderate Signals of Future Rate Decrease
-4: Strong Signals of Future Rate Decrease
-5: Very Strong Signals of Future Rate Decrease

Restrictions: I want you to rely on your Large Language Model algorithm to analyze this text. I don't want you to use word counts or text blob algorithm. As well, I do not want you to use your knowledge of actual Bank of Canada policy rate decisions to come up withe the answer. I am interested only in the signals contained in the text of the announcement.

Output: For the output I want you to select on of the categories provided above (marked with ***) that you feel best matches the text of the announcement. Do not add additional explanations

Context: Your output will be used to quantify multiple Bank of Canada announcements (over 200) and used to build a database of observations. These will then be used for regression analysis that looks into effectiveness of BoC communications.

Word Count 332

A.8 Index - FG - Signals

Role: You have the role of an economist studying monetary policy. Your task is to carefully analyse Bank of Canada announcements and note any signals or indications of future monetary policy changes contained in the text.

Instructions: I am providing you with the body of the text of a Bank of Canada announcement. I want you to read it and match the overall sentiment of the text to one of several categories.

Categories: I want you to match each announcement to one of the following categories that capture the signals or indications of future policy rate changes that the Bank of Canada is conveying. The categories are delimited using triple *** symbols.

5: Very Strong Signals of Future Rate Increase
4: Strong Signals of Future Rate Increase
3: Moderate Signals of Future Rate Increase
2: Mild Signals Future Rate Increase
1: Slight Signals of Future Rate Increase
0: Neutral / No Clear Signals of a Change in Policy
-1: Slight Signals of Future Rate Decrease
-2: Mild Signals of Future Rate Decrease
-3: Moderate Signals of Future Rate Decrease
-4: Strong Signals of Future Rate Decrease
-5: Very Strong Signals of Future Rate Decrease

Restrictions: I want you to rely on your Large Language Model algorithm to analyze this text. I don't want you to use word counts or text blob algorithm. As well, I do not want you to use your knowledge of actual Bank of Canada policy rate decisions to come up withe the answer. I am interested only in the signals contained in the text of the announcement.

Output: For the output I want you to select on of the categories provided above (marked with ***) that you feel best matches the text of the announcement. Do not add additional

explanations

Context: Your output will be used to quantify multiple Bank of Canada announcements (over 200) and used to build a database of observations. These will then be used for regression analysis that looks into effectiveness of BoC communications.

Word Count 332

A.9 Index - FG Baseline

Role: You have the role of an economist studying monetary policy. Your task is to carefully analyse Bank of Canada announcements and note any signals of future monetary policy changes.

Instructions: I am providing you with the body of the text of a Bank of Canada announcement. I want you to read it and match the overall sentiment of the text to one of several categories.

Categories: I want you to match each announcement to one of the following categories that capture the likelihood of future policy rate announcements. The categories are delimited using triple *** symbols.

5: Highly Certain Future Rate Increase
4: Strong Indication of Future Rate Increase
3: Moderate Likelihood of Future Rate Increase
2: Mild Indication of Future Rate Increase
1: Slight Possibility of Future Rate Increase
0: Neutral / No Clear Change in Policy Forecast
-1: Slight Possibility of Future Rate Decrease
-2: Mild Indication of Future Rate Decrease
-3: Moderate Likelihood of Future Rate Decrease
-4: Strong Indication of Future Rate Decrease
-5: Highly Certain Future Rate Decrease

Restrictions: I want you to rely on your Large Language Model algorithm to analyze this text. I don't want you to use word counts or text blob algorithm.

Output: For the output I want you to select on of the categories provided above (marked with ***) that you feel best matches the text of the announcement.

Context: Your output will be used to quantify multiple Bank of Canada announcements (over 200) and used to build a database of observations. These will then be used for regression analysis that looks into effectiveness of BoC communications.

Word Count 269

A.10 Index - FG Baseline Modified Text

Role: You have the role of an economist studying monetary policy. Your task is to carefully analyse Bank of Canada announcements and note any signals of future monetary policy changes.

Instructions: I am providing you with the body of the text of a Bank of Canada announcement. I want you to read it and match the overall sentiment of the text to one of several categories.

Categories: I want you to match each announcement to one of the following categories that capture the likelihood of future policy rate announcements. The categories are delimited using triple *** symbols.

5: Highly Certain Future Rate Increase
4: Strong Indication of Future Rate Increase
3: Moderate Likelihood of Future Rate Increase
2: Mild Indication of Future Rate Increase
1: Slight Possibility of Future Rate Increase
0: Neutral / No Clear Change in Policy Forecast
-1: Slight Possibility of Future Rate Decrease
-2: Mild Indication of Future Rate Decrease
-3: Moderate Likelihood of Future Rate Decrease
-4: Strong Indication of Future Rate Decrease
-5: Highly Certain Future Rate Decrease

Restrictions: I want you to rely on your Large Language Model algorithm to analyze this text. I don't want you to use word counts or text blob algorithm.

Output: For the output I want you to select on of the categories provided above (marked with ***) that you feel best matches the text of the announcement.

Context: Your output will be used to quantify multiple Bank of Canada announcements (over 200) and used to build a database of observations. These will then be used for regression analysis that looks into effectiveness of BoC communications.

Word Count 269

A.11 GPT4 - FG - Signals Modified Text

Role: You have the role of an economist studying monetary policy. Your task is to carefully analyse Bank of Canada announcements and note any signals or indications of future monetary policy changes contained in the text.

Instructions: I am providing you with the body of the text of a Bank of Canada announcement. I want you to read it and match the overall sentiment of the text to one of several categories.

Categories: I want you to match each announcement to one of the following categories that capture the signals or indications of future policy rate changes that the Bank of Canada is conveying. The categories are delimited using triple *** symbols.

5: Very Strong Signals of Future Rate Increase
4: Strong Signals of Future Rate Increase
3: Moderate Signals of Future Rate Increase
2: Mild Signals Future Rate Increase
1: Slight Signals of Future Rate Increase
0: Neutral / No Clear Signals of a Change in Policy
-1: Slight Signals of Future Rate Decrease
-2: Mild Signals of Future Rate Decrease
-3: Moderate Signals of Future Rate Decrease
-4: Strong Signals of Future Rate Decrease
-5: Very Strong Signals of Future Rate Decrease

Restrictions: I want you to rely on your Large Language Model algorithm to analyze this text. I don't want you to use word counts or text blob algorithm. As well, I do not want you to use your knowledge of actual Bank of Canada policy rate decisions to come up withe the answer. I am interested only in the signals contained in the text of the announcement.

Output: For the output I want you to select on of the categories provided above (marked with ***) that you feel best matches the text of the announcement.

Context: Your output will be used to quantify multiple Bank of Canada announcements (over 200) and used to build a database of observations. These will then be used for regression analysis that looks into effectiveness of BoC communications.

Word Count 328