The Government Bond Market and X-Based Public Optimism: During Pre-Rate Hike and Rate Hike Periods*

Kairan Chen Brigitte Granville Roman Matousek

kairan.chen@qmul.ac.uk b.granville@qmul.ac.uk r.matousek@qmul.ac.uk

Queen Mary University of London

Abstract

This paper aims to study fluctuations in the government bond market using macroeconomic information over a continuous time horizon. As economic fundamentals are released at scheduled intervals rather than continuously, it is necessary to obtain a continues quantitative indicator. We propose a "Deep Learning + Heuristic Rules" approach to measure optimism toward fundamentals by analyzing the stream of tweets on X. Validation results suggest that this new measure is a reliable and timely alternative to the survey-based measures. Empirical findings indicate that the derived optimism index is a strong daily predictor of gilt yields, showing a high positive correlation. This relationship aligns with existing arguments that, during periods of elevated mood, investors tend to be bullish about financial markets, which can lead to lower demand for bonds relative to speculative stocks. Lastly, our proposed optimism metric performs comparably to metrics based on large language models in tracking a survey benchmark and explaining bond market trends.

Keywords: Optimism, Twitter, X, The government bond, Rate hike

JEL Code: C02, C32, C55, E40, G12

* The authors declare that no funding was received for the research and preparation of this paper. The authors also have no financial relationships or other potential conflicts of interest to disclose.

1. Introduction

Government bond markets, a key component of investment portfolios, offer investors signals about interest rates and are closely monitored by both the public and private sectors. These markets are highly sensitive to a wide range of factors. A large body of literature has documented the impact of macroeconomic announcements—such as GDP, CPI, monetary policy, employment, and retail sales—on bond yields (e.g., Beechey and Wright 2009; Ludvigson and Ng 2009; Altavilla, Giannone, and Modugno 2017). Several have established causality link using high-frequency event study (e.g., Balduzzi, Elton, and Green 2001; Gürkaynak, Kısacıkoğlu, and Wright 2020). These findings demonstrate that macroeconomic news is the primary driver of market fluctuations, which offers valuable insights into yield movements around announcements. Yet, these news releases occur at discrete intervals, while yields often fluctuate outside of announcement days. It is not easy to explain these fluctuations by fundamental news over a continuous time horizon. Thus, this paper aims to address this challenge by analyzing the stream of tweets from X (formerly Twitter).

As an information platform, X significantly influences how investors access and disseminate news, such as central bank communications (Bianchi, Gómez-Cram, and Kung 2024), financial news (Bartov, Faurel, and Mohanram 2023), and economic fundamentals (Dergiades, Milas, and Panagiotidis 2015). A decade ago, Bloomberg began integrating tweets into its data service.[1] As of 2024, X remains the most popular microblogging platform with 355 million users. For market participants, watching X has gradually become as routine as watching Bloomberg.[2] From another perspective, behavioral finance has revealed that financial decisions are significantly driven by emotions (Nofsinger 2005; Kaplanski and Levy 2010; Hirshleifer, Jiang, and DiGiovanni 2020;

[1] Please see https://archive.nytimes.com/dealbook.nytimes.com/2013/04/04/twitter-arrives-on-wall-street-via-bloomberg/.

[2] Please see https://www.google.com.au/amp/s/amp.ft.com/content/c464d944-ee75-11e4-98f9-00144feab7de.

and Edmans et al. 2022). Given advances in textual analysis technology that enable the extraction of emotions from X, a growing body of papers has leveraged these extracted emotions to study asset pricing (e.g., Bollen, Mao, and Zeng 2011; Berardi 2022; Bartov, Faurel, and Mohanram 2023). Building on Consoli, Pezzoli, and Tosetti (2021), which emphasizes the role of sentiment extracted from news in predicting the yield spreads, we extract fundamentals optimism from X. This optimism is defined as the sentiment towards the three types of macroeconomic news highlighted by Beechey and Wright (2009): news about prices, the real economy, and monetary policy. Henceforth, we refer to these fundamentals as the Beechey-Wright fundamentals, while their corresponding optimism is referred as the Beechey-Wright optimisms.

To this end, a reliable way to gauge optimism is required. To our knowledge, no existing sentiment classifier has been specifically built for Beechey-Wright fundamentals. Conventional pre-trained classifiers are not suitable for extrapolation beyond their original training contexts (Loughran and McDonald 2011). The emerging large language models (LLMs) are applicable, yet their performance and usage vary considerably between models. Therefore, we propose a "Deep Learning + Heuristic Rules" (henceforth, DL+Rule) framework to measure the X-based optimism towards the Beechey-Wright fundamentals.

This suggested framework considers both contextual information and interpretability. To do so, (i) we first build a set of UK tweets by collecting those containing keywords related to the Beechey-Wright fundamentals. (ii) We use word2vec (Mikolov et al. 2013), a well-known neural network-based word embedding technique, to learn the meaning of each word and represent it as a distributed numerical vector. (iii) We separately compute the cosine similarity between a given word to the representative positive vector and negative vector. We further compute its net optimism score as the difference between its positive and negative similarity. Consequently, each

word is assigned a net optimism score, resulting in the creation of a lexicon (a list of word-score pairs) specifically focused on the Beechey-Wright fundamentals. (iv) Hutto and Gilbert's (2014) heuristic rules include a set of grammatical and syntactical rules. When computing optimism based on the new lexicon, we apply the heuristic rules to account for contextual characteristics.

Drawing on our proposed DL+Rule framework and a dataset of 5.7 million UK tweets related to Beechey-Wright fundamentals, we generate a daily X-based optimism index. We then validate it by comparing it with the GFK Consumer Confidence Index, a well-accepted survey in the UK. We find that our index aligns closely with this benchmark. Despite being derived from different data sources and methodologies, the two indices exhibit a strong correlation of *0.82*, surpassing the correlations of this survey with three LLM-based metrics: GPT3.5 (*0.75*), FinBERT (*0.44*), and Twitter-RoBERTa (*0.64*).

To examine how market fluctuations can be explained over continuous time, we conduct a regression analysis to assess the UK government bond market's response to our X-based optimism index across two distinct periods: before and after the Bank of England's ongoing interest rate hikes. By adding optimism to the baseline model for predicting UK gilt yields, we find that our novel optimism index increases the baseline model's explanatory power by *0.097* during the pre-rate hike period. This performance surpasses that of GPT3.5 (*0.047*), FinBERT (*0.015*), and Twitter-RoBERTa (*0.094*). The coefficients of this set of optimism indices are all significantly positive, indicating a positive correlation between public optimism regarding economic conditions and the level of government bond yields. Additionally, this result is also present in their first-order difference series, indicating a positive relationship between changes in optimism and changes in yields. These findings can be explained by existing studies: during periods of elevated mood, investors tend to be bullish about the financial markets (Qiu and Welch 2004), which can lead to

lower demand for bonds relative to speculative stocks (Baker and Wurgler 2012). On the other hand, during the rate hike period, our optimism index adds *0.184* to the model's explanatory power, which is slightly higher than GPT3.5 (*0.136*), FinBERT (*0.02*), and Twitter-RoBERTa (*0.08*). After controlling for the linear trend caused by continuous rate hikes, the positive relationship between yields and optimism remains evident.

This paper provides a fresh perspective on the outlined research problems in the following ways. First, this paper builds on the recent ground-breaking research by Li et al. (2021), which develop a measurement of corporate culture using word2vec and cosine similarity. We extend their approach by retaining certain key elements partially, while integrating heuristic rules. This allows for the construction of lexicons, enabling the measurement of optimism. In addition, this paper suggests a new framework for measuring textual sentiment, specifically for the economics and finance disciplines (e.g., Shapiro, Sudhof, and Wilson 2022). This framework tends to complement existing approaches by pursuing measurement accuracy and also considering interpretability. For instance, the embedded heuristic rules enable to capture contextual information via punctuations, negations, capitalizations, and degree modifiers. It thus considers context to a greater extent than previous dictionary-based approaches (e.g., García 2013; Bartov, Faurel, and Mohanram 2018). Additionally, it does not rely on a labelled dataset and can be applied to other disciplines, as it exploits word embedding to learn the characteristics of different corpora. Moreover, this paper may complement studies that utilize LLMs for sentiment classification (e.g., Lopez-Lira and Tang, 2023; Chen, Kelly, and Xiu, 2023) by offering an alternative that requires less computational power while showing performance comparable to that of LLMs.

Second, this paper enriches the body of empirical research that examines the relationship between sentiment and the government bond market. The impact of market mood on the bond

market is well-established, covering areas such as financial indicators (Baker and Wurgler 2012), special policy events (Milas, Panagiotidis, and Dergiades 2021), and fear-driven emotions (Da, Engelberg, and Gao 2015). Yet, given that sentiment changes based on the underlying subjects and measurement methods, we expand this literature by offering evidence of X-based optimism related to Beechey-Wright fundamentals. We also add to a growing body of literature (e.g., Balduzzi, Elton, and Green 2001; Ludvigson and Ng 2009; Beechey and Wright 2009; Altavilla, Giannone, and Modugno 2017; Gilbert et al. 2017; Gürkaynak, Kısacıkoğlu, and Wright 2020). These studies provide evidence from the time window surrounding official announcements. We expand on this by providing evidence from a continuous time horizon, including non-announcement days.

This paper is structured as follows: Section 2 reviews existing measurement approaches and develops our own. We present the validation results and compare our approach to LLMs. Section 3 describes the empirical methods and presents the results. We carry out robustness checks in Section 4. Section 5 concludes.

2. The Methodological Concept of X-based Measure of Optimism

2.1. Motivation for our Approach

Optimism is a type of sentiment. A crucial step for all sentiment-related studies is to measure sentiment, a latent variable that is not directly observable. Existing measurement methods broadly fall into four categories: the market-based, the survey-based, the search-based, and the text-based methods (see Table 1).

Regarding the market-based methods, Baker and Wurgler (2006) propose six market proxies for investor sentiment and construct a composite index using principal components analysis. It has been widely used in empirical studies (e.g., Stambaugh, Yu, and Yuan 2012; Bennani 2020) and

inspires a long stream of modelling work (e.g., Huang et al. 2015); Regarding the survey-based methods, many available surveys record the distinct subjective beliefs held by respondents. They offer scholars various proxies for gauging sentiment, including consumer confidence applied by Bergman and Roychowdhury (2008) and Chen, Lien, and Lin (2021), investor optimism index exploited by Adam, Matveev, and Nagel (2021), and return expectations applied by Greenwood and Shleifer (2014); Regarding the search-based methods, some studies exploit the Google search actives to measure the sentiment index of Fear (Da, Engelberg, and Gao 2015; Kostopoulos and Meyer 2018) and households (Gao, Ren, and Zhang 2020).

Table 1

Selected Papers for Sentiment Modelling

Approaches		**Selected Papers**	**Descriptions**
Market-based		Baker and Wurgler (2006)	Propose 6 proxies and construct a composite index using principal components analysis.
		Huang et al. (2015)	Exploit the information of BW's 6 proxies in a more efficient way using partial least squares.
Survey-based		Bergman and Roychowdhury (2008)	Track consumer confidence.
		Greenwood and Shleifer (2014)	Track return expectations.
		Adam, Matveev, and Nagel (2021)	Track Investor optimism.
Search-based		Da, Engelberg, and Gao (2015) Gao, Ren, and Zhang (2020)	Exploit Google search queries.
		Dergiades, Milas, and Panagiotidis (2015)	Exploit Twitter trend.
Text-based	Lexicon / Dictionary	Tetlock (2007)	Uses a general-purpose dictionary (Harvard-IV dictionary).
		Loughran and McDonald (2011)	Develop a finance dictionary.
		Hutto and Gilbert (2014)	Develop the Vader classifier by introducing heuristic rules.
		Shapiro, Sudhof, and Wilson (2022)	Create a new lexicon based the Vader sentence classifier.
	Machine Learning	Bartov, Faurel, and Mohanram (2018)	Use enhanced naïve Bayes method.
		Cookson and Niessner (2020)	Use maximum entropy-based method.
	Large Language Models	Lopez-Lira and Tang (2023)	Exploit GPT to classify the news headlines.
		Chen, Kelly, and Xiu (2023)	Exploit BERT, GPT, OPT, and GPT to classify the news articles.

Note: Selected papers only, not a comprehensive list.

While the three categories of methods above have been widely used in asset pricing research over the past decade, we are interested in textual analysis. This method has developed rapidly in recent years. Text information channels are diverse, including sources like news, announcements, and forums. Additionally, technological advances now allow us to extract valuable signals from these sources, which adds value to the baseline models. There are broadly two main approaches to textual sentiment analysis: lexicon-based and machine learning (ML)-based. Our proposed method is a hybrid approach derived from both methodological literature strands. In the following, we begin by describing the two approaches, followed by an overview of our modelling design.

The lexicon-based approaches use a pre-defined lexicon. A lexicon is a list of word-score pairs, with each word assigned a score indicating whether it is positive, negative, or neutral. The sentiment of a document is determined by counting the words in each category, which is commonly known as the Bag-of-Words (BOW) model. Earlier studies often conduct BOW model using the general-purpose dictionaries, such as the Harvard-IV dictionary (e.g., Tetlock 2007). However, BOW model has two mainly critiques: (i) Loughran and McDonald (2011) point out that while it is straightforward to implement, its accuracy suffers if an applied dictionary does not account for situations where certain words have entirely different meanings within a given discipline. For instance, a neutral word in the accounting discipline (e.g., “liability”) might be classified as a negative word by a general-purpose dictionary. (ii) BOW model ignores each word’s contextual characteristics, such as the sequence of words and the co-occurrence with other words. To separately address the two critiques, Loughran and McDonald (2011) build a financial dictionary that has been widely used (e.g., García 2013; Jiang et al. 2019), then Hutto and Gilbert (2014) develop the Vader model by incorporating the grammatical and syntactical rules to quantify the sentiment intensity. These two are subsequently incorporated into our modelling framework.

The ML-based approaches employ statistical learning algorithms to identify the relationship between words and sentiment. The common practice to achieve this is using supervised learning models (e.g., Bartov, Faurel, and Mohanram 2018; Cookson and Niessner 2020). They are trained on the text sets with sentiment labels, enabling them to learn the statistical relation between the input (texts) and output (sentiment classification). Nonetheless, applying these supervised models is not easy because of the requirement for labelled texts, which becomes extremely labor-intensive when conducted manually. In addition, the neural network—a special supervised learning model—has gained increased attention in computational linguistics because of its capability to learn the complex nonlinear relationships between inputs and outputs. Specifically, Large language models (LLMs), framed by deep neural networks, have received widespread attention in the past two years (e.g., Lopez-Lira and Tang 2023; Chen, Kelly, and Xiu 2023). While LLMs excel at processing text-related tasks, their handling of contextual information is largely unobservable, resulting in a limited understanding of their classification mechanisms.

Therefore, we propose a relative compromise: constructing a measurement framework that consists of both semantic extraction of single words and contextual information extraction of multiple words. This framework is referred to as "Deep Learning + Heuristic Rules" in this paper. In other words, deep learning techniques are used in upstream tasks to extract the semantic information of single words, while heuristic rules are applied in downstream tasks to extract contextual information. This design enables us to extract the semantics of individual words from a dataset and capture contextual information in a clear way.

To do so, as shown in Figure 1, we first build a new lexicon dedicated to UK Beechey-Wright fundamentals. This is achieved by combining the word2vec from Mikolov et al. (2013) and the cosine similarity technique. Word2vec is one of the best-known word representation

techniques built on the neural network. It parses the words in contexts deeply then represents each word by its corresponding numeric vector, thereby enabling to use the cosine similarity technique to calculate the similarity between any pair of words. In this way, we calculate the positive (and the negative) representative vectors as the mean of vectors of positive (and negative) words recognized by the Loughran-McDonald financial dictionary. Then, we employ the cosine similarity technique to separately computes to what extent a given word is close to the positive (and the negative) representative vectors, which then is denoted as its polarity score of positive (and negative). Next, we follow Shapiro, Sudhof, and Wilson (2022) to assign a word a net optimism score computed as the difference between its positive polarity score and negative polarity score. Their polarity scores are relied on the classification results of Vader, yet Vader is more applicable to the general context rather than economic domain. Unlike them, our polarity scores derived by cosine similarity can ensure that the resulting sentiment is anchored in the economics domain. After obtaining the lexicon, unlike a series of papers that apply BOW model to the downstream tasks, we exploit the heuristic rules of Hutto and Gilbert (2014) to compute the sentiment score in the tweet-level. Heuristic rules can capture the contextual characteristics via punctuations, negations, capitalizations, and degree modifiers. Hutto and Gilbert demonstrate that the rules-based lexicon approach performs exceptionally well in the social media domain including X. Finally, the aggregate optimism index for a given day is computed as the mean of the optimism of all tweets on that given day.

To conclude, the entire process is divided into upstream tasks (i.e., word transformation and lexicon construction) and downstream tasks (i.e., document optimism scoring). The appropriate techniques are assigned to them respectively, which capitalizes on the strengths of each technique. The advantages of the proposed framework are fourfold. For the upstream task: (i) it does not

require a labelled dataset, which saves labor-related costs; (ii) it can be applied to the corpus from other disciplines, whether labelled or unlabeled, because word2vec allows for learning the context of the corpus. For the downstream task: (iii) it considers the context more than conventional dictionary-based approaches by applying a heuristic rules-based lexicon approach. For the resulting index: (iv) since its inputs are social media feeds, researchers have the flexibility to select any time interval (from minutes to days) for data aggregation, thereby facilitating analysis at different levels of granularity.

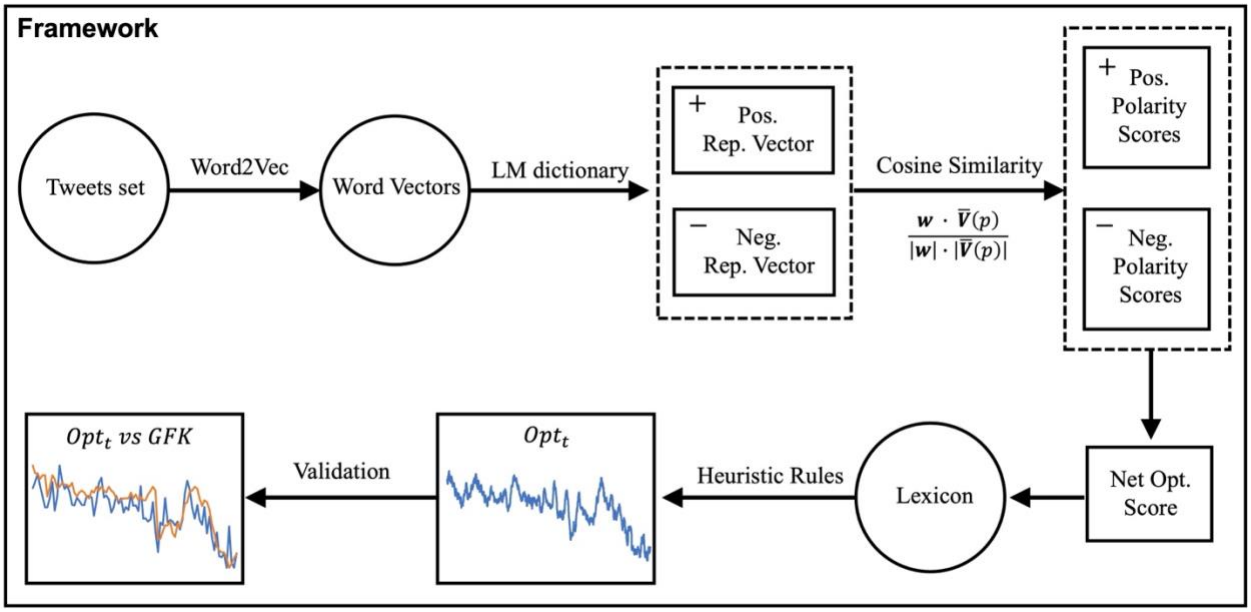


Fig. 1. Novel Framework Design

Note: This is the proposed framework for measuring public optimism about the Beechey-Wright fundamentals.

2.2. Measuring X-Based Optimism

In this section, we introduce a novel framework referred to as "Deep Learning + Heuristic Rules" (hereafter DL+Rule), designed to measure the X-based optimism and implemented in four steps. (i) Text collection and word embedding, converting each word into its corresponding distributed vector. (ii) Classifies positive and negative words using the LM financial dictionary and compute

the representative positive and negative vectors, respectively. (iii) Constructs a new lexicon dedicated to the domain of UK economy condition. (iv) Computes an aggregated X-based optimism.

2.2.1. Step 1: Text Collection and Word Embedding

A main method for constructing a corpus of tweets is to scrape tweets by keywords, ensuring that each tweet contains at least one word from the designated keywords list. A common practice in identifying a list of keywords is to include as many relevant keywords as possible related to the research topic.

Table 2

Beechey-Wright Fundamentals

Categories	Keywords	Description
Real Economy	"UK GDP", "UK GNP", "UK economy", "UK unemployment", "UK jobless"	Capture tweets discussing overall economic performance, growth, and the labor market
Price	"UK CPI", "UK inflation", "UK price level", "UK house price", "UK property price"	Capture tweets discussing inflation, general prices, and housing affordability
Monetary Policy	"UK monetary policy", "UK interest rate", "UK money supply"	Capture tweets discussing the policy-related events.

Given the vast volume of tweet data and related corpora, this paper does not aim to construct a perfect optimism index but rather to propose a reliable method for consideration. Due to resource constraints, the keyword selection in this study is intentionally focused on three categories outlined by Beechey and Wright (2009): news related to real output, prices, and monetary policy, as detailed in the Table 2. Additionally, we assume that sentiments expressed through different terms for the same concept on social media share similar trends or distributions. For instance, discussions of the "UK economy" are expected to show comparable sentiment patterns to those discussing the "British economy". As such, we have selected representative keywords to capture essential sentiment while accounting for the limitations of PC hardware. Building on Xu and Zhao's (2022)

work on the macroeconomic microblogs, we construct a keywords list for Beechey-Wright fundamentals. This list involves 13 keywords in total shown as below, which are highly relevant to the macroeconomic news studied by existing literature (e.g., Gürkaynak, Kısacıkoğlu, and Wright 2020).

All tweets are collected by using X official application programming interface (API). In addition to the textual content, each retrieved tweet contains its user's metadata, such as username, date, location, and description. Our dataset comprises nearly 5.7 million raw tweets (including retweets and replies) from close to 1.3 million users worldwide, spanning from January 1, 2016, to December 31, 2022. The words from these 5.7 million tweets form a corpus. We need to convert each word into its corresponding word vector. As a milestone breakthrough in textual analysis, word2vec exploits the neural network with a single hidden layer to predict the current word from a window of surrounding contexts words (or predict the surrounding contexts words from the current word). Beyond making predictions, a valuable by-product obtained during the training of word2vec is the distributed word vector. The process of using neural network models to learn contextual relationships and obtain word vectors is known in academia as "word embedding". The resulting distributed word vector has a much lower dimensionality than the traditional one-hot encoding vector, and each component of this vector carries specific semantic information "learned" from the corpus. When the distributed word vector is used in conjunction with the cosine similarity, the semantic similarities can be calculated (see Step 3 later).

We first utilize an algorithm from Genism—an open-source NLP Python package[3]—to train a word2vec model. Based on rule of thumb guidelines for setting hyperparameters, the word vector dimension is set to 300. Two words are defined as neighbors if they are within a 5-word distance

[3] See https://github.com/RaRe-Technologies/gensim.

in a sentence, and words appearing less than 5 times are omitted. Consequently, in addition to a word2vec model dedicated to discussions on X about the UK Beechey-Wright fundamentals, we also obtain the resulting word vectors for each individual word in our corpus.

2.2.2. Step 2: Representative Vectors for Positive and Negative Polarity

To extract features representing positive and negative polarity from the corpus, a benchmark is needed to differentiate between the two. A common practice for classifying words is to consult an established dictionary. We employ the Loughran-McDonald financial dictionary (hereafter, the LM dictionary) to ensure that the classification benchmark aligns with the economics and finance domain. The LM dictionary contains many words uncommon in spoken English, and these words are absent from our tweet corpus. Given the trained word2vec model can only produce a word vector for words within its training corpus (i.e., our tweets corpus), we utilize only the words that intersect between the tweet corpus and the LM dictionary.

The set of positive word vectors, $[\boldsymbol{w}_1^{\{pos\}}, \boldsymbol{w}_2^{\{pos\}},\ldots, \boldsymbol{w}_N^{\{pos\}}]$, is a series of word vectors from the intersected words between the tweet corpus and the LM dictionary's positive list. We obtain the representative vector for the positive sentiment by $\overline{V}_{pos} = \frac{1}{N}\sum_{\mathrm{i}=1}^{\mathrm{N}} [\boldsymbol{w}_1^{\{pos\}}, \boldsymbol{w}_2^{\{pos\}},\ldots, \boldsymbol{w}_N^{\{pos\}}]$. Also, the same paradigm is re-implemented on the set of negative word vectors to compute the representative vector for negative sentiment ($\overline{V}_{neg}$).

2.2.3. Step 3: Construction of a New Lexicon

A lexicon is composed of a series of word-score pairs, where a given word's sentiment extent is indicated by its score. In this step, we construct a novel lexicon specifically focused on discussions about UK Beechey-Wright fundamentals.

To do so, we need to measure the net positive score for each word. According to Eq. (1), we first employ the cosine similarity technique—a common practice to gauge the similarity between two vectors—to compute the extent to which a given word vector closes to the positive representative vector. Also, we implement Eq. (2) to gauge to which extent a given word vector closes to the negative representative vector.

$$PosScore_n = cosine\left(\boldsymbol{w}_n, \overline{V}_{pos}\right) = \frac{\boldsymbol{w}_n \cdot \overline{V}_{pos}}{|\boldsymbol{w}_n| \cdot |\overline{V}_{pos}|} \tag{1}$$

$$NegScore_n = cosine\left(\boldsymbol{w}_n, \overline{V}_{neg}\right) = \frac{\boldsymbol{w}_n \cdot \overline{V}_{neg}}{|\boldsymbol{w}_n| \cdot |\overline{V}_{neg}|} \tag{2}$$

where n denotes an individual word in our corpus of tweets and $\boldsymbol{w}_n$ denotes its numerical word vector. Secondly, built upon Shapiro, Sudhof, and Wilson (2022), we define the net optimism score $NetScore_n$ for a given word n according to Eq. (3).

$$NetScore_n = PosScore_n - NegScore_n \tag{3}$$

where $NetScore_n$ is then standardized. This approach assigns each word in our corpus a net optimism score, facilitating the construction of a novel lexicon focused on social media discussions about UK Beechey-Wright fundamentals. It can be easily extended to other fields, thereby reducing the significant labor costs associated with lexicon creation.

2.2.4. Step 4: Heuristic Rules

After obtaining the novel lexicon, a natural next step is to compute the optimism of a given word sequence (i.e., a tweet). Unlike previous work (e.g., García 2013; Bartov, Faurel, and Mohanram 2018), we account for the contextual characteristics of words within the text by applying the Hutto-Gilbert heuristics rules. It evaluates how punctuation, capitalization, degree modifiers, contrastive conjunctions, and trigrams, contribute to the overall meaning of the sentence. For example, the

words written by capitalization (e.g., “GOOD”), ended with exclamation point (e.g., “good !”), and followed by modifying adverbs (e.g., “very”) are assigned with higher magnitude of sentiment intensity. In addition, the contrastive conjunction (e.g., “but”) will inverse the sentiment polarity of the following statement.

By combining heuristics rules with the net score of individual words, we compute the optimism of an individual tweet *i* according to Eq. (4):

$$Opt_i = f\left(\left[NetScore_1, \ldots, NetScore_{N_i}\right]\right) \tag{4}$$

where $[NetScore_1, \ldots, NetScore_{N_i}]$ represents the sequence of scores for the series of words that make up the tweet, and $f(\cdot)$ denotes the Hutto-Gilbert heuristic rules. $Optimism_i$ is then standardized.

Finally, to construct a time series of optimism, we aggregate the $Optimism_i$ of all tweets on a daily basis by computing their average optimism for a given day *t*, shown as Eq. (5).

$$Opt_t = \frac{\sum_1^{N_t} Opt_{i,t}}{N_t} \tag{5}$$

where N_t denotes the volume of tweets posted on the day *t* in our dataset. This aggregated index measures the extent of public optimism regarding UK Beechey-Wright fundamentals. Since social media information fluctuates greatly on the daily frequency, we exploit the exponential moving average (EMA) of different windows to smoothen this index (please see Figure 2). It assigns greater weight to more recent data compared to older data, thereby capturing the more immediate impact of sentiment on markets.

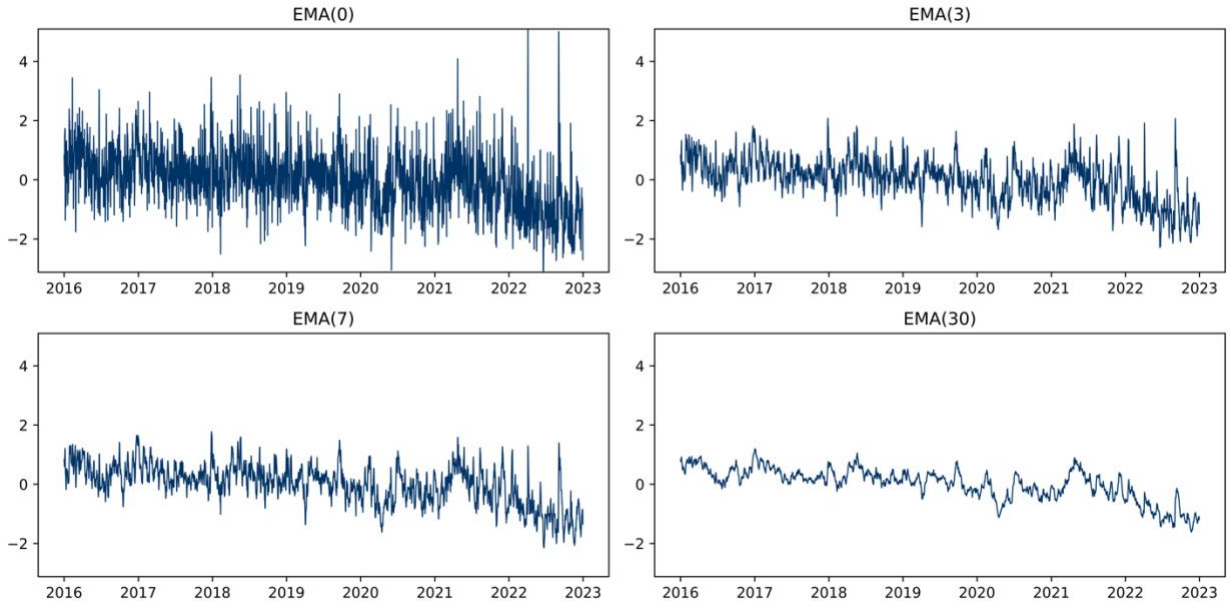


Fig. 2. X-Based Optimism

Notes: This diagram presents our X-based optimism using different exponential moving average windows: EMA 0 days, 3 days, 7 days, and 30 days. The EMA (0) represents the original series.

2.3. Validation

The lack of labelled data makes it unfeasible to use a confusion matrix to evaluate the sentiment classifier's performance. Therefore, we do not validate classification results on *individual* tweets. Instead, we validate the *aggregated* index derived from these results. In other words, our focus is not on the accuracy of assigning sentiment polarity to individual tweets, but rather on how well the final aggregated daily sentiment index aligns with the benchmark.

In doing so, we assess the efficacy of our index by comparing it with the GFK consumer confidence index that is a benchmark for measuring consumer confidence throughout the UK. The GFK index is published monthly and is available on Refinitiv Eikon. Since our X-based index is on the daily basis, we aggregate it to a monthly frequency to enable the overlay of both index curves on Figure 3.

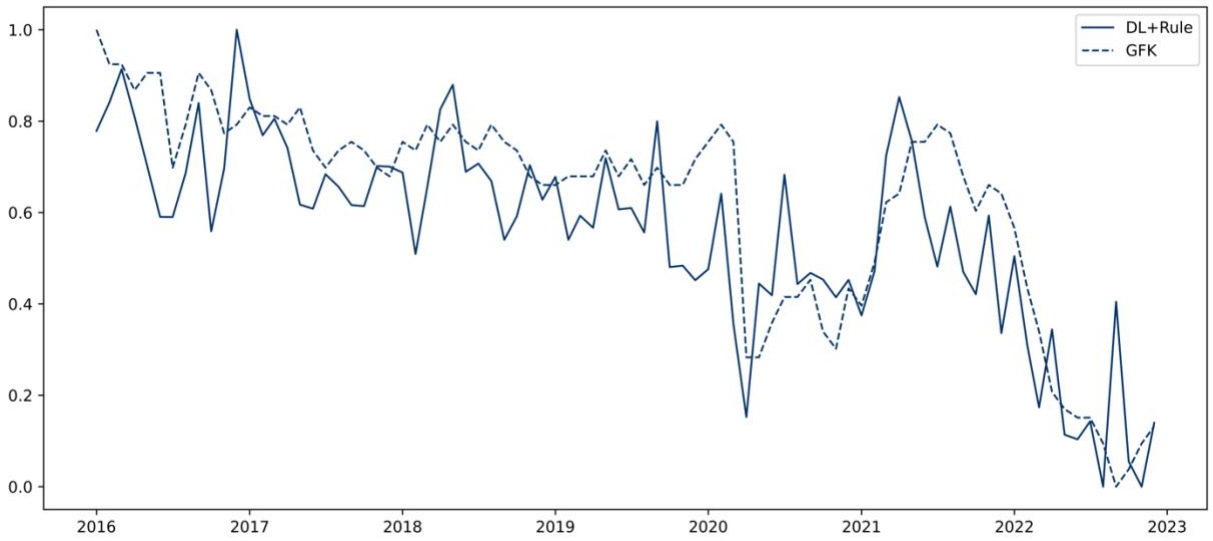


Fig. 3. X-based Optimism Versus GFK Consumer Confidence Index

Note: Both indices are normalized using min-max scaling, so the displayed indices take values from 0 to 1. The data spans from January 2016 to December 2022.

Visually, both indices closely align, and the X-based index exhibits slightly more volatility. We notice, however, that there are instances where our X-based index moves in the opposite direction to the survey-based metrics. Upon closer examination of the content of the tweets during these periods, we find that these discrepancies can be attributed to the high volume of certain specific discussion. For example, in September of 2022, our X-based index exhibits a peak, whereas GFK index displays a trough. It is observed that there are numerous tweets mentioning that India has overtaken the UK to become the fifth-largest economy in the world. Words that are classified as positive such as "largest", "fastest", and "growing" were mentioned repeatedly, leading to more tweets being assigned higher positive sentiment. This is a case of a false signal that our procedure failed to filter out.

Notwithstanding this, our X-based index is significantly correlated with the survey-based index. We regress the GFK index onto our X-based index according to Eq. (6) to assess the degree of fit of the two curves.

$$GFK_t = c + \beta \cdot Opt_t + \varepsilon_t \tag{6}$$

The R^2 value of *0.67* indicates that the X-based index explains approximately *67%* of the variance in the GFK index, while the correlation coefficient of *0.82* highlights a strong positive linear relationship between the two indices. The aim in reporting the regression results is solely to showcase their shared trend in the data only, rather than to imply causality. This suggests that the X-based optimism effectively tracks the benchmark trend. It is notable that, despite being derived from different metrics, datasets, and formats, both indices still display a meaningful level of correlation.

The originally constructed X-based index is on daily basis, yet information retrieved at this frequency from social media often contains substantial noise, such as irrelevant or repetitive news. To mitigate its observed high variance, we select the 30-day EMA index from a set of windows (e.g., 3-day, 7-day, and 15-day) as the standard for final index construction. The rationales for this decision are as follows. First, the 30-day EMA series can eliminate many fluctuations (please see Figure 2) and assign greater weight to more recent data compared to older data, thereby capturing the more immediate impact of sentiment on markets. Second, the validation through the GFK survey demonstrates the feasibility of utilizing monthly frequency windows.

2.4. Comparison with Large Language Models

To compare our approach with LLMs, we first measure three additional optimism indices using pre-trained LLMs: FinBERT, Twitter-RoBERTa, and GPT3.5. They are all built on one of two state-of-the-art models: Bidirectional Encoder Representations from Transformers (BERT) or Generative Pre-trained Transformer (GPT). We then compare our index with these in terms of their ability to track the GFK Consumer Confidence Index.

BERT was developed by Google in 2018 (Devlin et al. 2018), which has been widely used in textual analysis within economics and finance (e.g., Gorodnichenko, Pham, and Talavera 2023).

It was pre-trained on a large general-interest corpus, such as Wikipedia. To enhance the model's compatibility with our corpus, we utilize two fine-tuned BERT models. The first is FinBERT (Huang, Wang, and Yang 2023), a version of BERT fine-tuned on financial disclosure data. Its effectiveness in processing financial-related text has been well demonstrated (e.g., Curti and Kazinnik 2023; Cao et al. 2023). The second model is Twitter-RoBERTa (Barbieri et al. 2020), fine-tuned on tweets and social media data, making it well suited for our case.

GPT has been developed by OpenAI since 2018, with regular upgrades being released. In November 2022, OpenAI introduced ChatGPT, which is an interactive chatbot built on the GPT3.5 architecture. ChatGPT quickly gained popularity, attracting over 100 million monthly users within its first two months.[4] While more advanced versions, like GPT4, have since been released, GPT3.5 has proven effective in handling complex textual classification tasks (see Lopez-Lira and Tang, 2023; Hansen and Kazinnik, 2023). Most importantly, the pricing for GPT3.5 usage is nearly one-thirtieth that of GPT4.[5] When dealing with large amounts of data, as in this paper, the impact of this pricing difference is significant. Therefore, we utilize GPT3.5 to classify tweets.[6]

By separately applying pre-trained FinBERT, Twitter-RoBERTa, and GPT3.5 to our tweet set, we obtain emotion classifications for each tweet into one of three categories: positive, negative, or neutral. Thereafter, their three corresponding optimism indices, GPT_Opt_t, $FinBERT_Opt_t$, and $RoBERTa_Opt_t$, are simply calculated by the proportion of positive tweets on a daily basis, i.e., $N_t(Positive\ tweets) \ / \ N_t(tweets)$. Figure 4 illustrates the comparative performance of the three LLM metrics and our DL+Rule metric against the GFK index over time. Visually, the three LLM

[4] See https://lifearchitect.ai/%20chatgpt/

[5] See https://openai.com/api/pricing/

[6] The prompt is constructed as "You are an economics expert. Classify the sentiment for the following tweet: {tweet}. Answer in one word: positive, negative, or neutral."

metrics exhibit more volatility than our metric. To compare them statistically, we compute the correlation matrix among the metrics and then separately regress the GFK index onto each one. The results are presented in Table 3.

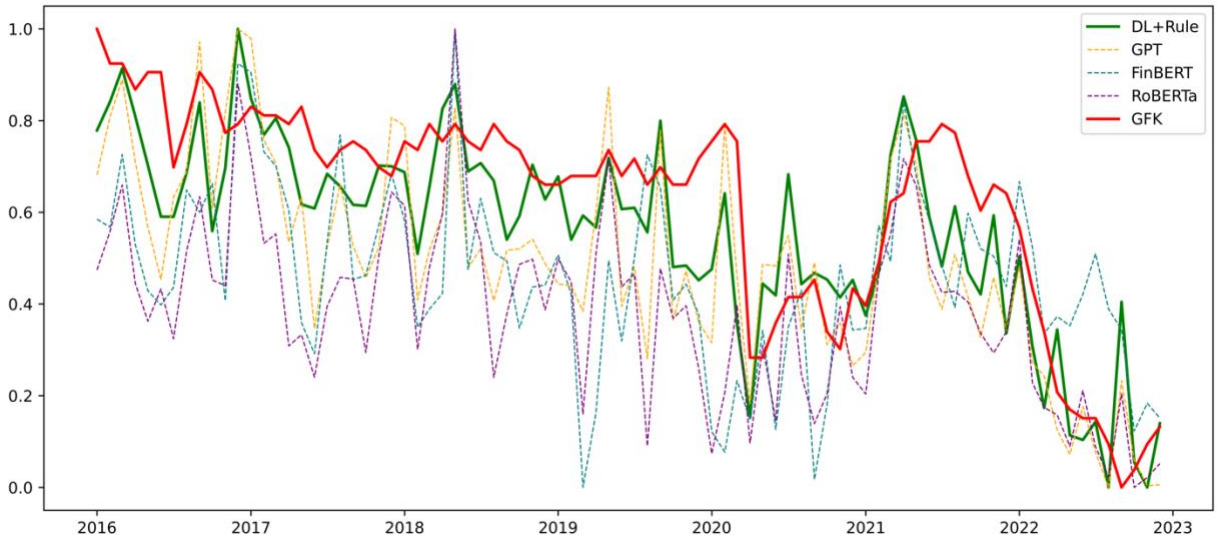


Fig. 4. Comparison among DL+Rule metric and Pre-trained LLMs

Note: This figure presents the monthly optimism indices measured using GPT, FinBERT, RoBERTa, and our DL+Rule metrics. The GFK confidence index serves as the survey-based benchmark. All indices are normalized using min-max scaling, so the displayed indices take values from 0 to 1. The data spans from January 2016 to December 2022.

Our findings suggest that the proposed DL+Rule metric generally perform better than three LLM metrics in tracking GFK index. It explains *67%* of the variance in the GFK index, which is higher than the variance explained by FinBERT (*19%*) and slightly higher than those explained by GPT (*56%*) and RoBERTa (*41%*). One might ask what would happen if we replaced word2vec with a LLM for word embedding in our DL+Rule framework. Unfortunately, this is not applicable because LLMs incorporate contextual information in word embedding, which is incompatible with the heuristic rule. To summarize, state-of-the-art LLMs like BERT and GPT leverage contextual information, allowing them to outperform context-free models like Word2Vec in tasks such as word embedding and classification. However, our results demonstrate that in specific cases,

simpler models like word2vec can perform comparably to LLMs when applied within an appropriate framework.

Table 3

Statistical Comparison

Correlation					
	DL+Rule	GPT	FinBERT	RoBERTa	GFK
DL+Rule	1				
GPT	0.89	1			
FinBERT	0.58	0.57	1		
RoBERTa	0.81	0.83	0.72	1	
GFK	**0.82**	0.75	0.44	0.64	1
Regression					
	DL+Rule	GPT	FinBERT	RoBERTa	
β	0.89	0.76	0.51	0.76	
R^2	0.67	0.56	0.19	0.41	/
Obs.	84	84	84	84	

Note: The regressions are implemented as $GFK_t = c + \beta X_t + \varepsilon_t$, where X represents measured optimism and is taken from DL+Rule, GPT, FinBERT, and RoBERTa, respectively.

3. The Relationship Between Gilt Yields and Public Optimism

With the establishment of a reliable X-based optimism index that gauges the extent to which people hold positive perceptions of UK Beechey-Wright fundamentals, we apply this index to analyse the UK government bond market. We aim to answer the questions: Whether the X-based optimism index helps to explain the bond market movements? In doing so, we concentrate on the response of the 5-year gilt market due to its data availability and its significant role in providing both liquidity and investment stability. To examine how 5-year gilt yields respond to UK public optimism, we incorporate our estimated X-based optimism index into the baseline regression model to evaluate its incremental explanatory power while controlling for a range of variables. These variables are grouped into two categories. The first consists of three common market factors: liquidity risk, global financial risk, and macroeconomic surprise. The second category represents

other optimism metrics, which include three LLM-measured indices (GPT, FinBERT, RoBERTa) and a survey-based benchmark for consumer confidence (the GFK index). All definitions are provided in Table 4.

Table 4
Variable Descriptions

y_t	5-year gilt yields.
Opt_t	The X-based optimism, measured by our DL+Rule metric and smoothed using a 30-day EMA. It quantifies public optimism regarding UK Beechey-Wright fundamentals.
GPT_Opt_t	Optimism index measured by GPT3.5, smoothed using a 30-day EMA.
$FinBERT_Opt_t$	Optimism index measured by FinBERT, smoothed using a 30-day EMA.
$RoBERTa_Opt_t$	Optimism index measured by Twitter-RoBERTa, smoothed using a 30-day EMA.
GFK_t	Survey-based UK consumer confidence index.
LR_t	The liquidity risk, derived from the bid-ask spread of the underlying gilt.
VIX_t	The global common risk, proxied by the VIX index, serves as an indicator of the market's expectations for future volatility.
$Supr_t$	The Citigroup Economic Surprise Index, available on Refinitiv Eikon database.

Initially, we explain the rationale for incorporating the controlled variables. (i) Liquidity risk is a well-accepted factor in asset pricing. Following a large body of literature (e.g., Milas, Panagiotidis, and Dergiades 2021), we exploit the bid-ask spread of the underlying gilt as a proxy for its liquidity risk. This spread represents the difference by which the ask price exceeds the bid price for the underlying gilt in the market. (ii) The VIX index—a measure of the market's expectations of future volatility—is served as the proxy for global common risk. (iii) A large body of papers highlights the significant impact of macroeconomic surprises on the government bond yields (e.g., Altavilla, Giannone, and Modugno 2017), in which the surprises are defined as the

sum of the difference between official economic results and forecasts. Because our X-based optimism is constructed based on the discussions on macroeconomic topics, it might correlate to macroeconomic surprises. To circumvent the potential endogeneity arising from omitted variables, we incorporate the Citigroup Economic Surprise Index, denoted as $Supr_t$, into the set of control variables.

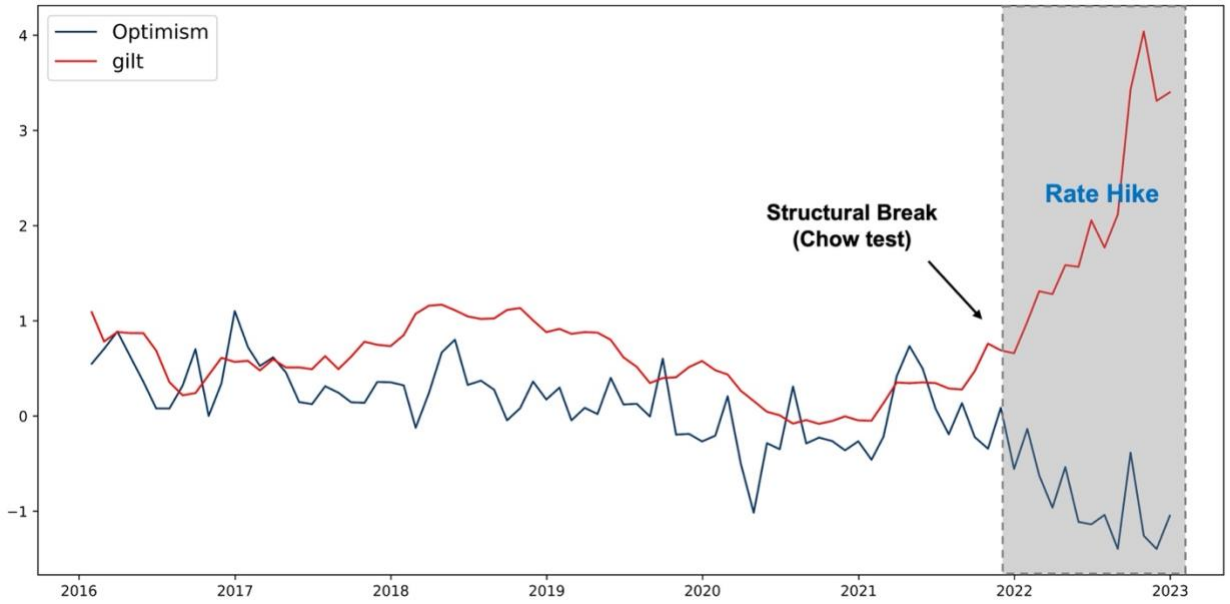


Fig.5. Optimism VS Gilt Yields

Note: This diagram displays both the X-based optimism index and the Gilt yield, using monthly data for the full sample period spanning from January 2016 to December 2022. The shaded area represents the period of rate hike in the UK, marked by the period between 2021-12-16 and 2022-12-31.

Next, we detail the time series characteristics of the data. As shown in Figure 5, there is a structural breakpoint over the entire span, where the optimism and gilt yields follow different patterns before and after this breakpoint. By conducting the Bai-Perron unknow structural changes test (Bai and Perron 1998), we identify 28 September 2021 as the potential breakpoint. Since this date, the UK has experienced a continuous 13-month rise in inflation. To curb it, the Monetary Policy Committee raised the policy rate from *0.10* percent to *0.25* percent on 16 December 2021. The rate then continued to rise linearly, reaching *5.25* percent by September 2023. This was a

significant move because the policy rate had been maintained at 0.10 percent for the previous 21 months. We eventually select the rate hike date—16 December 2021—as the breakpoint for three reasons. It is convincing to use the starting date of continued interest rate hikes as a breakpoint, given the direct significant impact of policy rate changes on the bond yields has been well documented (e.g., Piazzesi 2005). The distance between 28 September and 16 December in 2021 appears small on the entire 7-year timeline. It passes the Chow test at the 0.01 level.

The gilt yields experienced significant upward pressure due to the interest rate hikes. Using the Augmented Dickey-Fuller (ADF) unit root test, we find that: (i) the gilt series is stationary before the rate hikes, (ii) after the rate hikes, it shows trend stationarity (i.e., a stationary series around a deterministic linear trend) at the 0.1 level. To deal with structural breaks, many empirical papers utilize the piecewise regression to capture the changes in the regressors' coefficients within a fixed specification (e.g., D'Amico and King 2013). However, we encounter circumstances that differ from theirs. In this paper, the identified structural breaks not only reflect changes in coefficients but also modifications in the model's specification. Specifically, to remove the effect of rate hikes on the corresponding segment, it requires to include a linear time trend into the specification, which leads to a structural change in the regression specification. Therefore, we do not apply the same specification to the entire series to study the relationship between gilt and the X-based optimism. Instead, we are inclined to separately investigate that relationship before and after the structural breakpoint (initial hike date, 16 December 2021) based on the two following panels, "Panel A: pre-rate hike" and "Panel B: rate hike".

3.1. Panel A: Pre-rate Hike Segment

During the pre-rate hike period, Opt_t and y_t exhibit stationarity. Notably, y_t is approximately stationary, with a p-value of *0.12* for ADF test. We construct a baseline regression specification as Eq. (7) to study the effect of the X-based optimism on gilts yield.

$$y_t = c + \mu \cdot Opt_t + \sum_{i=1}^{3} \beta_i \cdot \boldsymbol{Factors}_{i,t} + \varepsilon_t \tag{7}$$

where ***Factors*** is a set of controlled variables including liquidity risk (LR_t), global financial risk (VIX_t), and macroeconomic surprise ($Supr_t$); μ measures the sensitivity of the gilt yields to people's net optimism.

Firstly, to assess the improvement in the explanatory power of the existing model due to the additional contribution of Opt_t, we conduct three sets of regressions: (i) A univariate regression of y_t on Opt_t. (ii) A regression of y_t on a set of control variables. (iii) The regression of y_t on the full set of regressors, including Opt_t, LR_t, VIX_t, and $Supr_t$. Table 5 presents the regression results. The coefficients of Opt_t in column [1] and [6] are significantly positive, indicating that the higher Opt_t correlates to the higher yields. This aligns with Qiu and Welch's (2004) argument that investors tend to be bullish about the financial markets when they are optimistic about the economy, and vice versa. Baker and Wurgler (2012) further support this by suggesting that high optimism may be indicated by periods of high demand for speculative stocks relative to demand for bond-like securities. Because mid-term government bonds serve as safe-haven assets, when investors are more optimistic about the economy (i.e., higher Opt_t), capital is more likely to shift from haven to the stock market or other financial sectors in pursuit of higher returns. This shift of capital can exert downward pressure on gilt prices, subsequently driving an increase in yield. In addition, by comparing columns [5] and [6], we demonstrate that adding our optimism index to the set of

control variables results in a significant increase in the adjusted R^2 from *0.318* to *0.415*. This indicates that our optimism index contains useful information for predicting movements in gilt yields. Furthermore, the significance and signs of the risk factors' coefficients remain unchanged. It indicates that the X-based sentiment is a valuable supplement to the existing model.

Table 5

Panel A: Regressions Results for the *Pre-Rate Hike* Based on Original Series

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Cons.	-0.139***	0.267***	0.327***	0.186***	0.991***	0.382***	0.727***	0.858***	0.681***
LR_t					1.483***	1.448***	1.454***	1.367***	1.268***
VIX_t					-1.186***	-0.637***	-0.969***	-0.993***	-0.870***
$Supr_t$					-0.778***	-0.863***	-0.889***	-0.823***	-0.838***
Opt_t	1.101***					0.923***			
GPT_Opt_t		0.523***					0.523***		
$FinBERT_Opt_t$			0.491***					0.290***	
$RoBERTa_Opt_t$				0.866***					0.720***
N	1482	1482	1482	1482	1482	1482	1482	1482	1482
Adj.R²	**0.181**	0.053	0.055	0.149	0.318	**0.415**	0.365	0.333	0.412

Note: This table reports the regression results of Eq. (7). *** p < 0.01, ** p < 0.05, * p < 0.1. The statistical significance of the coefficients reported in this table is based on heteroskedasticity and autocorrelation consistent (HAC) standard errors.

Secondly, to compare the performance of the optimism index across different metrics, we repeat the regression analysis using optimism generated by GTP3.5, FinBERT, and RoBERTa. By comparing columns [1], [2], [3], and [4] in Table 5, it is evident that the optimism index constructed using our DL+Rule metric provides higher explanatory power for gilt movement than the metrics based on the three LLMs. Furthermore, when comparing columns [6], [7], [8] and [9], our index consistently outperforms the others by increasing the model's adjusted R^2 by an additional *0.097*. This result highlights the potential contribution of our index to the model's overall explanatory power, even when considering other sentiment measures.

Table 6

Panel A: Regressions Results for the *Pre-Rate Hike* Based on Difference Series

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Cons.	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002
ΔLR_t					-0.0653***	-0.0666***	-0.0670***	-0.0642***	-0.0642***
ΔVIX_t					-0.1486***	-0.1457***	-0.1471***	-0.1482***	-0.1482***
ΔOpt_t	0.0308*					0.0277*			
ΔGPT_Opt_t		0.0371**					0.0367**		
$\Delta FinBERT_Opt_t$			0.0218					0.0192	
$\Delta RoBERTa_Opt_t$				0.0254*					0.0192
N	1481	1481	1481	1481	1481	1481	1481	1481	1481
Adj.R²	0.002	0.002	0.000	0.001	0.016	0.017	0.020	0.016	0.017

Note: This table reports the regression results of Eq. (7). *** p < 0.01, ** p < 0.05, * p < 0.1. For the daily frequency series processed with a 30-day EMA, applying a first-order difference would result in a substantial loss of volatility information. Thus, optimism indices are smoothened using 10-day EMA.

Thirdly, we also examine the relationship between changes in optimism and bond yields using their first-order differenced series. Given that the differenced series for macroeconomic surprises ($Supr_t$) is not statistically significant, we have excluded it from the set of control variables. The regression results are presented in Table 6. By comparing columns [1], [2], [3], and [4], we observe that all optimism indices, except for the differenced FinBERT optimism index, are statistically significant in explaining the changes in yields. The adjusted R^2 values remain quite low, which may be attributed to the daily frequency of the series, as it likely contains considerable noise. After incorporating changes in liquidity risk and global financial risk into the set of control variables, we find that only our DL+Rule optimism and GPT optimism remain statistically significant.

3.2. Panel B: Rate Hike Segment

During the rate hike period, while Opt_t exhibits stationarily, the gilt yields are recognized as a trend stationary series. Therefore, regression analysis on the original series is meaningless unless

the trend component is accounted for. Using differencing transformation to obtain a differenced series is a solution, but this would result in a significant loss of information. Alternatively, detrending (e.g., taking out the linear trend in original series) is a better approach for dealing with trend stationary series. For example, as early as forty years ago, Nelson and Plosser (1982) pointed out that many macroeconomic variables can be decomposed into a cyclical component and a growth component. Their opinion is still widely applied in recent empirical research (e.g., Ilut, Kehrig, and Schneider 2018; Busch et al. 2022; and Ferrante, Graves, and Iacoviello 2023). We follow those papers to apply detrending method for the following reasons. (i) Detrending allows the series-after-detrended to maintain the same dimensions as the original series. (ii) Gilts yields should not inherently exhibit a deterministic trend. The reason why they have developed an upward linear trend is due to the BoE's continuous and linear interest rate hikes. If the drivers of the policy rates have been specified, deviation of bond yields from their estimated trend should be stationary (Favero, Melone, and Tamoni 2023), which helps to mitigate the risk of spurious regression.

We implement Eq. (9) to isolate the cyclical (stationary) component and remove the deterministic growth component that is driven by monetary policy rates.

$$y_t = c + \delta \cdot trend + e_t \tag{8}$$

$$y_t^{DT} = c + \mu \cdot Opt_t + \sum_{i=1}^{3} \beta_i \cdot Factors_{i,t} + \varepsilon_t \tag{9}$$

where $trend$ is a linear sequence ($1, 2, \ldots, T$). The residuals series $\widehat{e_t}$ generated by Eq. (8) is served as the proxy of the detrended yields, denoted as y_t^{DT}, in Eq. (9).

After obtaining y_t^{DT}, we conduct the regression analysis on the Panel B as we do on Panel A. Table 7 presents the results, with some findings aligning with Table 5, which can be interpreted using the same aforementioned knowledge. We directly state the corresponding findings to avoid

redundancy. First, all the coefficients of Opt_t are significantly positive, indicating that our optimism index highly positively correlates to gilt yields, and it is a good predictor of gilt yields during the rate hike period. Second, our optimism index adds values to the baseline model by enhancing the model's explanatory power (R^2 increases by ranging from *0.257* to *0.441*). Furthermore, it consistently outperforms LLM-based indices, both when used in isolation and when combined with other common factors.

Table 7

Panel B: Regressions Results for the *Rate Hike* Based on Original Series

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Cons.	-0.221***	-0.163***	0.159*	-0.035	-1.256***	-1.517***	-1.536***	-1.643***	-1.554***
LR_t					1.044***	1.841***	1.769***	1.453***	1.706***
VIX_t					2.971***	2.643***	2.660***	2.968***	2.864***
$Supr_t$					0.844***	0.327	0.516**	0.977***	0.762***
Opt_t	0.855***					1.724***			
GPT_Opt_t		0.746***					1.747***		
$FinBERT_Opt_t$			-0.418**					0.655***	
$RoBERTa_Opt_t$				0.188					1.303***
N	264	264	264	264	264	264	264	264	264
Adj.R²	**0.061**	0.030	0.012	-0.001	0.257	**0.441**	0.393	0.277	0.337

Note: This table reports the regression results of Eq. (7). *** p < 0.01, ** p < 0.05, * p < 0.1. The statistical significance of the coefficients reported in this table is based on heteroskedasticity and autocorrelation consistent (HAC) standard errors.

Additionally, it is important to note that the meaning of intercept in Table 7 changes (comparing with that in Table 5) once the linear trend is removed from the original gilt yields. In this case, the intercept represents the expected value of the detrended gilt yields (y_t^{DT}) when all the regressors are set to zero, not the expected value of the gilt yields (y_t). The signs of intercepts change when comparing columns [1] versus [2] and [4] versus [5]. This change does not have significant economic implications; it results merely from data processing. To be specific, y_t^{DT} is defined as y_t minus a fitted straight line, which may lead to negative initial values ($y_{t=0}^{DT}$).

Table 8

Panel B: Regressions Results for the *Rate Hike* Based on Difference Series

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Cons.	0.011*	0.011*	0.011*	0.011*	0.011*	0.011*	0.011*	0.011*	0.011*
ΔLR_t					0.193**	0.199**	0.193**	0.195**	0.200**
ΔVIX_t					0.560**	0.558**	0.560**	0.558**	0.550**
ΔOpt_t	**-0.186****					**-0.192****			
ΔGPT_Opt_t		0.022					0.014		
$\Delta FinBERT_Opt_t$			-0.033					-0.045	
$\Delta RoBERTa_Opt_t$				-0.063					-0.077
N	264	264	264	264	264	264	264	264	264
Adj.R²	**0.012**	-0.004	-0.004	-0.003	0.026	**0.039**	0.022	0.023	0.024

Note: This table reports the regression results of Eq. (7). *** p < 0.01, ** p < 0.05, * p < 0.1. For the daily frequency series processed with a 30-day EMA, applying a first-order difference would result in a substantial loss of volatility information. Thus, optimism indices are smoothened using 10-day EMA.

Similarly, we also further investigate the relationship between changes in optimism and changes in bond yields by applying first-order differencing to all variables. Given that the differenced series for macroeconomic surprises ($Supr_t$) is not statistically significant, we have excluded it from the set of control variables. The results are presented in Table 8. Comparing columns [1], [2], [3], and [4], we observe that only our developed optimism index is significant in explaining the movement in Δy, while other LLM-based indices fail to reach significance. After controlling for changes in liquidity risk, global financial risk, and macroeconomic surprises, our DL+Rule optimism index remains statistically significant, indicating its robustness in explaining bond yield movements. The other optimism indices based on GPT, FinBERT, and RoBERTa, however, do not exhibit significant explanatory power in this context.

4. Robustness Checks

The X-based optimism index is an aggregate constructed from all tweets containing at least one of the selected keywords, resulting in a tweet set that includes some irrelevant content. To examine

whether our modelling results and empirical results are sensitive to the granular content of tweets, we conduct robustness checks using topic analysis using latent Dirichlet allocation (LDA) (Blei, Ng, and Jordan 2003). LDA is a generative probabilistic model for topic classification. It can be used to summarize the topic distribution across the given corpus, which is widely used in textual analysis studies including economics and finance (e.g., Li et al. 2021; Angelico et al. 2022).

We first incorporate LDA into the data preparation phase of optimism measurement, enabling to split the entire tweet set into different subsets in terms of their topics obtained by LDA. Specifically, LDA requires an ex-ante determination of number of topics. Considering that the content of these tweets was already very precise at the time of collection, we opt for 5 topics rather than more. Consequently, the entire tweet set is spited into 5 subsets. Table 9 reports the list of topics, their keywords of content, and the sizes of their subsets. Each topic is assigned a name based on the content of the tweets. We exclude topic 3 (foreign economy) and topic 5 (political economy) from the robustness checks. Topic 3 is omitted due to its relatively higher noise, which could introduce unnecessary variability into the analysis. Similarly, topic 5 is excluded because it contains a substantial amount of political insights, which may not directly align with the economic focus of the study. In this way, our smaller set includes topic 1 (economic growth), topic 2 (cost of living), and topic 4 (inflation), which collectively accounts for about *37%* of the total sample. This ensures that the analysis is not only focused on key economic themes but also assesses the stability of the findings across a smaller, more concentrated dataset.

In doing so, we first examine the robustness of measurement by regressing the GfK benchmark index onto Opt_t. We find that it can explain the *40%* variance of GFK index, and the correlation coefficient of them is *0.63*. It indicates that the proposed novel measurement of

optimism is not sensitive to more nuanced topics, as long as the text content is limited to macroeconomic fundaments.

Table 9

Topic Classification

Topics	(1) Economic growth	(2) Cost of living	(3) Foreign economy	(4) Inflation	(5) Political economy
Keywords	rate	economi	germani	inflat	economi
	year	bn	russia	price	brexit
	gdp	govern	itali	bank	tori
	high	cost	turkey	sunak	eu
	growth	loss	argentina	octob	whi
	sinc	make	gdp	england	trade
	show	new	economi	expect	would
	low	energi	area	rise	onli
	fall	food	euro	first	bad
	live	busi	world	want	countri
Volume	675,651	889,444	459,052	543,365	3,101,818

Note: This table displays the corresponding top 10 representative words for topics obtained by LDA. These representative words are shown in their stemmed forms. For instance, [economi] is the stemmed form of [economy]. Each topic is assigned a name based on the content of the tweets.

Second, we examine the robustness of the empirical results by re-implementing the same empirical methods as those applied in Section 3. The main empirical results are presented in Table 10. The coefficients of the variables in the robustness checks remain consistent with those found in the main empirical results. This suggests that the relationships identified in the primary analysis hold steady, even when applied to a more refined set of topics and a smaller sample size. The stability of the coefficients across both the main empirical analysis and the robustness checks indicates the reliability of our findings.

Additionally, the analysis based on the rate hike in Section 3.2 can essentially be treated as a form of robustness check. Although that is primarily presented in Section 3.2 rather than as part of the formal robustness checks, the coefficient of our optimism index remains consistent across

both periods. This consistency indicates that the positive relationship between gilt yields and optimism is robust, holding true during both the normal period and the continuous rate hike period.

Table 10

Robustness Checks

	Pre-Rate Hike			Rate Hike		
	[1]	[2]	[3]	[4]	[5]	[6]
	y_t	y_t	y_t	y_t^{DT}	y_t^{DT}	y_t^{DT}
Cons.	0.311***	0.982***	0.633***	-0.081	-0.781***	-1.221***
LR_t		1.565**	1.799***		1.073***	1.351***
VIX_t		-1.211***	-1.268***		2.673**	3.057***
$Supr_t$		-0.764***	-0.758***			
Opt_t	0.356**		0.498***	0.374		1.402***
$\boldsymbol{N}$	1472	1472	1472	264	264	264
$\boldsymbol{R^2}$	0.014	0.320	0.348	0.002	0.234	0.301

Note: *** p < 0.01, ** p < 0.05, * p < 0.1. The statistical significance of the coefficients reported in this table is based on heteroskedasticity and autocorrelation consistent (HAC) standard errors.

5. Conclusion

Given the constraints of discrete macroeconomic announcements, we propose a novel "Deep Learning + Heuristic Rules" framework to extract public optimism towards the Beechey-Wright fundamentals news from the stream of tweets on X. This allows us to study government bond market fluctuations by fundamentals news over a continuous time horizon.

The proposed framework may offer several potential advantages. It is a cost-effective approach that does not require labelled texts, which saves substantial labor-related costs. In contrast to other textual approaches, it gives more consideration to context and can be applied across disciplines. Moreover, it is timely, avoiding the publication lag common to survey-based indicators. Finally, it is free and requires less computational power, yet its performance has been shown to be comparable to state-of-the-art LLMs.

Drawing on 5.7 million tweets, we utilize this framework to compute a X-based optimism indicator and shows that it is a reliable proxy for public optimism. It tracks well with a known survey-based benchmark, the GFK confidence index. Despite these two indices are derived from different data types and methods, they exhibit a strong correlation of *0.82*, surpassing the correlations of this survey with three LLM-based indices: GPT3.5 (*0.75*), FinBERT (*0.44*), and Twitter-RoBERTa (*0.64*).

By applying the optimism index to study its impact on the UK government bond market, we find that it is a good predictor of the gilt yields and shows a high positive correlation with gilt yields in the pre-rate hike period. During the rate hike period, this correlation remains evident after controlling for the linear trend in yields due to rate hikes. It indicates that DL+Rule framework for optimism measurement and empirical findings remain robust when controlling for sub-samples across both pre-rate hike and rate hike periods. These findings are consistent with existing studies: investors tend to be bullish about the financial markets when they are optimistic about the economy (Qiu and Welch 2004), and this can lead to low demand for bonds relative to speculative stocks (Baker and Wurgler 2012). In addition, this optimism index is a supplement to the baseline model. It enhances its explanatory ability on the variance of gilt yields by *0.097* during the pre-rate hike period and *0.184* during the rate hike period, outperforming the three applied LLMs. By controlling the sample size and contents of tweets, we demonstrate that the modelling results and empirical results are robust as long as the content of tweets pertains to Beechey-Wright fundamentals.

Drawing upon these finding, this paper offers insights for policymakers, emphasizing the importance of considering the impact of optimism on the bond market when conducting monetary policy. In addition, this paper suggests an indicator that can gauge optimism even during non-announcement periods. Finally, in this paper, there are some limitations that require further

refinement in future research endeavors. For example, with regards to one part of measuring sentiment, we employ the heuristic rules to capture the contextual characteristics via punctuations, negations, capitalizations, and degree modifiers. However, given the irregularity of social texts and the diversity of contexts, rule-based algorithms may not perfectly capture all contextual characteristics. We leave this for future studies.

Glossary

Glossary	
Terms	Definitions
X	Twitter
Deep Learning	A type of artificial intelligence that uses layered neural networks to learn patterns from large amounts of data.
Heuristic Rules	A set of rules proposed by Hutton and Gilbert (2014) to capture context information.
LLMs	Large Language Models
GPT	Generative Pre-trained Transformer. An advanced AI language model developed by OpenAI.
BERT	Bidirectional Encoder Representations from Transformers. A large language model developed by Google that understands the context of words in text by looking at both the words before and after each term (bidirectional context).
Cosine Similarity	A measure of similarity between two non-zero vectors that calculates the cosine of the angle between them.
Word Embedding	A technique in natural language processing where words are represented as distributed vectors in a continuous space, capturing semantic relationships based on context.
Distributed Word Vector	A dense, continuous vector representation of a word, capturing its meaning based on given corpus.

References

Adam, K., D. Matveev, and S. Nagel. 2021. Do survey expectations of stock returns reflect risk adjustments?. *Journal of Monetary Economics* 117: 723–740.

Altavilla, C., D. Giannone, and M. Modugno. 2017. Low frequency effects of macroeconomic news on government bond yields. *Journal of Monetary Economics* 92: 31–46.

Angelico, C., J. Marcucci, M. Miccoli, and F. Quarta. 2022. Can we measure inflation expectations using Twitter?. *Journal of Econometrics* 228: 259–277.

Bai, J., and P. Perron. 1998. Estimating and testing linear models with multiple structural changes. *Econometrica*: 47–78.

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

Baker, M., and J. Wurgler. 2012. Comovement and predictability relationships between bonds and the cross-section of stocks. *Review of Asset Pricing Studies* 2: 57–87.

Balduzzi, P., E. J. Elton, and T. C. Green. 2001. Economic news and bond prices: Evidence from the us treasury market. *Journal of Financial and Quantitative analysis* 36: 523–543.

Barbieri, F., J. Camacho-Collados, L. Neves, and L. Espinosa-Anke. 2020. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. *arXiv preprint arXiv*:2010.12421.

Bartov, E., L. Faurel, and P. S. Mohanram. 2018. Can Twitter help predict firm-level earnings and stock returns?. *Accounting Review* 93: 25–57.

Bartov, E., L. Faurel, and P. Mohanram. 2023. The role of social media in the corporate bond market: evidence from Twitter. *Management Science* 69: 5638–5667.

Beechey, M. J., and J. H. Wright. 2009. The high-frequency impact of news on long-term yields and forward rates: Is it real?. *Journal of Monetary Economics* 56: 535–544.

Bennani, H. 2020. Central bank communication in the media and investor sentiment. *Journal of Economic Behavior and Organization 176*: 431–444.

Berardi, M. 2022. Uncertainty and sentiments in asset prices. *Journal of Economic Behavior and Organization 202*: 498–516.

Bergman, N. K., and S. Roychowdhury. 2008. Investor sentiment and corporate disclosure. *Journal of Accounting Research* 46: 1057–1083.

Bianchi, F., R. Gomez-Cram, and H. Kung. 2024. Using social media to identify the effects of congressional viewpoints on asset prices. *Review of Financial Studies*, forthcoming.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022.

Bollen, J., H. Mao, and X. Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2: 1–8.

Busch, C., D. Domeij, F. Guvenen, and R. Madera. 2022. Skewed idiosyncratic income risk over the business cycle: Sources and insurance. *American Economic Journal: Macroeconomics* 14: 207–242.

Cao, S., W. Jiang, B. Yang, and A. L. Zhang. 2023. How to talk when a machine is listening: Corporate disclosure in the age of AI. *Review of Financial Studies* 36: 3603–3642.

Chen, Y., B. T. Kelly, and D. Xiu. 2023. Expected returns and large language models. *Working Paper*.

Chen, Z., D. Lien, and Y. Lin. 2021. Sentiment: The bridge between financial markets and macroeconomy. *Journal of Economic Behavior and Organization* 188: 1177–1190.

Chow, G. C. 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica*: 591–605.

Consoli, S., L. T. Pezzoli, and E. Tosetti. 2021. Emotions in macroeconomic news and their impact on the European bond market. *Journal of International Money and Finance* 118: 102472.

Cookson, J. A., and M. Niessner. 2020. Why don't we agree? Evidence from a social network of investors. *Journal of Finance* 75: 173–228.

Curti, F., and S. Kazinnik. 2023. Let's face it: Quantifying the impact of nonverbal communication in FOMC press conferences. *Journal of Monetary Economics* 139: 110–126.

D'Amico, S., and T. B. King. 2013. Flow and stock effects of large-scale treasury purchases: Evidence on the importance of local supply. *Journal of Financial Economics* 108: 425–448.

Da, Z., J. Engelberg, and P. Gao. 2015. The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies* 28: 1–32.

Dergiades, T., C. Milas, and T. Panagiotidis. 2015. Tweets, Google trends, and sovereign spreads in the GIIPS. *Oxford Economic Papers* 67: 406–432.

Devlin, J., M. W. Chang, K. Lee, and K. Toutanova. 2018. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv*: 1810.04805.

Edmans, A., A. Fernandez-Perez, A. Garel, and I. Indriawan. 2022. Music sentiment and stock returns around the world. *Journal of Financial Economics* 145: 234–254.

Favero, C., A. Melone, and A. Tamoni. 2022. Monetary policy and bond prices with drifting equilibrium rates. *Journal of Financial and Quantitative Analysis* 59: 626–651.

Ferrante, F., S. Graves, and M. Iacoviello. 2023. The inflationary effects of sectoral reallocation. *Journal of Monetary Economics* 140: 64–81.

Gao, Z., H. Ren, and B. Zhang. 2020. Googling investor sentiment around the world. *Journal of Financial and Quantitative Analysis* 55: 549–580.

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

Gilbert, T., C. Scotti, G. Strasser, and C. Vega. 2017. Is the intrinsic value of a macroeconomic news announcement related to its asset price impact?. *Journal of Monetary Economics* 92: 78–95.

Gorodnichenko, Y., T. Pham, and O. Talavera. 2023. The voice of monetary policy. *American Economic Review* 113: 548–584.

Greenwood, R., and A. Shleifer. 2014. Expectations of returns and expected returns. *Review of Financial Studies* 27: 714–746.

Gürkaynak, R., B. Kısacıkoğlu, and J. H. Wright. 2020. Missing events in event studies: Identifying the effects of partially measured news surprises. *American Economic Review* 110: 3871–3912.

Hansen, A. L., and S. Kazinnik. 2023. Can ChatGPT decipher Fedspeak? *Working Paper*.

Hirshleifer, D., D. Jiang, and Y. M. DiGiovanni. 2020. Mood beta and seasonalities in stock returns. *Journal of Financial Economics* 137: 272–295.

Huang, A. H., H. Wang, and Y. Yang. 2023. FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research* 40: 806–841.

Huang, D., F. Jiang, J. Tu, and G. Zhou. 2015. Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies* 28: 791–837.

Hutto, C., and E. Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media* 8: 216–225.

Ilut, C., M. Kehrig, and M. Schneider. 2018. Slow to hire, quick to fire: employment dynamics with asymmetric responses to news. *Journal of Political Economy* 126: 2011–2071.

Jiang, F., J. Lee, X. Martin, and G. Zhou. 2019. Manager sentiment and stock returns. *Journal of Financial Economics* 132: 126–149.

Kaplanski, G., and H. Levy. 2010. Sentiment and stock prices: The case of aviation disasters. *Journal of Financial Economics* 95: 174–201.

Kostopoulos, D., and S. Meyer. 2018. Disentangling investor sentiment: Mood and household attitudes towards the economy. *Journal of Economic Behavior and Organization* 155: 28–78.

Li, K., F. Mai, R. Shen, and X. Yan. 2021. Measuring corporate culture using machine learning. *Review of Financial Studies* 34: 3265–3315.

Lopez-Lira, A., and Y. Tang. 2023. Can ChatGPT forecast stock price movements? Return predictability and large language models. *Working Paper*.

Loughran, T., and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66: 35–65.

Ludvigson, S. C., and S. Ng. 2009. Macro factors in bond risk premia. *Review of Financial Studies* 22: 5027–5067.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems* 2: 3111–3119.

Milas, C., T. Panagiotidis, and T. Dergiades. 2021. Does it matter where you search? Twitter versus traditional news media. *Journal of Money, Credit and Banking* 53: 1757–1795.

Nelson, C. R., and C. R. Plosser. 1982. Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics* 10: 139–162.

Nofsinger, J. R. 2005. Social mood and financial economics. *Journal of Behavioral Finance* 6: 144–160.

Piazzesi, M. 2005. Bond yields and the federal reserve. *Journal of Political Economy* 113: 311–344.

Qiu, L., and I. Welch. 2004. Investor sentiment measures. *NBER Working Paper*: 10794.

Shapiro, A., M. Sudhof, and D. J. Wilson. 2022. Measuring news sentiment. *Journal of Econometrics* 228: 221–243.

Stambaugh, R., J. Yu, and Y. Yuan. 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104: 288–302.

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

Xu, Y., and J. Zhao. 2022. Can sentiments on macroeconomic news explain stock returns evidence form social network data. *International Journal of Finance and Economics* 27: 2073–2088.