EconomiX

Capturing international influences in U.S. monetary policy through a NLP approach

Laurent Ferrara
Nicolas de Roux

2025-23 Document de Travail/ Working Paper

EconomiX - UMR 7235 Bâtiment Maurice Allais
Université Paris Nanterre 200, Avenue de la République
92001 Nanterre Cedex

Site Web : economix.fr
Contact : secreteriat@economix.fr
Twitter : @EconomixU

cnrs Université Paris Nanterre

Capturing international influences in U.S. monetary policy through a NLP approach

Laurent Ferrara

Skema Business School - University Côte d'Azur
laurent.ferrara@skema.edu

Nicolas de Roux *

QuantCube Technology & University Paris Nanterre - EconomiX
n.deroux@quant-cube.com

April 3, 2025

Abstract

Officially, the U.S. Federal Reserve has a statutory dual domestic mandate of price stability and full employment, but, in this paper, we question the role of the international environment in shaping Fed monetary policy decisions. In this respect, we use minutes of the Federal Open Market Committee (FOMC) and construct indexes of the attention paid by U.S. monetary policymakers to the international economic and financial situation. These indexes are built by applying natural language processing (NLP) techniques ranging from word count to built-from-scratch machine learning models, to OpenAI's GPT models. By integrating those text-based indicators into a Taylor rule, we derive various quantitative measures of the external influences on Fed decisions. Our results show that when there is a focus on international topics within the FOMC, the Fed's monetary policy generally tends to be more accommodative than expected by a standard Taylor rule. This result is robust to various alternatives that includes a time-varying neutral interest rate or a shadow central bank interest rate.

Keywords : Monetary policy, Federal Reserve, FOMC minutes, International environment, Natural Language Processing, Machine Learning

*We would like to thank Luca Barbaglia, Benoit Bellone, Oussama Belmejdoub, Menzie Chinn, Thomas Chuffart, Maxence David, Thomas Drechsel and Leif Anders Thorsrud for useful comments, as well as seminar participants at EconomiX - Université Paris Nanterre and researchers at QuantCube Technology for useful discussions.

1. Introduction

To manage monetary policy in the United States, the Federal Reserve (the Fed hereafter) has a dual statutory mandate to fulfill of promoting both maximum employment and stable prices, as well as moderate long-term interest rates. Clearly, those objectives are domestic ones and do not account, at least directly, for the international environment. In addition to this domestic focus, the U.S. economy is the archetype of a large, closed economy, suggesting thus that the Fed is generally considered to be largely indifferent to what's going on in the rest of the world. For example, in an influential paper, Barry Eichengreen was asking in 2013: *Does the Federal Reserve care about the Rest of the World?* (Eichengreen, 2013). However, having a first sight at the meeting minutes of the Federal Open Market Committee (FOMC, the committee in charge of the monetary policy within the Fed) shows that the international situation is often discussed among members during meetings, sometimes at length. Indeed, international factors are likely to play a role in the stability of the U.S. economy and financial system, through various mechanisms. The main channels of transmission are (i) the global integration of financial markets (which influences domestic market conditions), (ii) the role of international prices on domestic inflation (*imported inflation*, see Ha et al., 2023) and (iii) the potential spillbacks on the U.S. economy of U.S. monetary policy's influence on foreign economies (Obstfeld, 2020). Adverse economic conditions that could affect the U.S. economy can therefore lead the FOMC to raise or to cut the Federal Funds Rates (*FFR*), the main conventional tool of monetary policy (Ihrig and Wolla, 2020).

The influence of the international environment on U.S. monetary policy is sometimes explicitly evoked by policymakers themselves. In the midst of the global financial turbulence in 1998, marked by the Russian Federation's default on its debt, Fed Chairman Alan Greenspan remarked in a September 4th speech in Berkeley that *"it is just not credible that the United States can remain an oasis of prosperity unaffected by a world that is experiencing greatly increased stress"* (Greenspan, 1998). This take was widely seen by market participants as an early warning of the FFR cut decided later that month. Another example of this supposed attention to the international environment took place in August 2015, at a time when the Fed was expected to start hiking rates from the zero lower bound, where they had been stuck at for several years. At the same time, concerns over the rebalancing of China's growth model and the possible manipulation of the renminbi made international headlines. Vice Chairman Stanley Fischer declared in a Jackson Hole speech that the FOMC had to *"consider [...] the influence of foreign economies on the U.S. economy as [they] reach [their] judgment on whether and how to change monetary policy"*, explicitly referring to the Chinese slowdown in the same speech (Fischer, 2015). Some economic commentators interpreted this concern as the driving force behind the decision to shift the first rate hike since the global financial crisis to December 2015. Starting from this anecdotal evidence, our research question is to quantitatively assess the possible effects of the global economy on U.S. monetary policy decisions. Our empirical strategy relies on exploiting the textual information contained in the minutes of the FOMC, published after each meeting of the committee. In this respect, we construct various indicators of attention to the international environment with the help of Natural Language Processing methodologies. Then, we integrate those indicators into extended econometric equations building on the Taylor rule as initially put forward by John Taylor (Taylor, 1993, and Taylor, 1999).

Our work contributes to two major strands of the economic literature (see the literature review in the next section for more details). The first one considers the role of the U.S. Fed within the global economy. Generally, authors look at the spillovers of Fed monetary policy decisions to the rest of the world (see among others Dedola et al., 2017 or Miranda-Agrippino and Rey, 2020),

and to emerging countries more specifically. This literature has been highlighted by the famous words of Dilma Rousseff in 2012, then-President of Brazil, referring to the ultra-accommodative U.S. monetary policy as a driver of the *monetary tsunami* that was overwhelming emerging markets with large capital flows. However, spillbacks from the rest of the world to the U.S. monetary policy are less a concern in the literature (see however Breitenlechner et al., 2022). The other strand of the literature is the quantitative analysis of central bank speeches and especially Federal Reserve communication. This literature starts with the influential work of Romer and Romer, 2004, who dig into *Greenbook* documents prepared by Fed economists in order to estimate monetary policy shocks as the differences between FFR and forecasts of inflation, output and unemployment rate. With the development of Machine Learning tools and Natural Language Processing (*NLP*) techniques, authors have developed approaches to extract more efficiently textual information. For example, Aruoba and Drechsel, 2022 have recently proposed new NLP methods to estimate U.S. monetary policy shocks. Our work fits into both fields by using NLP approaches to compute from official Fed documents an index summarizing the attention of FOMC members to international economic conditions. In this respect, we extend and assess econometrically the idea put forward in the blog post by Ferrara and Teuf, 2018, who are simply counting some international words contained in FOMC minutes.

The rest of this work is structured as follows. Section 2 briefly reviews the literature related to our topics of interest. Section 3 presents the data and the methods we consider to create our international environment indices and discusses our results in light of major international events. Section 4 presents the framework under which we will consider our index to assess its significance on the conduct of U.S. Monetary Policy. Section 5 concludes.

2. Literature review

2.1. U.S. monetary policy and the rest of the world

The situation of the U.S. central bank as a pivot of the world economy has been well established in the literature. Relationships between the Fed and the rest of the world go in both ways: from the U.S. monetary policy to the rest of the world (*spillovers*) and from the rest of the world to the U.S. monetary policy (*spillbacks*). However, the literature appears somewhat asymmetric as the first direction has been by far the most studied. Indeed, most of the time, authors look at the spillovers of Fed monetary policy decisions to the rest of the world, and to emerging countries more specifically. Without being exhaustive, the works of Bruno and Shin, 2015, Dedola et al., 2017, or Miranda-Agrippino and Rey, 2020, provide a good overview of the concept of global financial cycle and of the risk-taking channel of monetary policy, while Degasperi et al., 2021, specifically study the effects of U.S. monetary policy tightening on advanced and emerging economies.

However, spillbacks from the rest of the world to the U.S. monetary policy are less a concern in the literature. Eichengreen, 2013, offers a very thorough history of the attention paid by the Federal Reserve to international considerations in its century of existence. He widely describes periods of lower attention and renewed interest in the international environment by U.S. monetary policy-makers. In the first half of the 1960s, as worries mounted over a possible devaluation of the dollar and the amount of U.S. foreign liabilities, the Fed paid "considerable attention" to the balance of payments, as argued by Bordo and Eichengreen, 2013, resulting in a tighter monetary policy than what would have been expected from inflation and output gap considerations, something Taylor,

1999, has called a *policy mistake*. Eichengreen, 2013, underlines that this influence has drastically diminished after the late 1970s, when the conjunction of the collapse of the Bretton Woods system in 1973, the Federal Reserve Reform Act of 1977 (which carved into stone the dual mandate) and the arrival of Paul Volcker (who made fighting inflation his top priority), made international situation far less important for the Fed. The share of the U.S. in world GDP (33 percent in 1985, its peak), the moderate rise in the Trade/GDP ratio (compared to the following decades) and weaker international finance links than in the subsequent years meant that in this period and through the 1980s, the Fed was able to decide its policy under the rough approximation that the U.S. could be considered as a closed economy. This situation of low attention to international environment (relative to what was seen before) has dominated ever since, even though the decades since the 1980s have seen an increasing coordination of the Federal Reserve and other central banks (for meetings, coordinated interventions like what was seen in 2008, or through dollar swap lines opened during crises such as the global financial crisis). But Eichengreen argues that these occurrences show that "what happens abroad doesn't stay abroad" and in a prospective part, contends that international considerations should be more explicitly taken into account by the FOMC in the future.

In a more analytical work, Obstfeld, 2020, explores the way the global economic situation can influence the U.S. economy, and in particular the economic variables monetary policymakers are interested in, especially inflation, employment, output and financial conditions. He notes that if global factors do not necessarily affect the ability of the central bank to exert a control on long-term price changes, they can have a strong influence in the short run. The author identifies the main channels of transmission through which these global factors can impact U.S. GDP. First of all, he highlights the role of international prices (such as commodities for example) but also of international competition in shaping U.S. inflation. The role of imported inflation is shown to be more and more important in setting domestic prices. Another channel originates from the integration of global financial markets, which affects the determination of asset returns (including r^*, the natural real rate of interest) and domestic financial stability. Overall, the author concludes that in a world made ever more complex by increasing cross-border interconnections, the U.S. economy is likely to be increasingly affected by global savings and investment trends, therefore requiring attention from U.S. monetary policy-makers. Those spillback effects have been highlighted in multiple studies, such as Sachs, 1985, who argues that higher interest rates mean a strong dollar which pushes import prices lower and favors disinflation, and extensively studied in Breitenlechner et al., 2022.

An overall result from empirical and theoretical works tends to acknowledge evidence of significant spillbacks, but of small magnitude. For example, Caldara et al., 2022, use the Sigma model in its dollar-dominance version, developed at the Fed Board, and show that a monetary tightening of 100 basis points in the foreign economies results in a drop of U.S. GDP of 0.15 percent at the trough, that is about half of the impact of U.S. monetary policy on foreign GDP.

2.2. Textual analysis of monetary policy material

Our work also fits into the literature on textual analysis of documents produced by monetary policymakers, and especially FOMC material. Within that common theme, researchers have used a large range of techniques to study widely different issues.

This literature starts with the influential work of Romer and Romer, 2004, who dig into *Greenbook* documents prepared by Fed economists in order to estimate monetary policy shocks as the differences between FFR and forecasts of inflation, output and unemployment rate done by the Fed staff. Those monetary policy shocks have been then extensively used in the macroeconometric field to assess their effects on macroeconomic and financial variables (see for example Coibion et al., 2017).
With the development of Machine Learning tools and Natural Language Processing (*NLP*) techniques, authors have developed approaches to extract more efficiently and more accurately textual information from monetary policy documents. Perhaps the first example of computational textual analysis of FOMC content is the paper by Boukus and Rosenberg, 2006 who are using *Latent Semantic Analysis* (LSA), a form of principal component analysis applied to text. This approach allows recognition of textual patterns in documents to decompose minutes into a set of themes, which are in turn linked to market moves. They conclude that market participants can gain a "multifaceted signal" from the minutes, paving the way for more computational analysis of minutes text data. The same LSA methodology has been also applied by Mazis and Tsekrekos, 2017 who identify themes in FOMC statements, among them credit spread, short or long end of the yield curve, with significant effects on the moves of the bond markets. Fligstein et al., 2014 apply another topic modeling technique, *Latent Dirichlet Allocation* (LDA, see Blei et al., 2003), to analyze how FOMC members recognize financial crises. Hansen and McMahon, 2016 also use a *LDA* to carry out an investigation of the effects of Fed communication on macro variables. Both LSA and LDA approaches are unsupervised topic modeling techniques, which sets them apart from the methods we use in section 3.2.3. These topic modeling techniques have in recent years become a *de facto* default for textual data analysis in this field, but several others have been used in the related literature or in other economics-related natural language processing papers. Aruoba and Drechsel, 2022 use NLP techniques to develop a novel method to identify monetary policy shocks by analyzing Fed *Greenbook* documents. In this respect, they construct multiple time series of the sentiment surrounding various economic concepts they identify in the documents ('aspect-based' sentiments) using the dictionary first put forward by Loughran and Mcdonald, 2011. This approach through dictionaries is close to what we study in section 3.2.1 and is in line with the initial work by Ferrara and Teuf, 2018.

Some research works have also been conducted in a *supervised* learning framework. For example, Handlan, 2022 applies a Neural Network that predicts change in Fed funds futures using FOMC statements to create a series of monetary policy shocks. Rohlfs et al., 2016, use a classification method (Max-Entropy Discrimination LDA or *MedLDA*) to classify whether a given topic leads to changes in interest rates. Then, they infer from this classification whether interest rates are likely to move in a given direction after the publication of the minutes of a meeting. Many other studies focus on the interactions between Fed decisions and financial markets. Cieslak and Vissing-Jorgensen, 2021, count market-related words in Fed minutes to show that Fed monetary policy pays attention to the stock market. Sharpe et al., 2022, use *Greenbook* documents to estimate a sentiment index which is in turn put into quantile regressions to show that sentiment can predict monetary policy surprises and stock returns. Shapiro and Wilson, 2021, estimate FOMC's preferences, including the implicit inflation target, from the tone of the language used in transcripts, minutes, and members' speeches using predefined dictionaries of words associated with sentiments ("Bag of Words" or "lexical" approach). Ochs, 2021, uses a measure of sentiment in FOMC documents to disentangle effects of a monetary policy shock between the actual effects of monetary policy and the reaction of private agents to the newly acquired information (measured by a text-extracted variable). Arseneau et al., 2022, analyze central banks' speeches with NLP approaches to understand

how central banks communicate about climate change. Recently, Granziera et al., 2025, assess to what extent FOMC members and regional Federal Reserve presidents are likely to influence private sector expectations. They show that speeches highlighting upcoming inflationary pressures tend to lead both households and professional forecasters to raise their inflation expectations.

Note that some other works have been conducted on other countries and central banks. Among others, we can quote the work by Shibamoto, 2016, who uses a text-mining approach to show that asset prices react to Bank of Japan communication, while Hubert and Labondance, 2021, find that the ECB statements' tone explains monetary surprises. Also, TerEllen et al., 2022 use NLP techniques to identify narrative monetary policy surprises in Norway by looking at the difference between official central bank communication and economic media coverage.

3. Computing International Attention Indexes

In this section, we describe data and methods used to compute our International Attention Indexes (IAI).

3.1. Data

Published three weeks after each FOMC meeting, FOMC minutes are a more complete but less timely overview of the mechanisms behind monetary policy decisions than the statement, published immediately after each meeting. As such, they are a major tool of transparency and accountability policy of the Federal Reserve. Despite this delayed publication, they can have a sizeable impact on financial markets. Nechio and Wilson, 2016, show that information contained in FOMC minutes can impact Treasury bond yields at the time of their release, and that these impacts are unsurprisingly larger when the tone of the minutes differs from the tone of the statement.
We choose to work on FOMC minutes (and not on speeches made by its members or press conferences by its chair) for three main reasons. First, our aim is to study whether discussions of international matters *within* the FOMC are likely to influence monetary policy decisions, and we therefore limit ourselves to materials reflecting communication within the Federal Reserve, not from the FOMC towards outsiders[1]. Second, among internal materials, minutes offer the best compromise between completeness and delay, since the more complete *Meeting Transcripts* are published with a longer lag[2] therefore preventing us from studying recent periods. Finally, minutes feature a general structure which is stable throughout the period under study (1993-2022), making comparisons between values at different times both easier and more relevant. Overall, we have access to two sets of FOMC minutes per quarter.

We retrieve text files of the minutes of FOMC meetings from the dedicated section on the Federal Reserve Board of Governors website. We include the minutes of conference calls held by the FOMC in case of emergency meetings (mainly during economic or financial crises) that are included at the end of the minutes of the following meetings. We also remove the list of all attendants of the meeting (included at the start of every document). Our dataset comprises 238 meetings,

[1] Additionally, while we recognize that all central bank communication is political communication, we assume that since these documents, including the minutes, reflect internal discussion and are released with a delay, they are less prone to suffer from the biases usually highlighted in central bank communication analysis, such as carefully chosen words.
[2] As of March 2024, the latest transcripts published date back to 2018.

around 30,000 paragraphs and 60,000 sentences of raw textual content that we process with **Python**.

All economic data time series were retrieved from the Atlanta Fed's Taylor Rule Utility website, which includes data from the Federal Reserve, Bureau of Economic Analysis and Congressional Budget Office, among others.

3.1.1 Textual Data Preprocessing

To efficiently use text data in computational textual analysis models, it is necessary to pre-process our raw text to "normalize" it. Our preprocessing [3] follows standard procedures for NLP projects, applied in much of the literature (see Thorsrud, 2020, or Handlan, 2022, for example).

Cleaning and stemming

First, stopwords (such as *the*, *a*, ...) words which do not bring any meaning to the sentence, are removed. We also remove numbers from the text. Words are then lowercased and tokenized. Finally, we stem all words, to reduce inflected words to their root form. For example, stemming transforms *dollars* into *dollar* or *fairly* into *fair*. To stem our text, we use the Snowball Stemmer (Porter, 2001), a stemmer which builds on the Porter stemmer and is widely accepted to be more efficient, for example at stemming adverbs. We provide examples of this 2-step process in Table 1 :

"A meeting of the Federal Open Market Committee was held."	meeting Federal Open Market Committee held.	meet feder open market committe held

Table 1: *Examples of the 2-step preprocessing of a sentence*

Vectorization

Since machine learning models are not capable of processing raw text but only numbers, we have to transform our (stemmed) sentences into vectors. For that, we select two methods whose results will be compared in our empirical analysis: Term Frequency - Inverse Document Frequency (*TF-IDF*) and Word2Vec.

TF-IDF vectorization, based on a given corpus D, transforms a given sentence d into a vector of size n (the number of different words in the corpus) by including for each word t in the corpus a measure of its importance in the sentence, denoted *tfidf(t, d, D)*. TF-IDF is the combination of two measures: Term Frequency, put forward by Luhn, 1958, and Inverse Document Frequency, first devised by Sparck Jones, 1972. For a term t in a sentence d in a corpus D, we define:

$$tfidf(t,d,D) = tf(t,d) \times idf(t,D) \tag{1}$$

- $tf(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$ is the relative frequency of term t within sentence d, where $f_{t,d}$ is the raw count of term t in sentence d.
- $idf(t,D) = log \frac{|D|}{|\{d \in D : t \in d\}|}$ is the inverse document frequency of term t in corpus D, where $|D|$ is the total number of documents in the corpus[4].

[3] A diagram of our preprocessing is featured in Figure 4, Appendix A.

[4] The denominator (number of documents where the term t appears) is commonly adjusted to $1 + |\{d \in D : t \in d\}|$ to avoid division by 0 if the word is not in the original corpus.

Word2Vec, published in Mikolov et al., 2013, is a word embedding method: it transforms words into vectors by using neural networks to infer word meanings from a corpus of text. The method generates a vector space, where, ideally, words are positioned to be close in the space (in terms of cosine similarity) if they are close in meaning. The vector space means that the representation can subtract or add vectors to each other. For example, we have: $king - man + woman \approx queen$.

Both vectorization methods need to be trained on a given corpus of textual data: TF-IDF derives frequency from a corpus and Word2Vec needs to learn word association. In our case, both trainings are performed on the corpus of cleaned, stemmed minutes from 1993 to 2018. Details on those vectorization methods are featured in Appendix C.

3.2. NLP Methods

We describe here the three NLP methods that we carry out after the vectorization step of the FOMC minutes, namely the *dictionary* approach, the *classification with supervised models* and the *classification with GPT 3.5*.

3.2.1 Dictionary count

In this first method, we track simple counts of international-related words contained in a given *dictionary*[5]. For each FOMC document, we count the number of occurrences of the dictionary's words contained in the document and divide it by the total number of words. This ratio is our first indicator of attention to foreign matters by the FOMC, shown in Figure 2 and defined by:

$$IAI_M = \frac{\sum_{w \in M} \mathbb{1}_{w \in D}}{|M|} \tag{2}$$

where M is the set of all words in stemmed minutes and $|.|$ denotes the cardinality of a set.
To obtain a quarterly time series, we perform this count on M_Q, the set of all words in stemmed minutes of a given quarter Q that includes two FOMC meetings.

Table 2 lists the words and groups of words we use in this dictionary count, organized in three different categories: *foreign-exchange related*, *trade-related*, *others*. Appendix E reports the number of times each search phrase appears in the corpus. Obviously, the choice of the dictionary is crucial in this exercise. In most of the literature on dictionary count, words are chosen according to economic intuition, in spite of recent efforts to develop new dictionaries (see for example Renault, 2017, Gardner et al., 2023, or Barbaglia et al., 2023). To justify our dictionary, we follow an algorithm inspired by Aruoba and Drechsel, 2022, detailed in Appendix B.

3.2.2 Classification with Machine-Learning models

In this second NLP method to construct international attention indicators from our textual data, we divide the text into sentences [6], which we then classify using Machine Learning models as either relating to international matters or not[7]. Once sentences are classified, we compute the

[5] While it is true that the index depends of the phrases selected, our methodology is robust to small changes in the dictionary. The effects of such changes are discussed in Appendix F.

[6] Extracted from the text using the `nltk` library in Python (Bird et al., 2009)

[7] We performed tests before the manual labeling to decide whether to use sentences or paragraphs as our reference atom for the study, and decided for sentences. Tests for paragraphs are featured in Appendix G.

Foreign Exchange		Trade	Others	
depreciation	dollar	international trade	international	china
currency	euro	imports	global	russia
foreign currencies	pound	exports	foreign economies	europe
foreign exchange	sterling	external sector	emerging market	india
	renminbi		EME	japan
	ruble		advanced economies	asia
			world	
			abroad	

Table 2: *Dictionary used in the study.*

ratio of the number of words in a text belonging to international-classified sentences on the total number of words in the text. This ratio is our international attention indicator for the minutes considered:

$$IAI_M = \frac{\sum_{s \in M} \mathbb{1}_{cl(s)=1} |s|}{\sum_{s \in M} |s|}, \tag{3}$$

where M is the set of all sentences in stemmed minutes, s is a sentence (set of words), cl is the sentence classifier that returns 1 if the sentence is classified as international-related and $|.|$ denotes set cardinality.
To obtain a quarterly series, we perform this count on M_Q, the set of all sentences in stemmed minutes of a given quarter Q.

The models we use to classify sentences fall into two categories that are detailed in the following subsections: supervised Machine-Learning models that we build specifically to perform this classification task (Section 3.2.3) and Large Language Models, multi-purpose models that we use to perform the same task (Section 3.2.4).

3.2.3 Classification with supervised models

Building classification-based indexes with supervised Machine-Learning models is essentially a four-steps process : (i) labeling data to train the algorithms, (ii) training algorithms on a *training set*, (iii) determining the most relevant ones using a *test set* (detailed in Section 3.2.5) and (iv) finally computing the classification of all sentences and the international environment index (whose results are shown in Section 3.3).

Data Labelling

In a first step, we manually label a total of 4,300 sentences to classify them between those which relate to the international situation (class 1, or "Positive") and those which do not (class 2, or "Negative"). To avoid labeling only data from a specific time period or specific part of the minutes structure, we shuffle the dataset and label sentences that have been randomly chosen. We also avoid inserting a forward-looking bias on recent events by only labeling data until 2018. Descriptive statistics of labeled data and total data are displayed in Table 3 below:
Since the dataset of labeled data is biased (there are 90.7% of non-international-related sentences and 9.3% of international-related sentences), we rebalance it by keeping only 1/9th of the negative (i.e. non-international) observations. Indeed, we find that even though the imbalance is not

	Labeled data	Total data
Total Number	4314	66557
Average Length (in words)	11.55	12.66
Number of positives	401 (9.3%)	

Table 3: *Descriptive statistics of sentences and labeled sentences. Positive means relating to international situation.*

extreme[8], the models trained on a dataset on which we do not apply this bias correction tend to exhibit a very low recall (percentage of positives well predicted by our model), i.e. many positive observations are likely to be classified as negative[9].

Classification algorithms training

We first divide the dataset of minutes sentences between a training dataset (two thirds of the dataset) and a testing dataset (the remaining third of the sentences). The attribution of a sentence to one set or the other is done randomly. The training dataset is the one the algorithms will estimate parameters on (train) while the testing dataset will be used to assess their performance (test). For every algorithm, the parameters are determined via a *k-fold Cross Validation* approach, as it is common in Machine Learning. The training dataset is divided into five subsets, and the algorithm is trained on four of them. It is then tested on the subset that was left out, and its performance on this subset is assessed. We retain the model version that performs the best on its testing subset. We also perform a grid search of hyperparameters, where every combination of hyperparameters is tested by cross validation.
Here we test two vectorization methods, TF-IDF and Word2Vec, and three machine learning classifiers : Naive Bayes, Logistic Regression, Random Forest[10]. That makes six pipelines (combinations of a preprocessing and an algorithm).

3.2.4 Classification with Large Language Models

We also classify sentences using a class of NLP models named Large Language Models (LLMs). This type of model sprung into the general public's view in late 2022, when OpenAI released ChatGPT, an online assistant based on its GPT 3.5 LLM, which quickly gained popularity (100 million users in two months). Based on neural networks and transformer architecture, these models have been trained on tens of terabytes of textual data to learn patterns in language, which are stored in hundreds of billions of parameters. Functionally, LLMs are able to predict the probability of a sequence of words[11], using its context. While the general public and most use cases have been focusing on their text generation capabilities, these models can efficiently perform numerous tasks including translation, summarization, and answering questions. An overview of the possible use cases for economists is featured in Korinek, 2023.

[8] Extreme imbalance would be defined in Machine Learning literature as 1, 2 or 5% of class 1 observations.

[9] We could have used a resampling technique to augment the number of positive observations in our unbalanced labeled dataset. Since the models we use do not require huge volumes of labeled data and more manual labeling is possible for our minutes sentences dataset, we decided against this option, which never reaches the results obtained with a non-augmented unbiased dataset, as shown by Baesens et al., 2022

[10] More information on these algorithms can be found in Appendix D

[11] For example, a LLM, faced with the sentence start "inflation today is", would assign a 0.4 probability of "low" as the next word, 0.2 of 'higher', but only 0.01 of "out-of control", etc.

What we set out to do is to compare the results of this class of state-of-the-art, multi-purpose models, which have not been built to perform the specific classification under study, to that of models that are far less complex but built specifically for classifying sentences according to their interest in international matters, as described above. The characteristics of LLMs (size of the models, computational power needed for training and running) make them extremely complex and expensive to build from the ground up (like we did for previous models). To use this state of the art to perform classification of sentences, we decided to use an existing model and decided on **OpenAI's GPT 3.5**[12].

The exact parameterization being out of our scope[13], the use of such a model, for our case, relies mainly on the prompt, i.e. the input we give it to explain what we want it to return. After initial tests, we conduct a large-scale evaluation of three options, which differ by the richness of the prompt. A step by step explanation of our different prompts is featured in Table 4. The core of our prompt is the following sentence: *Act as a classifier : classify if a sentence is about international macroeconomy or not*. We then add a "role" for the classifier (1) and some more context (2) to create our *"basic prompt"*. Adding a few examples (3) creates our *"examples prompt"* and by adding more examples (4) we obtain what we call *"advanced prompt"*.

(1)	*You are a macroeconomics expert.*
(2)	*Act as a classifier : classify if a sentence is about international macroeconomy or not, placing yourself in a US context and classifying as international macroeconomy all economic matters that are not specifically domestic to the US (including trade between the US and other countries, and events currently happening in other countries).*
(3)	*For example, "labor force participation rate and the employment-to-population ratio increased" is not related to international, "inflation decreased in most major economies" is related to international.*
(4)	*For example, "labor force participation rate and the employment-to-population ratio increased" **"financial markets downturn is affecting households"** are not related to international, "inflation decreased in most major economies" or **"disruptions in China impacted supply chains"** are related to international.*

Table 4: *Prompt*

We obtain a classification for each sentence in our dataset, and especially on the testing set, for which we report results in Section 3.2.5.

3.2.5 Results of classification pipelines

To select the most relevant pipelines, we will measure their ability to classify data that we labeled but that they have never seen in their training phase (*i.e.* the *testing set*). We take a look at the

[12]Since the limitations of the web interface (including size limitations) do not allow us to perform this classification in ChatGPT itself, we use OpenAI's Application Programming Interface (API) which enables to "query" the underlying model (much in the way a user would query ChatGPT), and retrieve its answer, in an automatized fashion which enables us to send, by batch of 100 sentences, the more than 65,000 sentences in our dataset.

[13]By using the API, we can only have an influence on a few parameters of the model, the most interesting one in our case being the *temperature*, which controls the randomness of the model's answer, and that we set to the lowest possible value. Higher temperature yields more creative answers, at the expense of replicability. Such parametrization is not possible for web interface users.

performance (on the testing set) of the six Supervised Learning classifiers, as well as the three GPT classifiers. We add a tenth pipeline: *Dictionary-based naive classification*, where we classify as international-related sentences that contain at least one word or group of words belonging to the dictionary described in Table 2.
We therefore compare the results of ten pipelines, using standard classification metrics: Precision, Recall, F1 Score. For every class, *Precision* is the proportion of true positives (i.e. rightly classified in this class) on all predicted positives, *Recall* is the proportion of actual positives that have been predicted as such, and *F1 Score* is the harmonic mean of precision and recall. Table 5 reports a macro average of every metric for the whole dataset (not just one class)

Vectorization	Algorithm	Precision	Recall	F1-Score
Word2Vec	Logistic Regression	0.3952	0.4749	0.3754
	Naive Bayes	0.2841	0.5000	0.3624
	Random Forest	0.2841	0.5000	0.3624
TF-IDF	Logistic Regression	0.8987	0.8843	0.8892
	Naive Bayes	0.8427	0.8378	0.8191
	Random Forest	0.9077	0.8971	0.9011
	Dictionary	0.8692	0.8532	0.8582
	GPT 3.5 Basic Prompt	0.8897	0.8841	0.8864
	GPT 3.5 Examples Prompt	0.8867	0.8882	0.8874
	GPT 3.5 Advanced Prompt	0.9057	0.9065	0.9061

Table 5: *Results of the ten pipelines*

Results show that the models using Word2Vec for vectorization perform poorly on our test dataset. One explanation could be that the *training dataset* of the Word2Vec algorithm, that is our corpus, is not large enough for the neural network to learn words' association. As a consequence, we will only focus on the TF-IDF vectorization approach in our results: performance for Naive Bayes seems to be lower than other algorithms, as it performs worse in *Precision* and *Recall* criteria, and Random Forest seems to perform best, boasting the best results in all metrics.
Our Naive Dictionary classifier has performance on par with these well-performing models, which validates this dictionary-based approach: the phrases selected in our dictionary are relevant to detecting the occurrences of attention to the international situation in the FOMC minutes.

When it comes to the three GPT classifications, we note that the performance of the basic prompting classification is close to the best models, and that performance improves with prompt refinement. The best GPT 3.5 classification is the one using the most advanced prompt, vindicating our efforts on prompt engineering. It scores slightly higher than the best supervised learning-based model in terms of F1-score, which shows the potential of this state-of-the-art framework.
The fact that this is only a small improvement shows that less advanced techniques can have a role to play for this kind of highly specialized task. When trained properly and on high-quality data (in our case, thousands of hand-labeled sentences), a "simple" pipeline might rival the performance of more advanced NLP algorithms. The trade-off to make between the two is no simple one: LLMs can be easier to implement, while build-from-scratch Machine Learning models require more effort and technical NLP knowledge to achieve satisfactory results since they require labeling and fine-tuning work. Yet, the latter offers greater control and replicability than "black-box" LLMs,

which can prove useful in the context of economic analysis.

We therefore choose to build and test indexes with the two following pipelines: Random Forest with TF-IDF vectorization and GPT 3.5 with advanced prompting, which we add to the dictionary count.

3.3. International Attention Indexes

In this subsection, we present and discuss the various versions of our International Attention Index (IAI). Figure 1 displays indexes stemming from our dictionary count approach described in section 3.2.1 and the two classification methods described in section 3.2.2.

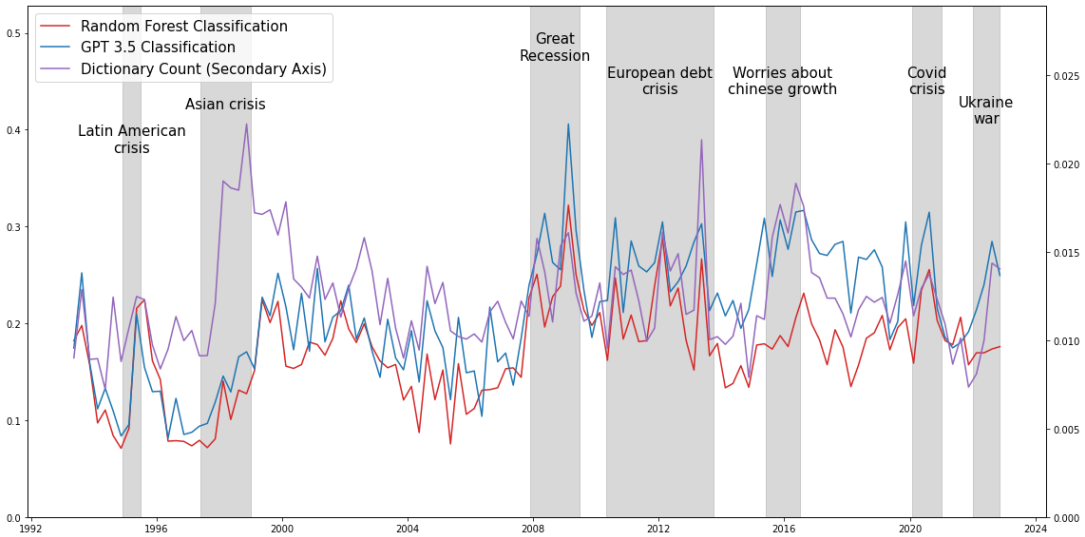

Figure 1: *International Attention Indexes obtained with dictionary count and classification methods.*

By eye-balling those graphs, we note that spikes in the indexes, that is, periods of accrued international attention in FOMC minutes, match the major international events of the past three decades. In particular, there is a good fit with crises of any type outside the U.S.: we point out spikes at the time of the Latin-American crisis of the mid 1990s, the Asian crisis, the 2008-09 Great Recession, the European debt crisis in the 2011-13 period, the concerns about the Chinese economy in 2015-2016, as well as both recent crises, namely the Covid pandemic and the Russian invasion of Ukraine.

The size of these spikes differs depending on the version of the index. The relatively small rise seen during the Great Recession in dictionary count (relative to the Machine Learning-based models) might come from a different type of attention in these years (through swap lines with foreign central banks, for example) that is not as well captured in our dictionary. Similarly, Machine Learning-based models are able to catch a significant spike during the Covid crisis while dictionary count does not move much.
Interestingly, the IAIs based on GPT 3.5 Classification and Random Forest Classification have a very similar behavior. The most notable difference in value between the two is exhibited during the 2015-2016 interrogations on Chinese growth, where the GPT-based model detects higher

levels of attention than the Random-Forest-based one. A look at the 2018-2022 period shows another difference in behavior during the Ukraine war period, where GPT-based IAI rises sharply, but it should be noted that GPT 3.5 has been trained on textual data from this period[14], while random-forest classifier has seen no training data after 2018 (we only labeled sentences from 1993 to 2018).

All indexes tend to show an asymmetric behavior in the sense that discussions about the international environment seem to take place mainly when there is negative news on foreign economic activity or markets[15].

One of the advantages of the dictionary approach is that we can split international words into categories that sum up to the global index. Thus we decompose the dictionary-based IAI into three main categories, namely trade, currency, and others (see Table 2). Figure 2 shows this decomposition of the IAI into specific channels through which foreign events might affect the Fed monetary policy discussions.

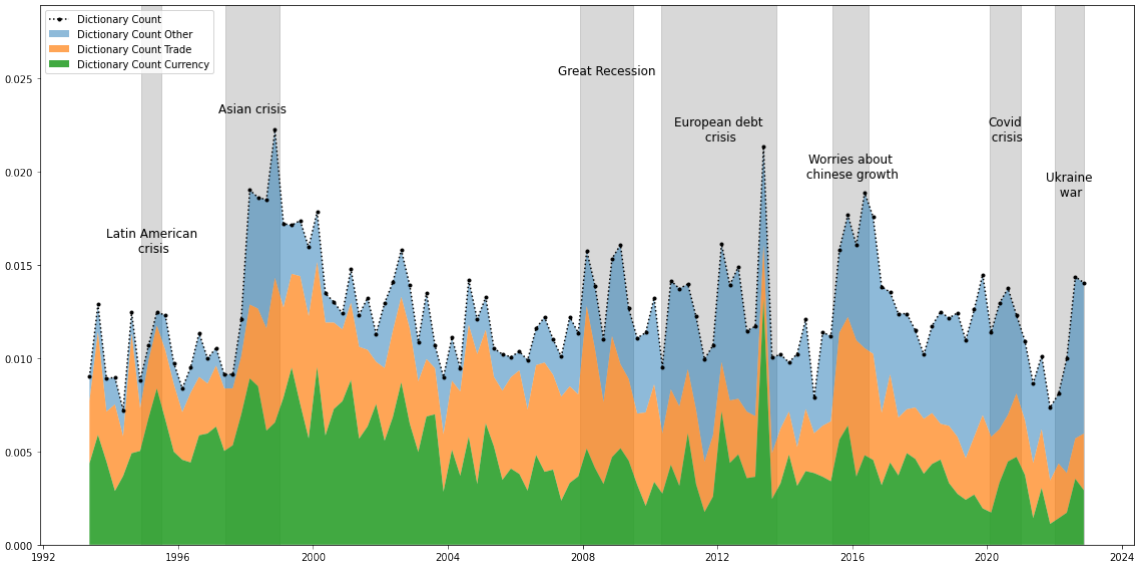

Figure 2: *Value of the International Attention Index for dictionary count method and sub-dictionaries counts.*

We first note a long-run decline in the component related to currencies. This was the topic of discussions before the 2000s, maybe related to the introduction of the euro as a competitor for the dollar, but since then it does seem to be an issue among FOMC members. Let's just note the spike in the first quarter of 2013 at the end of the European debt crisis, a long period of dollar over-evaluation with respect to the euro. In comparison, the trade component tends to show more spikes, the last one being between mid-2015 and the beginning of 2016 amidst concerns on the Chinese economic growth and the renminbi. Interestingly, the trade war between the U.S. and China/Europe started by the Trump's administration in 2018 didn't affect discussions among FOMC members. The increase in tariffs from imports of those countries was not perceived as a

[14]The last training update of GPT 3.5 being in early 2022, it has no knowledge of the full-blown invasion of Ukraine by Russia, but has been trained on data regarding the build-up of tensions that preceded it.

[15]Appendix H investigates this point by studying the positive or negative sentiments of international-related words, but we do not conclude on the systematic negativity of attention to international matters.

concern for monetary policy by Fed's officials. The last component, named *others* is driving the headline value of the international attention index since the mid 2000s. This is likely due to the fact that most of the mentions of foreign crises link the name of the involved countries with broad economic concepts (growth, inflation, etc.), which we could not include in our dictionary because these concepts are also associated with separate domestic news in other parts of the documents that are not of direct interest for our research.

4. Effects of International Attention Indexes on monetary policy

This section empirically investigates if the attention to the international environment by the FOMC (as measured by the different indexes detailed in section 3.3) can have an influence on monetary policy decisions. In this respect, we follow a framework in the spirit of Cecchetti et al., 2000, by estimating a Taylor rule augmented with a given International Attention Index (IAI). This framework enables us to assess whether our indexes provide significant information on Fed decisions. The estimation sample goes from 1994q1 to 2022q3. We assume in this section that despite the wider range of tools policymakers have now at their disposal (forward guidance, asset purchases ...), the Fed Funds Rate remains the main monetary policy tool during the period under study. To account for the zero-lower bound period, we will also consider the U.S. shadow rate put forward by Wu and Xia, 2016 in order to account for unconventional monetary policy tools launched since 2009 (Section 4.3.2).

4.1. Framework

The so-called Taylor rule, initially put forward by Taylor, 1993, estimates the response of the Fed interest rate to macroeconomic conditions. In its basic form, it is the endogenous response of monetary policy to domestic macroeconomic conditions. The first version of the Taylor rule (Equation (4) below) linearly relates the nominal Fed Funds Rate r_t, to core PCE inflation rate π_t, the inflation target π_t^*, the output gap g_t and a Gaussian white noise ε_t in the following way:

$$r_t = \alpha + \gamma(\pi_t - \pi_t^*) + \beta g_t + \varepsilon_t \tag{4}$$

where α can be interpreted as the nominal neutral interest rate that prevails when inflation is at the target and the output gap is closed (i.e. $\alpha = \pi_t + r_t^*$, where r_t^* is the real neutral interest rate).

In this work, we use an extended version of the Taylor Rule given by Equation (5) below, which includes ρ, a smoothing term for the FFR and i_t, our International Attention Index (IAI)[16].

$$r_t = \rho r_{t-1} + (1-\rho)(\alpha + \gamma(\pi_t - \pi_t^*) + \beta g_t + \delta i_t) + \varepsilon_t \tag{5}$$

The underlying idea is that the core part of equation 5 describes the reaction function of the Fed to domestic matters (γ its reaction to domestic inflation, β to output gap, α to the nominal neutral interest rate, ρ to the fact that FFR moves are incremental). Adding i_t, the attention paid by the FOMC to international matters, which is supposed to be exogenous and not strictly related to the Fed's domestic mandate, is therefore a way to assess the significance of this factor in explaining

[16] Another version of the Taylor Rule, with a time-varying r^*, is discussed in Appendix I.

FOMC decisions (depending on the value and significance of δ).

We chose to work on data provided by the Atlanta Fed on its Taylor Rule Utility website and not on Greenbook forecasts as for example, Istrefi et al., 2021, since those latter forecasts are no longer available on the Fed website from 2010 onward, and we would also like to study the influence of our index in the 2010s and early 2020s. We choose to use as π_t^* the four-quarter core PCE inflation, constructed by the U.S. Bureau of Economic Analysis (BEA), and as g_t the output gap derived from the Congressional Budget Office (CBO) estimate of potential real GDP. All variables are expressed in percent, and our index i_t is standardized before entering the equation. We estimate this Taylor Rule by Non-Linear Least Squares and get the results featured in the following sections.

4.2. Main empirical results

Table 6 presents the results of our various estimated regressions. Estimated parameters of the standard Taylor rule (column 0) are quite close to the usual calibration based on Taylor's initial results. The smoothing parameter is highly significant, underlying the strong persistence of FFR. The intercept is close to 4%, meaning that the real neutral interest rate r^* can be proxied by a constant around 2% over the 1994-2022 period. We will later relax this assumption in line with recent research showing time-variation in r^* (see Laubach and Williams, 2003). The estimated weight of the inflation gap is above 1 (1.28) but doesn't appear significantly different from zero. In opposition, the estimated coefficient for the output gap is clearly significant. We now integrate our various IAIs into the Taylor equation (5). First, we note that the integration of the IAIs doesn't change much the other coefficients, highlighting the stability of the equation. Then, we get that the estimated IAI coefficients are always negative, and generally significant, except for model (1) (dictionary count).
The similar values of the i_t coefficient for the IAIs based on machine-learning models ((2) and (3)) is also consistent with the fact that the methods perform similarly in terms of out-of-sample classification (Table 5) and their IAIs exhibit relatively similar behavior (Figure 1).
The version of the IAI that yields the best results in terms of accuracy, based on AIC and MSE criteria, is model (2), that is the one based on Random Forest classification of sentences. With this Random Forest-based IAI (2), the (*standardized*) IAI has a coefficient of -8.15, which means that when the (normalized) index is equal to 0.1, the marginal effect of the international environment on FFR is about 8 basis points on average.

4.3. Robustness checks

Let's focus on Model (2), which can be considered the best model according to all criteria, and assess some robustness checks.

4.3.1 Control variables

Let's first introduce some additional control variables in equation (5) to ensure that our IAI provides information beyond that of a simple measure of international financial instability.
In this respect, we estimate the following augmented equation (6)):

$$r_t = \rho r_{t-1} + (1-\rho)(\alpha + \gamma(\pi_t - \pi_t^*) + \beta g_t + \delta i_t + \phi s_t) + \varepsilon_t. \quad (6)$$

where s_t is an index of global conditions. We consider two indexes: the MSCI World (equity) index and the Geopolitical Risk (GPR) index of Caldara and Iacoviello, 2022 (both taken in differences).

Dependent variable			FFR		
Model	(Cal)	(0)	(1)	(2)	(3)
r_{t-1}	0.85	0.915***	0.913***	0.903***	0.884***
		(0.024)	(0.024)	(0.023)	(0.026)
α	4.0	4.17***	4.191***	3.529***	3.63***
		(0.557)	(0.548)	(0.459)	(0.398)
$PCEGap_t$	1.5	1.285	0.889	0.922	0.692
		(0.794)	(0.777)	(0.615)	(0.521)
g_t	1.0	1.209***	1.251***	0.79***	0.865***
		(0.244)	(0.246)	(0.206)	(0.179)
i_t			-3.06	-8.149***	-5.443***
			(2.506)	(2.639)	(1.761)
Observations		115	115	115	115
Period		1994-2022	1994-2022	1994-2022	1994-2022
AIC		115.637	115.95	102.444	109.279
MSE		16.872	16.626	14.784	15.689

Table 6: *Taylor rule regression results — Equation 5 specification*

Note: The table shows Taylor Rule (Equation 5) results for Non-Linear Least Squares Regression, with different inputs as the international attention index i_t: (0) is no IAI, (1) the dictionary count-based index, (2) the index from random forest classification of sentences, (3) the index from GPT classification of sentences. The (Cal) column features the calibrated parameters (from Atlanta Fed)
* p<0.1; ** p<0.05; *** p<0.01.

Results are reported in Table 7. The coefficient for the global equity index does not appear statistically significant when we control for our IAI, but becomes significantly positive when it is integrated alone in the basic Taylor rule. This may reflect the fact that the IAI and the MSCI are somewhat negatively correlated, suggesting that a drop in the MSCI is associated with higher international attention by the FOMC.

In opposition, the geopolitical risk as estimated by the GPR index appears always negatively significant in the Taylor rule, independently of the inclusion of the IAI. This indicates that the two indicators are not necessarily correlated, suggesting that the IAI doesn't capture geopolitical tensions. From a policy perspective, this result suggests that the international attention of the FOMC is more focused on international financial markets than on geopolitical tensions between countries. This is in line with the absence of movement in the IAIs during the trade war between U.S. and China, or during full-blown conflicts such as the war in Iraq.

These results enable us to confirm that our NLP-based measure of international attention provides information for FFR changes beyond the effect of international market-based financial or geopolitical indicators.

Dependent variable			FFR			
Model	(Cal)	(2)	(2.1.1)	(2.1.2)	(2.2.1)	(2.2.2)
r_{t-1}	0.85	0.903***	0.912***	0.926***	0.909***	0.92***
		(0.023)	(0.023)	(0.024)	(0.022)	(0.023)
α	4.0	3.529***	3.397***	3.927***	3.657***	4.329***
		(0.459)	(0.507)	(0.618)	(0.476)	(0.594)
$PCEGap_t$	1.5	0.922	1.154	1.698*	1.161*	1.586*
		(0.615)	(0.722)	(1.004)	(0.672)	(0.882)
g_t	1.0	0.79***	0.769***	1.167***	0.774***	1.186***
		(0.206)	(0.224)	(0.27)	(0.211)	(0.248)
i_t		-8.149***	-8.111***		-8.115***	
		(2.639)	(2.862)		(2.707)	
$MSCI_t$			7.538	13.566*		
			(5.689)	(8.068)		
$Risk_t$					-1.619**	-2.062**
					(0.733)	(0.959)
Observations		115	115	115	115	115
Period		1994-2022	1994-2022	1994-2022	1994-2022	1994-2022
AIC		102.444	101.967	112.32	96.663	109.004
MSE		14.784	14.469	16.109	13.816	15.652

Table 7: *Taylor rule regression results - with financial situation control variables*

Note: The table shows Taylor Rule regression with additional variables to control for the global financial environment All models use Non-Linear Least Squares Regression, with the same international attention index. Model (2) features no control, (2.1.1) and (2.1.2) the Quarter over Quarter evolution of the MSCI World Equity index, respectively with and without the IAI index included (Equations 17 and 18), (2.1.1) and (2.2.2) the Quarter over Quarter evolution of the GPR index, respectively with and without the IAI index included (Equations 19 and 20). The (Cal) column features the calibrated parameters (from the Atlanta Fed).
* p<0.1; ** p<0.05; *** p<0.01.

4.3.2 Shadow interest rates

Second, let's consider some robustness checks on interest rates. Since the monetary policy of the late 2000s and early 2010s has been marked by rates that stayed for a long time at the Zero Lower Bound, we want to investigate in this section whether it affected the potential accuracy of the Taylor rule, and therefore of our measure of the influence of international attention on Fed decisions. To offset the potential hurdles posed by the monetary policy of this period and account for the Zero Lower Bound, we test another dependent variable in the Taylor rule by replacing the Fed Funds Rates with the Shadow Rates put forward by Wu and Xia, 2016.

It turns out that monetary policy tools have changed since the Global Financial Crisis (GFC) and now unconventional tools, such as quantitative easing or forward guidance, have joined the toolbox of policy-makers. To account for this, some researchers have proposed a modified version of central bank interest rates that reflects additional accommodation in the monetary policy stance. In this respect, Wu and Xia, 2016, provide a *shadow* interest rate for the U.S. economy. That new measure has the advantage of taking into account unconventional monetary measures

(quantitative easing) in the period when additional stimulus through the conventional channel of monetary policy (lowering interest rates) was not available. Accounting for these unconventional monetary policy tools leads to some negative values for the shadow FFR. That effectively allows us to provide a measure of the effect of Fed IAI on monetary policy decisions for the Zero Lower Bound period (when our index moves but FFR stays at zero).
Table 8 presents the results obtained when the Fed Funds Rate is replaced as the dependent variable by this Wu-Xia *shadow rate*.

Dependent variable		FFR			Shadow Rates		
Model	(Cal)	(0)	(2)	(3)	(0)	(2)	(3)
r_{t-1}	0.85	0.915***	0.903***	0.884***	0.921***	0.916***	0.905***
		(0.024)	(0.023)	(0.026)	(0.026)	(0.025)	(0.028)
α	4.0	4.17***	3.529***	3.63***	4.204***	3.421***	3.706***
		(0.557)	(0.459)	(0.398)	(0.856)	(0.76)	(0.723)
$PCEGap_t$	1.5	1.285	0.922	0.692	1.979	1.615	1.425
		(0.794)	(0.615)	(0.521)	(1.286)	(1.1)	(1.017)
g_t	1.0	1.209***	0.79***	0.865***	1.459***	0.945***	1.145***
		(0.244)	(0.206)	(0.179)	(0.374)	(0.34)	(0.328)
i_t			-8.149***	-5.443***		-10.483**	-5.101*
			(2.639)	(1.761)		(4.449)	(3.056)
Observations		115	115	115	115	115	115
Period		1994-2022	1994-2022	1994-2022	1994-2022	1994-2022	1994-2022
MSE		16.872	14.784	15.689	33.279	30.622	32.553
Taylor Version		9	12	12	9	12	12

Table 8: *Results of Taylor Rule Regression with Shadow Rates as the dependent variable*

Note: The table shows Taylor Rule (Equations 4 & 5) results for Non-Linear Least Squares Regression, with two possible dependent variables (Fed Funds Rates, Shadow Rates) and different inputs as the international attention index i_t: (0) is no international environment index, (2) the index from random forest classification of sentences, (3) the index from GPT classification of sentences. The (Cal) column features the calibrated parameters (from the Atlanta Fed).
* p<0.1; ** p<0.05; *** p<0.01.

We conclude from Table 8 that when used to explain variations in the Shadow Rates of Wu and Xia, 2016, our Fed IAI remains significant and the conclusions of Section 4.2 still hold : attention to the international situation has a negative effect on Shadow Rates. Against this background, results are stronger in the sense that the estimated coefficient of the IAI is highly significant and increases to -10.48 (compared to -8.15 when using the standard FFR). A rise in the Fed's attention to the global economy translates into a more accommodating monetary policy (conventional and unconventional).

4.3.3 Variation over time

Third, we study the variation over time of δ, the estimated coefficient of our IAI, in search of possible shifts during some specific periods. We carry out a recursive analysis: starting from the period 1994-2005, and progressively extending the window by adding a new data point, we

re-estimate the augmented Taylor rule. For each point in time, we report the estimated value of δ, the coefficient of i_t, and its 68% confidence interval. Figure 3 shows the evolution of $\hat{\delta}$ over time.

We observe that the estimate is always largely significantly different from zero, underlining the robustness over time of the model. The international environment was playing negatively on central bank interest rates before the Global Financial Crisis (GFC), but there is a clear downward shift in 2008 due to the cross-border propagation of the crisis and its spillback effect on U.S. monetary policy as the financial crisis went really global. In the wake of the GFC, the estimated coefficient stayed at a low level, reflecting an always strong sensitivity of policy-makers to external developments. Additional results involving Model (2) are reported in the Appendix, for example when taking the unemployment gap instead of the output gap.

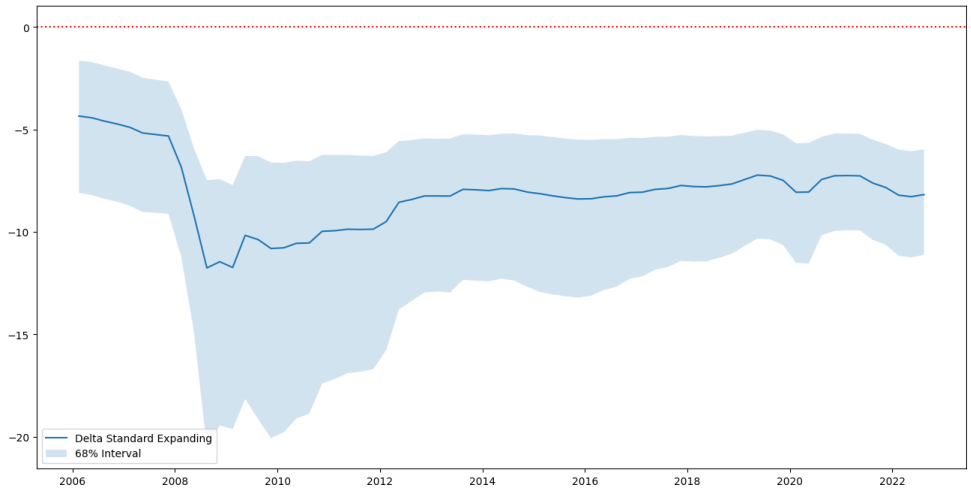

Figure 3: *Time-varying estimated value $\hat{\delta}$, the IAI coefficient, and its 68% confidence interval in an expanding window.*

4.3.4 Time-varying r*

Fourth, another stylized fact from the monetary policy literature is that there is clear evidence of a long-run decline in r_t^*, the real natural interest rates of the economy. This is mainly due either to structural issues related to population aging or to the secular stagnation hypothesis (see, for example, Laubach and Williams, 2003). Appendix I explores other versions of the augmented Taylor rule that incorporate the time-varying r_t^* estimated by Laubach and Williams, 2003 and updated by the Federal Reserve Bank of New York.

When r_t^* is integrated into a standard Taylor rule, both coefficients of the inflation gap and output gap sharply decline, but the one related to the output gap stays significant. There is no change as regards the IAI estimated coefficient that stays at -8.03 (*vs* -8.15 in the standard version without r_t^*). So results appear extremely robust to the inclusion of the time-varying neutral interest rate.

5. Conclusions

The U.S. Federal Reserve has officially domestic objectives of price stability and full employment and is often considered to pay little attention to the rest of the world. Yet, economic analysis and anecdotal evidence suggest that the international situation is likely to play a role in shaping U.S. monetary policy, or at least is somewhat considered by policy-makers. To test this hypothesis, we develop a quantitative approach to assess both the attention paid by members of the Federal Open Market Committee (FOMC) to foreign economic conditions and its potential effects on monetary policy decisions from 1993 to 2022.

In this respect, we deploy various Natural Language Processing (NLP) methods of computer-based textual analysis to construct International Attention Indexes (*IAIs*) from the minutes of FOMC meetings. Those indexes are proxies of the attention paid by FOMC members to the economic situation abroad. We focus on the FOMC minutes as they reflect internal discussions within the committee and offer a good trade-off between publication delays, on the one hand, and completeness in terms of policy discussions, on the other hand. After pre-processing this large textual dataset, we include it into both dictionary-based models and Machine Learning classification models. As an output, we get various IAIs reflecting the main economic events of the last decades.

The inclusion of two very different Machine Learning frameworks for classification, namely built-from-scratch supervised models based on Random Forest and Large Language Models in the form of OpenAI's GPT 3.5, and the observation that they yield similarly interesting results, allows us to draw conclusions regarding the strengths of these two approaches for very specific tasks such as this one. While state-of-the-art LLMs are easier to implement and demonstrate very good performance, they remain somewhat opaque in their exact operation. In contrast, less refined classification models, whose performance can be very close for our case, offer greater control and replicability than "black-box" LLMs but require more effort to achieve satisfactory results, as they necessitate additional labeling and fine-tuning work.

In a second step, we use our IAIs to augment standard Taylor rules in order to assess Fed monetary policy reaction to both U.S. domestic and international conditions. Overall, we get that those indexes have a significantly negative effect on the level of Fed Funds Rates. In other words, those indexes reinforce a dovish pattern with respect to a standard Taylor rule. We interpret those results as evidence that FOMC members tend to take global conditions into account when reaching a decision. Those results turn out to be robust to various changes in modeling, such as the integration of a shadow central bank interest rate or the use of time-varying real neutral interest rates.

One possible continuation of this work could be to conduct the same type of study on a longer time frame to try to identify the longer-term shifts put forward by Eichengreen, 2013. While very appealing, this potential continuation runs into data-linked obstacles, namely the availability of comparable documents. *Minutes* of FOMC meetings are available from 1993 onward, but only *Minutes of Actions* exist from 1967 to 1992, and *Historical Minutes* from 1936 to 1967. If historical minutes are pretty comparable to current-format minutes, the *Minutes of Actions* are much shorter in length and feature significantly less discussion.

References

Algaba, A., Ardia, D., Bluteau, K., Borms, S., & Boudt, K. (2020). Econometrics Meets Sentiment: An Overview Of Methodology And Applications. *Journal of Economic Surveys*, *34*(3), 512–547.

Arseneau, D. M., Drexler, A., & Osada, M. (2022). Central bank communication about climate change. *Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System*, (No. 2022-031).

Aruoba, B., & Drechsel, T. (2022). Identifying monetary policy shocks: A natural language approach. *CEPR Discussion Papers*, *17133*.

Baesens, B., Banulescu-Radu, D., Hurlin, C., Kougblenou, Y., & Verdonck, T. (2022). Benchmarking state-of-the-art resampling techniques for classification models. *mimeo*.

Barbaglia, L., Consoli, S., & Manzan, S. (2022). Forecasting with economic news. *Journal of Business and Economic Statistics*, *forthcoming*.

Barbaglia, L., Consoli, S., Manzan, S., Pezzoli, L. T., & Tosetti, E. (2023). Sentiment analysis of economic text: A lexicon-based approach. *SSRN*.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python: Analyzing text with the natural language toolkit*. O'Reilly Media, Inc.

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 993–1022.

Bordo, M., & Eichengreen, B. (2013). Bretton Woods and the Great Inflation. In *The Great Inflation: The Rebirth of Modern Central Banking* (pp. 449–489). National Bureau of Economic Research, Inc.

Boukus, E., & Rosenberg, J. V. (2006). The information content of FOMC minutes. *Federal Reserve Bank of New York*, (Working Paper).

Breitenlechner, M., Georgiadis, G., & Schumann, B. (2022). *What goes around comes around how large are spillbacks from US monetary policy?* European Central Bank.

Bruno, V., & Shin, H. S. (2015). Capital flows and the risk-taking channel of monetary policy. *Journal of Monetary Economics*, *71*(100), 119–132.

Caldara, D., Ferrante, F., & Queralto, A. (2022). *International spillovers of tighter monetary policy* (FEDS Note No. 26609). Board of Governors of the Federal Reserve System.

Caldara, D., & Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review*, *112*(4), 1194–1225.

Cecchetti, S., Genberg, H., Lipsky, J., & Wadhwani, S. (2000). Asset prices and central bank policy. *The Geneva Report on the World Economy*.

Chakraborty, C., & Joseph, A. (2017). Machine learning at central banks. *SSRN Electronic Journal*.

Cieslak, A., & Vissing-Jorgensen, A. (2021). The economics of the fed put. *The Review of Financial Studies*, *34*(9), 4045–4089.

Coibion, O., Goronidchenko, Y., Kueng, L., & Silvia, J. (2017). Innocent bystanders? monetary policy and inequality. *Journal of Monetary Economics*, *88*, 70–89.

Dedola, L., Rivolta, G., & Stracca, L. (2017). If the Fed sneezes, who catches a cold? *Journal of International Economics*, *108*(S1), 23–41.

Degasperi, R., Hong, S. S., & Ricco, G. (2021). The global transmission of u.s. monetary policy. *Sciences Po. OFCE Working Paper*, (9), 49.

Eichengreen, B. (2013). Does the federal reserve care about the rest of the world? *Journal of Economic Perspectives*, *27*(4), 87–104.

Ferrara, L., & Teuf, C.-E. (2018). *International environment and US monetary policy: A textual analysis* [Banque de france blog].

Fischer, S. (2015). *Speech on u.s. inflation developments* [Board of governors of the federal reserve system].

Fligstein, N., Brundage, J. S., & Schultz, M. (2014). *Why the federal reserve failed to see the financial crisis of 2008: The role of "macroeconomics" as a sense making and cultural frame* (Institute for Research on Labor and Employment, Working Paper Series). Institute of Industrial Relations, UC Berkeley.

Gardner, B., Scotti, C., & Vega, C. (2023). Words Speak as Loudly as Actions: Central Bank Communication and the Response of Equity Prices to Macroeconomic Announcements. *Journal of Econometrics, forthcoming*.

Granziera, E., Larsen, V., Meggiorini, G., & Melosi, L. (2025). *Speaking of Inflation: The Influence of Fed Speeches on Expectations* (CEPR Discussion Papers). C.E.P.R. Discussion Papers.

Greenspan, A. (1998). Question: Is there a new economy? *Remarks at Haas Annual Business Faculty Research Dialogue, University of California, Berkeley, September 4*.

Ha, J., Kose, M. A., Ohnsorge, F., & Yilmazkuday, H. (2023). *Understanding the Global Drivers of Inflation: How Important are Oil Prices?* (CEPR Discussion Papers). C.E.P.R. Discussion Papers.

Handlan, A. (2022). Text shocks and monetary surprises: *mimeo*.

Hansen, S., & McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, *99*, S114–S133.

Hubert, P., & Labondance, F. (2021). The signaling effects of central bank tone. *European Economic Review*, *133*, 103684.

Ihrig, J., & Wolla, S. A. (2020). The feds new monetary policy tools. *Page One Economics*, 11.

Istrefi, K., Odendahl, F., & Sestieri, G. (2021). *Fed communication on financial stability concerns and monetary policy decisions: Revelations from speeches.* (Working Paper No. 2110). Banco de España.

Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*, *61*(4), 1281–1317.

Laubach, T., & Williams, J. C. (2003). Measuring the natural rate of interest. *The Review of Economics and Statistics*, *85*(4), 1063–1070.

Loughran, T., & Mcdonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, *66*(1), 35–65.

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, *2*(2), 159–165.

Mazis, P., & Tsekrekos, A. (2017). Latent semantic analysis of the fomc statements. *Review of Accounting and Finance*, *16*(2), 179–217.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *International Conference on Learning Representations*.

Miranda-Agrippino, S., & Rey, H. (2020). U.s. monetary policy and the global financial cycle. *The Review of Economic Studies*, *87*(6), 2754–2776.

Nechio, F., & Wilson, D. J. (2016). How Important Is Information from FOMC Minutes? *FRBSF Economic Letter*.

Obstfeld, M. (2020). Global dimensions of u.s. monetary policy. *International Journal of Central Banking*, *16*(1), 73–132.

Ochs, A. C. R. (2021). *A new monetary policy shock with text analysis* (Working Paper). Faculty of Economics, University of Cambridge.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Porter, M.-F. (2001). *Snowball: A language for stemming algorithms*.

Rehurek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, *3*(2).

Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, *84*(100), 25–40.

Rohlfs, C., Chakraborty, S., & Subramanian, L. (2016). The effects of the content of FOMC communications on US treasury rates. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2096–2102.

Romer, C., & Romer, D. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, *94*(4), 1055–1084.

Sachs, J. D. (1985). The dollar and the policy mix: 1985. *Brookings Papers on Economic Activity*, (1), 117–197.

Shapiro, A., & Wilson, D. (2021). Taking the fed at its word: A new approach to estimating central bank objectives using text analytics. *The Review of Economic Studies*, *89*(5), 2768–2805.

Sharpe, S. A., Sinha, N. R., & Hollrah, C. A. (2022). The power of narrative sentiment in economic forecasts. *International Journal of Forecasting*.

Shibamoto, M. (2016). *Empirical Assessment of the Impact of Monetary Policy Communication on the Financial Market* (Discussion Paper Series DP2016-19). Research Institute for Economics & Business Administration, Kobe University.

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, *28*(1), 11–21.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, *39*, 195–214.

Taylor, J. B. (1999). A historical analysis of monetary policy rules. In *Monetary policy rules* (pp. 319–348). University of Chicago Press.

TerEllen, S., Larsen, V. H., & Thorsrud, L. A. (2022). Narrative Monetary Policy Surprises and the Media. *Journal of Money, Credit and Banking*, *54*(5), 1525–1549.

Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, *38*(2), 393–409.

Wu, J. C., & Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, *48*(2), 253–291.

Appendices

A. Preprocessing

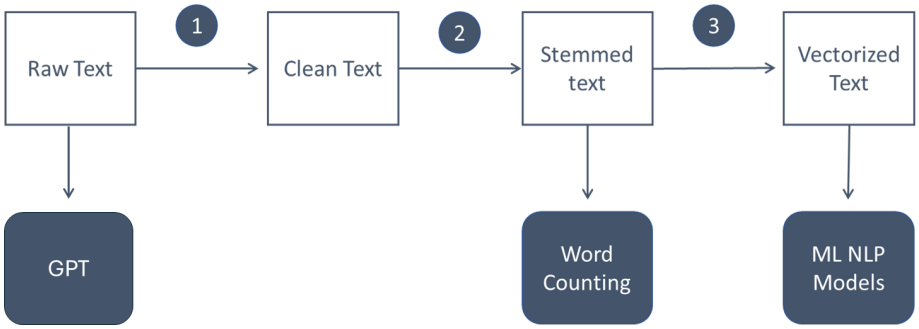

Figure 4: *Preprocessing of our text data*

(1) Cleaning. Stopwords and numbers removal
(2) Stemming. The text is ready to be used as input for Word Counting models.
(3) Vectorization. The text is ready to be used as input for Machine Learning models.

B. Selecting international-related phrases

1. We sort all words, doubles (sets of two words) and triples (three words) in the documents by their number of occurrences.
2. We only keep those related to international economic matters.
3. We limit ourselves to phrases above a frequency threshold (above 460 occurrences for words, 230 for doubles, 115 for triples).
4. For phrases contained in or containing other phrases, we select either the containing phrase or the contained: *foreign economies*, for example, is contained in *advanced foreign economies* and contains *foreign*, and we settle on including *foreign economies*.

C. Vectorization Methods

In this section, we give more information on the workings of the two vectorization methods we use, TF-IDF and Word2Vec. Both methods are implemented in standard python libraries [17].

C.1. TF-IDF

TF-IDF vectorization based on corpus D transforms a sentence d into a vector of size n (the number of different words in the corpus) by including for each word t in the corpus a measure of its importance in the sentence: its *tfidf(t, d, D)*. An example of this process can be found in Table **??**. *tfidf(t, d, D)* is the combination of two measures: Term Frequency, put forward by Luhn, 1958 and Inverse Document Frequency, first devised by Sparck Jones, 1972. For a term t in a document d, in a corpus D, $tfidf(t,d,D) = tf(t,d) \times idf(t,D)$.

- $tf(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$ is the relative frequency of term t within document d, where $f_{t,d}$ is the raw count of term t in document d.
- $idf(t,D) = log \frac{|D|}{|\{d \in D : t \in d\}|}$ is the inverse document frequency of term t in corpus D, where $|D|$ is the total number of documents in the corpus[18].

As a consequence, for a given word, high weight in TF-IDF is reached by a high term frequency (in the given document) and a low document frequency of the term in the whole collection of documents; TF-IDF vectorization tends to filter out common terms.

Sentence 1: Inflation today is very high and lasting.
Sentence 2: Inflation today is not high and is temporary.
Sentence 3: Inflation today is entrenched and dangerous.

Term	Count			TF			IDF			TF-IDF		
	1	2	3	1	2	3	1	2	3	1	2	3
Inflation	1	1	1	1/7	1/8	1/6	0	0	0	0	0	0
today	1	1	1	1/7	1/8	1/6	0	0	0	0	0	0
is	1	2	1	1/7	1/4	1/6	0	0	0	0	0	0
very	1	0	0	1/7	0	0	0.477	0.477	0.477	**0.068**	0	0
high	1	1	0	1/7	1/8	0	0.176	0.176	0.176	**0.025**	**0.022**	0
and	1	1	1	1/7	1/8	1/6	0	0	0	0	0	0
lasting	1	0	0	1/7	0	0	0.477	0.477	0.477	**0.068**	0	0
not	0	1	0	0	1/8	0	0.477	0.477	0.477	0	**0.060**	0
temporary	0	1	0	0	1/8	0	0.477	0.477	0.477	0	**0.060**	0
entrenched	0	0	1	0	0	1/6	0.477	0.477	0.477	0	0	**0.080**
dangerous	0	0	1	0	0	1/6	0.477	0.477	0.477	0	0	**0.080**

Table 9: *TF-IDF Vectorization example*

[17] We use the sklearn package of Python, developed by Pedregosa et al., 2011 for TF-IDF and the gensim package of Python, developed by Rehurek and Sojka, 2011 for Word2Vec.

[18] The denominator (number of documents where the term t appears) is commonly adjusted to $1 + |\{d \in D : t \in d\}|$ to avoid division by 0 if the word is not in the original corpus.

C.2. Word2Vec

Word2Vec, published in Mikolov et al., 2013 is a word embedding method: it transforms words into vectors by using a neural network to infer word meanings from a corpus of text. The method generates a vector space (generally of several hundreds dimensions), where, ideally, words representations (*embeddings*) are positioned to be close in the space (in terms of cosine similarity) if they are close in meaning.

In the vector space, it is possible to infer the semantic proximity of two words from the distance between their representations in the vector space, usually determined with cosine similarity, i.e. the cosine of the angle between the two vectors. The vector space also means that the representation can substract or add vectors to each other: in that representation, *king - man + woman = queen*.

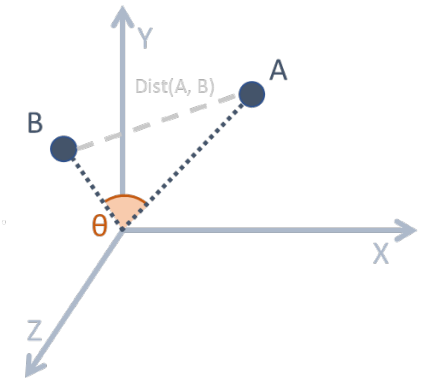

Figure 5: *Distance between two vectors in Word2Vec representation: an example*

Here, cos(θ) is the distance between A and B.

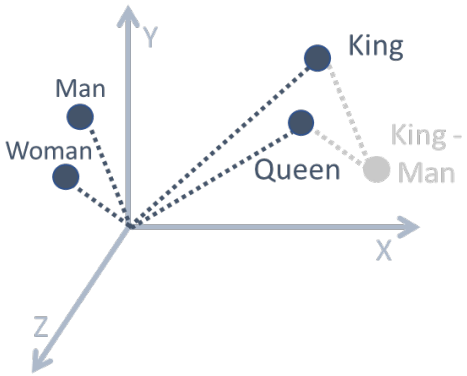

Figure 6: *Vector composition in Word2Vec representation: an example*

The Word2Vec technique can use one either continuous bag-or-words (CBOW) or continuous skip-gram to produce the representation of a word in the vector space. Both methods use a 3-layer neural network, and consider both individual words and the context words surrounding them. They differ in that in the CBOW method the model predicts the current word from its surrounding words (without taking into account the order of the context words, hence *bag-of-words*) while the continuous skip-gram uses the current word to predict its surrounding context words (and weights context words differently based on their place in the text).

Words which are semantically similar (and should be used in similar contexts), will usually share a large part of their context words and have similar influences on the relative word probabilities in the model, meaning that they will therefore be positioned in similar embeddings. That explains that semantically similar words are represented by similar vectors.

D. Supervised Machine Learning algorithms

In this section, we give an overview [19] of the theory behind the machine learning models we used in Section 3.2.3, namely Naive Bayes Classifier and Random Forest Classifier.

Both models are supervised learning, which means that the target Y is given. If we represent the model by h(·), and assume that the objective function to minimize is the Mean Squared Error (MSE), training the model consists in determining the parameters β minimizing the error.

$$\begin{aligned} \min_{\beta} \quad & \text{ERR}(X, Y, \beta) \\ \text{(MSE)} \quad & = \frac{1}{m}\sum_{i=1}^{m}(h(x_i, \beta) - y_i)^2 \end{aligned}$$

In the case of a classification algorithm, Y is a vector of 0s and 1s, and the function h(·) returns 0 or 1.

D.1. Naive Bayes

The Naive Bayes Classifier, often taken as a baseline to evaluate more complex methods (Chakraborty and Joseph, 2017), is based on applying the Bayes' rule of conditional probability to determine the probability that observation c_i is a part of class c_i. In computing the probability, the assumption that all features are independent from each other is made, hence the *Naive* denomination. In a situation with C classes (in our problem, C = 2), $c_1, ..., c_C$, the classifier is based on the maximising-a-posteriori (MAP) decision rule :

$$\begin{aligned} h(x_i) &= \underset{c}{\text{argmax}}\, P(c|x_i) = \underset{c}{\text{argmax}}\left(\frac{P(x_i|c)P(c)}{P(x_i}\right)) \\ \textit{(Independence)} &= \underset{c}{\text{argmax}}\, P(c)\prod_{j=1}^{n} P(x_{i,j}|c) \end{aligned}$$

$P(c_i)$ is the probability of being assigned to a class (and therefore is constant).
The $P(x_{i,j}|c)$ are the class-conditional likelihood factors for x_i being labeled c according to feature j [20].

D.2. Random Forest

The random forest classifier operates by constructing multiple decision trees and aggregating their output.

The idea behind the working of a decision tree is to divide the dataset X based on the x_i, j features of the x_i to minimize the entropy $H(Y|X)$ within areas of Y defined by the tree, which acts as the objective function in classification problems. This is done through successive identifications

[19] Among multiple others, a more complete explanation of the working of these models can be found in Chakraborty and Joseph, 2017, Section 3.

of the feature x of X which, when used as the split criterion, leads to the highest information gain $I(Y|x)$. The decision tree model consists of the sequence of splitting rules and the final value assignment defined.

$$I(Y|x) = H(Y|X) - \sum_{v} \in x \frac{|X_v|}{|m|} H(Y|X_v)$$

$$H(Y|X) = -\sum_{c=1}^{C} p(Y = c|X) log(p(Y = c|X))$$

$p(Y = c|X)$ is the relative frequency of class c observations in X.

While they can be very powerful and can handle complex relations within data, on top of having the advantage of being relatively interpretable, decision trees can be prone to severe over-fitting. Indeed, the learning method can lead the model to learn very specific patterns of the training dataset, which leads to high variance in out-of-sample testing, and therefore large errors.

Random forest models, developed to alleviate this issue, are based on separately training a number of roughly independent tree models from randomly selected subsamples of X, and make the classification decision based on majority voting of all different trees. Figure 7 illustrates this process.

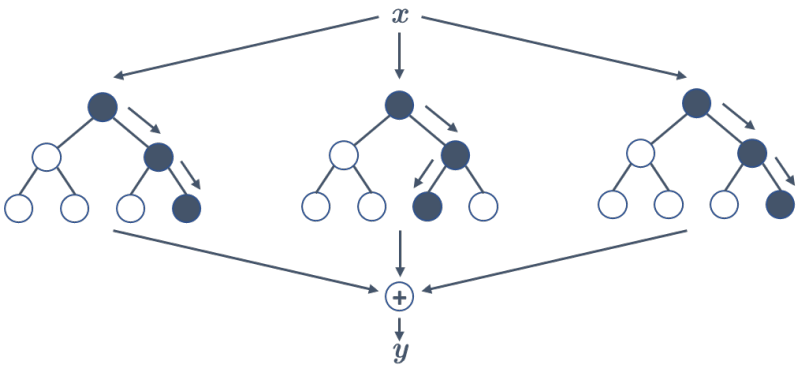

Figure 7: *Random Forest*

Even though trees in the forest are not independent from each other (as they as constructed from subsets of the same data), the variance of the final prediction is expected to decline compared to a single decision tree, therefore making random forest a reliable choice for classification.
While they are less interpretable than decision trees[21], as a consequence of the randomness inherent to the construction of the model, random forests can be seen as a manifestation of the *wisdom of the crowd* phenomenon, putting crowd intelligence into artificial intelligence.

[21] It can be possible, by studying which branch splittings are performed by the trees, to extract information about the most important features, and even economic rationale (Chakraborty and Joseph, 2017).

E. Counts by phrase in the international dictionary

	Number of Occurences
dollar	1415
currency	932
foreign exchange	488
foreign currencies	355
euro	304
depreciation	206
pound	36
renminbi	28
sterling	17
ruble	3
imports	1782
exports	1207
international trade	179
external sector	43
global	693
international	459
abroad	377
japan	363
foreign economies	314
china	278
emerging market	262
europe	201
EME	188
asia	147
world	111
russia	54
advanced economies	43
india	17

Table 10: *Counts by phrase.*

F. Dictionary count index with variations in the dictionary

To ensure that our results are robust to small changes in the dictionary, we study whether the random exclusion of 3 of the 28 phrases in the dictionary yields different indexes.

	Missing Phrases
Alternative 1	dollar, asia, EME
Alternative 2	abroad, foreign currencies, currency
Alternative 3	sterling, emerging market, EME
Alternative 4	sterling, foreign exchange, ruble
Alternative 5	yuan, india, advanced economies

Table 11: *Missing phrases for each of our alternative dictionaries*

Table 12 shows the Pearson correlation between indexes obtained with the reference dictionary and the 5 alternatives. What is shown is that all correlations are high (in the 90s), and even extremely high (above 98 %) for alternatives 3 to 5. Alternative 1 has the lowest correlation to other indexes, but this stems from the fact that *dollar*, the phrase with the second-highest count (Table 10), has been excluded, as well as *Asia*, which explains why this index is not able to catch the start of the late 90s Asian crisis while the other alternatives manage to do it, as shown in Figure 8.

	Basis	1	2	3	4	5
Basis	1.000000	0.922090	0.934825	0.982972	0.986841	0.999289
1	0.922090	1.000000	0.860623	0.911276	0.926918	0.919711
2	0.934825	0.860623	1.000000	0.904087	0.938426	0.931466
3	0.982972	0.911276	0.904087	1.000000	0.952997	0.986321
4	0.986841	0.926918	0.938426	0.952997	1.000000	0.983821
5	0.999289	0.919711	0.931466	0.986321	0.983821	1.000000

Table 12: *Pearson correlation of headline dictionary count (Basis) index and its 5 alternatives*

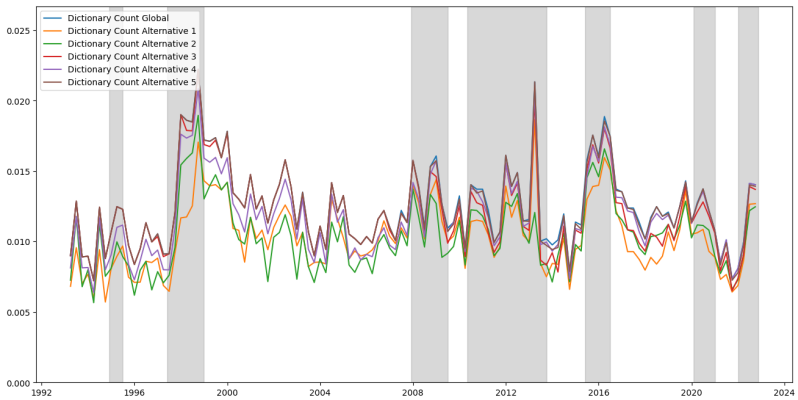

Figure 8: *Headline dictionary count index and its alternatives*

We conclude from Table 12 and Figure 8 that our dictionary count index is robust to small changes in the dictionary, even though more significant changes (exclusion of words with higher counts) can lead to slightly more different results.

G. Classification of paragraphs

We test the naive dictionary classification method on paragraphs, i.e. we classify paragraphs as positive or negative based on whether or not they contain one of the phrases in the international dictionary described in Table 2. We then compute the ratio of words in paragraphs classified as positive on the total number of words, and obtain the index featured in Figure 9.

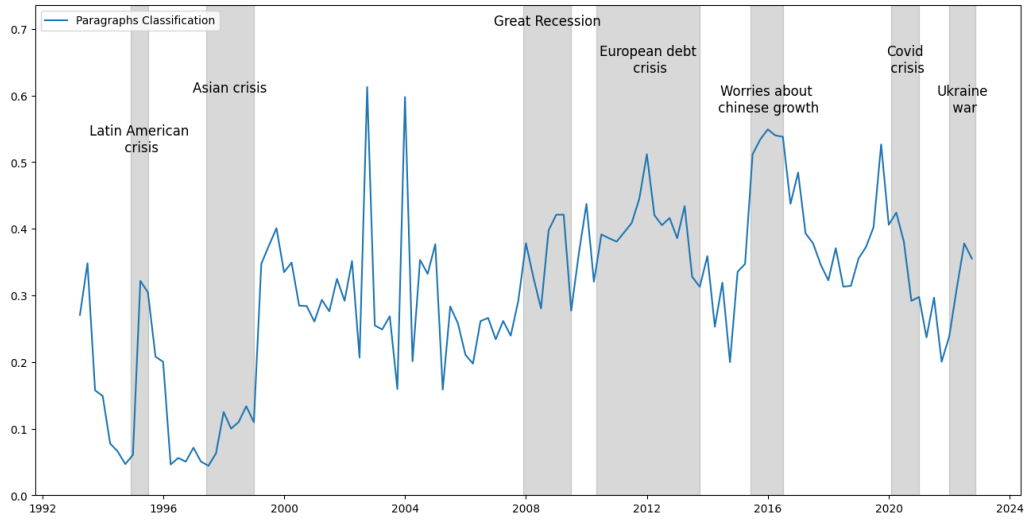

Figure 9: *Index of Fed attention to international, computed via paragraphs classification.*

A quick comparison with the naive dictionary classification on sentences shows that this classification option is less promising. Indeed, the highs of the index are close to 60%, with a mean value around 30%, which would mean that about half of the minutes conversation is directly related to international matters (something that contradicts basic assumptions about the role of the Fed). This might be due to the size of minutes paragraphs, which can start with discussion of domestic matters and end with an international situation briefing. Therefore, classifying the whole paragraph as related or not to international matters does not make economic sense. We also observe shocks in the value of the index, for example at the start of 1999. These jumps do not seem to be related to international events, but rather to structural breaks in minutes structure and paragraph size (since we use the paragraphs delimitations provided on the Fed website).

H. Sentiment around international-related words

Against our background, it also tempting to extract a sentiment around some international topics evoked during FOMC meetings. Sentiment indexes have proved extremely useful for economists to assess the tone of large textual databases (see Algaba et al., 2020), including for economic forecasting (Barbaglia et al., 2022). Here, we investigate evidence of a negative bias of Fed attention to the international economic situation, in the sense that FOMC members are likely to discuss more about international situation during crises. In this respect, we follow a methodology used by Aruoba and Drechsel, 2022: for a given set of phrases (our international dictionary in Table 2), we isolate the 10 words [22] surrounding each occurrence of any of the phrases and compute the "sentiment" (positivity or negativity) of these words using the Loughran and Mcdonald, 2011, dictionary. The sentiment around a phrase is estimated as the number of positively connoted words around its vicinity minus the number of negatively connoted ones. *In fine*, for a given FOMC meeting, a final international sentiment score is obtained by summing up the sentiment scores of all phrases contained in the related document. Given the structure of our dictionary in Table 2, it is also possible to compute a sentiment index for the three categories, namely *Currency*, *Trade* and *Others*. Our estimated sentiment indexes are presented in Figure 10.

By eyeballing Figure 10, we note that the global international sentiment estimated from FOMC minutes seems to be relatively well balanced until the Great Recession. However, starting from 2009, the sentiment index is clearly tilted to the downside. We observe large negative spikes during various phases such as the European sovereign crisis, the period of rising concerns about the Chinese economy or the Covid pandemic. The sequence of international crises seems to have introduce a negative bias in the attention of FOMC members.

[22] As a robustness check, we test the size of the "halo", i.e. the number of surrounding words taken into account. Results remain extremely similar.

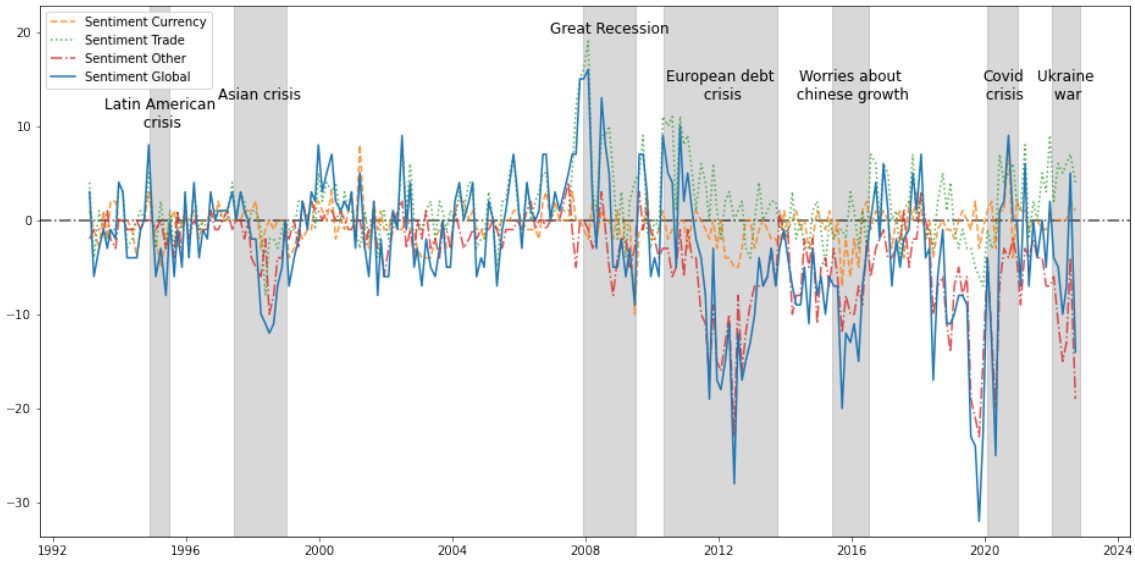

Figure 10: *Sentiment around international-related words, for the global dictionary and each of our 3 channels (10 words surrounding each international-related word).*

Breaking down the global sentiment into the three categorical sentiments, also featured on Figure 10, shows that the sentiment related to category *Others* is driving the global dynamics. The currency-related international sentiment exhibits much less volatility and seems overall balanced. However, we note negative values during the period of concern about Chinese economic growth in line with rising uncertainties at that time on the composition of the panel of currencies against which the renminbi was managed by Chinese authorities. The trade-related international sentiment does show a clear pattern and appears well balanced over time. Interestingly, the trade war launched in 2018 by the Trump's administration seems to positively affect the sentiment index.

I. Results of Taylor Rule Regression with time-varying r^*

In this section, we estimate a different version of the Taylor Rule which differs from Equation 5 in that α is replaced by a time varying neutral interest rate.

$$r_t = \rho r_{t-1} + (1-\rho)((r_t^* + \pi_t) + \gamma(\pi_t - \pi_t^*) + \beta g_t + \delta i_t) \tag{7}$$

We take as r_t^* the Laubach-Williams model 1-sided estimate, which is based on Laubach and Williams, 2003 and updated by the Federal Reserve Bank of New York.

Using the same non-linear least square method as in Section 4, we obtain the following results.

Dependent variable	FFR	FFR (TVR)	
Model	(Cal)	(0)	(2)
r_{t-1}	0.85	0.913***	0.919***
		(0.027)	(0.025)
α	4.0		
$PCEGap_t$	1.5	0.567	0.537
		(0.808)	(0.82)
g_t	1.0	0.828***	0.604***
		(0.191)	(0.197)
i_t			-8.034**
			(3.569)
Observations		115	115
Period		1994-2022	1994-2022
AIC		117.029	107.8
MSE		17.377	15.76
Taylor Version		10	13

Table 13: *Results of Taylor Rule Regression with time-varying R^**

Note: The table shows time-varying R^* Taylor Rule (Equation 7) results for Non-Linear Least Squares Regression, with different inputs as the international attention index i_t: (0) is no international environment index, (2) the index from random forest classification of sentences. The (Cal) column features the calibrated parameters (from the Atlanta Fed).
* p<0.1; ** p<0.05; *** p<0.01.

The results are globally on par with those in Table 6: the coefficient our IAI is negative and significant (although less than in the standard version of the Taylor rule) and the conclusions of Section 4 hold in terms of the effect of international attention. Yet, Table 13 shows that the PCE gap is never statistically significant.

J. Results of Taylor Rule Regression with unemployment gap

Dependent variable		FFR	
Model	(Cal)	(2)	(2.3)
r_{t-1}	0.85	0.899***	0.935***
		(0.026)	(0.024)
α	4.0	3.592***	2.94***
		(0.478)	(0.715)
$PCEGap_t$	1.5	1.142	2.244
		(1.05)	(1.724)
g_t	1.0	0.75***	
		(0.231)	
i_t		-8.067***	-13.955**
		(2.771)	(5.656)
u_t			0.092
			(0.209)
Observations		104	104
Period		1994-2019	1994-2019
AIC		86.072	91.334
MSE		12.413	13.057
Taylor Version		12	15

Table 14: *Results of Taylor Rule Regression with unemployment gap*

Note: The table shows Taylor Rule regression (Equations 12 and 15) results for Non-Linear Least Squares Regression, with the same international attention index and two options for the economic slack variable : (2) uses the output gap, (4.3) the unemployment gap (from the Congressional Budget Office). The (Cal) column features the calibrated parameters (from the Atlanta Fed).
* p<0.1; ** p<0.05; *** p<0.01.

Results in Table 14 show that including the unemployment gap as the slack measure in an augmented Taylor Rule does not lead to improve performances in terms of AIC and MSE. In such a model, the unemployment gap does not seem to be statistically significant.

K. Different versions of Taylor Rule Regression Used

In the equations below:

- r_t is the Fed Funds Rate (or the Shadow Rate if relevant)
- π_t is the (core PCE) inflation rate
- π_t^* is the inflation target
- g_t is the output gap
- ρ is a smoothing term for the FFR
- α is a constant playing the role of the neutral interest rate
- R_t^* is the time-varying neutral interest rate
- i_t is our International Attention Index (*IAI*)
- u_t is is the unemployment gap
- s_t is a variable derived from an equity index
- v_t is a variable derived from a risk index

K.1. Non-augmented versions

$$r_t = \pi_t + R_t^* + \gamma(\pi_t - \pi_t^*) + \beta g_t \tag{8}$$

$$r_t = \rho r_{t-1} + (1-\rho)(\alpha + \gamma(\pi_t - \pi_t^*) + \beta g_t) \tag{9}$$

$$r_t = \rho r_{t-1} + (1-\rho)((R_t^* + \pi_t) + \gamma(\pi_t - \pi_t^*) + \beta g_t) \tag{10}$$

K.2. Augmented versions

$$r_t = \pi_t + R_t^* + \gamma(\pi_t - \pi_t^*) + \beta g_t + \delta i_t \tag{11}$$

$$r_t = \rho r_{t-1} + (1-\rho)(\alpha + \gamma(\pi_t - \pi_t^*) + \beta g_t + \delta i_t) \tag{12}$$

$$r_t = \rho r_{t-1} + (1-\rho)((R_t^* + \pi_t) + \gamma(\pi_t - \pi_t^*) + \beta g_t + \delta i_t) \tag{13}$$

K.3. With Unemployment Gap

$$r_t = \pi_t + R_t^* + \gamma(\pi_t - \pi_t^*) + \beta u_t + \delta i_t \tag{14}$$

$$r_t = \rho r_{t-1} + (1-\rho)(\alpha + \gamma(\pi_t - \pi_t^*) + \beta u_t + \delta i_t) \tag{15}$$

$$r_t = \rho r_{t-1} + (1-\rho)((R_t^* + \pi_t) + \gamma(\pi_t - \pi_t^*) + \beta u_t + \delta i_t) \tag{16}$$

K.4. With additional control variables

$$r_t = \rho r_{t-1} + (1-\rho)(\alpha + \gamma(\pi_t - \pi_t^*) + \beta g_t + \delta i_t + \sigma s_t) \tag{17}$$

$$r_t = \rho r_{t-1} + (1-\rho)(\alpha + \gamma(\pi_t - \pi_t^*) + \beta g_t + \sigma s_t) \tag{18}$$

$$r_t = \rho r_{t-1} + (1-\rho)(\alpha + \gamma(\pi_t - \pi_t^*) + \beta g_t + \delta i_t + \mu v_t) \tag{19}$$

$$r_t = \rho r_{t-1} + (1-\rho)(\alpha + \gamma(\pi_t - \pi_t^*) + \beta g_t + \mu v_t) \tag{20}$$