Name: Predict Financial Market by analyzing speech by central banks

Team Members:
Keerthan Ugrani (ugranikeerthan6@gmail.com)
Anusha Chattopadhyay (anusha.chattopadhyay@stud.uni-heidelberg.com)
Renuka Sahani (renuka.sahani@stud.uni-heidelberg.com)
Shruti Ghargi (shruti.ghargi@gmail.com)

Project Source: https://challengedata.ens.fr/participants/challenges/70/
GitHub repo:

Motivation:

Since the last global economic crisis, central banks around the world have given a briefing to the public regarding the financial and economic situation (Ianthi). These briefings are primarily followed by financial actors, and therefore have a strong influence on the financial markets and also the overall economy. Central bank speeches can impact heavily on key economic factors such as interest rates, monetary policy, credits, debts, and inflation. Ability to decipher and understand the briefing made by the central banks has become a key area of research for all financial analysts and also for the quick analysis of how financial markets react to the briefings made by the central banks.

Hence, we want to solve the problem of analyzing the speeches and understanding how financial markets react when central bankers deliver official speeches. The project is primarily relevant to the data science topic in text analysis as we are using the BERT-transformed data of the speeches delivered by the central banks, and giving input to the problem. The output is the mean price evolution of a collection of 39 different time series; these time series correspond to 13 different markets measured at 3 different time scales. The exact dates of the speech and the contents of the speech are kept anonymous for obvious reasons.

Research topic summary:
Central bank communication either through their statements or by their speeches aims at increasing the financial stability by informing the markets about the current economic situations. They also share their medium term and short term predictions on the growth of economy. These speeches will impact in the financial markets on exchange rate and interest rates or can provide signals about the evolution of stock market indices.

In recent years with the advancement of machine learning and data Natural Language Processing is gaining ground as a tool for performing analysis. In this study we research through the processes transformed data along with the time series data of financial markets. There are number of research on the analysis of speeches to predict the sentiment of the financial market of limited countries, but here in this study of research we aim to provide the global sentiment index aggregating speeches from around 39 central banks around the world and also the impact of those speeches on the financial markets around the world, providing a global surveillance tool for monitoring the evolution of the global economy.

In recent years, numerous studies have performed text-mining analysis in order to define the trends in the official central bank's communications and respective financial market variables (Hamza #). There are some set of studies where central bank's communication reaction is quantified into dummy classification based on the subjective assessment of its content by the researcher. (Graeme and Wright and Julian #) use central bank communications to show how central bank statement can be used to implement monetary policy in New Zealand. The communication is classified based on the subjective assessment of the authors. Show that the average market return and sharpe ratio are significantly higher on important macroeconomic announcement days, that is the analysis is based on the importance of the speech based on dummy variable for the presence or absence of the speech.

The issues or the drawbacks with the current and above research works are, the classification of the communication is subjective and can thus vary depending upon the researcher as well as on the objective of the study, and the study focuses only on the event of speech ignoring the actual content of the speech and its importance. We propose a way to classify objectively based on the quantifying the sentiments. We hope the proposed methodology and the problem statement marks the improvement in the current methods of textual analysis in finance. In this study we explore the application of machine learning algorithms on the natural language processing, which will quantify the sentimental analysis and creating the global sentimental index We afterwards use this index in addressing the impact on the markets in the next two weeks of central bank speeches.

Project Description:

The contribution to the literature about central bank communications and their impact on the financial markets are, we provide the methodological contribution by directly quantifying the information in communication using NLP and machine learning techniques. We build a pipeline of models and techniques that fully automate the process of text data cleansing, transforming, filter and provide adaptive global index by machine learning models that assign relevance and sentiment score. We can also provide some sub-sentimental indices for capturing several events and index using a set of machine learning and comparing them in based on the metrics provided by the dataset provider and find the best one.

The sub-tasks of the project includes the understanding as well as processing the given data. The BERT transformer does all the preprocessing like conversion of text to the same case, removing the punctuations, removing all the stop words. Then we can use this to train the model using the machine learning techniques such as XGBoost or random forest in classification of the speech.

The data is taken from one of the challenge hosted by Natixis on Challenge data platform (https://challengedata.ens.fr/participants/challenges/70/). They have provided the data such that the speeches are processed by a pre-defined BERT-style transformer, and this gives the input of the problem. The output is the mean price evolution of a collection of 39 different time series; these time series corresponds to 13 different markets measured at 3 different time scales. They have computed the difference between the closing prices of these 13 markets at 3 different maturities and the price of these markets at the closing time of the date of speech. The data is free from short time effects (betwwen the beginning of the speech and the closing of the same day) and leakingb effects (trading occurring because of information leakage before the beginning of the speech). An indication that if a speech has effect on the markets, it

seems to intervene before the end of 2 weeks lag to measure the possible effects on the markets based on few tests. The 13 markets are the following:

1. VIX
2. V2X
3. EURUSD
4. EURUSDV1M
5. SPX
6. SX5E
7. SRVIX
8. CVIX
9. MOVE
10. USGG2YR
11. USGG10R
12. GDBR2YR
13. GDBR10YR

The training data consists in 2000 transformed speeches and their subsequent market moves. More precisely:

1. x_train.csv is a (2024, 768) array. Each line represents one speech issued at some date by some central banker (FED or ECB); for obvious reasons, they are anonymous. They were transformed into a feature vector of dimension 768 using a Bert-style transformer. These transformer are known to be extremely successfull for conveying the semantics of a natural language text.
2. y_train.csv is a (2024, 39) array. Line i represents the market variations consecutive to the i-th speech whose transformed version is in the i-th line of the x_train file. These variations are relative.

The test data is as follows:

1. x_test.csv is a (415, 768) array. Each line represents one speech, at dates that are different than for the x_train

The metric data has been provided and has to be used for assessing the model.

Works Cited

Graeme and Wright, Guthrie, and Julian Julian. “Open mouth operations.” *Journal of Monetary Economics*, vol. 46, no. 2000, 2000, 489--516. *Open mouth operations*, https://www.sciencedirect.com/science/article/pii/S0304393200000350.

Hamza, Bennani. “Does People's Bank of China communication matter? Evidence from stock market reaction.” *Emerging Markets Review*, vol. 40, no. 2019, 2019, p. 100617. *Does People's*

Bank of China communication matter? Evidence from stock market reaction, https://www.sciencedirect.com/science/article/pii/S1566014118304539.

Ianthi, Vayid. *Central Bank communications before, during and after the crisis: from open-market operations to open-mouth policy*. 2013. *vayid2013central*, Bank of Canada Working Paper, https://www.econstor.eu/handle/10419/95684.