Sentiment Analysis of Federal Reserve Speeches for Bitcoin Price Prediction

Group: Citadel 2

FAN Sunyan,	3036017798
LI Yutong,	3036020214
LIU Ruiyang,	3036021440
PAN Haoyu,	3036021567
TAM Karim,	3035876745

Chapter 1: The Scope and Methods of Study

Introduction

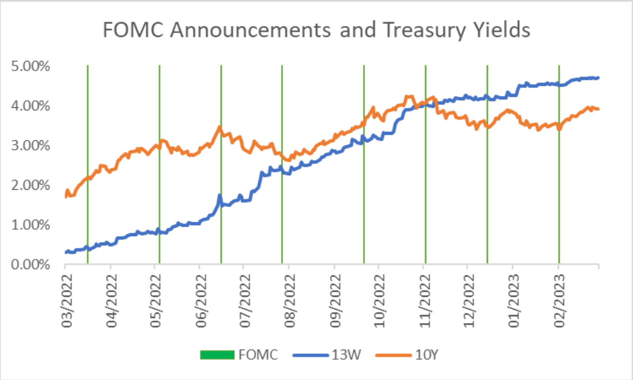
The cryptocurrency market has expanded rapidly, from 5 million owners in 2016 to over 300 million owners by 2021. However, it involves risk factors to invest in cryptocurrencies as the market can be subject to extreme volatility (Rejeb *et al*., 2021). For instance, in 2022, Bitcoin lost more than 60% of its value, putting investors in a difficult position. The cryptocurrency market has no central governing authority (Grant and Hogan, 2015), so the prices can fluctuate based on various factors, such as the sentiment of the public, natural disasters in a country, and global news. However, there are two benefits of cryptocurrencies that are important for potential investors, namely, diversification and inflation protection. As claimed by Corbet *et al.* (2019), cryptocurrencies can offer investors diversification from conventional financial assets such as stocks and bonds since the prices are not associated with other markets, making them a suitable source for portfolio diversification. Investing in a portfolio of assets with low price correlation will help investors produce more consistent returns. In addition, Bitcoin and other cryptocurrencies can provide protection against inflation. There is a hard cap for Bitcoin on the total number of coins ever minted (Farell, 2015). Thus, with the growth of the money supply exceeding the growth of the Bitcoin supply, the price of Bitcoin is supposed to rise. Accordingly, it is meaningful to study the price movements of Bitcoin.

The aim of the project is to investigate whether the sentiment of Federal Reserve speeches can produce accurate predictions about the future price fluctuations of Bitcoin. In order to address it, recurrent neural networks (RNN) and a dual-model approach are employed in the project. Specifically, the first step is to apply pre-trained datasets (FiQA and FinancialPhraseBank) for sentiment analysis to determine the sentiment score. Then, Bitcoin price fluctuations are determined by incorporating sentimental factors, which are positive and negative scores.

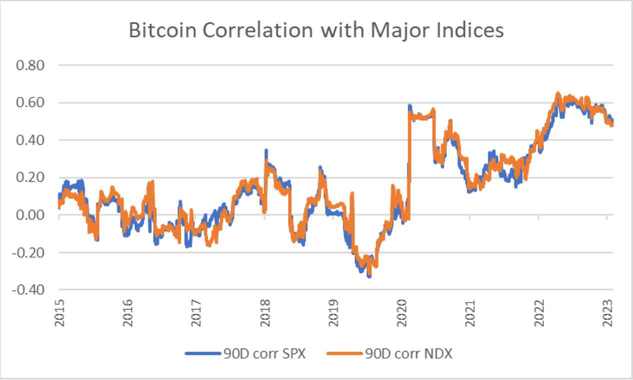
Background

It has long been established in both academia (Bernanke and Kuttner, 2005) and by practitioners that there is significant premium in trading around FOMC announcements. This is because such announcements contain valuable information on market factors, such as macroeconomic data and guidelines on upcoming policy. Such information is important to markets in many ways; for example, rates decisions affect the yield curve significantly, which is itself a basis upon which lending decisions are made and valuations are conducted.

Following an extended period of low interest rates throughout much of the past decade, and as economic conditions become volatile in the past few years, markets have been anticipating and trading around expected Fed policy decisions – this is evident in the 2021-2023 taper tantrum, and which has been a major contributing factor to the recent US equity market weakness. As non-FOMC meeting Fed speeches also often contain information regarding officials' expectations of macroeconomic and market conditions, and by extension, implicit policy guidelines, we expect that they also carry a similar premium as FOMC announcements. It is thus our goal to research whether this hypothesis stands true.



We believe there are two possible chains of reasoning where Bitcoin is affected by sentiments expressed in Fed speeches. Bitcoin has been touted by early adopters as an inflation hedge, as it was designed with a 21-million coin limit – meaning no new coins will be issued afterwards. Its "money supply" (and "monetary policy") is thus considered relatively contractionary compared to traditional monetary policies, as global money supply generally trends upwards. And up to 2020, this seemed true as there is empirical evidence that Bitcoin gains in positive shocks to expected inflation (Choi & Shin, 2022). After 2020, outside of idiosyncratic shocks, cryptocurrencies have exhibited increasing correlation with US equities – particularly the higher beta growth tech stocks, which are sensitive to inflation expectations and respond negatively to expected inflation shocks due to the increased discounts to their future earnings from higher rates.



Despite the seemingly contradictory nature of the two possible chains of reasoning, regardless of whether we believe Bitcoin is an inflation hedge, or that the trend of increasing correlation between cryptocurrencies and equities will continue, we should be able to analyze and forecast Bitcoin returns under a framework based on the implicit inflation and policy expectations in Fed speech sentiments.

Chapter 2: Data Collection and Manipulation

Data Collection and Pre-Processing

Official transcripts are uploaded to the Federal Reserve website after each speech. The website is simple in structure, and we were able to use the newspaper3k, BeautifulSoup and Selenium packages to scrape the speeches. As a source of text data, the concise and formal nature of the speeches meant that little cleaning was required.

Nonetheless, we still inspected and cleaned the data in case the downloaded text unexpectedly deviated from the actual speeches. We dropped references from the speech text, then removed useless characters and converted all letters to lowercase. Next, we summarized the article by removing stop words, calculating the relative frequencies of the words, then ranking each sentence by relative importance to remove insignificant sentences from the speech. We then split the passage into sentences using the NLTK sentence tokenizer, and removed any outlier or error before putting the data through the natural language processing models.

Our cleaned text data is as follows:

```
index article_time                                       Sentences_c
    0   2023-02-27  while we have all experienced prices going up,...
    1   2023-02-27  for example, the share of medical services is ...
    2   2023-02-27  as a result, when health-care services or hous...
    3   2023-02-27  the substitution effects captured by the pce p...
    4   2023-02-27  the inflation outlook for this nonhousing cate...
  ...          ...                                                ...
 5055   2014-11-07  however, the board also recognizes that limite...
 5056   2014-11-07  attention to a bank's practices in making its ...
 5057   2014-11-07  in addition, we revised our consumer complianc...
 5058   2014-11-07  billion in total consolidated assets.a second ...
 5060   2014-11-07  billion bank begin to develop capital stress t...
```

We also needed accurate cryptocurrency price data. Through the Yahoo Finance! API, we obtained the USD denominated Bitcoin prices (‘BTC-USD’) from Sep 2014 to Feb 2023, which aligns with the Fed speech text data. The API has multiple benefits – there is complete OHLC price data at daily intervals, the data is easy to obtain via Python (with the specific purpose-built package yfinance), and it is available for free. To control for USD price changes, we also collected USD index data from yfinance, which captures the strength of the dollar. Finally, we produced the data for other variables that we may use in our price prediction model, such as the 3-day moving average price for Bitcoin.

Chapter 3: Sentiment Analysis by RNN

Model Overview

To build a reproducible method for our applications on the cryptocurrency market, we decided to investigate recurrent neural networks (RNN) for modeling sequential data. Therefore, the normalized data is then split into two parts. One set, which accounts for 80% of the total, will be used for training the model while another set, the other 20%, will be used for validation.

From our perspective, simple models perform better than mixture of several other models because simpler models are less susceptible to overfitting.

As a robust type of neural network, RNNs can modelize dynamic processes, for example time series, so it is best suited for data like text, speech, financial data, etc. By using recurrent connections, it takes data from previous input-output pair to predict the output of the next one and is not subject to limited size of context. The key essence to RNNs’ success in sequential problems lies in that hidden layers can retain memory that represents contextual information. At each step, RNNs do a series of calculations before producing an output, in the form of the hidden state, which is then combined with the next input in the sequence to produce another output. This process continues until the model is programmed to finish or the input sequence ends.

In our work, the system pre-processes the training and validation set to RNN, which serves as the sentiment analyzer to provide sentiment score and sentiment prediction.

Input dimension	81
Embedding dimension	50
Hidden layer	128
Number of layers	2
Output dimension	3

Our model consists of 2 RNN layers, the word embedding of 50, hidden layer of 128 features and the max sequence length of 81 which is the median length of the FiQA training set. The embedding dimension is the word vector length due to Google's pretrained word2vec model. The label of training set is categorized into the 3 groups so that the output dimension of neural network should be 3 to make it a multiple classification problem. The number of layers is 2 and the number of hidden layers is higher than the embedding dimension. It means that there is more hidden information or features in a sentence that are not easily observed by interpreting the individual word meanings.

Model Training

Since using text representations that were pretrained on a vast corpus can reduce overfitting of the model, we applied pretrained datasets (FiQA and FinancialPhraseBank), to sentiment analysis.

The training set consists of 2 parts which are the sentences and the corresponding sentiment. The sentences are simply cleaned by the filter based on regular expression. The sentiment is transformed into one-hot vector in which [1, 0, 0] indicates the positive sentiment, [0, 1, 0] indicates the neutral sentiment and [0, 0, 1] indicates the negative sentiment.

The sentences are transformed into a one-dimensional vector in which every element is the index of word in the word-to-vector dictionary. The sentences are either cut or filled with '0' which correspond to the 'unknown' word vector.

After the sentences are passed into the model, an embedding function is applied to transform the index of word to a word vector. Therefore, all input sentences should be two-dimensional matrixes of shape {81, 50}. Every hidden information of the former word is passed to the next word until the end of the sentence. Considering both training set and testing set are longer and formal sentences than the simple sentence structure that would appear in a twitter comments, it is more reasonable to design a many-to-many RNN model rather than a many-to-one model that uses the last word information as the output. Finally, all hidden information is aggregated by a fully connected layer to generate the output in the form of the probability of each sentiment.

Throughout the training process, two main metrics are monitored: accuracy and loss. Accuracy measures the percentage of correct classifications and loss measures how far off the model's prediction is from the ground truth. We then calculated the gradient of the loss function and updated the weights accordingly so that future computations with the input data will produce more accurate results.

The validation accuracy and loss are monitored throughout the training process to oversee how well the model performs on new data, and optimizer and scheduler are used to keep the learning rate low with increasing epoch and to make sure the model does not overfit the training data. Finally, we calculated and plotted the accuracy and loss statistics of the model.

Model Validation Results

The result is satisfactory because it achieves 85% accuracy after 70 epochs training on CPU. It proves that the result is robust enough for the FiQA and Financial Phrasebank dataset. It is observed that the loss curve is very smooth and the variance (fluctuation) of loss is kept low all the time. Also, the curve is not ultimately steep or flat which means that the model is learning at an appropriate rate. The model may be further improved by using bidirectional RNN model that will do once more backwards training for each sentence. However, it will at least double the training time and the efficiency-cost trade-off should always be considered.

Chapter 4: Price prediction by RNN

Choice of Variables

The Bitcoin data is daily data so the input for price prediction should be the frequently changing features that would infer the change the Bitcoin price. The sentimental factors are the positive and negative scores. The reason for not choosing neutral sentimental is the assumption neutral sentiment only gives a little positive impact on price. The positive sentiment indicates the future act of federal reserve will be positive so investors will be more optimisitic. The negative snetiment indicates the past performance has not reached expectations so that the federal reserve may also take positive action. The previous price is a benchmark for today's price. The volatility and trading volume to some extent reflects the limited attention of investors to Bitcoin's news, the willingness to invest in Bitcoins and the cost of arbitrage to withdraw from the current position in Bitcoin.

Model Overview

Input dimension	30
Embedding dimension	6
Hidden layer	64
Number of layers	1
Output dimension	1

The RNN structure for price prediction is the simpified version to that in sentiment prediction. The input shape is {30, 6} which means from t-30 to today's information is used to generate a price prediction for tomorrow (t+1). The past prices have two oppositive impact on the model's performance, it gives a more direct view of the price trend. And because of that, the model can easily be overfit on the past prices so that it may block all the information by setting weights on others features to zero and the prediction will be a straight line. This is due to the autocorrelation that the model fit on the residuals/noises that are not explained by the features. Therefore, after several trials we have increased the features from the embedding size of 6 to 64 hidden layers which is not extremely large to induce the curse of dimensionality.

Chapter 5: Strategy

The price prediction is not the end of the project as the investors invest in the return. Therefore, a naïve strategy is built based on the predicted price. If the predicted price for tomorrow is higher than the threshold (the percentage higher) of today's real price, then the investor should long the position in the Bitcoin, otherwise short the position. The benchmark is the long-term holding of Bitcoin from 01/07/2021 (dd/mm/yyyy) to 27/02/2023. It is observed from the cumulative return that no strategy beat the benchmark before 11/2021. After that and up to 01/2023, we saw Bitcoin prices drop, and during this period the zero-threshold strategy outperforms all other strategies – including the benchmark.

We believe the strong performance of the threshold = 0 strategy can be attributed to the fact that we are able to catch most of Bitcoin's moves. However, this means that the strategy will have a higher turnover, and exhibits higher volatility. In comparison, the threshold = 5 strategy is more conservative in nature, and only trades for significant predicted moves in Bitcoin prices. Naturally, this strategy will have a lower turnover and volatility.

Chapter 6: Further Research Ideas

Data Granularity

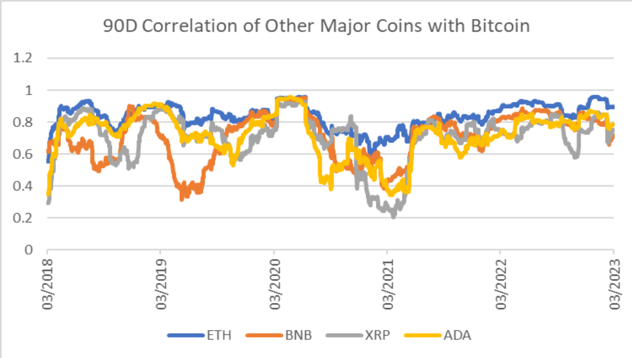
In the context of this group project, our aim is to investigate whether Bitcoin returns are correlated with fed speech sentiments in the short-medium term, and Yahoo! Finance data is enough for us to achieve our aim. Ultimately, we could benefit from data with a higher granularity; for example, if we are to build a trading strategy, immediate price action following each speech would be necessary. On this note, if we are to run this strategy in an institutional context, we must also incorporate considerations (and thus data) of trading volume and market liquidity into our algorithm. Is there enough liquidity in the market for this strategy to run profitably? Will slippage be a significant issue? These are some of the issues to be addressed.

Volatility Concerns

Predictions on the cryptocurrency market often encounter liquidity problems and can be quite problematic during crypto bubbles. For example, 2017 witnessed a boom on the cryptocurrency market where Bitcoin price started the year off by crossing $0.1k and culminated by hitting a record $20k. Also, we all remember the 2022 crypto crash where Bitcoin and other cryptocurrencies dropped significantly. We succeeded in avoiding overfitting in the project, but it would still be interesting to see how our model performs in a bearish market.

To avoid deviations caused by the volatile nature of cryptocurrencies, we suggest adjusting position sizing based on changing market conditions, or to train only on short periods before and after major price fluctuations, to understand these bubbles.

Underlying Choice



It is established that Bitcoin prices are affected by Fed speech sentiments, possibly due to its characteristics as an inflation hedge and high correlation with US equities. Using Bitcoin as the underlying asset, we have built a successful trading strategy around this discovery. Despite not sharing all the aforementioned characteristics with Bitcoin, many other major (non-stable coin) cryptocurrencies by market cap are all quite correlated with Bitcoin.

It is thus worth investigating whether these other cryptocurrencies' returns also respond to Fed speech sentiments, and if they do so with the same timing, direction and magnitude as Bitcoin. Given the fact that other cryptocurrencies are not studied to the same level as Bitcoin in academia and in industry, there is possibly extra premium by using other cryptocurrencies as the strategy underlying.

Generalization

The cryptocurrency market is very dynamic. Certain major cryptocurrencies may disappear, and some ill-known ones may become dominant. The market structure is very likely to have changed years later when someone refers to this report. Therefore, it is advisable to update the research object in relation to the evolution of the market structure.

Chapter 7: Conclusions and recommendations

In this study, we have developed, compared and selected deep learning-based Bitcoin price prediction models using Federal Reserve speeches and Bitcoin historical return. More specifically, we tested the state-of-the-art deep learning models such as recurrent neural networks (RNN), long short-term memory (LSTM) models, convolutional neural networks (CNN), and their combinations. It's heartening to find out in practice that our RNN model outperformed the other models by taking far less training time while ensuring high accuracy. We presented a general introduction on RNN architecture, defined features to feed into the model algorithm, and translated the resulting predictions into trading strategies. However, although RNN seems to predict the Bitcoin price very well, it is still premature to solely use it for algorithmic Bitcoin trading.

To build a more applicable portfolio trading strategy, we may need to further consider another type of machine learning called reinforcement learning, which simulates a real artificial intelligence by finding a balance between exploration and exploitation.

There are several research directions for future work. First, model selection involves multiple trials and errors, which need important computation power and time that we did not have in this project. Otherwise, other types of models can be tested. Besides, to improve the prediction accuracy, we could incorporate other variables such as major currency exchange rates, price fluctuations of major cryptocurrencies such as Ethereum and Ripple, as well as social media comments and replies.

In the future, we will perform sentiment analysis on other articles and speeches made by policymakers and explore their correlation with cryptocurrency price.

References

Corbet, Shaen, et al. "Cryptocurrencies as a financial asset: A systematic analysis." *International Review of Financial Analysis* 62 (2019): 182-199.

Farell, Ryan. “An Analysis of the Cryptocurrency Industry”, Wharton Research Scholars (2015), 130.

Grant, Gerry, and Robert Hogan. "Bitcoin: Risks and controls." *Journal of Corporate Accounting & Finance* 26.5 (2015): 29-35.

Rejeb, Abderahman, Karim Rejeb, and John G. Keogh. "Cryptocurrencies in modern finance: a literature review." *Etikonomi* 20.1 (2021): 93-118.

Bernanke, Ben S., and Kenneth N. Kuttner. "What explains the stock market's reaction to Federal Reserve policy?." *The Journal of finance* 60.3 (2005): 1221-1257.

Choi, Sangyup, and Junhyeok Shin. “Bitcoin: An inflation hedge but not a safe haven.” *Finance research letters vol. 46* (2022): 102379