

Does FED Communication cause immediate and abnormal returns stock market?

A sentiment analysis on the S&P 500 Stock Market Index

Matthias Farngruber, 11801413 Filip Lukijanovic, 11776896
Peter Prllehi, 11776041

Contents

Abstract	2
1. Literature Review	3
2. Data	3
3. Methodology	3
3.1 Sentiment Extraction - VADER Method	4
3.2 Sentiment Extraction - FinBERT Method	5
3.3 Financial Data Processing	5
3.4 Regression Sepcifications	5
4. Bayesian Model Averaging	15
4.1. A brief excursion into the stability of the Tone Effect over time	17
5. Potential for Future Research	18
6. Conclusion	19
7. Bibliography	20

Abstract

This research project conducts a topical analysis of sentiment analysis techniques applied to central bank statements, with a focus on the Federal Open Market Committee (FOMC) announcements. Initially, our approach mirrors the regression methodology of Möller and Reichmann (2021). However, we diverge from their examination of macroeconomic factors, shifting our attention to market factors that influence abnormal returns. Our investigation yields several key insights. First, we find a marked enhancement in the explanatory power when using heuristics-adjusted dictionary methods like VADER, especially in more uniform texts. Second, we emphasize the importance of considering the mechanical aspects of FOMC statements, such as interest rate announcements and their unexpected components, to avoid skewed analyses of central bank communications. Additionally, we uncover a nuanced sensitivity to these surprises during periods of high leverage. Ultimately, our research serves as a foundation for further studies on the intersection of natural language processing and financial markets, especially in the context of event studies.

1. Literature Review

The following research project is mostly based on the work done by Möller and Reichmann (2021) in the field of sentiment analysis. In their paper “ECB Language and Stock Returns - A Textual analysis of ECB Press Conferences” they explore the impact of the language used by the ECB in their regular press conferences on stock returns in the Euro Area.

They achieve this by mining the statements and ranking the sentiments of each statement based on general tone, uncertainty and constraint. Once this is classified, the researchers then cross-check the high frequency intraday data for Euro Area stock returns on each statement day, by employing the technique of an event study. This allows them to see how stock returns reacted to statements by the ECB at whatever point they wanted to examine them.

As mentioned, the researchers focused on tone, uncertainty and constraining language for the sentiment analysis. Tone captures the overall language - or the overarching narrative - of a statement, uncertainty measures how ambiguous a statement may be and constraining language quantifies how constraining the ECB communicates to be in the future. Of course the researchers did not read through every single statement by the ECB, instead, they employed a dictionary-based sentiment mining approach that considered grammatical and syntactical cues to analyze the sentiment expressed in ECB press conferences. Afterwards they scored each statement with regards to each category by employing heuristic-adjusted sentiment scores based on word lists used in previous studies. Möller and Reichmann (2021a) The Authors had used a plethora of controls in their regression analyses, an approach that we were more reluctant to follow due to our slightly smaller sample size. The main reason behind that is that our sample could hardly accommodate the sheer number of variables used in the initial analysis. Below, we delve into the subset of data used, and into our methodology for modelling Möller and Reichmann’s (2021) sentiment methodology. In addition, we explore whether transformer-based approaches that are bespoke to Federal Reserve Statements can outperform heuristics-adjusted sentiment approaches using dictionaries, i.e. our proprietary approach. We do so by investigating the explanatory power of regressions using each sentiment approach.

2. Data

The data used for this project consists of the following:

- Federal Reserve (FED) Meetings: The U.S. American FED holds regular meetings multiple times per year where they talk about the current economic situation and what the plan is going forwards. This leaves us with 74 statements over 10 years from January 30th 2013 until July 26th 2023.
- Standard & Poor’s 500 stock market index pricing: We cross-check the statements and their nature with the price of the S&P 500 stock market index on each day of a statement. This index is of particular interest since it incorporates 500 U.S. american companies representing a large part of the whole market itself. This means that we can use the S&P 500 as a proxy for how the market behaves at a given day. Alongside that we obtained weighted debt to equity ratios of the index to see if higher leverage leads to stronger reactions to sentiment. A third and final element of S&P 500 data was lagged returns that were used in some of our analysis
- Federal Funds Futures data: the rate or price equivalent of the interest rate that the market prices in for the next 30 days was obtained as well, as a proxy for market expectations on interest rates.

All of our data, save for the FOMC meeting minutes were obtained via the Bloomberg Terminal data services, whereas the FOMC minutes were manually extracted.

3. Methodology

Our analysis of equity returns following FOMC sentiment levels can be partitioned in two. The first part is sentiment extraction, with our approach closely following that of Möller and Reichmann (2021), as well as an implementation of Google’s FinBERT NLP transformer for an alternative gauge of the tone of FOMC minutes. The second part focuses on the construction of the variables needed in our regression, both core

variables as are equity returns and sentiment, as well as various controls inspired by the paper from Möller and Reichmann.

3.1 Sentiment Extraction - VADER Method

Our first approach to modelling the sentiment of FOMC minutes closely resembles that of Möller and Reichmann (2021), who use the NLTK VADER package in Python to extract sentiment. The VADER package is special insofar as it adjusts sentiment for heuristics, which gives it an advantage over conventional dictionary-based NLP methods (Hutto, 2014) Hutto and Gilbert (2014). Furthermore, sentiment can be modeled for intensity as well, with Sentiment $S \in [-4, 4]$. Möller and Reichmann (2021) expand the inbuilt dictionary of words and corresponding stand-alone sentiment scores by financial terms that correspond to mere Tone, Uncertainty, as well as Constraining language, so as to be able to capture different dimensions of speech.

Our approach to this was to update the lexicon in VADER as well, whereby our extraction of Tone merely necessitated the updating of the pre-existing lexicon. For Unc and Con, however, we used the libraries by Bodnaruk (2015) Bodnaruk, Loughran, and McDonald (2015) for constraining language and Loughran and McDonald (2011) Loughran and McDonald (2011) to construct scores for uncertain language and for extending the VADER dictionary by financial terms. To extract the nuances of uncertain and constraining language, the sentiment scores of all tokens not included in the respective external libraries received reduced sentiment scores, to 20% of their initial sentiment score. Meanwhile, the tokens from the libraries were weighted regularly, i.e. with sentiment scores $S \in \{2, -2\}$, as 2 is equivalent to moderate intensity in the VADER package. Adjusting every word's sentiment score manually for its intensity was not deemed feasible for the scope of a project.

Once these transformations took place, the FOMC data was tokenized into sentences, which were then tokenized for each constituent word. The sentiment score for every sentence is weighted by the word count in each sentence relative to the total word count in the statement. The sum of each sentence's weighted sentiment then constitutes the sentiment of the entire FOMC statement at a given date.

Finally, we constructed the three sentiment elements and saw their trajectory move as follows:

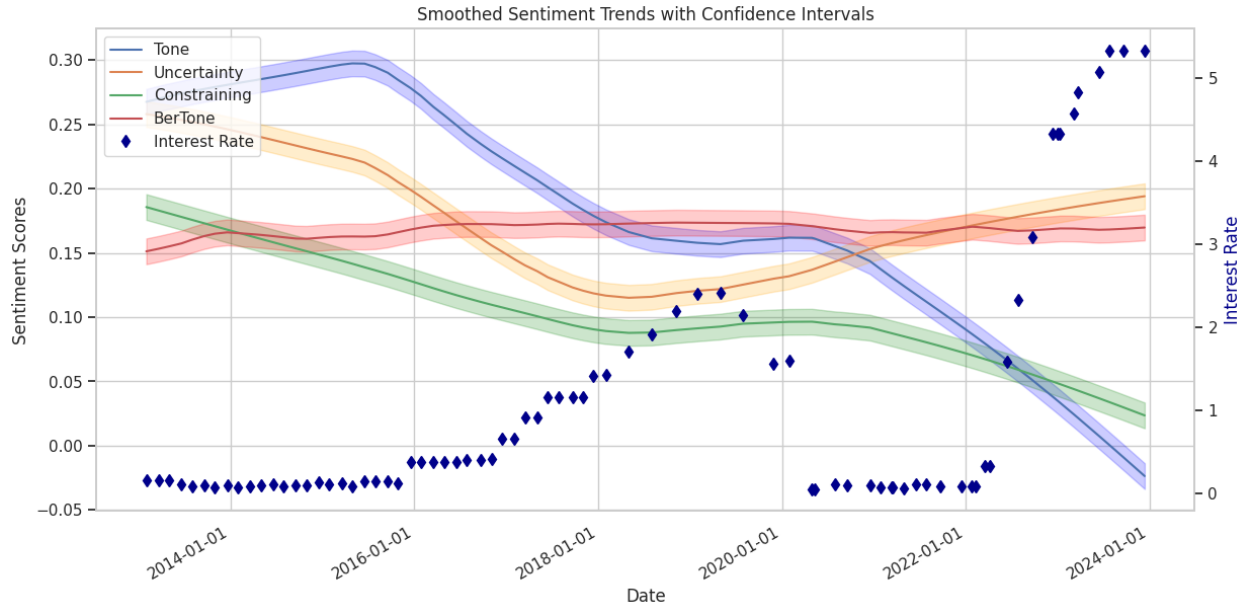


Figure 1: Sentiment Trajectory

In the plot above, we can see the trajectory of each sentiment element of FOMC minutes over time, along with the Fed's interest rate, offering some interesting insights into the interplay between monetary policy, forward

guidance and sentiment. Notably, as sentiment in FOMC minutes deteriorated in the periods of 2020, mostly attributable to the adverse impacts of COVID as well as in the period of 2021 to 2023, mostly attributable to the flare-ups of armed conflict around the world, a rise in interest rates would follow a drop in sentiment. This is a very interesting development, as historically a deterioration in economic outlook, i.e. during the time of the great financial crisis, lead to a decrease in interest rates. However, one must discriminate between sentiment deterioration lead by financial instability shocks which necessitate an injection of liquidity into capital markets, and the contemporary drivers of sentiment deterioration, which are geopolitical woes and inflationary pressures resulting from them. Whether it be Russian aggression or Houthi attacks on global commerce, these factors incessantly are inflationary phenomena from a monetary perspective. When looking at the graph, the amount of uncertainty around monetary policy seems to rise more poignantly as well, whereby the Fed seems to have turned to a less binding language, in line with that of the ECB, that points to more data-driven decision making and allows for more flexibility, so as to not cause volatility due to policy deviations and surprises, an occurrence forward guidance was meant to alleviate in the first place.

Constraining language, interestingly, does seem to exhibit some form of relationship with tone, which could indicate that as economic conditions deteriorate, the Federal Reserve's decision set is considerably reduced.

3.2 Sentiment Extraction - FinBERT Method

Another route we took to sentiment construction was that of FinBERT. FinBERT is a transformer-based natural language processing model that can analyze phrases contextually by looking at both preceding and succeeding words, i.e. it is bidirectional, which our VADER approach is not. FinBERT is an adaptation of BERT which was a generalized sentiment analysis tool developed by Google, whereby Araci (2019) Araci (2019) trained it on an extensive corpus of financial data. Luckily for us, Chen et al. (2023) Chen et al. (2023) took it one step further and trained BERT on FOMC statement sentiment, which seemed like an ideal approach to pit our tone construction against in our analysis. Similarly to the approach with VADER, we constructed sentiments by BERT on a sentence by sentence basis and weighted every sentence by its relative word count.

As a next step, we had to transform and prepare all of our financial data for analysis, which we delve into below.

3.3 Financial Data Processing

For every data point of our financial data, the last available datapoint was carried forward to the day of the FOMC minute issue, whereby returns were constructed via log-returns, in order to facilitate the computation of average returns in order to calculate abnormal returns. The abnormal returns, our independent variable of interest, was constructed using the model of MacKinlay (1997) MacKinlay (1997), who computed abnormal returns as the of cumulative returns after an event over average returns over a period, in our case 3 years.

Furthermore, we constructed a variable which captures the surprise effect of interest rate changes, in line with Möller and Reichmann's (2021) method in order to truly isolate the effect of sentiment from the mechanical effect of interest rate levels. This variable was constructed via the difference of implied interest rate expectations from federal funds rate futures one day before the announcement and the actual interest rate that was announced.

The point where we depart from the methodology of Möller and Reichmann (2021) is the fact that we also investigate the effects on returns for heterogeneity based on general debt levels of companies, essentially posing the question if returns react to negative sentiment more in times of high indebtedness.

3.4 Regression Sepcifications

```
setwd(dirname(rstudioapi::getSourceEditorContext()$path))
path<-dirname(rstudioapi::getSourceEditorContext()$path)

Core <- read_excel("./SENTIMENTS_FOMC_FINAL_with_BERT.xlsx")
```

```

Core$Date <- as.Date(Core$Date, format = "%d/%m/%Y") # converting the date column
FilterDates<-c(Core$Date)
Core <- Core[, -2]

#####
Fed_Futures <- read_excel("./FED_FORWARDS.xlsx")
Fed_Futures$Date <- as.Date(Fed_Futures$Date, format="%Y-%m-%d %H:%M:%S")

Fed_Futures<- Fed_Futures %>%
  mutate(Date = as.Date(Date)) %>%
  complete(Date = seq.Date(min(Date), max(Date), by="day"))

Fed_Futures<-na.locf(Fed_Futures, fromLast = FALSE)
Fed_Futures<- Fed_Futures[Fed_Futures$Date %in% FilterDates,]

#####

IR <- read.csv("./DFF.csv")
IR$Date <- as.Date(IR$DATE, format = "%Y-%m-%d") # converting the date column

IR<- IR %>%
  mutate(Date = as.Date(Date)) %>%
  complete(Date = seq.Date(min(Date), max(Date), by="day"))

IR <-na.locf(IR, fromLast = FALSE)
IR <- IR[IR$Date %in% FilterDates,][,-1]

#####

SPX_DE <- read_excel("./SPX_DE.xlsx")

SPX_DE$Date <- as.Date(SPX_DE$Date, format="%Y-%m-%d %H:%M:%S")

SPX_DE<- SPX_DE %>%
  mutate(Date = as.Date(Date)) %>%
  complete(Date = seq.Date(min(Date), max(Date), by="day"))

SPX_DE<-na.locf(SPX_DE, fromLast = FALSE)
SPX_DE<- SPX_DE[SPX_DE$Date %in% FilterDates,]

#####
SPX_price <- read.csv("./HistoricalPrices.csv", header = TRUE)
SPX_price <- SPX_price[,c(1,5)] # choosing only the closing price each day

SPX_price$Date<-as.Date(SPX_price$Date, format = "%m/%d/%y")# converting the date column

SPX_price<-na.locf(SPX_price, fromLast = FALSE)

SPX_return <- SPX_price %>%
  arrange(Date) %>% # Make sure the data is sorted by date in ascending order
  mutate(Daily_Return = log(Close) - log(lag(Close)))

days_in_year <- 252

```

```

# Calculate rolling 3-year average return
SPX_return$Rolling_3Y_Avg_Return <- rollapply(SPX_return$Daily_Return,
                                             width = 3 * days_in_year,
                                             FUN = mean,
                                             by.column = TRUE,
                                             fill = NA,
                                             align = 'right')

SPX_return$abnormal_returns<-SPX_return$Daily_Return-SPX_return$Rolling_3Y_Avg_Return

SPX_return$lagged_return<- lag(SPX_return$Daily_Return)

SPX_return<- SPX_return %>%
  mutate(Date = as.Date(Date)) %>%
  complete(Date = seq.Date(min(Date), max(Date), by="day"))

SPX_return<- SPX_return[SPX_return$Date %in% FilterDates,]

##### SURPRISES #####

IR_Lead <- read.csv("./DFF.csv")
IR_Lead$Date <- as.Date(IR_Lead$DATE, format = "%Y-%m-%d") # converting the date column
IR_Lead$DFF<- lead(IR_Lead$DFF)

IR_Lead<- IR_Lead %>%
  mutate(Date = as.Date(Date)) %>%
  complete(Date = seq.Date(min(Date), max(Date), by="day"))

IR_Lead <-na.locf(IR_Lead, fromLast = FALSE)
IR_Lead <- IR_Lead[IR_Lead$Date %in% FilterDates,][,-1]

IR_Surprises<- IR_Lead$DFF-Fed_Futures$FORWARD

##### CONSTRUCTION OF CORE DATASET FOR REGRESSIONS #####

Core$return <- SPX_return$Daily_Return
Core$lagged_return <- as.numeric(SPX_return$lagged_return)
Core$abnormal_return <- as.numeric(SPX_return$abnormal_returns)
Core$IR<- IR_Lead$DFF
Core$Surprise<-IR_Surprises
Core$debt_equity<- SPX_DE$Net_Debt_Share

```

Looking at Table 1, the intuition seems to be correct. We can see neither of those sentiment variables are significant in our baseline regression. Model fit also seems to be an issue as the R2 and adjusted R2 are no way near 1. Therefore, we can shortly conclude, that only (for example) constraining messages or uncertainty messages from the FED will not impact abnormal returns by themselves. Regressing all sentiment variables in one model also does not seem to be the correct approach, as those sentiment variables might be correlated due to their nature. Also, the authors of the original paper stayed away from such an approach. Furthermore,

```

simple_tone <- lm(abnormal_return ~ Tone, data = Core)
simple_unc <- lm(abnormal_return ~ Unc, data = Core)

```

```

simple_con <- lm(abnormal_return ~ Con, data = Core)
simple_bert <- lm(abnormal_return ~ Bert, data = Core)

stargazer(simple_tone, simple_con, simple_unc, simple_bert, column.labels = c("Tone", "Unc", "Con", "Be

```

Table 1:

	<i>Dependent variable:</i>			
	abnormal_return			
	Tone	Unc	Con	Bert
	(1)	(2)	(3)	(4)
Tone	0.013 (0.010)			
Con		0.009 (0.024)		
Unc			0.018 (0.019)	
Bert				0.001 (0.043)
Constant	-0.002 (0.002)	-0.001 (0.003)	-0.003 (0.004)	-0.00003 (0.041)
Observations	76	76	76	76
R ²	0.023	0.002	0.011	0.00000
Adjusted R ²	0.010	-0.011	-0.002	-0.014
Residual Std. Error (df = 74)	0.010	0.010	0.010	0.010
F Statistic (df = 1; 74)	1.774	0.156	0.859	0.0002

Note:

*p<0.1; **p<0.05; ***p<0.01

the potential for confounding factors that cloud the effect that sentiment could have on returns is not taken into account here, which, given the diverse nature of the information set that capital markets absorb concurrently, does seem relevant.

Because the sentiments alone do not seem to have the a significant standalone impact on abnormal returns, we start looking at different economic circumstances which could extract the partial impact of sentiments on abnormal returns. By adding a different variables one step at a time, we can fully grasp what impacts abnormal returns on “FED announcement days” and prevent overfitting, which could be an issue when looking at the fully specified model immediately. Firstly, a good variable to start off with is interest rates. Interest rates will paint the picture of monetary conditions in the economy at the moment. We therefore include IR into our models.


```

simple_tone_IR <- lm(abnormal_return~Tone + IR, data = Core)
simple_unc_IR <- lm(abnormal_return~Unc + IR, data = Core)
simple_con_IR <- lm(abnormal_return~Con + IR, data = Core)
simple_bert_IR <- lm(abnormal_return~Bert + IR, data = Core)

stargazer(simple_bert_IR, simple_bert_IR, simple_bert_IR, simple_bert_IR,
          column.labels = c("Tone", "Unc", "Con", "Bert"), header = F)

```

Table 2:

	<i>Dependent variable:</i>			
		abnormal_return		
	Tone	Unc	Con	Bert
	(1)	(2)	(3)	(4)
Bert	0.009 (0.045)	0.009 (0.045)	0.009 (0.045)	0.009 (0.045)
IR	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Constant	-0.007 (0.043)	-0.007 (0.043)	-0.007 (0.043)	-0.007 (0.043)
Observations	76	76	76	76
R ²	0.006	0.006	0.006	0.006
Adjusted R ²	-0.021	-0.021	-0.021	-0.021
Residual Std. Error (df = 73)	0.010	0.010	0.010	0.010
F Statistic (df = 2; 73)	0.211	0.211	0.211	0.211

Note:

*p<0.1; **p<0.05; ***p<0.01

Looking at Table 2, we unfortunately can see that there is still no significant impact on abnormal returns, when regressing the sentiments alongside interest rates. This would generally make sense as the prevailing interest rate is known to the market. It also can be argued that those sentiment scores and interest rates could possibly correlated a bit. High constraining language, by theory, could easily go hand in hand with higher interest rates, and so on. Model fit, i.e. R² and adjusted R² seemed to have also deteriorated by adding interest rates. The latter part is to be expected, as adjusted R² takes adding variables into account via a penalty term. However, we still stick by our step-wise addition of variables in order to in the end arrive at a feasible model, where interest rates will likely not be included. We therefore built upon this model and added the lagged returns on in our model to account for any autoregressive impact of past returns on abnormal returns. We now regress abnormal returns on Tone, Con, Unc, as well as on IR and lagged_return.

```

simple_tone_LR <- lm(abnormal_return ~ Tone + IR + lagged_return, data = Core)
simple_unc_LR <- lm(abnormal_return ~ Unc + IR + lagged_return, data = Core)
simple_con_LR <- lm(abnormal_return ~ Con + IR + lagged_return, data = Core)
simple_bert_LR <- lm(abnormal_return ~ Bert + IR + lagged_return, data = Core)

stargazer(simple_tone_LR, simple_con_LR, simple_unc_LR, simple_bert_LR, column.labels = c("Tone", "Unc"

```

Table 3:

	<i>Dependent variable:</i>			
		abnormal_return		
	Tone	Unc	Con	Bert
	(1)	(2)	(3)	(4)
Tone	0.018 (0.014)			
Con		−0.001 (0.034)		
Unc			0.016 (0.020)	
Bert				0.010 (0.046)
IR	0.001 (0.001)	−0.001 (0.001)	−0.0003 (0.001)	−0.001 (0.001)
lagged_return	−0.090 (0.162)	−0.060 (0.163)	−0.067 (0.162)	−0.062 (0.162)
Constant	−0.003 (0.004)	0.001 (0.005)	−0.002 (0.004)	−0.008 (0.043)
Observations	76	76	76	76
R ²	0.030	0.007	0.016	0.008
Adjusted R ²	−0.011	−0.034	−0.025	−0.034
Residual Std. Error (df = 72)	0.010	0.011	0.010	0.011
F Statistic (df = 3; 72)	0.736	0.172	0.382	0.188

Note:

*p<0.1; **p<0.05; ***p<0.01

Unfortunately, looking at the results from table 3, there are still no significant impacts on abnormal returns, which is a bit disheartening. However, model fit seems to have improved a bit. We still have a lot more variables which could help our ordeal though. We now add a special variable which could be a key contributor in the process. The other variables are self-explanatory (for the most part) – but how was the Surprise variable constructed? It is, as mentioned above, the difference between the Federal Funds rate on announcement day and Fed Funds Futures as a gauge for market expectations, i.e. the surprise effect of the interest rate change on the market. This means that, this variable will account for unexpected announcement results from the fed and discounts it.

```

simple_tone_SUR <- lm(abnormal_return ~ Tone + IR + lagged_return + Surprise, data = Core)
simple_unc_SUR <- lm(abnormal_return ~ Unc + IR + lagged_return + Surprise, data = Core)
simple_con_SUR <- lm(abnormal_return ~ Con + IR + lagged_return + Surprise, data = Core)
simple_bert_SUR <- lm(abnormal_return ~ Bert + IR + lagged_return + Surprise, data = Core)

stargazer(simple_tone_SUR, simple_unc_SUR, simple_con_SUR, simple_bert_SUR, column.labels = c("Tone", "Unc", "Con", "Bert"))

```

Table 4:

	<i>Dependent variable:</i>			
	Tone	Unc	Con	Bert
	(1)	(2)	(3)	(4)
Tone	0.024* (0.014)			
Unc		0.012 (0.020)		
Con			0.013 (0.034)	
Bert				0.015 (0.045)
IR	0.0004 (0.001)	−0.001 (0.001)	−0.001 (0.001)	−0.001 (0.001)
lagged_return	−0.047 (0.160)	−0.023 (0.162)	−0.018 (0.162)	−0.019 (0.162)
Surprise	0.022** (0.011)	0.017 (0.011)	0.019* (0.011)	0.018* (0.011)
Constant	−0.005 (0.004)	−0.001 (0.004)	−0.001 (0.005)	−0.013 (0.043)
Observations	76	76	76	76
R ²	0.084	0.049	0.046	0.046
Adjusted R ²	0.033	−0.004	−0.007	−0.008
Residual Std. Error (df = 71)	0.010	0.010	0.010	0.010
F Statistic (df = 4; 71)	1.635	0.919	0.862	0.855

Note:

*p<0.1; **p<0.05; ***p<0.01

Analysing table 4, one can now start to understand why we think highly of the surprise variable. By adding surprise, we can now start to see how abnormal returns react to FED Statements. In our model now, Tone is significant. According to the authors, this indicates that a more positive (negative) language of the FED is associated with higher (lower) intraday abnormal returns. This is in line with the findings of Schmeling and Wagner (2019)Schmeling and Wagner (2019), who found that negative changes in the fraction of negative words within an introductory statement of an ECB press conferences have a positive influence on stock prices and a negative influence on volatility risk premia and credit spreads. We also find that, Surprise is significant

in every model other than Unc. This can be explained, intuitively, as any information not included in the market price is bound to move it once it becomes available to its participants, which in the case of interest rates is on the day of the announcement.

Last but not least, we include the debt to equity ratio of the index into our model. This variable tells us the rate of debt those firms have in relation to their equity. The higher this ratio is the higher the leverage of its firm, the more it might react to interest rates changes. This will now complete our intended model.

```
simple_tone_DE <- lm(abnormal_return ~ Tone + IR + lagged_return + Surprise + debt_equity, data = Core)
simple_unc_DE <- lm(abnormal_return ~ Con + IR + lagged_return + Surprise + debt_equity, data = Core)
simple_con_DE <- lm(abnormal_return ~ Con + IR + lagged_return + Surprise + debt_equity, data = Core)
simple_bert_DE <- lm(abnormal_return ~ Bert + IR + lagged_return + Surprise + debt_equity, data = Core)

stargazer(simple_tone_DE, simple_unc_DE, simple_con_DE, simple_bert_DE, column.labels = c("Tone", "Unc"
```

Table 5:

	<i>Dependent variable:</i>			
	Tone	abnormal_return Unc	Con	Bert
	(1)	(2)	(3)	(4)
Tone	0.032** (0.015)			
Con		0.025 (0.037)	0.025 (0.037)	
Bert				0.003 (0.050)
IR	-0.0003 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
lagged_return	-0.057 (0.159)	-0.019 (0.162)	-0.019 (0.162)	-0.016 (0.163)
Surprise	0.022** (0.011)	0.019* (0.011)	0.019* (0.011)	0.017 (0.011)
debt_equity	0.00003 (0.00002)	0.00002 (0.00002)	0.00002 (0.00002)	0.00001 (0.00002)
Constant	-0.016** (0.008)	-0.008 (0.009)	-0.008 (0.009)	-0.005 (0.045)
Observations	76	76	76	76
R ²	0.113	0.057	0.057	0.051
Adjusted R ²	0.050	-0.010	-0.010	-0.017
Residual Std. Error (df = 70)	0.010	0.010	0.010	0.010
F Statistic (df = 5; 70)	1.792	0.853	0.853	0.756

Note:

*p<0.1; **p<0.05; ***p<0.01

As seen in Table 5, we still have a significant value of tone, but the significance of surprise has changed in

some variables. Instead of our FinBERT sentiment being significant, uncertainty has now become significant. Generally an overall underperformance of the BERT approach seems to become evident, something that we will delve into later. A model which could become interesting is our model with interaction terms. The idea is that by adding interaction the effect of “Tone”, “Unc”, “Con”, “Bert” or “Surprise” on abnormal returns is moderated by the company’s leverage. In other words, the impact of these sentiment on returns may vary depending on the leverage ratio. This is intuitive as a company with higher leverage will be much more sensitive to changes in its cost of capital, and hence, sensitive to indications that such a cost of capital change could come. Such an indication is the sentiment of FOMC statements.

```
interactions_tone <- lm(abnormal_return ~ Tone * debt_equity + IR + lagged_return + Surprise * debt_equity + debt_ratio)
interactions_unc<-lm(abnormal_return~Unc*debt_equity + IR + lagged_return + Surprise*debt_equity + debt_ratio)
interactions_con<-lm(abnormal_return~Con*debt_equity + IR + lagged_return + Surprise*debt_equity + debt_ratio)
interactions_bert <- lm(abnormal_return~Bert*debt_equity + IR + lagged_return + Surprise*debt_equity + debt_ratio)
stargazer(interactions_tone, interactions_unc, interactions_con, interactions_bert, column.labels = c("Tone", "Uncertainty", "Confidence", "BERT", "Surprise"))
```

We can see in Table 6 that, the interest rate surprises are highly significant by itself but also within its interaction term. This can be explained rather intuitively. If a company has high leverage it will have a significant impact on its performance by a “surprise” in interest rates, as they are more vulnerable to interest rates. However, again none of the base sentiment scores are significant on their own and in combination with the debt to equity ratio. This is a bit surprising, as one could have made an argument that firms with an higher (lower) leverage could have been more (less) vulnerable to FED statements as they are at the mercy of the monetary policy in place. Generally, it could be interpreted that companies with high leverage are especially susceptible to both interest rates as well as sentiment, as forward guidance tends to affect their projected cost of capital and their enterprise value more vigorously.

Before we delve into our conclusion and our findings regarding the applicability of BERT vs. bespoke VADER, the interplay between indebtedness and the ensuing reaction of abnormal returns to monetary policy, we want to extend our analysis by one step that Möller and Reichmann (2021) did not delve into, namely the question of how feasible our model is given the dataset that we have created.

Therefore, we now ensure the soundness of our models and the inclusion of our variables, especially of the interest rate variable which prompted a negative R2, via a bayesian model averaging algorithm. By cycling over all potential combinations of models, we want to ascertain the relative importance of each variable. The bayesian model averaging algorithm does so by determining the post-inclusion probability for each of our variables, including their interaction terms. We do so with the sentiment model that has proven most reliable so far, namely the model including Tone.

Table 6:

	<i>Dependent variable:</i>			
	Tone	abnormal_return		Bert
		Unc	Con	
	(1)	(2)	(3)	(4)
Tone	0.082 (0.053)			
Unc		0.143 (0.094)		
Con			0.147 (0.126)	
Bert				-0.095 (0.207)
debt_equity	0.0001 (0.00004)	0.0001 (0.0001)	0.0001 (0.0001)	-0.0002 (0.001)
IR	-0.001 (0.002)	-0.001 (0.001)	-0.001 (0.002)	-0.001 (0.001)
lagged_return	-0.154 (0.158)	-0.166 (0.163)	-0.116 (0.161)	-0.086 (0.165)
Surprise	0.305*** (0.102)	0.356*** (0.107)	0.310*** (0.103)	0.279*** (0.103)
Tone:debt_equity	-0.0001 (0.0002)			
Unc:debt_equity		-0.0003 (0.0003)		
Con:debt_equity			-0.0003 (0.0004)	
Bert:debt_equity				0.0002 (0.001)
debt_equity:Surprise	-0.001*** (0.0002)	-0.001*** (0.0002)	-0.001*** (0.0002)	-0.001** (0.0002)
Constant	-0.032** (0.015)	-0.041* (0.022)	-0.029 (0.019)	0.082 (0.193)
Observations	76	76	76	76
R ²	0.205	0.190	0.158	0.135
Adjusted R ²	0.123	0.106	0.071	0.046
Residual Std. Error (df = 68)	0.010	0.010	0.010	0.010
F Statistic (df = 7; 68)	2.500**	2.272**	1.819*	1.514

Note:

*p<0.1; **p<0.05; ***p<0.01

4. Bayesian Model Averaging

After regressing everything and getting the results we do a final check via Bayesian Model Averaging (BMA), where we can see which of the Variables from our Core data set truly are important to abnormal returns. This methodology is to check which variables have the highest comparative importance in a plethora of potential models that include all of our variables, along with interaction terms and both of their parent terms. According to Tang (2003), Bayesian Methods are often used in the field of Finance, as those eliminate the uncertainty of a potential misspecified model, which BMA would eliminate. BMA methods can pick up true predictors even if the model is not linear, as it can find true predictors. We therefore feel confident, we can pick up the variable importance of the variables given.

```
# New BMA

corebma <- na.omit(Core[, -c(1,3,4,5,6,7)])

corebma <- mutate_all(corebma, function(x) as.numeric(as.character(x)))

# Ensure 'abnormal_return' is the first column
corebma <- corebma[c("abnormal_return", setdiff(names(corebma), "abnormal_return"))]

independent_vars <- setdiff(names(corebma), "abnormal_return")

# Create all two-way interactions between independent variables
for(i in 1:(length(independent_vars)-1)){
  for(j in (i+1):length(independent_vars)){
    interaction_term <- paste(independent_vars[i], independent_vars[j], sep="#")
    corebma[[interaction_term]] <- corebma[[independent_vars[i]]] * corebma[[independent_vars[j]]]
  }
}

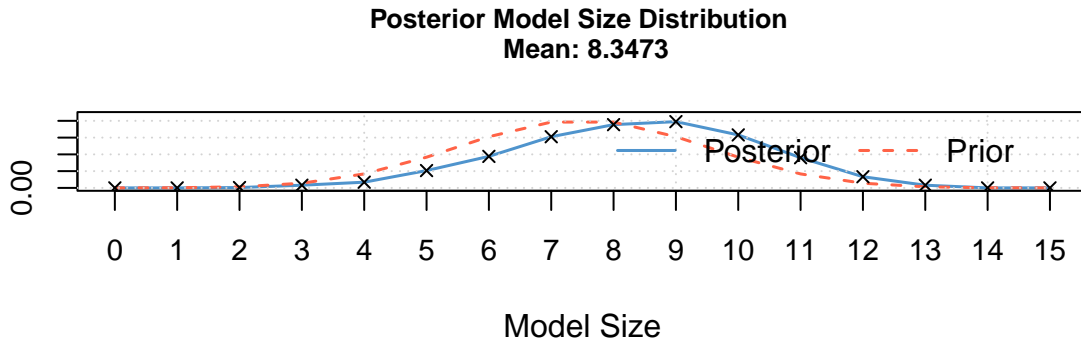
# Now corebma includes 'abnormal_return' as the first column and all two-way interaction terms
# Run bms with the expanded corebma
bma <- bms(corebma, burn = 1000, iter = 3000, g = 1, mprior = "uniform", mcmc = "bd.int")
```

##	PIP	Post Mean	Post SD	Cond.Pos.Sign
## Surprise	0.9943333	2.023044e-01	2.115725e-01	0.99798860
## Tone	0.9556667	2.034416e-02	2.499649e-02	1.00000000
## IR	0.8850000	-8.751226e-04	4.666906e-03	0.59924670
## debt_equity	0.8753333	1.375868e-05	2.290114e-05	0.99047982
## lagged_return	0.7996667	-6.317295e-02	4.195235e-01	0.38891205
## Tone#Surprise	0.6440000	-2.251100e-01	2.811725e-01	0.04037267
## IR#Surprise	0.6030000	-1.118064e-02	1.287336e-02	0.00000000
## Surprise#debt_equity	0.5363333	-3.670018e-04	5.229840e-04	0.13051585
## Tone#lagged_return	0.3690000	-4.419063e-01	8.344297e-01	0.00000000
## Tone#debt_equity	0.3090000	-7.857989e-06	6.893019e-05	0.23732470
## lagged_return#Surprise	0.3083333	1.462018e-01	8.015716e-01	0.74054054
## Tone#IR	0.2976667	-5.557422e-04	7.235516e-03	0.26539754
## IR#debt_equity	0.2723333	2.791912e-06	9.283195e-06	0.97307222
## lagged_return#IR	0.2490000	4.681927e-03	5.595935e-02	0.71084337
## lagged_return#debt_equity	0.2486667	2.667182e-04	1.020589e-03	0.99329759
##	Idx			
## Surprise	4			
## Tone	1			
## IR	3			
## debt_equity	5			

```

## lagged_return          2
## Tone#Surprise          8
## IR#Surprise            13
## Surprise#debt_equity   15
## Tone#lagged_return      6
## Tone#debt_equity        9
## lagged_return#Surprise  11
## Tone#IR                 7
## IR#debt_equity         14
## lagged_return#IR        10
## lagged_return#debt_equity 12
##
## Mean no. regressors      Draws      Burnins      Time
##      "8.3473"           "3000"      "1000"      "0.7307508 secs"
## No. models visited      Modelspace 2^K      % visited      % Topmodels
##      "2720"           "32768"           "8.3"           "85"
##      Corr PMP           No. Obs.      Model Prior      g-Prior
##      "0.3824"           "76"      "uniform / 7.5"      "numeric"
##      Shrinkage-Stats
##      "Av=0.5"
##
## Time difference of 0.7307508 secs

```



Of the variables we included in the final model on table 6, the interaction term of tone and debt to equity falls visibly below our threshold, prompting us to rely mostly on the analysis on table 5 for interpretation, i.e. our slight extension of Möller and Reichmann (2021). This threshold is 50% and refers to the post-inclusion probability of a variable. Looking back at table 5, we can see one very rudimental thing, namely the fact

that the variation in abnormal returns seems to be best explained by the model with Tone as its sentiment variable. Conversely, throughout the majority of specifications, the BERT language processor seems to exhibit lackluster performance. This may be due to the fact that the VADER approach may be better-suited for analyzing text that in its content and structure may be similar, as opposed to heterogeneous news or rather Tweets, which BERT was initially designed around. Furthermore, as intuition would suggest as well, market participants and followers of the FOMC do not try and dissect fed speeches into their respective sentiment components, but rather try to eke out the change in tone, or the mood which the speech is conveying. Based on that, transactions are made and hence, Tone would be a plausible front-runner in explaining returns out of the three VADER-based sentiment components.

4.1. A brief excursion into the stability of the Tone Effect over time

As a mild robustness check, we wanted to see, as best possible with our limited data, whether the tone variable can be considered stable over time, i.e. whether sentiment effects have temporary dependence. To that end, we investigated the last 50 statement effects at the earliest possible date and arrived at the following visualization, where the coefficient for tone along with its standard error bands are plotted.

```
window<- 50

RegressionDates<- FilterDates>window:length(FilterDates)]
Tone_SE<- c()
Tone<- c()

for (i in RegressionDates){
  # this part is tricky - how to filter the right window of data??
  index_end <- which(FilterDates==i)

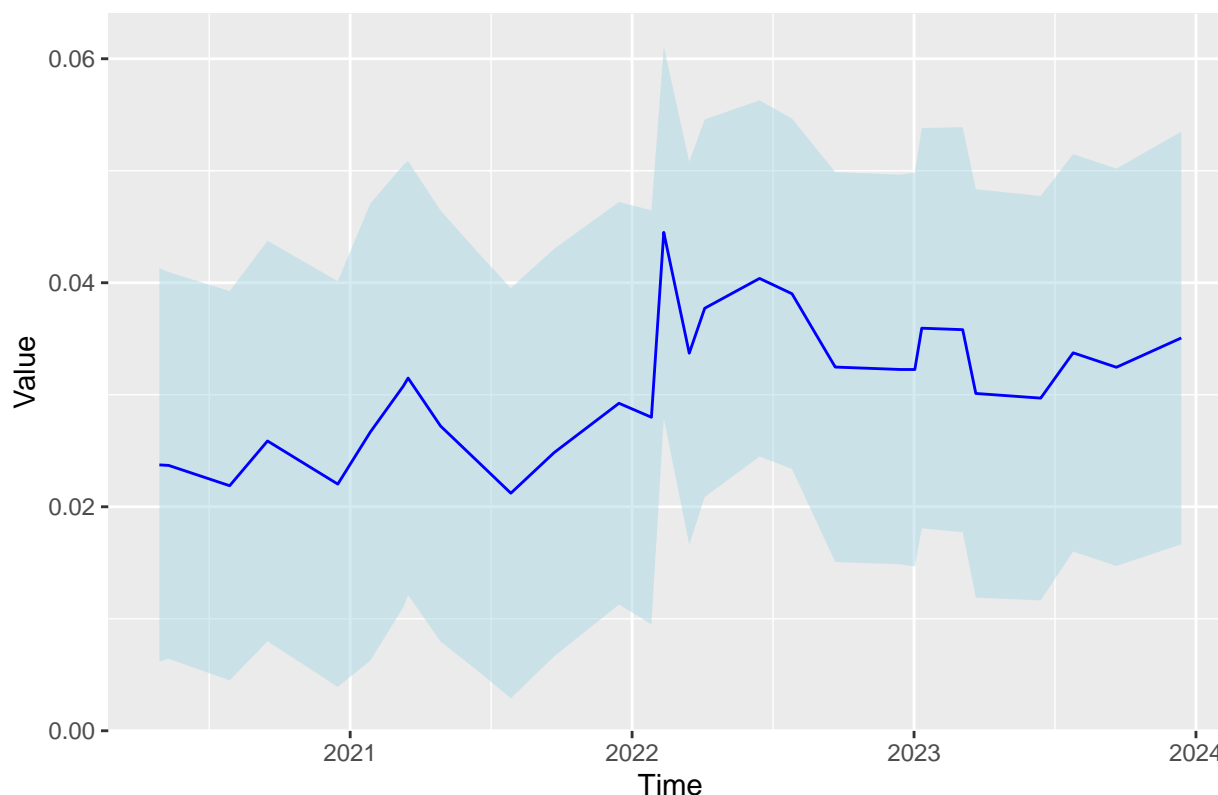
  index_start <- index_end - window

  data<- Core[index_start:index_end,]
  reg <- summary(lm(abnormal_return ~ Tone + IR + lagged_return + Surprise + debt_equity, data = data))
  Tone<- c(Tone,reg$coefficients[2])
  Tone_SE<- c(Tone_SE,reg$coefficients[8])
}

upper_Tone<- Tone+Tone_SE
lower_Tone<- Tone-Tone_SE
plotbands<-data.frame(cbind(Tone, lower_Tone,upper_Tone))
plotbands$date<-as.POSIXct(RegressionDates)

ggplot(plotbands, aes(x = RegressionDates, y = Tone)) +
  geom_ribbon(aes(ymin = lower_Tone, ymax = upper_Tone),
            fill = "lightblue", alpha = 0.5) +
  geom_line(color = "blue") +
  labs(title = "Effect of Tone over time, regressions on a 50 Statement Window",
       x = "Time", y = "Value")
```

Effect of Tone over time, regressions on a 50 Statement Window



Below, we will further delve into the potential avenues for future research, however, this may be one element to consider in any future endeavor as well, and may open up an interesting research question of investigating heterogeneous market sensitivities to FOMC sentiment, e.g. during periods of high uncertainty versus “regular” ones. In any case, we do see the coefficient of Tone remaining roughly within the band of 0.03 to 0.04, thereby exhibiting a rather stable behavior over time, albeit with the caveat of a limited sample of only 26 dates, which, along with their 49 preceding statements were regressed over to construct the one of the 26 datapoints in the plot.

5. Potential for Future Research

Considering the fact that FED statements have been forced to consider significant humanitarian crises such as the outbreak of the COVID-19 pandemic (First Lockdown in the U.S. began on March 12th, 2020), the invasion of Russia into Ukraine (February 24th, 2022 - immediately after economical from COVID began the lessen) and the Hamas terrorist attack on Israel and the war following after (since October 7th, 2023), there exists an avenue for future research.

While the U.S. interest rate policy must inevitably take into account global geopolitical events, it can only incorporate them to a limited extent. This means that a discrepancy between the U.S.-focused and more “fact-based” interest rate and the more “human” statements will likely occur, as shown in the graph in Chapter 3.3. Future research could seek to find a way to quantify the difference between economically relevant word-combinations by refining linguistic analysis methods, adjusting dictionaries, and filtering out word combinations that may focus on factors irrelevant to immediate, upcoming policy, thereby emphasizing a more financially centered context.

This research potential especially becomes apparent when observing the disparity between the overall trend the SPX shows since the “COVID-Crash” in March 2020 - particularly the clear positive trend since the beginning of 2023 - and the language employed by the FED in its statements since the beginning of the

recorded statements in the data used. As mentioned, future research may seek to refine the dictionaries used and isolate sentences or word-combinations related to unforeseen geopolitical events, that cannot be ignored in a statement by a governing body.

Additionally, we have seen Tone have a significant effect and being a valuable variable in explaining abnormal stock returns. This opens up the avenue of including sentiment as an additional factor in equity return models. Another promising avenue of research would be a deeper dive into the model differences between VADER and BERT, especially when it comes to assigning sentiment to texts that are less heterogeneous, as are FOMC statements or other texts with more standardized structure and language.

6. Conclusion

Our analysis allows us to make a plethora of conclusions. First, we can see an almost twofold increase of explanatory capability when it comes to Tone vs. BERT, indicating that more homogeneous texts lend themselves better to sentiment analysis via heuristics adjusted dictionary approaches as is VADER. Secondly, we can also see that the mechanical component of FOMC statements, i.e. the interest rate announcement and the surprise it entails, is a non-negligible factor that should be included in central bank statement analyses to avoid biased estimates. Third, we could observe an ever so slight sensitivity of higher leverage periods with regards to said surprises. Finally, this research laid the foundations for us to further investigate the highly intriguing intersection between natural language processing and financial markets in an event-study context.

7. Bibliography

- Araci, Dogu. 2019. “Finbert: Financial Sentiment Analysis with Pre-Trained Language Models.” *arXiv Preprint arXiv:1908.10063*.
- Bodnaruk, Andriy, Tim Loughran, and Bill McDonald. 2015. “Using 10-k Text to Gauge Financial Constraints.” *Journal of Financial and Quantitative Analysis* 50 (4): 623–46.
- Chen, Ziwei, Sandro Gössi, Wonseong Kim, Bernhard Bermeitinger, and Siegfried Handschuh. 2023. “FinBERT-FOMC: Fine-Tuned FinBERT Model with Sentiment Focus Method for Enhancing Sentiment Analysis of FOMC Minutes.”
- Hutto, Clayton, and Eric Gilbert. 2014. “Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.” In *Proceedings of the International AAAI Conference on Web and Social Media*, 8:216–25. 1.
- Loughran, Tim, and Bill McDonald. 2011. “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks.” *The Journal of Finance* 66 (1): 35–65.
- MacKinlay, A Craig. 1997. “Event Studies in Economics and Finance.” *Journal of Economic Literature* 35 (1): 13–39.
- Möller, Rouven, and Doron Reichmann. 2021b. “ECB Language and Stock Returns – a Textual Analysis of ECB Press Conferences.” *The Quarterly Review of Economics and Finance* 80 (May): 590–604. <https://doi.org/10.1016/j.qref.2021.04.003>.
- . 2021a. “ECB Language and Stock Returns – a Textual Analysis of ECB Press Conferences.” *The Quarterly Review of Economics and Finance* 80 (May): 590–604. <https://doi.org/10.1016/j.qref.2021.04.003>.
- Schmeling, Maik, and Christian Wagner. 2019. “Does Central Bank Tone Move Asset Prices?”