What does the Swedish Riksbank say and does it affect the stock market?

Group 35: Description of who is in charge of the different sections:

Annalisa Tassi (lfr292): 2, 3.3, 3.5, 4.2, 6.1
Mark Drejer Jensen (hxz518): 3, 3.1, 5.1.1, 6.3
Sebastian Agergaard Andersen (hcz942): 3.4, 4.1, 5.1, 5.1.2, 5.2
Matìas Piqueras (bvj349): 1, 3.2, 3.6, 6.2, 7

Social Data Science - Summer School 2019

Contents

1 Introduction

Within the field of Financial Economics, a substantive body of literature has been concerned with understanding and predicting the impact of public announcements on financial markets. This phenomenon, commonly referred to as *announcement effect*, assumes that the behavior of the financial system– and actors in it -correlates with the nature of the announcement. Consequently, spurring either a negative or positive market reaction. In this context, central bank announcements of policy rates are often seen as particularly poignant. Generally speaking, a lowered policy rate is associated with an increase in stock prices and, conversely, if the policy rates increase, then stock prices are liable to fall. Although stock price indices rarely constitute the main metric of which central banks use to determine inflation, they are – undoubtedly – the fastest to react to any changes in monetary policy. Communication is therefore a question of salient and increasing importance for central banks, further empathized by the added financial stability mandate (in addition to monetary policy) many central banks received in the wake of the global financial crisis of 2008 (Blinder et al., 2008). Moreover, central banks very deliberately use communication as a tool for achieving inflation targets (Woodford, 2005). This means that when trying to understand central bank announcement effects, one must also consider the content and sentiment of the communication as variable in itself.

The aim of this paper is two-fold. First, we attempt to create a variable that captures the general sentiment towards the economic outlook in the policy rate announcement of the Swedish Riksbank between January 2000 and July 2019. This is done using a simple bag-of-words model, comprising a lexicon of positive, negative, and neutral words specifically coded by researchers at the Federal Reserve to analyze central bank communication (Correa et al., 2017). For analytical purposes, the corpus we use consists of the minutes from the monetary policy meetings. After a careful reading of a random sample of minutes, and based on the fact that the policy rate is the outcome from the monetary policy meeting, we postulate that this will serve as a proxy for the actual announcement with the added benefit of containing a satisfactory amount of words for computational text analysis.

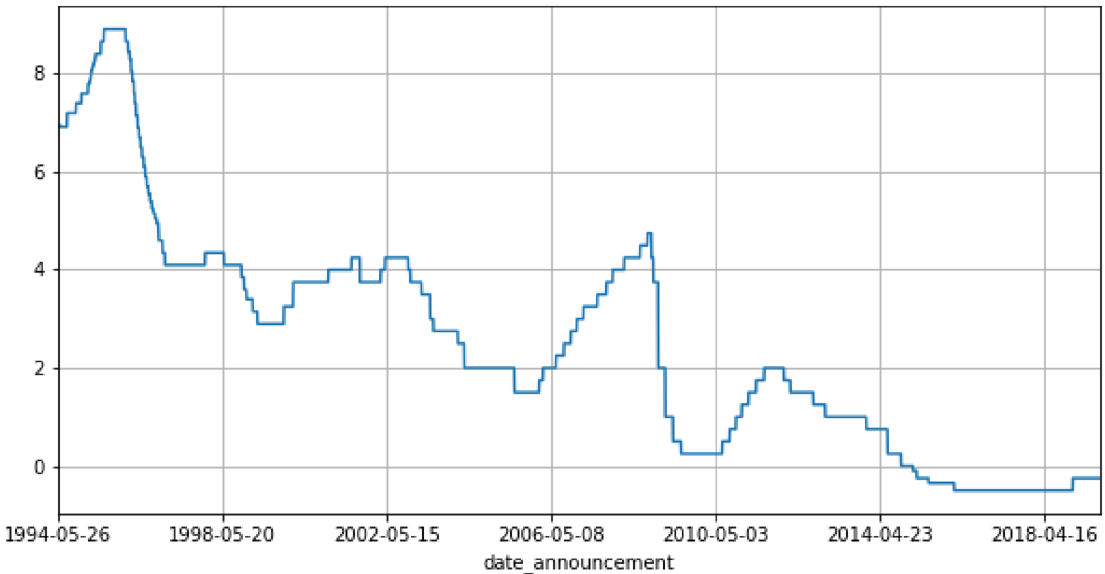
Our second aim is to investigate if we can use the sentiment variable to make better predictions on how the stock market reacts to policy rate announcements. For this purpose, we use a linear regression and a lasso machine learning model that use two variables: (1) The sentiment variable (2) The announced policy rate. For measuring the stock market, we use data on the opening and closing market capitalization of the OMX Stockholm 30. Data on the policy rate, announcement– and effective date is provided by the Swedish Riksbank. We have decided to try to implement machine learning models following Elson (2015), as stock market data are highly frequent.

The rest of the paper is structured as follows: First, we report the theoretical background underlying our research question; second, we describe the data we use; third, we write about the methodology we have implemented, and we present the results and our methodological and ethical considerations in Section 6. The conclusion (Section 7) wraps up our main findings.

2 Theoretical background

The repo rate is the rate at which banks can deposit or borrow funds from the central bank over a seven-days period, i.e. a short-term interest rate (Swedish Central Bank, 2019). Thus, decisions on monetary policy are taken by adjusting the repo rate in order to meet the inflation target. The underlying mechanism can be explained as follows. The repo rate influences the overall nominal interest rates' level, which has effects on the market as actors adjust their demand and supply of goods and (financial) services - among others. In turn, market interactions lead to a new "equilibrium" with a specific price level, reaching the inflation target.

Figure 1: Repo rate (1994-2019).



Source: Swedish Central Bank

Note: on the x-axis, we plot the date, on the y-axis, the repo rate (i.e. 7 corresponds to 7%)

In Figure 1, we plot the level of the repo rate in the period 1994-2019. Since February 2015, the repo rate has been negative. This is a so-called *unconventional expansionary monetary policy.*[1] The rationale behind this policy is the protracted effect of the economic crisis. Sweden is a small open economy and has to adapt its policies to other countries' policies: after the crisis, the ECB initiated a regime of negative interest rates and Sweden aligned its policies in order to avoid currency appreciation and related negative market shocks. The fact that the repo rate is negative does not mean that all interest rates are negative as well. This policy mainly affects direct parties of the central bank, i.e., commercial banks. Even though this unconventional monetary policy might be perceived as a *crisis tool*, there have been no signs of falling demand in the economy, and unemployment has gone back to pre-crisis levels (Swedish Central Bank, 2016).

[1]The arguments against the adoption of a negative repo rate are the creation of uncertainty which might cause a decrease in demand, and the increase of housing prices and debt due to overall lower interest rates. The increase of households' debt is monitored with particular interest.

The announcement effect. The efficient market hypothesis states that the security prices in the financial market fully reflect all available information. According to this hypothesis, positive -or negative- announcements (about a company, for instance) do not have effects on stock prices as the information is already reflected in the price. Moreover, only new and unexpected information is expected to have effects on the stock market as soon as it is available (Mishkin and Eakins, 2012). Fama (1998) and other empirical studies show that, in some cases, announcements have effects protracted over time. In what follows, we focus on short-term announcement effects (same day opening and closing value) as additional variables and more complex models would be needed in order to isolate the sentiment effect. Nonetheless, we perform robustness checks for changes in the stock market variable up to 30 days after the announcement.

2.1 Literature Review

Moniz and Jong (2014) perform a text analysis on the minutes of the Bank of England. They combine text analysis, which is more traditional in the financial context, with Machine Learning. In this way, they are able to enrich commonly used dictionaries with expressions from different economic topics (i.e., inflation, exchange rates, etc.) and their perception. In addition to that, they include expression *additioners* and *diminishers* that can modify the meaning of a sentence: for example, they are able to classify "low inflation" as positive. Their main findings are that if a central bank communication emphasizes positive economic growth and discusses an increase of the interest rate, investors' expectations of future interest rates will rise by 3%. On the other hand, if the central bank's tone towards economic growth is low, it discusses declining bank lending and has a negative tone towards interest rates, investors will reduce their expectations of future interest rates by 4%. This analysis helps understand which aspects have stronger effects on investors' expectations, and whether negative or positive words and tones have a greater impact.

Karadi and Jarociński (2018) study the announcement effects of the FED with a similar approach. In contrast with previous studies, they disentangle the announcement effect into a monetary policy and economic outlook components. The two effects are found to move in opposite directions: in the short run, the stock market reacts negatively to announcements of expansionary policy, because it signals the need to boost the economy. On the other hand, indexes go up in the long run as the economy recovers. This confirms the fact that central bank's communications are two-fold and release information which is not publicly available.

3 Data collection and measurements

This section deals with the data collection and measurements of data, used throughout this paper. The aim of the section is to describe the composition and use of each of the variables included, as well as describing the process of collecting the data. The variables used in this study are the OMX Stockholm 30 stock market index (the dependent variable), the general sentiment, extracted from the minutes publications of the Swedish

Central Bank, and the repo rate (explanatory variables). Further, the dictionary used to extract the sentiment of the minutes is described.

3.1 Minutes data

We have downloaded all the minutes of the Executive Board meetings of the Swedish Central Bank (*Riksbank*) from the online archives and the website of the Swedish Central Bank and we have formatted all the PDF files so that they were named after the date of the meeting. We have collected all minutes from January 2000 until July 2019.[2] The meetings were held more frequently in the period 2000-08, and it is also worth mentioning that in 2016 extraordinary meetings took place in order to discuss interventions on the foreign exchange market, as a complementary monetary policy measure to safeguard the rise in inflation.[3] Since 2009, the executive Board meets six times per year (every second month) and deliberates about the **repo** rate. Around 25 documents needed manual editing as the original formatting could not be read by the software. The total number of texts collected is 135. The earlier texts contained less than 10 pages, while the most recent may have up to 25 pages of text. We notice how the communication becomes more precise and detailed over time. Even though the minutes are published 2 weeks after the Central Bank's press release, the content of the texts is richer and, thus, there is a greater likelihood of computing a relevant sentiment measure.

3.2 Sentiment dictionary

The dictionary which is used for the sentiment analysis has been coded by Correa et al., 2017 at the Federal Reserve, based on the language of the financial stability reports that the Federal Reserve produces. This means that the words coded as either positive or negative are so with regards to the context of central bank communication. The dictionary contains 96 positive and 295 negative words and 1093 words that could potentially convey sentiment but were either too neutral or descriptive (Correa et al., 2017). The frequency of the words from the different classes in our corpus is presented in Table 1. It can be noted that the positive words only comprise a small fraction of the sample. This in it self does not pose a problem as the small amount just represents the fact that the coders did not identify more words that could reliably be classified as positive. It implies however that the sentiment might be skewed negatively. This can also be taken with some ease since the sentiment is computed as a ratio and for our purpose what matters is that the measurement can say something about the relative difference between observations, i.e. if one minute is more or less negative/positive than the others.

[2]The executive board meetings were held since 1994, but no minutes are available in the period 1994-99. The minutes have been published online since 1999, but due to lack of files in some instances of 1999, we start from complete data availability (2000).

[3]We have removed the two additional minutes as they did not contain any information about the repo rate decision.

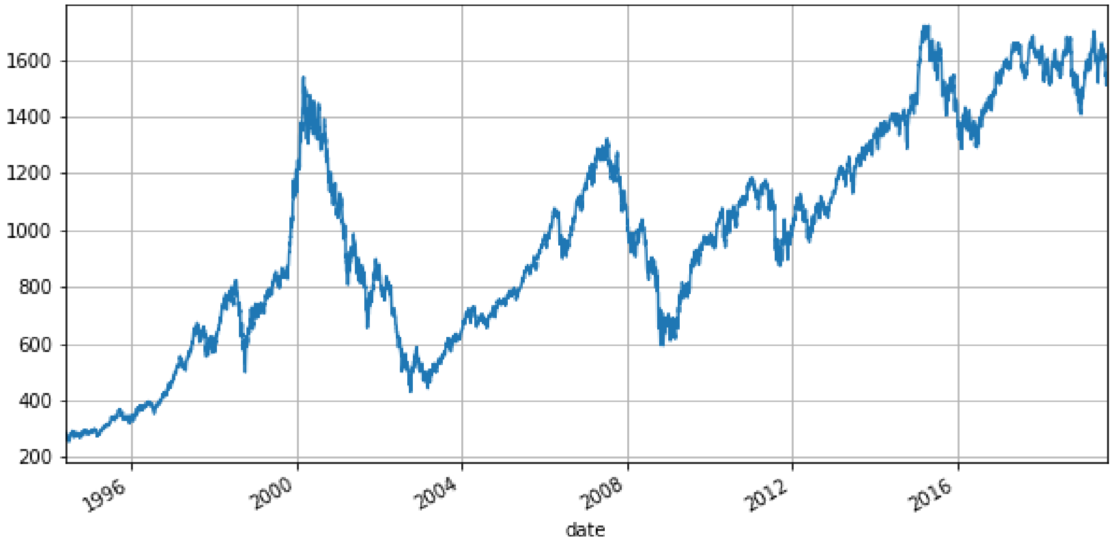
Table 1: Dictionary and Word Frequency

	Words in Dictionary	Word Frequency	Word Frequency (percent)
Positive	96	6,372	0.5%
Negative	295	10,692	0.9%
Neutral	1093	65,940	5.4%

Note: Word frequency (percent) is calculated as the total amount of positive/negative/neutral divided by total amount of words in the complete corpus.

3.3 Dependent variable: Stock market

Figure 2: OMX Stockholm 30 (1994-2019), daily.



Source: NASDAQ OMX NORDIC

Note: on the x-axis, we plot the date, on the y-axis, the stock market index

In order to measure the market reaction to the announcements, we use the OMX Stockholm 30 index. The index closely follows the 30 most traded securities in the Swedish market and it is a market weighted price index. Its composition is revised twice per year (Bloomberg, 2019). The data is available at high frequency (on a daily basis for the past 25 years), as shown in Figure 2. Given the nature of the index, we argue that its movements are a reliable measure of the average market returns and reactions. To be more specific, the available information is the closing value of the index and the respective date. Based on that, we are able to compute the opening value, which is equal to the closing value of the previous day. The percentage change of the OMX30 index is computed as

$$\Delta\%OMX30 = \frac{\text{closing value} - \text{opening value}}{\text{opening value}} \tag{1}$$

In the first column of Table 2, we report the values of the percentage change of the stock market index for our sample, at the announcement date.

The mean (and the median) values of this variable are small and positive, meaning that, on average, the market closes positively after the announcement by the central bank. This phenomenon might be related with the positive effect of forward guidance from the central bank. Moreover, in absolute terms, positive changes are larger than slumps, in our sample.

3.4 Explanatory variable: Sentiment

In the fourth column of Table 2, we report the sentiment that we have computed. The index is small as it is computed as

$$\text{Sentiment} = \frac{\text{No. of positive words} - \text{No. of negative words}}{\text{Total amount of words}}, \tag{2}$$

but it reports the predicted sentiment of the overall discussion: were the board members concerned about the economic outlook, or rather optimistic? The average sentiment is small but negative.

3.5 Explanatory variable: Repo rate

The repo rate change is computed as

$$\text{Repo-rate change} = \text{effective rate} - \text{announced rate}. \tag{3}$$

In this way, we are able to take into account the direction of the monetary policy: a positive change is signalling the implementation of restrictive monetary policy, while a cut of the repo rate implies the implementation of an expansionary policy.[4] In our sample, we observe an average cut of the repo rate of 0.02 percentage points, meaning that our average observation is in times of expansionary monetary policy. For the extreme values, we observe changes of up to 1.75 percentage points in absolute terms.

3.6 The complete data set

As a last step in the data-collection process, we merged our data into one data frame using the *inner-join* function on the date key. The main variables used in our results and discussion section are presented in Table 2. The complete data frame containing all the variables and descriptive statistics is included in Appendix A.

[4]For a theoretical discussion on the monetary policy transmission, refer to section 2.

Table 2: Descriptive statistics

	Δ% OMX daily	Δ% OMX +14	Δ% OMX +30d	Sentiment	Repo rate change
Minimum	-0.0346	0.8489	0.8013	-0.0134	-1.7500
Mean	0.0018	1.0041	1.0023	-0.0039	-0.0243
(SD)	0.0135	0.0609	0.0849	0.0034	0.2526
Median	0.0010	0.9964	0.9956	-0.0033	0.00
Maximum	0.0559	1.3090	1.3394	0.0042	0.50

Note: N = 135, Decimals are rounded to the 4th position.

4 Methods

4.1 Bag of Words: Applying and visualizing

Using the method called "Bag of words" (BoW) is a simple approach to create an overview of the sentiment in the textual content. As the name of the method implies, we have a large amount of words, which is allocated depending on whether it is signalling a positive or a negative meaning.

Our *BoW* should not only indicate that we are dealing with a positive or negative matter, it should also take into account that the text is written by the Central Bank of Sweden. Therefore, a standard list of positive words would be inadequate in measuring the sense of positiveness throughout the minutes.

Simple approaches can often be followed by issues or minor challenges, and this case is no exception. The *BoW* method has a lack of accuracy, hence the number of positive words measured in a file will not always be the exact indicator of "how positive" a text is. First of all, the list might not include all words, which implies an impact in the sense of positiveness. Even though the list is quite thorough, the *BoW* method will rarely be able to cover up every hint of positive or negative imprints. Second of all, only taking singular words leaves out the possibility to include sentences and phrases, that clearly points to a somehow sentiment direction, and which might not contain a word from the specific lexicon, or even might contain words from the opposite lexicon. This issue will be discussed further in Section 6.1.

Applying both the positive and negative lexicons on the text from the minutes, we are able to calculate the average tone through each published minutes of the executive board of monetary policy. The average tone ratio is computed as in Equation (2).

As words from the minutes are measured and counted, we are able to create a list of them all. And from this list, we can create a dictionary that has the word itself as key, and the frequency of the word as value. From here the possibilities of visualization are only limited by your imagination.

4.2 Machine learning

For the scope of our analysis, and given the dimension of our data set, we have opted for the use of simple machine learning algorithms: *LinearRegression* and *Lasso* models. The idea behind using machine learning is to fit a subset of the observations (training data) in order to estimate the parameters that generalize to the *true* model. In addition to that, a Lasso model adds a λ parameter in the estimation, which is trained to minimize the (root) mean squared error, i.e. reduces the variation not explained by the independent variables. The RMSE is computed as follows,

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}. \tag{4}$$

The results one may want to achieve is the convergence of RMSEs of the training and the test sets. This implies that the model generalizes well, and can be systematically used.

With an appropriate data set, we would apply a supervised model in order to predict the outcome/future, i.e. we would have a target variable (the change in the market stock index) whose variation could be predicted by two variables, the sentiment index, that we have developed, and the announced repo rate. Since our predicted variable is continuous, our algorithm would fit in the subcategory of *regression* models.

Following Elson (2015), predicting stock market changes well adapts to machine learning for the subsequent reasons: there are experts in the field, the theoretical and mathematical tool are well understood and developed, *but* the data changes very frequently, and it is thus relatively inefficient to use other models. Even though we initially keep only stock market changes on the announcement days, we follow this strand of literature.

5 Results

5.1 Content analysis

The following subsection shows the results obtained in the process of analyzing the contents of the minutes. We have looped over the whole text written by the Central Bank of Sweden and, depending on the specific purpose, the collected data is visualized in various ways.

5.1.1 Sentiment analysis

Before implementing a machine learning model and elaborating on the output, it is relevant to look at the development in the use of positive and negative words over time, in order to get a sense of the composition of the sentiment variable. Figure 3 plots the ratio of positive and negative words to the total number of words for each year.

Figure 3: Ratio of positive and negative words by year

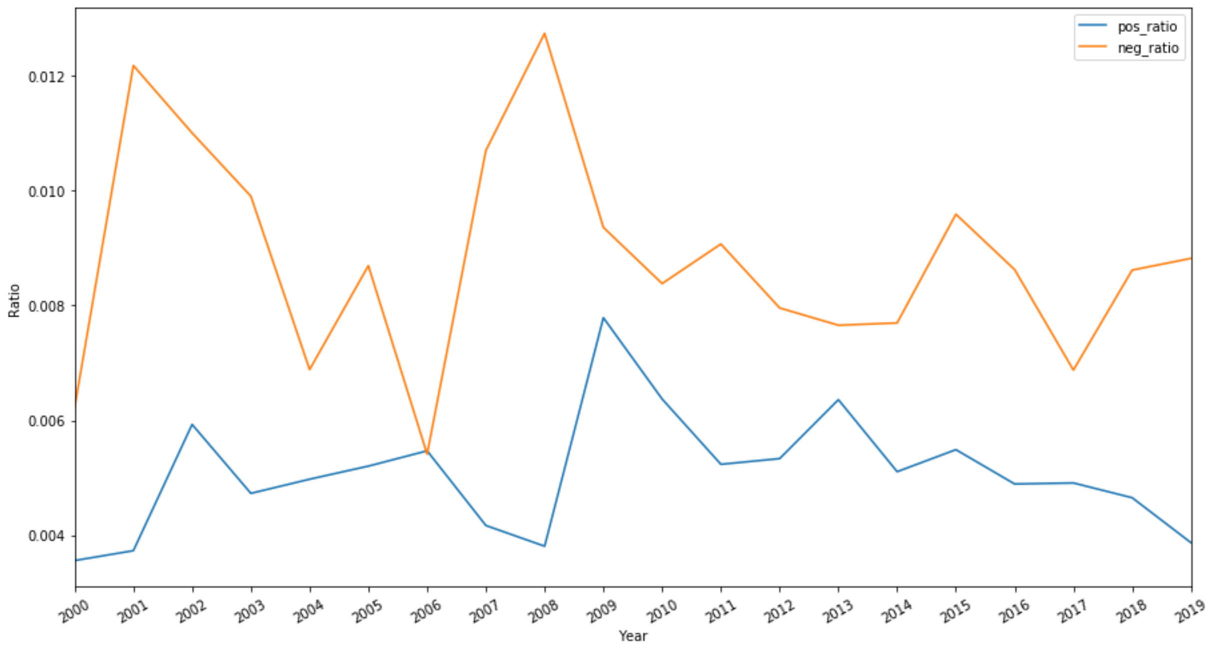
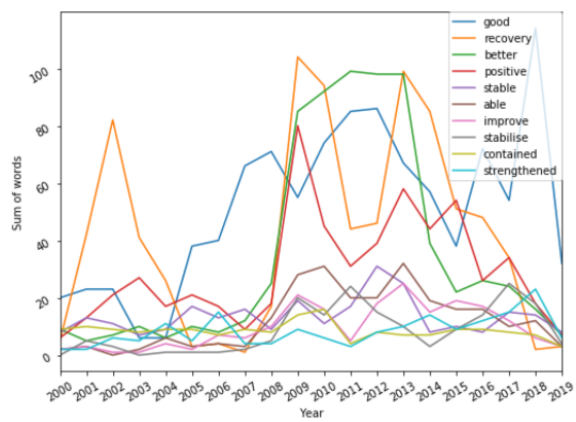


Figure 3 clearly shows that the amount of negative words identified in the minutes is greater than the amount of positive words identified, which is also confirmed by Figure 5. Further, the use of positive and negative words appears to follow the same general trend, except for the period of the financial crisis, where the ratio of negative words rises substantially for a period of around three years (2007 until 2010), whilst the ratio of positive words declines for a period of two years (from 2007 until 2008), followed by a sharp increase in the ratio in 2009. Including the rise in the ratio of negative words in 2001, when the world feared to dot-com bubble, it appears that the classification of positive and negative words is capturing the sentiment towards the economy.

In Figure 4, the top ten positive and negative words used in the minutes are plotted for each year:

Figure 4: Top ten positive (left) and negative (right) words by year



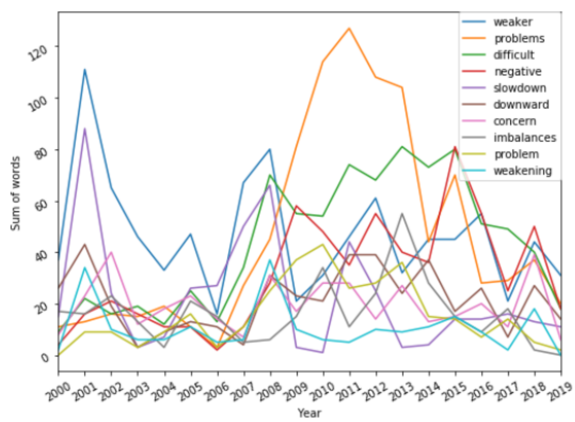
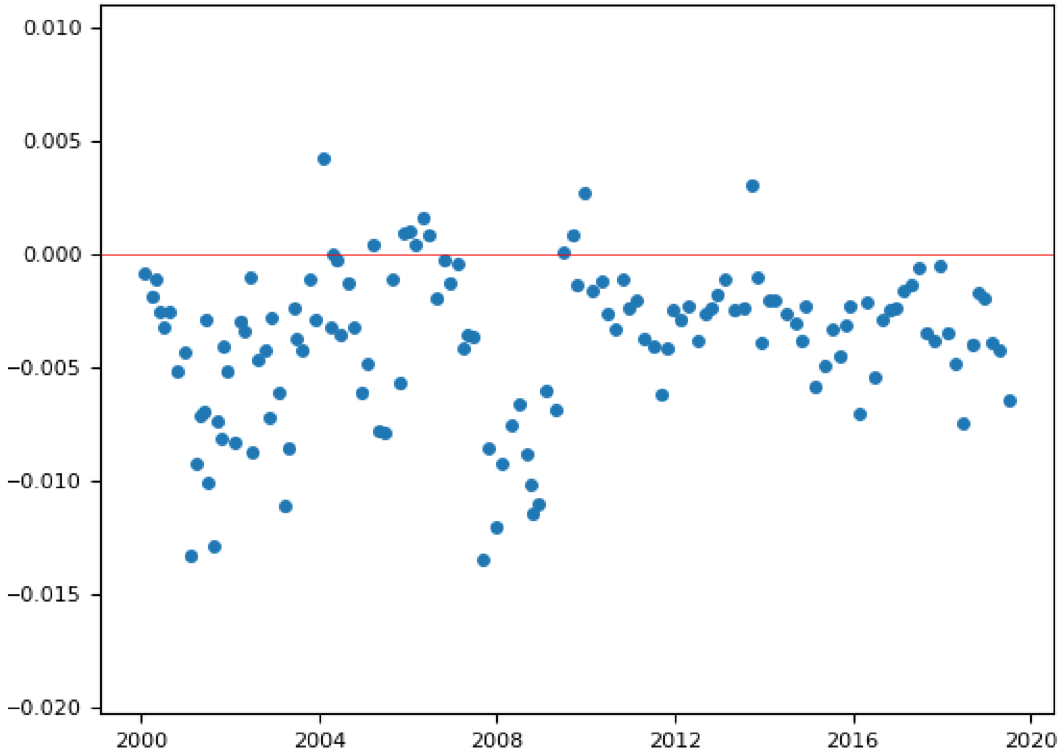


Figure 4 shows that positive words such as “recovery”, “strengthened”, “stabilise”, and “positive” and negative words such as “imbalances”, “concern”, “slowdown”, and “negative” are classified as expected. This finding provides some sense of proof that the right words are picked up by the sentiment analysis.

Figure 5 plots the average sentiment of each publication of minutes.

Figure 5: Average sentiment by publication of minutes.



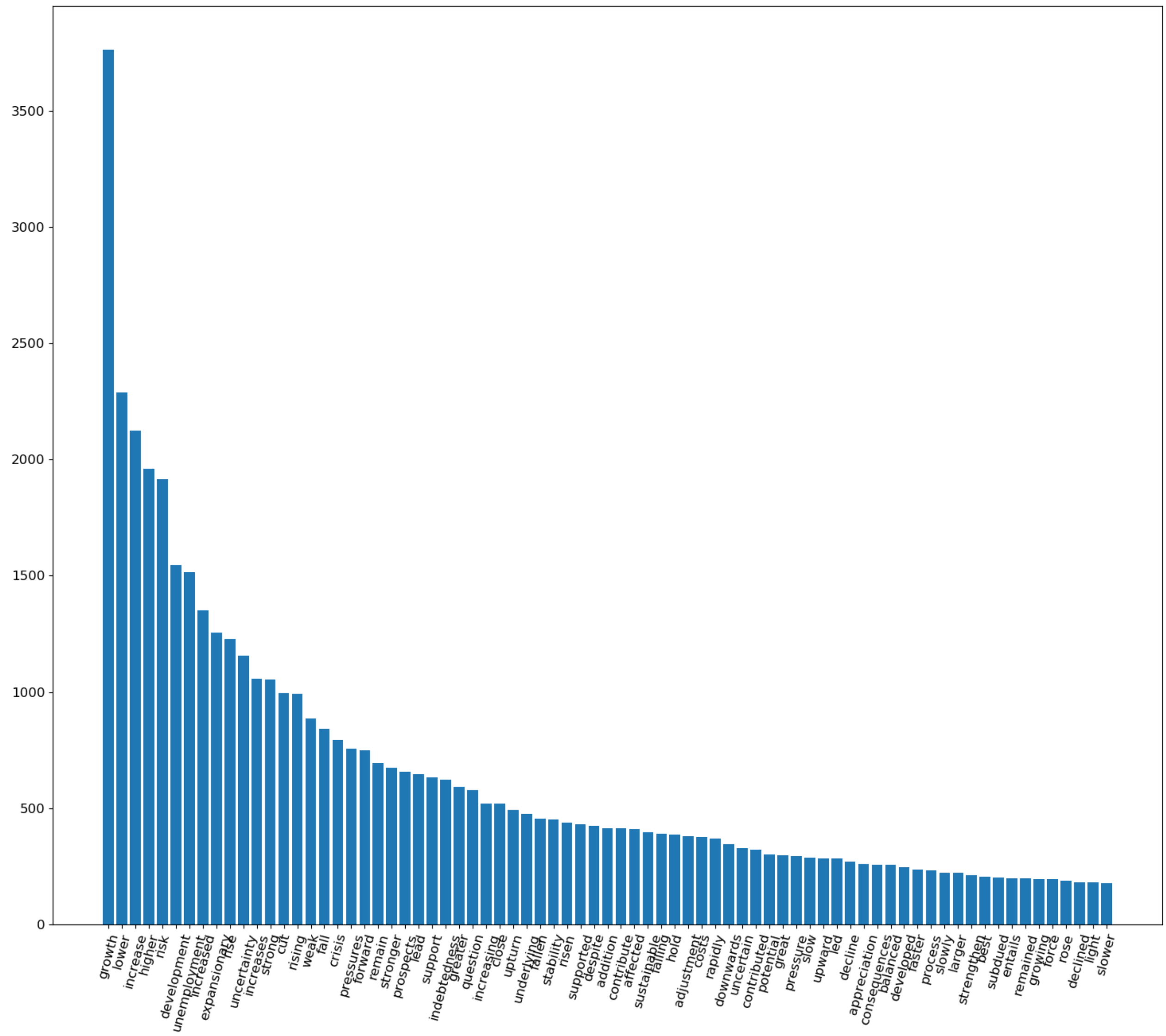
It is clear from Figure 5 that only a smaller number of the minutes are being classified as positive. The classifications of Figure 5 are not completely comparable with Figure 3, as in Figure 3, the rates are computed by year, whereas in 5, the rates are computed for each publication of the minutes, using the approach described in Section 3.4. Contrary to the expectations formed by Figure 3, the minutes, which are classified as positive, are mainly found in the period leading up to the financial crisis of 2007. This raises some concerns about the classification of words, as the opposite result was expected. However, in the same period, a large number or minutes is classified as negative, counteracting the concerns.

5.1.2 Word frequency

As expected, some words naturally appear more often in the minutes. Talking about the change of interest rate, it is not uncommon to mention words like "inflation", "growth", "unemployment", and "stability". By looping through every single word of every single file, we are able to catch the number of times certain words are mentioned. Take for instance the the word "growth". This particular word appears 3,762 throughout the files, which is an average of almost 28 times per analyzed document.

We applied the list of neutral words to the *json* file containing a dictionary with file names as keys, and every word in the particular file as values. We then took the 75 most common words, and plotted the outcome, which is shown in Figure 6. This particular bar plot has the advantage of the observer being able to approximate the absolute frequency of some words.

Figure 6: Frequency of most common words



Note: The bar plot shows the frequency of the 75 most common words from the neutral words lexicon.

Taken into to account that this number is not necessarily meaningful, and the words themselves are more important, we created a *wordcloud*, (Vu, 2018), see Figure 7. This plot has the advantage of highlighting which words are frequent.

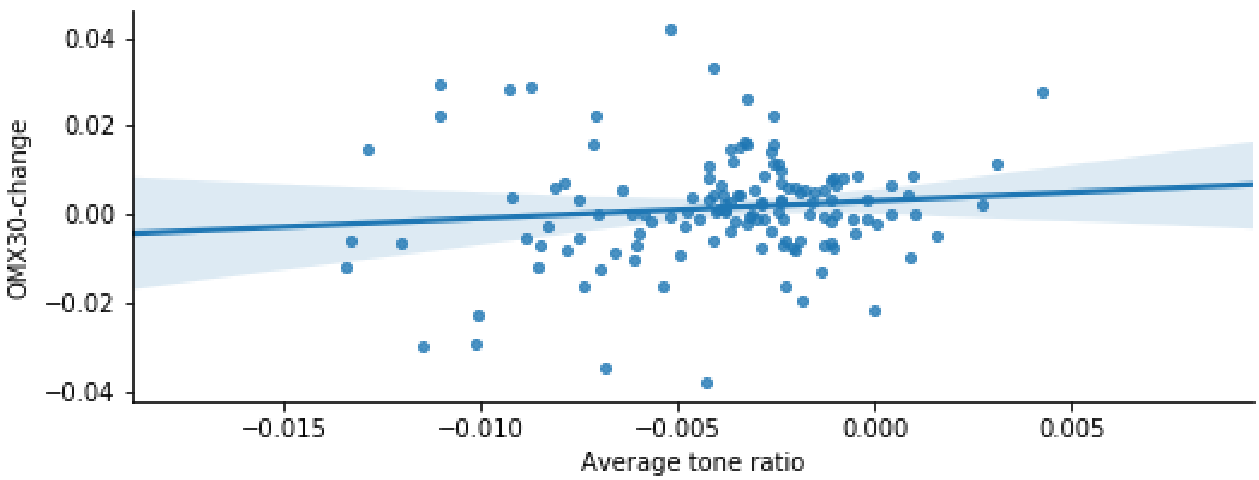
Figure 7: Wordcloud of the most common words

Note: The wordcloud indicates how frequent the 75 most frequent neutral words are compared to each other.

5.2 Prediction of the stock market value

Having thoroughly discussed the creation of the sentiment variable, we turn to the implementation of the machine learning techniques. First, we plot the relationship between the OMX30 and our sentiment variable. As shown in Figure 8, the linear relationship is close to 0, i.e., given the graphical representation, we shall not expect the two variables to correlate.

Figure 8: Correlation between OMX30 and sentiment



Note: N = 135.

In Table 3, we present the results of the algorithm. Specification (1) is the simplest model, as we try to predict the stock market change by using the average change of the stock market index. In specification (2), we include the announced repo rate as explanatory variable. Last, in specification (3) we include both the announced repo

Table 3: Root mean squared errors

	(1)	(2)	(3)
Linear Regression	0.009934	0.009879	0.0106166
Lasso	0.009934	0.010023	0.0101702

Note: $N = 135$, divided into three equal subsets of training, validation, and test data set. The RMSE presented are those of the test set.

rate and the sentiment variable. We apply *PolynomialFeatures* to specifications (2) and (3). From the results, it is clear that the Lasso model does not outperform the Linear Regression model. Moreover, no gains for adding explanatory variables are found: the test set's RMSEs are almost equivalent across the three specifications. We could conclude that we are possibly just capturing noise, and that our model has no predictive power, as discussed in reference to Figure 8.

6 Discussion and Analysis

6.1 Validation of sentiment model

Despite the remarks made in the previous sections on the validity of the classification into positive, negative and neutral tone, we have decided to perform a random sampling of the sentences, by looking for specific keywords. We repeat the exercise for four of the most frequent words: (-) problems and difficult, (+) recovery and better. The following two sentences contain the word "problems":

(1) there have been no general problems in the American banking sector,

(2) these problems stem from the us sub-prime market.

We observe that (2) is correctly classified as negative: problems did stem from the US sub-prime market. On the other hand, in (1), "problems" is not used negatively, as the speaker says that there are no general problems. In our opinion, this is due to the fact that we are *tokenizing* words, and therefore we cannot read "no problems" as a positive statement.

(3) Predicting the krona exchange rate has been very difficult.

(4) The financial sector will in future find it difficult to contribute significantly to restoring growth.

The examples of the word "difficult" were more in line with our expectations, as sentences (3) and (4) show. In addition to that, we noticed that this word is frequently used in the context of assessment and prediction.

(5) [...] had recently signalled that uncertainty over future policy could delay the recovery.

(6) mr ingves noted that the recovery in sweden has to date mainly benefited from increased consumption.

The same considerations as for “problems” apply in this case; delayed recovery is no good news for the reader. Turning to the last selected word, “better”, we report the two sentences:

(7) [...] and slightly better than expected with regard to sweden.

(8) this means that even if the markets are functioning better, [...].

Even though (8) might imply a negative outcome elsewhere, it is well predicted as a positive tone, since the markets are working well.

6.2 The use of Machine Learning

In this section, we discuss the use of machine learning model and possible reasons as to why our model yielded results suggesting model rejection, but let us begin by stating that the use of Machine Learning methods are not always fruitful and highly dependent on the research question and the type (and amount) of data available. Embarking on the task of predicting announcement effects on the stock market using Machine Learning, we knew that we faced some potential methodological and theoretical issues that could affect the interpretability and validity of our models. This being said, the research question we posed in the introduction is inherently a prediction problem. Secondly machine learning algorithms are widely used to predict stock markets because of the stock markets nature (Elson, 2015). Furthermore, we believe that testing models can be valuable even if results suggest model rejection as they could potentially provide some guidance for future research. Against the background of uncertain model validity, we deemed it necessary to test our models to the simplest prediction model possible. The model we opted for took the mean value of the change in opening and closing stock market value from our training set to predict the stock market.

Our main results (Table 3) suggested that there is no benefit of going from the most simple and parsimonious model to a more complex model using the sentiment variable. We have identified some possible reasons for these results, which shall now be discussed in further detail.

The announcement effect. It is possible that the the announcement effect – both in terms of actual repo rate and sentiment – has a only a marginal or no effect on the OMX30. Referring back to our theoretical background, according to the efficient market hypothesis only new and unexpected information is expected to have effects on the stock market (Mishkin and Eakins, 2012). A plausible explanation could therefore be that the outcome of the announcements made by the Swedish Riksbank are already expected by the markets. This holds some theoretical ground as the Riksbank aims at being as predictable and transparent. In this sense, the fact that our models do not yield any predictive power could rather be interpreted as a testament to the Riksbank success in communicating- and following the predicted repo rate and inflation trajectories. Following the same logic, it is also possible that other influential actors such as analysts and chief economists at Swedish financial institutions have been successful in predicting the Riksbanks decisions.

Limited data. Another problem with our models could be that the amount of observations was too limited for using a machine learning model. The problem of having a small data set is that both the training and test sets have too few observations to be correctly trained. The model will simply not be generalizable or possible to fit on an independent or unseen data set. Although there is no definite answer on how many observations are needed, it is likely that 135 are too few.

The stock market measurement. We rely on the daily change of the stock market index, but it is possible that the main effects are visible just in the hour after the announcement. This was not possible for us as data on the stock market index is not so detailed. Conversely, it is also possible that the a longer time horizon is needed to detect announcement effects. Although not part of our main results, we were also able to create a models with the outcome variable being the change in stock market value between the announcement date and 14 and 30 days after. Here we calculated the the change as:

$$\Delta\%14/30\text{days OMX30} = \frac{\text{closing value } +14/30 \text{ days} - \text{opening value}}{\text{opening value}}. \tag{5}$$

From the results presented in table 4, we can conclude that model (2) and (3) start performing slightly better than the simplest model (1). Adding the sentiment variable does however not increase the predictability of the model.

Table 4: Root mean squared errors

	(1)	(2)	(3)
Δ% OMX +14			
Linear Regression	0.044256	0.042339	0.0414817
Lasso	0.044256	0.042806	0.0416279
Δ% OMX +30			
Linear Regression	0.060311	0.056952	0.0561393
Lasso	0.060311	0.058264	0.0573672

Note: N = 135, divided into three equal subsets of training, validation, and test data set. The RMSE presented are those of the test set.

The sentiment variable One last possibility is that the sentiment variable did not manage to capture the sentiment of the minutes well enough to make predictions. As was noted in our section describing the data collection, only a small portion of the corpus contained words that were classified as either positive or negative.

As for other machine learning models that could be implemented these are the *Cross Validation* and *Time Series* models. The first method consists of dividing the data set into k subsets. The development data set is inserted in a loop repeated k times, where every time a different subset of the data is selected into validation data set, depending on the number of bins. In this procedure, usually the more bins there are, the better it is, as more fits can be performed, but everything is relative to the data availability. Even though this model is likely to have a good performance as compared to other models, we have decided not to implement it as our data set is too small to define an implementable number of bins (10, for example), having already excluded the test data.

In time series, the customary procedure is to use the oldest data to predict the evolution of the latest observations. In this case as well, we have preferred not to use this technique due to the small data set and to the fact that we do not include stock market indexes in the period between two announcements. From a theoretical point of view, we cannot consider our data set as a time series, especially because we do not try to predict in the trend of the OMX30, but rather its prediction conditional on a sentiment analysis, which occurs only at some points.

6.3 Ethical considerations

In light of the rapid expansion of online data sources available to researchers, an important and often overlooked aspect of research concerns ethics regarding the use of these data sources. To avoid the various pitfalls, Salganik (2017) proposes four main principles for researchers to consider, when performing analysis based on (big) data sources (Salganik (2017), pp. 294-301). These are based on the Belmont Report, which composes the concepts of ethics for pre-21st century research, and the Menlo Report, which can be viewed as a 21st century expansion of research ethics, adapted to the digital aspects of research in the modern age (Ibid). The four principles, consisting of respect for persons, beneficence, justice, and respect for law and public interest, can be summarized as follows:

Respect for persons: Respect for persons consists of two main principles – autonomy of individuals and protection of individuals with diminished autonomy. In summary, the two principles suggest that researchers should not do things to people without their consent (Ibid). Seen in relation to the data chosen for the analysis of this paper, the data consists of macro-level time series, and hence is not affecting any individual, from whom consent must be given. For this reason, the analysis is viewed to be in compliance with the principle of respect for persons. Moreover, we are not distorting or manipulating the opinions expressed by the board members in the minutes.

Beneficence: As with respect for persons, the principle of beneficence is comprised of two main elements – no harm, and maximization of possible benefits and minimization of possible harms. Beneficence as a main principle is the result of the tradition of human trails in medicine, and is broadly speaking generalizable to all experiments involving participation of individuals. Besides the consideration of potential harm to individuals, there is also a consideration of potential harm to overall systems or patterns (Ibid). By applying the principle to this paper, it is concluded that the principle of beneficence is not violated, for the same reason that the principle of respect for persons is not violated – data does not consist of individual-level observations, but on macro-level time series. Potentially providing proof that sentiment of publications from the Swedish Central Bank can be used to improve predictions of outcomes in the Swedish stock market is unlikely to cause disruption of the entire investment market in Sweden. Therefore, in the case of this paper, it is possible to maximize possible benefits without causing harm to any individual or to the general system.

Justice: The principle of justice refers to the problem that some groups of a study may carry all the potential risks, whereas one or more groups gain all the potential benefits. Stated differently, the principle of justice is meant to secure that researchers do not pray on the powerless (Ibid). In this case, no one is being prayed on, as the data consists of macro-level time series. However, questions can be raised concerning the groups that could potentially benefit from the results of the paper. It is likely, that a certain level of knowledge is required to apply text analysis to a model, which then predicts outcomes on the stock market. With this in mind, the group most likely to benefit from the results of this paper are stock market investors, who are in turn thought to belong to the wealthiest part of society. However, as the Swedish welfare state collects taxes at a high marginal rate, any potential distorting effects are partly offset by the system. As is it stated in the principle of justice, that appropriate compensation is a way to minimize potential harm to specific groups, the Swedish tax system can be viewed as part of the solution (Ibid).

Respect for law and public interest: Whilst the principle of beneficence focuses on respect for persons, the principle of law and public interest can be viewed as an extension to the principle of beneficence, in the sense that researchers are encouraged to also consider the law as well as guidelines of any given system or data source used. In the Menlo Report, it is stated that the principle of respect for law and public interest consists of two main elements – compliance and transparency-based accountability. Compliance is of special interest in this study, as it states that researchers should consider laws, contracts, and terms of service, for example when scraping websites for data. Transparency-based accountability means that researchers should avoid using people's data to do research, if consent was not given, or the data collection was not intended for that purpose. Further, researchers should always be willing to publish their results and research procedures in order to give both the research community and the public the chance to review it. In this way, researchers are less likely to perform unethical research and/or experimentation, as they are accountable to everyone who could be considered as a stakeholder (Ibid). In this paper, the considerations regarding compliance are relevant to include, as a fair amount of data is from the web page of the Swedish Central Bank. Even though the data is not scraped or parsed through an API or an HTML parsing module, it should still be a concern whether or not the data collection is in compliance with the general terms of service. Shiab, 2015 suggests checking the *robots.txt* file of any web page from which a researcher might be scraping. A check of the robots.txt file from the Swedish Central Bank's web page yields the following lines (The Swedish Riksbank, Robots.txt, 2019):

"User-agent: *

Disallow: *.doc

Disallow: *.docx

Disallow: *.xls

Disallow: /*?query=*

Disallow: /*/?year=*".

From the above, little knowledge can be extracted. What is apparent, however, is that it is disallowed to make queries from the web page, which tries to extract everything (*), and to make queries, that extracts everything for all years (*/?Year=*). Given these specifications, the data collection for this paper is thought to be in compliance with the overall guidelines for scraping of the Swedish Central Bank.

7 Conclusion

A few conclusions can be drawn from our work. First of all, our study shows that the simple bag-of-words method in conjunction with a context specific dictionary as suggested by Correa et al., 2017, can be a valuable way to analyze both the sentiment and the content of central bank communication. The method has some obvious limitations, including accuracy and the inability to process complete sentences. Notwithstanding, when validating our model by looking at how it had classified a random selection of sentences, we found that the sentiment was well captured most cases. As for the second aim of our study, we can conclude that the machine learning models were not able to make any convincing predictions of the stock market change. We discussed a couple of reasons as to why this could be the case, such as: limited data, marginal announcement effects and that our measurement of the stock market may not be viable. For future studies we recommend the use of an enriched dictionary and of a more comprehensive data set.

References

Blinder, Alan S et al. *Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence.* Working Paper 13932. National Bureau of Economic Research, 2008. DOI: 10.3386/w13932. URL: http://www.nber.org/papers/w13932.

Bloomberg. *OMX Stockholm 30 Index.* 2019. URL: https://www.bloomberg.com/quote/OMX:IND (visited on 08/26/2019).

Correa, Ricardo et al. *Constructing a Dictionary for Financial Stability.* IFDP Notes 2017-06-28. Board of Governors of the Federal Reserve System (U.S.), 2017. DOI: 10.17016/2573-2129.33. URL: https://ideas.repec.org/p/fip/fedgin/2017-06-28.html.

Elson, Torran. *When should Machine Learning be used?* 2015. URL: https://www.smithinst.co.uk/insights/when-should-machine-learning-be-used/ (visited on 08/27/2019).

Fama, Eugene. "Market Efficiency, Long-Term Returns and Behavioral Finance". *Journal of Financial Economics* (1998), 283–306.

Karadi, Peter and Marek Jarociński. "The transmission of policy and economic news in the announcements of the US Federal Reserve". *VOX* (2018).

Mishkin, Frederic S. and Stanley G. Eakins. *Financial Markets and Institutions.* 7th. Prentice Hall, 2012.

Moniz, Andy and Franciska de Jong. "Predicting the impact of central bank communications on financial market investors' interest rate expectations". In: *The Semantic Web: ESWC 2014 Satellite Events.* Ed. by Troncy R. Sack H. Papadakis I. Tordai A Presutti V. Blomqvist E. Cham: Springer, 2014.

Salganik, Matthew J. *Bit by Bit: Social Research in the Digital Age.* 1st. Princeton University Press, 2017.

Shiab, Nael. "On the Ethics of Web Scraping and Data Journalism". *J-Source.ca* (2015).

Swedish Central Bank. "Perspectives on the negative repo rate" (2016).

– *Repo rate, deposit and lending rate.* 2019. URL: https://www.riksbank.se/en-gb/statistics/search-interest--exchange-rates/repo-rate-deposit-and-lending-rate/ (visited on 08/25/2019).

The Swedish Riksbank, Robots.txt. *Robots.txt.* 2019. URL: https://www.riksbank.se/robots.txt (visited on 08/27/2019).

Vu, Duong. *Generate WordClouds in Python.* 2018. URL: https://www.datacamp.com/community/tutorials/wordcloud-python (visited on 08/27/2019).

Woodford, Michael. *Central Bank Communication and Policy Effectiveness.* Working Paper 11898. National Bureau of Economic Research, 2005. DOI: 10.3386/w11898. URL: http://www.nber.org/papers/w11898.

Appendix A: Complete Dataframe

Table 5: Appendix A: Complete Data Frame

	repoann	repoeff	repo_ch	closing value	opening value	dif abs	dif frac	dif +30d	dif +14d	N positive words	N negative words	Average tone ratio
mean	1.9034	1.9280	-0.0246	1074.9286	1073.5903	1.3383	0.0018	1.0023	1.0041	47.5303	79.9621	-0.0039
std	1.6941	1.6966	0.2526	336.8038	337.9245	11.7090	0.0135	0.0849	0.0609	31.0897	43.3391	0.0034
min	-0.5000	-0.5000	-1.7500	476.2000	465.9800	-27.0600	-0.0346	0.8013	0.8489	4.0000	12.0000	-0.0134
25%	0.2500	0.2500	0.0000	807.8600	794.6525	-6.7000	-0.0053	0.9549	0.9682	22.7500	37.7500	-0.0059
50%	2.0000	2.0000	0.0000	1055.7700	1052.6300	1.0950	0.0010	0.9956	0.9964	42.5000	81.0000	-0.0033
75%	3.7500	3.7500	0.0000	1374.4325	1369.9975	7.7925	0.0073	1.0265	1.0267	64.2500	110.5000	-0.0018
max	4.7500	4.7500	0.5000	1700.3500	1695.3700	36.6600	0.0559	1.3394	1.3090	129.0000	171.0000	0.0042