Lorena Gril, David Rackl

Financial tweets and the stock market

Investigating the effects of twitter sentiments on stock prices

Alpen-Adria-Universität Klagenfurt
Department of Mathematics

Klagenfurt, May 17, 2021

Abstract

In this paper we will investigate the effects of financial tweets about certain stocks on their respective stock prices. For this task we will use some data [1] obtained via the StockerBot API [2] in which each tweet has undergone sentiment analyis to determine wheter we expect to see a postive or negative impact on stock price [3]. This data is then binned (daily) and a combined sentiment level for all tweets within that day is calculated. Lastly we convolute the (appropriately scaled) resulting function with the differenced stock price. This forms the basis of our investigation.

Contents

1 Introduction

1.1 Objectives

Our main objective will be to determine wheter financial news (financial tweets) and their corresponding sentiments have any meaningful impact on the respective stock prices. At this point it should be noted that if correlation is found, this not necessarily implies causality as financial news come in a variety of flavours such as advice, rumours, factual news and others. The latter is particularly tricky here since a potential drop in the stock price will likely be influeced mainly by the original action of the company itself instead of the reporting about it. However, we find that nonetheless such an investigation may produce interesting results even if they should be taken with a large grain of salt.

1.2 Overview of the dataset

The dataset we used originated from data generated by the StockerBot API [2] - a program that tracks and records tweets from a variety of sources in the financial sector such as @GoldmanSachs, @Forbes and @jimcramer. This data was then subjected to sentiment analyis to determine wheter a tweet had a postive, neutral or negative sentiment [1]. We then combined this with daily historical data of the stocks mentioned in a specific tweet [3]. To make everything a little more accessible, we filtered this data to make sure every stock we looked at had sufficient information that would warrant an investigation. The main criterion for this filtering was that a stock had to be mentioned a sufficient number of times over a given time frame. This left us with 13 viable candidates, among which we can find companies like Flex(FLEX) and LyondellBasell(LYB).

2 Methods

For each stock we looked at the timeframe from its first mention to its last. Mentions were then grouped by day to make it match the daily stock data. Sentiments were then assigned the following values

$$s = \begin{cases} 1 \text{ if the sentiment was postive} \\ 0 \text{ if the sentiment was neutral} \\ -1 \text{ if the sentiment was negative} \end{cases} .$$

Using this we can calulate the mean sentiment of a given day by

$$\mu_s = \frac{1}{n}\sum_{i=1}^{n} s_i,$$

where n is the number of tweets that stock was mentioned in that day. If a tweet was not mentioned in a given day we assign μ_s the value 0, i.e. we consider no mention the same as net neutral sentiment. This forms a daily time series of sentiments, which will be used later.

Next we need to access the historical stock data for the stock we are investigating. Since historical stock data is usually given in the form of a high and low value for each day, we first computed the average of the high and low values of the stock to obtain a single representative for each day. For the purpose of our investigation it will be useful to look at the differenced stock data, since we want to investigate how the stock changes and constant trends will not be relevant. This diffenced time series forms the second component of our approch. Before we move on to the next stage, we will also scale our time series of sentiments by the absolute value of the differenced stock price series at each point to make them match in magnitude.

Lastly we will use the convolution of both time series (however we do not reverse the second series) to visualize how the stock price is affected by the sentiments of tweets about it for different lags (although only postive lags are really relevant here).

3 Results

We will discuss only two examples in this section due to space constraints, however if the reader is interested more examples can be obatined by executing the R notebook attached. The following examples were picked because we believe that they illustrate the possible behaviours quite well.

3.1 IHS Markit Ltd. (INFO)

For IHS Markit Ltd. we begin by plotting the daily average of the stock price for July 2018. For each day we added a dot to indicate the mean daily sentiment μ_s.

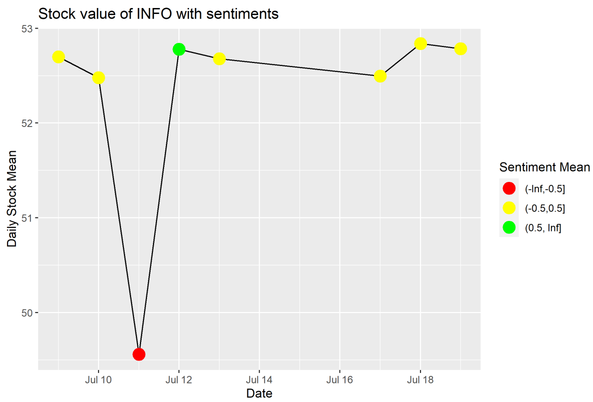

Figure 3.1: INFO daily stock price with sentiments

In this instance we can already see that the sentiments appear to match the behaviour of the stock perfectly. We can further observe this behaviour by taking the convolution as described above.

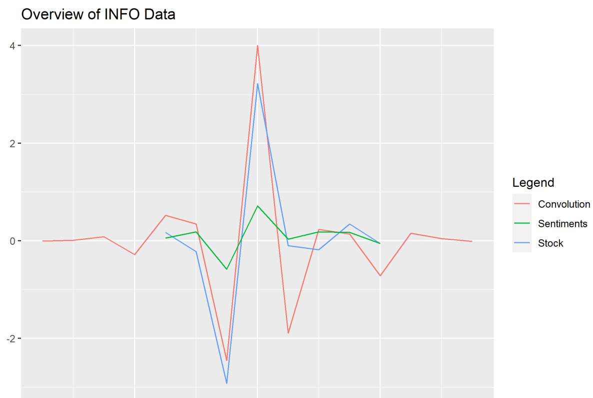

Figure 3.2: INFO diffenced stock with sentiments and convolution

We observe a large positive spike in the convolution of both sequences for shift 0 indicating that their behaviour is almost perfectly synchronous. This behaviour could potentially be used for computational detection in larger datasets. This spike indicates correlation between the chnage of the stock itself and the news about it, even when there is no obvious major event such as a crash. However not all sequences we tested behaved as well as this one as we will see in the next two examples.

3.2 Flex Ltd. (FLEX)

As the fundamental methods haven't changed since the last example we will just present the results without further discussion.

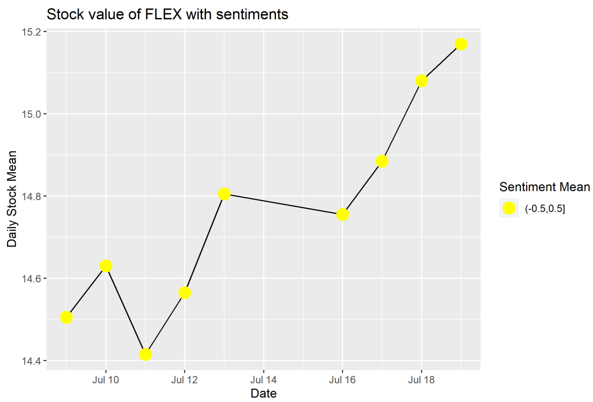

Figure 3.3: FLEX daily stock price with sentiments

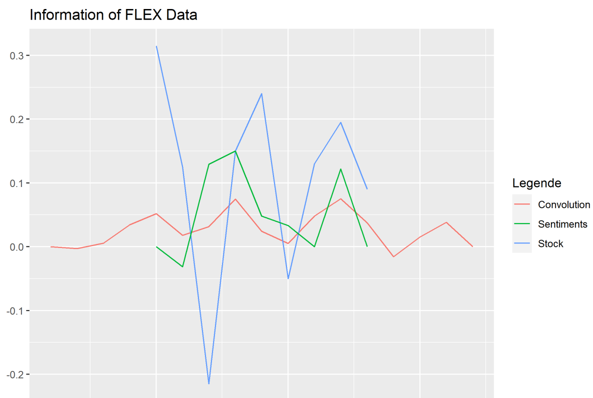

Figure 3.4: FLEX diffenced stock with sentiments and convolution

For Flex Ltd. we cannot find a large spike in the convolution indicating strong correlation. However if we look a little closer we can see that do look quite similar. The convolution alone, does not work as well in this case as it always shifts the second sequence by the same amount. However the stock and sentiment sequences don not match perfectly in response times in this example - the first sentiment spike and the second sentiment spike provoke responses in the stock at different speeds. This matches our real world expectation that news and their responses do not always propagate at the same speed — some news spread faster, others slower. To fix this we could allow for warping of the sequences before convoluting them.

Lastly, not all stock show observable responses to twitter news as our last example demonstrates.

3.3 International Paper Company (IP)

Again, as the fundamental methods haven't changed we will just present the results.

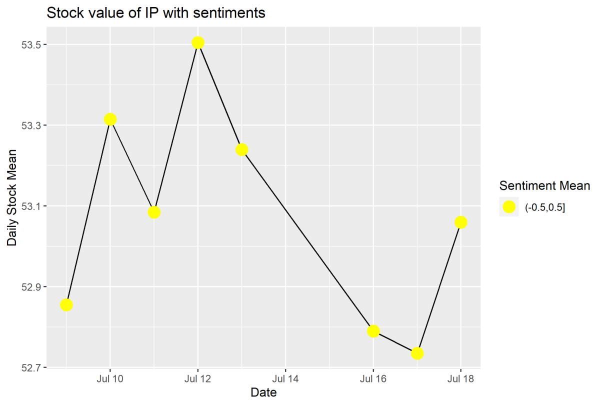

Figure 3.5: IP daily stock price with sentiments

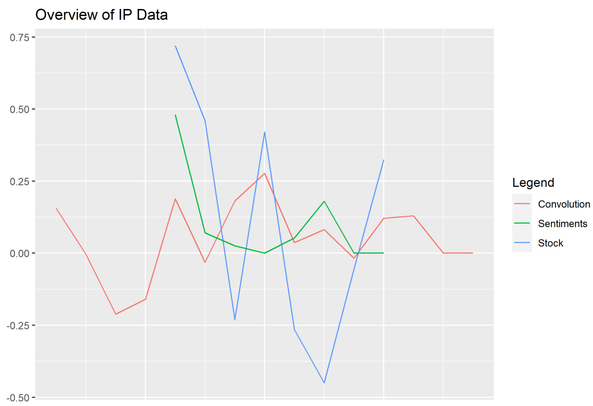

Figure 3.6: IP diffenced stock with sentiments and convolution

In the case of the International Paper Company stock we can find little to similarity between both sequences, with the convolution telling a similar story. We believe that allowing for warping of the sequences would not help much either. In this last example the stock price change and twitter news about it appear to be largely uncorrelated.

4 Final thoughts

After looking at our examples, we may find ourselves asking if we can even give a clear answer to our research question. We personally believe that we cannot. While many examples do indicate correlation, there also some that do not. With all that being said there also many factors we did not account for, such as the news source or the type of news itself (rumour, fact, speculation, opinion, ...). All these may play an additional role when it comes to evaluating the impact of news on stock prices.

Even though our investigation was not able to produce a clear result, we do believe that both the research question itself and the methods discuss have potential as in some cases we were able to observe some promising results with FLEX and INFO (and some others not mentioned here). With some further research and additional tweaking it may one day be used for modelling and predicting stock prices.

Bibliography

[1] Rathi, Vivek, *Sentiment Analysis on Financial Tweets*. kaggle, 2019. [Online]. Available: https://www.kaggle.com/vivekrathi055/sentiment-analysis-on-financial-tweets

[2] Wallach, David, *StockerBot*. GitHub, 2018. [Online]. Available: https://github.com/dwallach1/StockerBot

[3] yahoo finance, *Historical Stock Data (daily)*. yahoo, 2021. [Online]. Available: https://finance.yahoo.com