Impact of News Events on the Financial Markets

Miha Torkar
Jožef Stefan International Postgraduate School
and
Artificial Intelligence Laboratory,
Jožef Stefan Institute,
Jamova 39, 1000 Ljubljana,
Slovenia
miha.torkar@ijs.si

Dunja Mladenič
Jožef Stefan International Postgraduate School
and
Artificial Intelligence Laboratory,
Jožef Stefan Institute
Jamova 39, 1000 Ljubljana,
Slovenia
dunja.mladenic@ijs.si

ABSTRACT

In this work we investigate how news events can be used to predict the financial markets. Namely we built a time series model that includes features obtained from the news and investigated whether the changes in volume of traded shares can be predicted more accurately with this information. The time series model that was built is of an ARMA-GARCH type, because we wanted to account for any clustering of the volatility that is normal for the financial markets. The models were evaluated with the Akaike and Bayesian Information Criterion, while also being compared to the baseline model that did not include any features from the news. Overall our results show that there is an improvement in the model when the information from the news is used and hence show a promising avenue for future research work.

1. INTRODUCTION

The predictability of future movements of financial markets is a well researched area in the literature and offers many interesting theories. Financial institutions and investors have been using various sources of information in order to increase the accuracy of their predictions and consequently outperform others. There are several approaches to building the best predictive model, but in general we can divide them into two categories: technical analysis and fundamental analysis. On one side we have models that are based on the historical market data and believe that the past movements will repeat themselves. This is the so called technical analysis approach to modelling markets and believes that an experienced observer can detect the repetition of a pattern from a graph of market data. The effectiveness of this approach was tested by [9].This approach however does not offer any reasons why market movements would repeat, and in order to incorporate more fundamental believes, a second approach named fundamental analysis was developed. These models use data that is available from multiple sources, which ranges from company's balance sheet data, financial market data like company's index, financial data about government activities to data about political or geographical circumstances presented in news.

From the variety of sources for fundamental data, we will focus on how news can be used in modelling financial markets. There have been various studies on this topic, which differ by the extent of the analysis of textual data describing news story. Initially research focused on the impact of frequency of news stories on the market movements (see [6] or [7]). In these papers authors found some correlation between increased number of published news stories and larger market movements. To extend this approach, researchers also analysed the content of the news stories, which led to determining the sentiment of the news articles and consequently determining the impact on the market on the basis of whether it is positive (upward trend predicted) or vice versa (see [3] or [4]). Our approach is similar to the second one, but instead of determining sentiment for each news event, we use the effect that the past similar events had on the market as a proxy for the impact of the current event.

We define an event as a collection of news articles from different sources about the same story in the news. From this collection of articles we will be able to extract the topic, date, location, social score (how trending was this event on social media) and all of the entities involved in the event, which will add to the complexity of our dataset. One possible source of such dataset is a system called Event Registry (see [5]), which automatically extracts events from news articles. Using this type of data source is novel for this research area and hence serves as an additional contribution from this work.

In order to see the impact that these events had on the financial market, we looked at how the volume of traded shares changes on the days of the event. We obtained several features from the news events and checked whether they would allow us to improve the time series model for the volume change.

2. DATA

For the historical market data of the company, we collected the following values (prices): Open, High, Low, Close, Volume. In addition to the market values of a company we also used the value of the market volatility index VIX (closing price). Secondly with the use of the Event Registry system we obtained all of the news events relating to the company. The general description above was intentional as it can be applied to any publicly listed company, but for our specific example we will use data about investment bank Goldman Sachs (GS). In Figure 1 we present the dataset we will be modelling and predicting, where it can be seen that there are large spikes at certain time periods. Moreover the volume change graph demonstrates that we have clusters of periods with high volatility.

2.1 Data Description

Our dataset spanned from 2.12.2013 to 30.12.2016, which offered us 777 trading days on which we were able to collect historical market data. On the other hand the number of news events that occurred in that time period was significantly higher. When we singled out events using 50 as the relevance threshold we obtained 4336 events (details of how Event Registry calculates the weights of concepts in

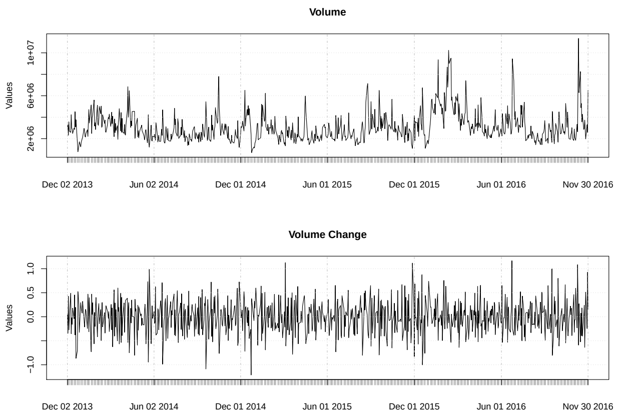

Figure 1: Volume Dataset

the event are presented in [5]). For testing purposes we split our dataset into two parts, where we allocated 757 observations for training and testing, while we used the remaining 21 observations for the out of sample prediction.

For each event we also obtained related events to it. These related events are obtained by computing the TF-IDF weights on the concepts present in the event and then by using cosine similarity measure other events with similar concept weights are found (see [5]). These past similar events formed a crucial dataset, because we could link them to the market movements and deduce their impact.

2.2 Data preprocessing

In order to measure the impact of the event we will be considering the change in volume that occurs between the closing value of today and the day after the event. Specifically we will be analysing and predicting the value of

$$(\text{Volume Change})_t \equiv VC_t = \frac{V_{t+1} - V_t}{V_t}, \tag{1}$$

where V_t is the value of volume at time t. In this formulation we used future values, so that we can observe the impact an event has on future volume. It should be noted that this value is calculated in the same manner as that of the returns of shares, which we will also use in our analysis. The formula is identical except that we replace values of volume with those of closing price P_t. Hence we write

$$(\text{Returns})_t \equiv r_t = \frac{P_{t+1} - P_t}{P_t}. \tag{2}$$

Changes in the volatility index VIX were calculated with the same formula. Additionally we also added rolling 5 and 10 day moving average of volume change to the feature set, so that we would have another measure of impact an event has on value. Hence the complete list of all of the features that were obtained from the stock market is:

- Returns
- Volume Change
- Open
- High
- Low
- Close
- Volume
- VIX Close
- VIX Change
- Rolling mean 5 days
- Rolling mean 10 days
- Rolling EMA 5 days
- Rolling EMA 10 days

In order to reduce noise in the dataset of events we selected only the most important ones. Namely we set a limit, which determined the lower boundary for the relevance of the events to the given company. Hence if an event was not relevant enough we discarded it. This naturally raises the issue of selecting the appropriate boundary for the relevance of the events and after testing several values we selected 89 out of 100. After this we were left with 424 events.

What should be noted here is that many events occurred on the days when the markets were closed (weekends, holidays), so they had to be linked with the next possible trading day that followed. Another issue was also that in many cases multiple events occurred on the same day and hence we were not able to isolate the effect of a single event, but rather looked at the average effect of all the events on a specific day. Therefore we are also relying on the fact that there are no duplicates in the our dataset.

3. METHODS

3.1 Predictions from Similar Events

In order to build our time series model, which uses data from news events, we examined the effects of past similar events on the market for each new event separately. This was done by first extracting the market information on the days of the similar events and combining them with the additional information about the event (social score, relevance to the company, correlation to current event, number of articles). On average each event had between 5-17 similar events from our dataset.

Illustration of this processes for an event E *"Goldman Profit Rises 74% as Bond Trading Beats Estimates"* with two similar events (SE) is represented in Figure 2. On the two days of the SE_1 and SE_2 we then collect all of the available stock market values for the company and index VIX. It should also be noted that the time difference between event and past similar events is not limited or predetermined. In this case we have events that are years apart.

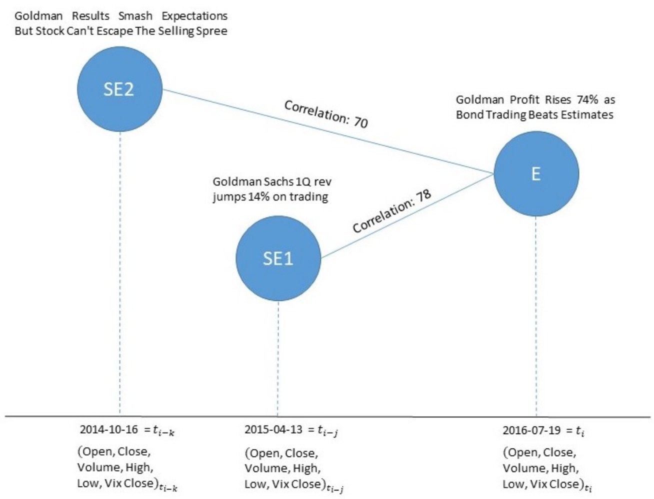

Figure 2: Similar Events

Once the dataset of market impacts from past similar events was obtained an Ordinary Least Squares (OLS) model was fit to the dataset. In the Figure 2 this would be all of the market variables available at times t_{i-k} and t_{i-j}. This dataset then allowed us to use data from the market and current event to form a prediction of what the future market movement could be. This procedure was done for both volume change and returns. Several choices of external regressors that were used in OLS were tested, where the main issue arose when there were fewer similar events than features to be fit. Namely we fitted two models, one with all of the market information available and one where only Returns, VIX Change and concept weight were used to predict Volume Change. Similarly two models were fit to predict Returns, where Volume Change, VIX Change and concept weights were used as features in the second one. In addition we also calculated the average of volume changes and returns from dates of similar events and used it as a feature. In order to take into account the correlation to the similar events we created an extra feature that represented the value of average of volume changes and returns multiplied by the normalized correlation (correlation/100). All together the following features were added to the model (all of which are predictions from similar events):

- Number of Events that day
- Returns (all regressors)
- Returns (selected regressors)
- Volume Change (all regressors)
- Volume Change (selected regressors)
- Average Returns
- Average Returns ∗ Correlation
- Average Volume Change
- Average Volume Change ∗ Correlation

With the OLS model trained on the dataset of similar events we were then able to predict the impact of the original event on the volume change. It should also be noted that the OLS model was selected in order to avoid over fitting an ARMA type model. The predictions that were obtained in this way were then used in the next step when we were building our regression model.

3.2 ARMA-GARCH MODEL

The main model in our analysis will be of the ARMA-GARCH type, because this formulation allows us to capture the effects of values from previous periods and account for clustering in volatility.

The ARMA model is in its general form specified as:

$$\begin{aligned} X_t &= \mu_t + Z_t, \\ \mu_t &= \sum_{i=0}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \omega_j Z_{t-j} \\ Z_t &= \sigma\epsilon \Rightarrow Z \sim N(0, \sigma^2) \end{aligned} \tag{3}$$

where X_t is the target variable at time t, μ_t is the equation for the mean at time t,$\epsilon \sim N(0, 1)$ is an iid normally distributed noise term and σ is the variance at time t. The first step in modelling is to determine the best values for (p, q) according to the evaluation criteria.

The above model however assumes that the variance (σ) is constant. This assumption is dropped due to the clustering in the volume change (periods of high changes are followed by lower ones). This type of models are called Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models ([2]) and have been shown across literature to improve the models (see [8]). They allow us to model $\sigma = \sigma_t$ by an additional non linear model. Namely general formulation of the problem is the following:

$$\sigma_t^2 = \beta_0 + \sum_{i=1}^{r} \beta_i Z_{t-i}^2 + \sum_{j=1}^{s} \gamma_j \sigma_{t-j}^2 \tag{4}$$

So the current value of the variance also depends on the previous values of the variance. From this general formulation one has to determine the value of (r, s) that is most suitable for the model. Finally we can add external regressors (features) to the equations 3 as an additional sum term in the first line of the equation.

3.3 Evaluation Criteria

In order to asses which model was most suitable for the given dataset we have used a variety of different tests. In order to be able to assume that the time series is stationary we first ran Augmented Dicky Fuller test and KPSS test. In both cases the p-value was below and above the 5% threshold respectively. To differentiate between variety of different model we analysed the performance of each model by the Akaike information criterion (AIC) and Bayesian information criterion (BIC), which are the classical information criteria used across literature that also penalize model for high complexity ([1]). Finally in order to determine the significance of the features we will also build a model without the predictions from the similar news events. This will serve as our baseline model and the difference in performance will be then the main measure of how significant these features are.

4. RESULTS

Our first step in determining which features are significant for our analysis was running a t-test selection procedure for all regressors. From this analysis it was determined that only the following features were relevant for our model:

- VIX Close price
- Rolling mean 10 days
- Rolling mean 5 days
- Rolling EMA 5 days
- Rolling EMA 10 days
- Prediction of Returns

Hence our first finding was that the predictions that we obtained from the similar events were significant, but instead of using the predicted volume changes, the predicted returns turned out to be more significant for our model.

With this set of regressors the best ARMA model was of order (2,2), so two lagged terms in both variables were included. This model preformed according to the evaluation criteria shown in the table 1. As mentioned before in order to capture clustering in the dataset we modelled the variance term even further with a GARCH type model. Again the same evaluation criteria were used and a grid search was preformed to find the best coefficients (r, s) for order of the model. This resulted in a GARCH (5,1) model, with the evaluation criteria presented in the same table 1.

Comparison of results yields what improvements have been made by including the feature. This table also serves as a comparison to the baseline model. Since the AIC is merely a heuristic, differences between its values are important. So the predicted feature from similar events improves AIC value by 10.22, while the improvement of our final model is 32.69.

One way of interpreting this difference is also in terms of relative likelihood, which is defined as $exp((AIC_{min} - AIC_i)/2)$ for model i, where AIC_{min} represents the lowest AIC value from all models. Hence the baseline model is $7.97 * 10^{-8}$ times as probable as the best model to minimize the information loss, while the ARIMA(2,2) with feature is $1.06 * 10^{-5}$ times as probable. However we can see that due to additional complexity of our final model, BIC criterion is actually the highest for our chosen model. Hence an optimisation by the BIC criterion would result in a different model.

	AIC	BIC
Baseline	-925.57	-874.66
ARMA(2,2)	-935.35	-879.81
ARMA(2,2)-GARCH(5,1)	-958.26	-869.96

Table 1: Evaluation criteria train set

It should be noted that we have also tested the assumption about the distribution of the standardized residuals. So instead of using normal distribution we fitted the model with student t distribution with and without skew. Some minor improvements in AIC value were obtained, but not really significant. Hence we kept the distribution as normal.

Final test included out of sample predictions, where our best chosen model was ARMA(2,2)-GARCH(5,1). The time period for prediction was 1 month (December 2016) with 21 trading days. We also calculated errors when making these predictions, where our best model from above scored a value of 0.1344944 for Mean Absolute Error and 0.1565652 for Root Mean Squared Error. Figure 3 shows how the model predicted future values, where the middle line represent predictions for the mean value. Additionally the plot also included intervals of upper and lower 95 and 80 quantile range.

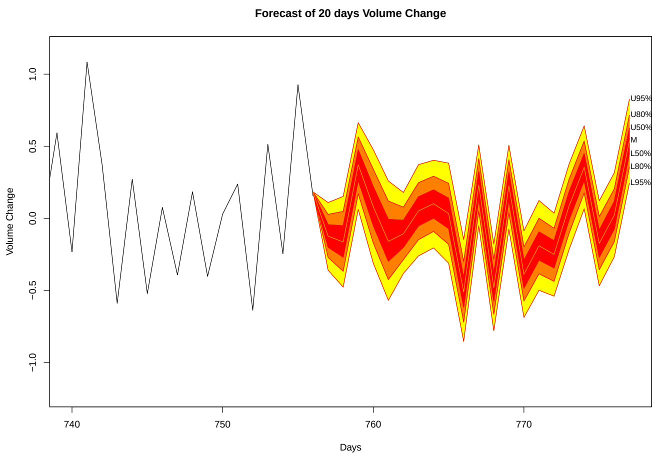

Figure 3: Out of Sample Forecast

5. CONCLUSION

In this paper we investigated one possible approach of determining the impact that news have on the financial markets. This is a growing research area since nowadays any model that wishes to capture market dynamics has to account for the effect of world events. In order to extend previous work in this area, we demonstrated how a more complex text data can be used to obtained relevant features for modelling changes in financial markets. Our data consisted of news events that were automatically extracted from the news articles. For each event we then collected past similar events and observed how the market reacted to those events. On the basis of these reactions we built various features vectors that helped us improve our model. Results show that when predicting change in volume, the predicted returns from similar events served as a useful feature. Additionally we tested the performance of our improved times series model on the out of sample dataset for the period of 1 month. Future work will be done in this direction, where we will look for further similarities between news events and possibly obtain new features that could be used for modelling financial markets.

6. ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675044.

7. REFERENCES

[1] H. Akaike. *Information Theory and an Extension of the Maximum Likelihood Principle*, pages 199–213. Springer New York, New York, NY, 1998.

[2] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307 – 327, 1986.

[3] X. Ding, Y. Zhang, T. Liu, and J. Duan. Deep learning for event-driven stock prediction. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 2327–2333. AAAI Press, 2015.

[4] R. Fehrer and S. Feuerriegel. Improving Decision Analytics with Deep Learning: The Case of Financial Disclosures. *ArXiv e-prints*, August 2015.

[5] G. Leban, B. Fortuna, J. Brank, and M. Grobelnik. Event registry: Learning about world events from news. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, pages 107–110, New York, NY, USA, 2014. ACM.

[6] M. L. Mitchell and J. H. Mulherin. The impact of public information on the stock market. *The Journal of Finance*, 49(3):923–950, 1994.

[7] D. Shen, W. Zhang, X. Xiong, X. Li, and Y. Zhang. Trading and non-trading period internet information flow and intraday return volatility. *Physica A: Statistical Mechanics and its Applications*, 451:519 – 524, 2016.

[8] R. Tsay. *An Introduction to Anaysis of Financial Data with R*. John Wiley & Sons, New Jersey, 2013.

[9] H. Yu, G. V. Nartea, C. Gan, and L. J. Yao. Predictive ability and profitability of simple technical trading rules: Recent evidence from southeast asian stock markets. *International Review of Economics & Finance*, 25:356 – 371, 2013.