The Effect of News on Intra-Day Stock Prices & their Volatility

Luc Bellintani, Junior Sophistor

The exact effect of news on stock prices is something that is of great interest to many people in the business world. This paper, by Luc Bellintani, looks at the effect of news on both stock prices and volatility, using almost 7.5 million total rows of data to draw its conclusions. First, much of the existing research and its limitations are discussed, which is followed by his own analysis. In particular, he builds several variables to try to "adjust" news sentiment and news relevance prior to his analysis. The paper quantifies the effect of news on stock prices, finding negative news had a negative effect, while positive and neutral news had a positive effect, with the previous close having a large effect.

Introduction

The most visible piece of financial information about a company is arguably its stock price. Its growth or decline represents not only how the company is faring, but the perception of its present-day performance, as well as its future viability. The recent cryptocurrency bubble has illustrated how important market sentiment is to the price of a stock or commodity. Bitcoin, an electronic currency with little in terms of use, rose and fell sporadically on a daily basis. The crests and troughs of its existence seemed to follow in the footsteps of its mentions in the mainstream news and Twitter (Galeshchuk, Vasylchyshyn & Krysovatyy, 2018).

News is a vital contributor to market sentiment. A string of bad news could affect the way an investor perceives the position of a company. A positive market sentiment regarding a company can see its stock price rise, whereas a negative sentiment could see the company's worth be wiped in a day. Whether or not the sentiments are well founded in information and logic or a simple guess, they can decide the fate of a company.

Articles give a summary of what has happened in relation to the company and speculate on the potential effects that said happening will have on the company as a result. Figuring out a potential method to the apparent madness of a stock market could be the key to more guaranteed and sustainable returns for investors, effectively helping to remove the gambling aspect the markets are so widely likened to.

The question this essay aims to answer is simply how news affects the price of a stock. The topic has been theorised by many in mainstream media, coupled with those advertising the "Buy low, sell high" method of investing. It has been looked at academically using newspaper articles, finding GDP and unemployment mentions in newspapers tended to affect the stock market (Birz & Lott Jr., 2011).

It was found that news articles that are marked as either positive, negative or neutral in terms of the way a company is mentioned in said article had little effect on the day's close price of the stock versus its opening price. The results are not only significant for investors, but also for economic theory; William Sharpe writes that a change in investor preferences (in this case, as a result of news) could mean that numerous parts of models need to alter, to accommodate the preference changes (Sharpe, Risk-Aversion In The Stock Market: Some Empirical Evidence, 1965).

Background

Previous literature has seen researchers look at newspapers for example; it was also found that conditional volatility of stocks was lower on days that preceded scheduled policy announcements from 1994 (Birz & Lott Jr., 2011). Other research has seen Twitter sentiments look at stock prices correlation to tweets. Researchers found they could predict the polarity of the change in stock prices with high accuracy based on that day's tweets (Bollen, Mao, & Zeng, 2011).

A group from Kuwait looked at the effect of news on 7 Kuwaiti oil and gas companies over 93 days in 2012 and found that positive and negative sentiment news did, in fact, affect the stock prices of these companies (Al-Augby, Majewski, & Nermend, 2013). In response to this, I raise the following argument; oil and gas companies are paramount to the success of the Kuwaiti economy. As such, the public will likely place more importance on these companies because of their juggernaut status in the local economy. Also, the sample size in question is minute and potentially may fail to capture other reasons the stocks may have been fluctuating in price during that time period. As such, the data may fail to capture other sources of change and variance within the model.

A 2001 study from the Financial Analysts journal found that there was no correlation between the postings of stocks on a message board for the 10 months starting April 1999, and any change in their stock prices (Whitelaw, 2001). However findings in 2013 also show that it is likely that movements in financial markets and movements in financial news are intrinsically interlinked (Merve Alanyali, 2013). Research was done on the effect of earnings news on the price of stocks. A group from Queen's University, Ontario looked at a set of pharmaceutical stocks for a 6 month period when mentioned in the news at a 30 minute interval. They found they could predict the direction the stock price would move to a high degree of accuracy (Dev Shah, 2018).

The goal of the analysis performed for this essay is to provide a much broader analysis and examination as to what effect news articles mentioning a company have on the price of the stock from the beginning of the day until it closes. The above-mentioned research tackles facets of the question at hand, while all of them fail to illustrate a true picture of the situation. This essay aims to provide said picture and can be used as a springboard for other research to be carried out as well.

Analytical Approach

As this research aims to find the link between the price of a stock and the news that may affect it, I have chosen a multi-regression model with eleven independent variables and one dependent variable. This was chosen over other models such as a clustering analysis as there is little in terms of predictive power for a cluster analysis. A principal components analysis would have only analysed the variance of the whole model, and not predict the potential changes in the daily stock price from the other data provided.

The independent variables aim to show what can affect the price of the stock. These will be the following; Previous market adjusted returns, sentiment class of the article (1 = Positive, -1 = Negative, 0 = Neutral) & the positive, negative and neutral sentiment scores of the article.

I have added variables that should not affect the analysis to act as control variables, which include the articles word count and the number of companies mentioned in the article.

The regression equation is listed below:

$$\begin{aligned} Y_{Delta_i} = \; & X_{Lagged\ Variable_i}\beta_{Lagged\ Variable_i} + X_{1\ Day\ Market\ Return_i}\beta_{1\ Day\ Market\ Return_i} \\ & + X_{1\ Day\ Market\ Close_i}\beta_{1\ Day\ Market\ Close_i} + X_{Article\ Length_i}\beta_{Article\ Length_i} \\ & + X_{Article\ Urgency_i}\beta_{Article\ Urgency_i} \\ & + X_{Number\ of\ Companies\ Mentioned_i}\beta_{Number\ of\ Companies\ Mentioned_i} \\ & + X_{Article\ Relevance_i}\beta_{Article\ Relevance_i} + X_{Sentiment\ Class_i}\beta_{Sentiment\ Class_i} \\ & + u_i \end{aligned}$$

Intuition would tell us that as the positivity of a news article increased, it would have a positive effect on the outlook for the company mentioned, and would cause the stock to increase in price. On the other hand, a negative article should decrease the confidence investors and the public have in a company, therefore decreasing the price of the stock.

DESCRIPTION OF THE DATASET

The data comes in two portions. The market data from Intrinio and the news data from Thomson Reuters accessed through a Python kernel on Kaggle.

Looking first at the market data; it runs from February 2007 until December 2016 for a group of roughly 1700 stock tickers per day for 2277 trading days. Basic information such as open & close prices along with volume counts was included. Following from that were more advanced tools such as both raw and market-adjusted returns for that stock for the previous 1 & 10-day horizons along with the next 1 & 10 days. The dataset numbers roughly 4,000,000 rows, which is more than enough market data to be relevant.

The news data is the key to this analysis. It consists of roughly 3.5 million rows of data. Each row consists of the headline, date, body size and stock ticker mentioned. Each news article was analysed by roughly seventy readers and given a sentiment score, days with multiple articles and sources were aggregated into one singular article. Python was used to clean the data, merge news articles and merge the news data with the stock price data.

Graphing the data, Figure 1 (located in the Appendix) suggests that the stock deltas are normally distributed with mean value μ=0.0055 and a standard deviation of σ= 0.941. This shows us that from 2007 to 2016, stocks tended to increase in price by $0.0055 per day with a variance of $0.94 per day. One can also surmise from the graph that it has an extremely high kurtosis, implying that the point on the graph is sharp and its tails are extremely tight to the mean. This is highlighting the fact that intra-day prices tended to centre themselves close to the mean (and 0, given how close μ is to 0).

Empirical results

Multiple regression analyses were performed using Ordinary Least Squares (OLS) at a 5% level of significance. The dependent values tested using two separate regressions; first the stocks daily delta, with and without news in a day. Second the returns in the 10 days after a piece of news (market adjusted). The sentiment parameters used are the relevance adjusted figures.

Looking firstly at Figure 2, where the stocks deltas were regressed against numerous independent values, one notices a few interesting points; The adjusted negative sentiment did indeed have a negative effect on the size of the delta of stocks with a $\beta_{\text{Adjusted Negative}}=-0.082$. The logic behind the positive sentiment stayed true to theory, where the model produced $\beta_{\text{Adjusted Positive}}=0.017$. Adjusted Neutral sentiment had a positive effect on the price of a stock with $\beta_{\text{Adjusted Neutral}}=0.011$. Neutral news had a slightly positive effect on the company as investors potentially thought "No real news is good news". All of the above figures had $P<0.01$ values which mean they are all statistically significant to the model.

The main constituent to the predictive power from the model came from the previous day's non-market adjusted return, which is the Lagged Variable in the study. With $\beta_{\text{Lagged Variable}}=17.8$. This means that as the close price of the previous day increased, the close price of the next day tended to increase too.

Looking at the model as a whole we see that we have R^2=R_adj^2=0.22 which says the model accounts for 22% of the total variation in the data. With something as intricate and complicated as the stock market, this number is totally expected and acceptable. The F-test had a P=0, which means the model is statistically significant, the likelihood of the β coefficients being equal to zero, is zero.

The Durbin-Watson statistic measures the degree of autocorrelation in the model. Autocorrelation is the measure of the correlation of a variable with past values of the variable (i.e. the correlation between yesterday's stock price with today's stock price). The model produced had a d=1.83, which indicates a near complete absence of autocorrelation within the model, adding to its validity.

A Dickey-Fuller test was performed with the following Hypotheses; H_0:There is a unit route for the series and H_1:There is no unit route for the series,therefore stationary. The value obtained had a P-value of P=0. This indicates the model displays no variance with time.

The Jarque-Bera test establishes whether or not the distribution of the data is normal (one of the assumptions of the regression model). The model produced a $JB=1.2*10^{13}$ and prob(JB)=0 which indicates from the test that there is no normality in the sample. However, this test can be inaccurate for any sort of outlier. As sample size as this dataset is ~3.75 million rows, there are numerous outliers (in comparison to the rest of the data) that can nullify the test (Brys, Hubert, & Struyf, 2004). Another argument against the accuracy of the Jarque-Bera test is the Central Limit Theorem (CLT). The CLT states that the sampling distribution of the sample means tends to a normal distribution as the sample size increases — regardless of the shape of the population distribution. This tells us that with a sample as large as the one being used in the model, it is more than likely normally distributed.

The model presented was chosen over other models, seen in Figure 3. For example, the non-adjusted sentiment ratings (Negative Sentiment, Neutral Sentiment and Positive Sentiment) mostly had P>0.05 so were statistically insignificant. This is seen in Figure 3. The purpose of adjusting the variables was to see how much the relevance of the news article affected the change in stock price. This adjustment standardises the effect a piece of news had on a stock. The R^2 values remained the same between the models.

Figure 4 looks at how news would affect the volatility of a stock as a result of a piece of news. The logic behind this is that the 10 Day Market Return is a measure of the market adjusted returns of the stock 10 days after the news is published. We can see from Figure 4 that the model $R^2=0.004$ which is admittedly low, but that the adjusted returns depend on multiple extraneous market variables that are completely unaffected by one company as a whole. Many of these variables simply cannot be included in a model such as the one presented, this is due to the amount of unknown variables that could affect a stock each day compounds to a point where a model simply can't reasonably account for them all. Knowing this, we see that news has a negligible effect on the volatility of the stock over the next 10 days of trading.

POSSIBLE EXTENSIONS

In an attempt to add more predictive power to the model, some market variables could be investigated as to whether or not they have an effect on the movement of a stock price. This could include; general sentiment towards the world and local markets and general sentiment towards a company and sector.

This could give an idea of how much an investor is likely to pay attention to a particular article in relation to a company.

Copious extra regressions could be performed to establish a company index on the effects of news on the stock price of a company. A cluster analysis could be performed on this to establish if there are any clusters of companies (for example, companies in the oil sector) that are all affected by news in a similar way. This could show allow for more intelligent trading into the future.

Conclusions

The goal of the analysis performed was to investigate the effect of news pieces on stock prices. The model provided has given the following results; positive, negative or neutral news has a negligible, but statistically significant, effect on the price of a stock at an intra-day level. Positive, negative or neutral news has almost nil effect on the volatility of the stock in the 10 days following publishing.

The stock market is very much elastic in its response to news shocks as the vast majority of price changes due to news seem to return to normal by the end of the day. However, this could have ramifications for the Capital Market equilibrium as described by William Sharpe. As the tastes of investors change, the market equilibrium price could change too (Sharpe, 1964),

Knowing the way investor preferences change in reaction to news could allow for the development of more intelligent trading and investment strategies. It has been found that investors and financial professionals do not always make rational decisions (Chandra, 2009). The potential to take the human randomness out of stock trading could potentially revolutionise the face of algorithmic stock trading into the future.

Reference List:

1. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal Of Computational Science, 1-8.
2. Brys, G., Hubert, M., & Struyf, A. (2004). A robustification of the jarque-bera test of normality. Compstat, 754-760.
3. Chandra, A. (2009). Decision-making in the stock market: incorporating psychology with finance . New Delhi.
4. Dev Shah, H. I. (2018). Predicting the Effects of News Sentiments on the Stock Market. 2018 IEEE International Conference on Big Data. New York City: IEEE.
5. Merve Alanyali, H. S. (2013). Quantifying the Relationship Between Financial News and the Stock Market. Scientific Reports 3.
6. R.LottJr., G. &. (2011). The effect of macroeconomic news on stock returns: New evidence from newspaper coverage. Journal of Banking & Finance, 2791-2800.
7. Salam Al-Augby, S. M. (2013). Reaction to News of Oil and Gas Sector Companies in Kuwait . International Journal of Business and Social Science , Vol. 4 No. 5.
8. Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. The Journal of Finance,, 425-442.
9. Sharpe, W. F. (1965). Risk-Aversion In The Stock Market: Some Empirical Evidence. The Journal of Finance, 416-422.
10. Svitlana Galeshchuk, O. V. (2018). Bitcoin Response to Twitter Sentiments. ICTERI 2018, 199-208.
11. Whitelaw, R. T. (2001). News or Noise? Internet Postings and Stock Prices. Financial Analysts Journal.

Appendix

Figure 1(Distribution of Stock's Intra-day prices)

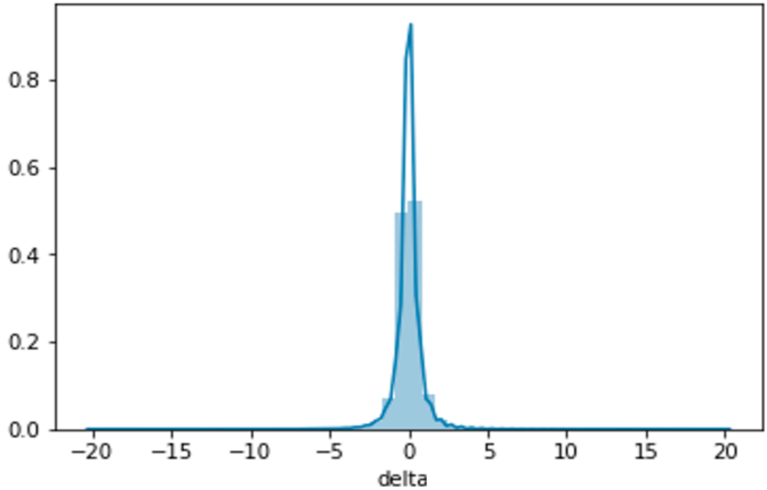

FIGURE 2

OLS Regression Results

Dep. Variable:	Delta	R-squared:	0.221
Model:	OLS	Adj. R-squared:	0.220
Method:	Least Squares	F-statistic:	8.124e+04
Date:	Sat, 02 Feb 2019	Prob (F-statistic):	0.00
Time:	13:02:47	Log-Likelihood:	-4.8837e+06
No. Observations:	3733316	AIC:	9.767e+06
Df Residuals:	3733303	BIC:	9.768e+06
Df Model:	13		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Lagged Variable	17.8335	0.027	672.649	0.000	17.782	17.885
1 Day Market Return	-7.1779	0.031	-233.168	0.000	-7.238	-7.118
1 Day Market Close	-0.0003	0.000	-1.762	0.078	-0.001	3.89e-05
Article Urgency	0.0014	0.001	1.769	0.077	-0.000	0.003
Article Length	-6.897e-07	2.4e-07	-2.868	0.004	-1.16e-06	-2.18e-07
Number of Companies Mentioned	6.192e-06	0.000	0.043	0.965	-0.000	0.000

Sentiment Class	-0.0003	0.002	-0.123	0.902	-0.005	0.004
Adjusted Negative	-0.0818	0.007	-12.496	0.000	-0.095	-0.069
Adjusted Neutral	0.0111	0.004	2.646	0.008	0.003	0.019
Adjusted Positive	0.0173	0.005	3.256	0.001	0.007	0.028

Omnibus:	9024802.074	Durbin-Watson:	1.833
Prob(Omnibus):	0.000	Jarque-Bera (JB):	12072648679852.670
Skew:	-23.356	Prob(JB):	0.00
Kurtosis:	8812.544	Cond. No.	2.60e+05

FIGURE 3

OLS Regression Results

Dep. Variable:	Delta	R-squared:	0.220
Model:	OLS	Adj. R-squared:	0.220
Method:	Least Squares	F-statistic:	8.799e+04
Date:	Fri, 08 Feb 2019	Prob (F-statistic):	0.00
Time:	15:15:22	Log-Likelihood:	-4.8838e+06
No. Observations:	3733316	AIC:	9.768e+06
Df Residuals:	3733304	BIC:	9.768e+06
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>ltl	[0.025	0.975]
Lagged Variable	17.8347	0.027	672.740	0.000	17.783	17.887
1 Day Market Return	-0.0006	0.000	-3.078	0.002	-0.001	-0.000
1 Day Market Close	-7.2063	0.030	-240.376	0.000	-7.265	-7.148
10 Day Market Return	0.0011	0.000	5.894	0.000	0.001	0.001
Article Urgency	0.0030	0.002	1.745	0.081	-0.000	0.006
Article Length	-6.208e-07	2.43e-07	-2.554	0.011	-1.1e-06	-1.44e-07
Number of Companies Mentioned	1.865e-05	0.000	0.118	0.906	-0.000	0.000
Article Relevance	-0.0087	0.003	-2.591	0.010	-0.015	-0.002
Sentiment Class	0.0032	0.003	0.959	0.338	-0.003	0.010
Negative Sentiment	-0.0401	0.009	-4.715	0.000	-0.057	-0.023
Neutral Sentiment	0.0054	0.006	0.832	0.406	-0.007	0.018
Positive Sentiment	0.0098	0.009	1.134	0.257	-0.007	0.027

Omnibus:	9026934.298	Durbin-Watson:	1.833
Prob(Omnibus):	0.000	Jarque-Bera (JB):	12092532544465.189
Skew:	-23.370	Prob(JB):	0.00
Kurtosis:	8819.796	Cond. No.	2.57e+05

FIGURE 4

OLS Regression Results

Dep. Variable:	10 Day Market Return	R-squared:	0.004
Model:	OLS	Adj. R-squared:	0.004
Method:	Least Squares	F-statistic:	1358.
Date:	Fri, 08 Feb 2019	Prob (F-statistic):	0.00
Time:	13:46:38	Log-Likelihood:	-8.7345e+06
No. Observations:	3733316	AIC:	1.747e+07
Df Residuals:	3733304	BIC:	1.747e+07
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>ltl	[0.025	0.975]
Lagged Variable	-0.1324	0.079	-1.682	0.093	-0.287	0.022
1 Day Market Return	0.0678	0.001	127.467	0.000	0.067	0.069
1 Day Market Close	-0.0052	0.085	-0.062	0.951	-0.171	0.161
Article Urgency	0.0062	0.002	2.699	0.007	0.002	0.011
Article Length	-5.966e-07	6.74e-07	-0.885	0.376	-1.92e-06	7.25e-07
Number of Companies Mentioned	-0.0003	0.000	-0.871	0.384	-0.001	0.000
Article Relevance	-3018.1934	8791.640	-0.343	0.731	-2.02e+04	1.42e+04
Sentiment Class	0.0022	0.006	0.355	0.723	-0.010	0.014
Adjusted Positive	3018.1768	8791.640	0.343	0.731	-1.42e+04	2.02e+04
Adjusted Negative	3018.1816	8791.640	0.343	0.731	-1.42e+04	2.02e+04
Adjusted Neutral	3018.1870	8791.640	0.343	0.731	-1.42e+04	2.02e+04
Delta	0.0086	0.001	5.899	0.000	0.006	0.011

Omnibus:	26297150.539	Durbin-Watson:	2.000
Prob(Omnibus):	0.000	Jarque-Bera (JB):	97852700392756032.000
Skew:	803.835	Prob(JB):	0.00
Kurtosis:	793131.873	Cond. No.	4.15e+10