Market Response to News with Sentiments

Project Report
High Frequency Finance & News Analytics
Mentor - Prof Ashok Banerjee

Submitted By - (Group 5)
Bodhisattwa Prasad Majumder, 15BM6JP11
Keshav Sehgal, 15BM6JP20
Rachit Tripathi, 15BM6JP33
Robin Singh, 15BM6JP36
Abstract

The project aims to investigate whether the sentiment of news released by the providers has any effect on the stock market. We concentrate on the Indian Stocks market taking the news released from various news providers with their time stamp data. It is hypothesized that a news with a strong sentiment has a greater effect on the stock movement. We document that positive sentiment news has a greater and significant effect on abnormal returns as compared to negative sentiment news.
Hypothesis and Experiment

The recent years have seen a dramatic increase in the algorithmic trading and high frequency finance. The trading window has reduced from intraday to minutes, seconds and even nanoseconds. Such high frequency trading is majorly based out of tracking some technical indicators or prices predictions based on models. And none of it is based out on the information coming in form of news.

Through the years trading has been conducted on the arrival of news in the market. The profits in such cases were made from the asymmetry of arrival of information that exist in the market. With the advent of internet this time has reduced from days to minutes. A new information in market is absorbed very fast by the traders and the opportunity window of trading is very less.

The news has been traded in market on the basis of perceived sentiment. It is hypothesised that high sentiment news has a tendency to abnormally move the market against the expected returns. Sentiment score for such an analysis can be calculated through Natural language Processing on the the news articles released by the providers.

Methodology

![Diagram showing the methodology]

- **Equity Prices (from Bloomberg)**
- **Returns on equities and SENSEX.**
- **Expected returns from CAPM Model,** 
  \[ R_a = R_f + \beta \times (R_m - R_f) \]
- **Cumulative Abnormal returns** 
  \[ CAR = R_e - R_a \]
- **Equity News data (from Factiva)**
- **Extraction of Time, Headlines, Source and content (in Beautifulsoup)**
- **Parsing of news articles in R**
- **Sentiment analysis based on Sentiment words and polarised context cluster**
- Regress CAR against the sentiment score for:
  1. Positive sentiment News
  2. Negative Sentiment News
  3. Combined with dummy
- **Results**
Experiment:

Two sets of data are required for this analysis: One is the news data released from various news providers throughout the world. The data is extracted for those news providers for which the minute time stamp is also available. The second data set is the Equity price data of the scrips under consideration. The data is collected at a minute level to study the news sentiment effect at a very high frequency.

The effect on the market price is calculated in terms of the Cumulative Abnormal Returns (CAR). CAR is different from the realised returns as it takes into effect the expected return for the stock and looks at the deviation from that level. The expected returns in this case are calculated from the CAPM model. For calculating the Expected Market returns we note the Index data on the which that stock is listed. For our case it has been taken as the SENSEX.

CAPM Model equation :

\[ Ra = Rf + Beta \times (Rm - Rf) \]

Ra = Expected Return  
Rf = Risk free rate  
Rm = Market rate of return  
Beta = Beta of the security

Cumulative Abnormal Returns = Re − Ra

Re = Actual return of equity

The return is calculated for the following windows (in mins):

\((-15, 60)\)

Note: Returns before the event are also considered as leakage can be there in the market in form of insider information or grapevine.

Next the CAR is regressed against the news sentiment scores for checking the correlation and significance of correlation. Objective is to test the Hypothesis that market reacts in response to News having sentiment.
Sentiment Calculation:

The sentiment calculation for all of the news and headlines have performed adapting the method described in the paper by Hu and Liu (2004). The method uses the sentiment dictionary to tag the polarized words in the query text. Each paragraph of news has been broken into element sentences. Each sentence is then broken into an ordered bag of words. Punctuation has been removed except the pause punctuations (commas, colons, semicolons) which are considered a word within the sentence to under the break/change of sentiment.

The words in each sentence are token searched and compared to the dictionary of polarized words. Positive and negative words are tagged with +1 and −1 respectively. After this, a polarised cluster, subset of the query sentence have been formed to weighted the polarity of the sentiment of the word in experiment. The polarized cluster of words is formed picking from the neighbourhood of polarized word where the neighbourhood contains 4 words before and 2 words after the word in question. These set is called valence shifters. The words from the polarized cluster are tagged as neutral, negator, amplifier, or de-amplifier. Neutral words don’t affect the polarity. Each polarized word is then weighted based on the number of the valence shifters directly surrounding the positive or negative word. Pause locations (commas, colons, and semicolons) are indexed and considered as the upper and lower bounds in the polarized cluster. Amplifiers increase the polarity by 1.8. Amplifiers become de-amplifiers if the polarized cluster contains an odd number of negators and de-amplifiers decreases the polarity. Negation acts on amplifiers/de-amplifiers in the same way but flip the sign of the polarized word by multiplying -1 to the power of the number of negators. Simply, this is a result of a belief that two negatives equal a positive, 3 negatives mean a negative and so on. The "but" conjunctions (i.e., 'but', 'however', and 'although') before a polarized word up-weights the cluster effect whereas presence of it after the polarized word down-weights the effect of cluster. Finally all weights have been added and normalized by the square-root of the number of words to achieve an unbounded sentiment score for each sentence. After getting the polarity for all the sentences in a news body, a simple average has been taken to normalize the overall sentiment score for the whole news.

Sample News 1: “0637 GMT [Dow Jones] HDFC Bank (500180.BY), India's second-biggest private sector bank, has substantial levers in place for sustained profitability, says Prabhudas Lilladher. The bank has now increased its reach through spending in technology to get boost its digital presence, the broker says. Overall asset quality remains best in class with one of the lowest non-performing assets level in both retail and corporate spaces, it adds. HDFC Bank has improved its retail share in deposits to 80% in FY16 from 67% in FY10. Retail deposits are low-cost deposits compared with other products. The broker expects the bank's net profit to rise at an annual compounded rate of 24% in FY16-FY19. It maintains a buy rating on the stock and
raises its price target by 5.6% to 1,425 rupees. The stock is trading flat at 1,286 rupees. (debiprasad.nayak@wsj.com)” Company - HDFC

Calculated Sentiment of the News - 0.1008

Sample News 2: “Reliance Jio Infocomm and AirHop Communications have been nominated for the Small Cell Forum 2016 HetNet Management Software and Service Award for their JioSON powered by AirHop's eSON that employs a multi-vendor, multi-technology, real-time Self-Organizing Network (SON) solution based on scalable and virtualized software platform with real-time optimization and big data analytics algorithms. "This nomination highlights the successful outcome and recognition of Reliance Jio's commitment to build a more advanced LTE HetNet through innovation," said Joe Thome, Sr. Director of Product Marketing at AirHop Communications. "This advanced SON solution based on a virtualized eSON platform has delivered significant benefits such as low cost template based network planning, capacity and coverage improvements, end-user experience improvements, and scalability using a common SON platform across their HetNet." The Small Cell Forum industry award winners will be announced on May 11th at the gala dinner as part of the 3-day SCWS World exhibition and conference.” Company - Reliance

Calculated Sentiment of the News - 0.2582

Sample News 3: Aug 1 (Reuters) - "Indian shares edged lower on Monday, retreating from earlier gains as investors booked profits and as ICICI Bank slumped after posting disappointing results. ICICI Bank lost 5.1 percent as its results fell short of estimates late on Friday. Other lenders also fell, with State Bank of India ending down 0.9 percent. The broader NSE index fell 0.02 percent to 8,636.55, after earlier rising as much as 0.84 percent to its highest level since April 16, 2015. The benchmark BSE index lost 0.17 percent to end at 28,003.12, after earlier hitting its highest since Aug. 10, 2015." Company - State Bank of India

Calculated Sentiment of the News : -0.397

Finally we have removed sentences with zero (or neutral) sentiments and also few outliers bearing very high sentiments but zero abnormal returns around the date of news which denotes glitches and inability of the algorithm to find proper sentiments.

**Linear Regression Analysis:**

We have considered applying regression on sentiments to estimate abnormal returns. We tried applying regression on both positive and negative sentiments points as well as separately on them.
Regression Combined:

For combined regression we created a dummy variable for positive and negative sentiment points denoting 1 for positive sentiments points. We employ regression keeping abnormal returns as response and corresponding sentiments and dummy as explanatory variable considering all the points.

Regression with Positive Sentiment News:

We segmented the data keeping only the positive sentiment points and apply regression using sentiments as the explanatory variable.

Regression with Negative Sentiment News:

Similarly, we kept only the negative sentiment data points and apply regression using sentiments as the explanatory variable.

**Data Description**

The following firms have been considered for the analysis:

<table>
<thead>
<tr>
<th>Capitalization</th>
<th>Company Name</th>
<th>Number of News in last 6 months</th>
<th>Market Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>ACC Cement</td>
<td>15</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>HDFC</td>
<td>29</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>Reliance</td>
<td>81</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>State Bank of India</td>
<td>69</td>
<td>1.76</td>
</tr>
<tr>
<td></td>
<td>Airtel</td>
<td>99</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Larsen and Toubro</td>
<td>29</td>
<td>1.86</td>
</tr>
<tr>
<td></td>
<td>TCS</td>
<td>46</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Dr. Reddy Laboratory</td>
<td>6</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>ITC</td>
<td>52</td>
<td>0.58</td>
</tr>
<tr>
<td>Mid</td>
<td>GMR Infrastructure</td>
<td>6</td>
<td>2.17</td>
</tr>
<tr>
<td>Company</td>
<td>Size</td>
<td>P/E</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>PC Jewellers</td>
<td>10</td>
<td>1.84</td>
<td></td>
</tr>
<tr>
<td>JSW Energy</td>
<td>14</td>
<td>1.45</td>
<td></td>
</tr>
<tr>
<td>Canara Bank</td>
<td>18</td>
<td>2.19</td>
<td></td>
</tr>
<tr>
<td>Wockhardt</td>
<td>10</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>MRF</td>
<td>3</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Havells</td>
<td>4</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>Arvind Limited</td>
<td>13</td>
<td>1.96</td>
<td></td>
</tr>
<tr>
<td>Shoppers Stop</td>
<td>9</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>VRL Logistics</td>
<td>14</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Raymond Limited</td>
<td>5</td>
<td>1.26</td>
<td></td>
</tr>
<tr>
<td>JK Tyres</td>
<td>9</td>
<td>1.58</td>
<td></td>
</tr>
<tr>
<td>Chennai Petroleum Corporation Limited</td>
<td>3</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>NBCC</td>
<td>6</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td>GHCL</td>
<td>4</td>
<td>0.8</td>
<td></td>
</tr>
</tbody>
</table>

**Price Data**
- 6 Months minute price data for all the mentioned equities
- BSE SENSEX has been used as market index for market return calculation
- 10 year govt bond data for risk-free rate calculation

**News Data**

![Search builder of Factiva](image)
Results
The descriptive analysis done on sentiments have been presented here:

1. Histogram before removing neutral and outlier sentiments
2. Histogram after normalizing all the sentiments and taking which lies between the radius of 1 sigma.
Histogram of News Sentiment
with Outliers and zero magnitude points removed

Histogram of Points with strong sentiment
Linear regression on all data points:

The combined regression shows the significance of news sentiment on abnormal returns as the p-value suggest with 90% significance.
Linear Regression on Positive Points

> fit <- lm(CAR ~ NewsSentiment, data = positive_points)
> summary(fit)

Call:
  lm(formula = CAR ~ NewsSentiment, data = positive_points)

Residuals:
   Min     1Q Median     3Q    Max
-0.05545 -0.016463 -0.004744  0.022521  0.078907

Coefficients: 
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.03812  0.01509  -2.526  0.0148 * 
NewsSentiment 0.06965  0.02613   2.666  0.0104 * 

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 . ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03183 on 49 degrees of freedom 
Multiple R-squared:  0.1267,  Adjusted R-squared:  0.1089 
F-statistic: 7.108 on 1 and 49 DF,  p-value: 0.01037
The regression suggests even better significance of News Sentiment when we only considered positive news and it shows the effect of sentiment with 95% significance (p-value 0.0137).

**LR on Negative Points**

![Linear Regression on Only Negative Sentiments with Abnormal Returns](image)

```r
> fit<-lm(CAR~NewsSentiment,data=negative_points)
> summary(fit)

Call:
  lm(formula = CAR ~ NewsSentiment, data = negative_points)

Residuals:
    Min     1Q Median     3Q    Max
-0.087919 -0.023729 -0.003227  0.014274  0.130749

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.002969    0.019094  0.155    0.877
NewsSentiment -0.010991    0.053449 -0.206    0.838

Residual standard error: 0.04356 on 43 degrees of freedom
Multiple R-squared:  0.0009824, Adjusted R-squared:  -0.02225
F-statistic: 0.04229 on 1 and 43 DF,  p-value: 0.838
As the plot and the result show very less significance with p-value 0.838 which is sufficient to accept the null which means the News sentiments do not affect negative returns.

Future Scope:

The sentiment analysis based on various other algorithms and considering Finance related vocabulary the analysis can be extended. Also, analysis based on only Large, Mid or Small cap companies is also another direction to extend this work. For the interest of the project, we conclude that the news sentiments affect abnormal returns when there is positive sentiment in the news.

References

- Factiva Database, Dow Jones and Company
- News Sources
  - Reuters
  - Dow Jones
- R Package for Sentiment analysis: sentimentr https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf

Appendix

CAR Code in Python:

```python
def CAR(before, after, time):
    acc_ix = np.argmin(np.abs(ACC2.index.to_pydatetime() - time))
    sen_ix = np.argmin(np.abs(SENSX2.index.to_pydatetime() - time))
    rf = RF.Price[np.argmin(np.abs(RF.index.to_pydatetime()-time))]/100
    Ret = math.exp(ACC2.log_ret[acc_ix - before : acc_ix + after + 1].sum()) - 1
    mar_ret = math.exp(SENSX2.log_ret[sen_ix - before : sen_ix + after + 1].sum()) - 1
    CAR = Ret - (rf + (beta * (mar_ret-rf))

    return CAR
```