Sentiment impact on stock prices of news with selected topic codes: Part One

Daniel Lam
Quantitative Research
Bloomberg L.P.

February 2018

Bloomberg

Sentiment impact on stock prices of news with selected topic codes: Part One

Abstract

We use Bloomberg's machine-readable newsfeed to explore the contemporaneous impact of news sentiment on daily stock returns, with a focus on understanding how sentiment impact varies across news stories with different topic codes. We find that news associated with certain topic groups (such as key equity news and analyst actions/comments) demonstrates strong sentiment impact, while those associated with more controversial topic codes (such as those relating to legal disputes) demonstrate much weaker sentiment impact. These results are stable over time, with some of the effect being more pronounced for stocks with smaller market capitalizations.

Contents

Introduction

It has long been recognized that investor sentiment can drive stock prices (Baker & Wurgler, 2007). Investor sentiment may be reflected in many forms – of which the tone of news stories has been shown to impact stock returns in an episodic way that varies across firms and over time. (Khurshid, Han, Hutson, Kearney, & Liu, 2016). Previously, we used the Bloomberg machine-readable newsfeed to study the sentiment-price relationship (on both a contemporaneous and lagged basis) and demonstrated three different types of equity trading strategies. (Cui, Lam, & Verma, 2016) In this study, we further investigate how sentiment impact varies across news stories tagged with various "topic codes." A topic code identifies the origin, subject matter or other characteristics of a news story and may be added by a human reporter or editor or an algorithm. One might expect that topic codes could carry information about the importance and/or relevance of a news story and thus help distinguish between news stories with more or less impact on stock returns.

We measured the contemporaneous impact of news sentiment on stock returns of large-cap U.S. stocks in 2013-2016, filtering news stories by selected topic codes. The main findings are:

- Topic codes that identify "Key Equity News" and analyst actions and comments are associated with news stories with stronger sentiment impact. This effect is consistent over time and is more pronounced for smaller cap stocks.

- Topic codes that identify "controversial" topics such as legal disputes and ESG controversies are associated with news stories with weaker impact. This effect is consistent over time and across size groups among large-cap stocks.

We venture some speculative hypotheses to explain these results:

- Stories tagged as "Key Equity News" and stories reporting on analyst actions are explicitly about specified stocks and so are more likely to be directly relevant to stock prices. This is especially the case for non-mega-cap stocks whose fortunes are less likely to be affected by information in more general stories not explicitly about their stocks.

- The "sentiment" of news stories reporting on "controversial" topics may reflect the viewpoint of one or another side in some dispute and may not be directly relevant to the bottom line of the companies mentioned in the stories.

Data & method

Our data source is the End-of-Day Story-Level News and Analytics File for four calendar years 2013-2016. We focus on news stories tagged with tickers of large-cap U.S. stocks.

The End-of-Day Story-Level News and Analytics File

The End-of-Day Story-Level News and Analytics File contains one entry for each news story. The fields that are relevant to our analysis are:

- Ticker: This is the ticker of a security identified as being a subject of the story.

- TimeOfArrival: The original timestamp of each story.

- Headline: The headline of the story.

- DerivedTopics: A list of topic codes that identify the origin, subject matter or other characteristics of the story. These may be coded by a human reporter or editor or an algorithm.

- Sentiment_Score: A classification of the sentiment of the story as being positive (+1), negative (-1), or neutral (0).

- Sentiment_confidence: A score in the range 0-100 that is a measure of the confidence of the classification encoded in Sentiment_Score.

See *EDF Research Note – Social Sentiment* (Bloomberg, 2016) for more details on the sentiment scoring. We analyze each calendar year separately to check results for stability across time.

Measuring "Sentiment Impact" of news stories on stock prices

We restrict our analysis to larger cap U.S. stocks, those that are current members of the Russell 1000 Index. For each calendar year, we measure the "sentiment impact" of news stories on each stock, ticker-by-ticker, starting with the largest stocks and proceeding to progressively smaller stocks until we have obtained results for 500 stocks. We now describe this procedure more precisely.

For each ticker and each calendar day, we identify all "relevant" news stories for that ticker on that day by filtering as follows:

- The Ticker field must match the ticker we are considering.

- The TimeOfArrival field must match the calendar day we are considering, with any timestamp after 20:00 GMT assigned to the next calendar day to roughly match the U.S. trading day.

- The Headline field must be non-empty. This excludes Twitter posts in particular.

- The Sentiment_Score field must be non-empty and nonzero. Thus we consider only polar stories with positive or negative sentiment and exclude all stories classified as having "neutral" sentiment.

- The DerivedTopics field may be required to include or exclude certain topic codes. By varying this requirement we can investigate which topic codes tend to be associated with more or less impactful stories. See below for further details on the exact filters tested.

Having thus identified the "relevant" stories, we define the "daily average sentiment" for the ticker on the calendar day as the average Sentiment_Score weighted by Sentiment_Confidence among all relevant stories. The daily average sentiment always ranges between -1 and +1.

The "sentiment impact" of news stories on a stock during a calendar year is then defined as the coefficient in a univariate OLS regression of the daily excess return of the stock over the S&P 500 index on the daily average sentiment for that stock on that day. We include only days on which the market is open and there is at least one relevant news story. We emphasize that sentiment impact is a measure of contemporaneous correlation and should not be interpreted as either a causal or a predictive factor.

Filtering on topic codes

In preliminary exploration of the data, we observed two possible regularities that we now proceed to test more systematically:

- Certain topic codes, namely EQUITYKEY (Key Equity News) and codes that identify stories reporting analyst actions and commentary seem to be associated with stories with higher sentiment impact. It is not hard to see why stories carrying such codes, which are explicitly related to equity news, may be expected to be more directly relevant for the ticker concerned.

- Certain topic codes identifying topics that may be referred to as "controversial" – law, litigation, environmental, social and governance issues – seem to be associated with stories with lower sentiment impact. A plausible rationale is that, in such stories, "sentiment" as derived from textual analysis may reflect the viewpoint of one or another side of a dispute and may not be directly relevant to a company's bottom line.

Thus, our working hypothesis is that one group of codes (analyst-related codes plus EQUITYKEY) is associated with stronger sentiment impact, while another group of codes (the "controversial" codes) is associated with weaker sentiment impact.

To test this more systematically across stocks and across time, we will use 7 filters on the DerivedTopics field:

- "All" means no filter at all. The topic codes are ignored and all stories are included.

- "inaek" includes all stories that are tagged with at least one of these 13 topic codes: EQUITYKEY, ANA, ANAMOVES, ANACHANGE, ANATGTCHG, ANACUT, ANAHOLD, ANABUY, ANARESU, ANATGTUP, ANARAISE, ANANEW, ANATGTDWN. All other stories are excluded.

- "noaek" excludes all stories that are tagged with at least one of these same 13 "AEK" topic codes. All other stories are included.

- "incon" includes all stories that are tagged with at least one of these 10 topic codes: ESGCONTROV, LAW, ESGRES, LITIGATE, LAWPRAC, LAWSUITS, IP, PATENT, CLASS, CALVPOSS. All other stories are excluded.

- "nocon" excludes all stories that are tagged with at least one of these same 10 "controversial" topic codes. All other stories are included.

- "inaek_or_nocon" is the union of "inaek" and "nocon."

- "inaek_and_nocon" is the intersection of "inaek" and "nocon."

Under our hypothesis, we would expect "inaek" and "nocon" to display generally stronger sentiment impact, and "noaek" and "incon" weaker.

Results

The results are generally in line with our hypothesis. We present them segmented by calendar years and by size groups.

Across calendar years

Figure 1 shows the average sentiment impact across 500 stocks for each calendar year from 2013 to 2016. The results are in accordance with our hypothesis and are stable across time. The two specifications that are expected to lead to weaker sentiment impact, "noaek" and "incon," do, in fact, have weaker impact than "All" in all years. Conversely, "inaek" and "nocon" both have stronger impact than "All" in all years, as expected. In all four years, the strongest impact was with "inaek_and_nocon," i.e., using only those stories that both included at least one code from the "analyst plus EQUITYKEY" group and did not include any from the "controversial" group.

Across size groups

Next we explore how sentiment impact varies among stocks of different sizes. Figure 2 shows the sentiment impact averaged across all four calendar years for groups of 100 stocks ranked by their current size. The results are generally consistent across size groups with the small exception of the largest group of 100 stocks, where "inaek" has only very slightly stronger impact than "noaek" and both are weaker than "All." Across all the size groups, the gap between "inaek" and "noaek" increases monotonically and steadily as you go from larger to smaller

stocks. This suggests that the impact of EQUITYKEY and analyst-related stories is more powerful for smaller stocks. This is consistent with the previous finding (Womack, 1996) that "the market reaction [to changes in analysts' recommendations] associated with smaller-capitalization firms is significantly larger than that associated with larger-capitalization firms." By contrast, the gap between "incon" and "nocon" actually shrinks a bit as you go from larger to smaller stocks.
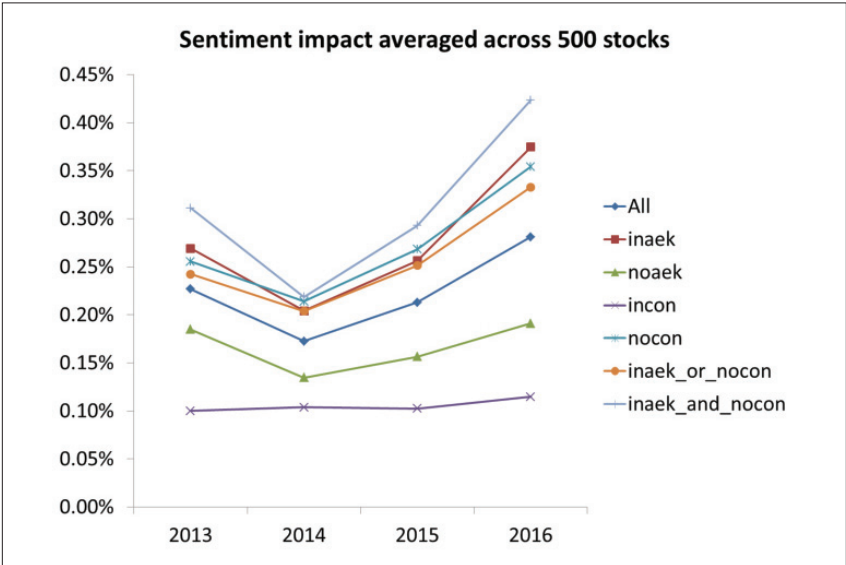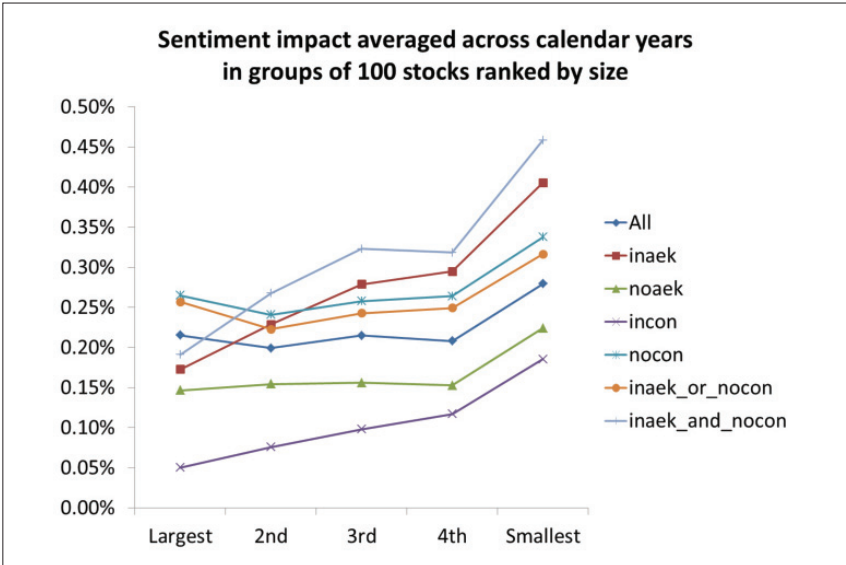
Conclusion

We have found that sentiment as reflected in news stories in the Bloomberg End-of-Day News and Analytics File has an impact on the prices of large-cap U.S. stocks, with the impact differing across stories depending on the set of topic codes with which they are tagged. Topic codes that identify "Key Equity News" and analyst actions and comments are associated with news stories with stronger impact. This effect is consistent over time and is more pronounced for smaller stocks. Topic codes that identify "controversial" topics such as legal disputes and ESG controversies are associated with news stories with weaker impact. This effect is consistent over time and across size groups.

Although the current study demonstrates that news topics can meaningfully influence the sentiment impact of stock prices, the test is based on a handful of selected topic codes from anecdotal observation. A following study will present an analytical framework to systematically identify principal vectors in the high-dimensional topic code space and offer significantly enhanced interpretability.

References

Baker, M., & Wurgler, J. (2007, Spring). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives 21(2)*, 129-151.

Bloomberg. (2016). *EDF Research Note – Social Sentiment*, DOCS 2080750.

Cui, X., Lam, D., & Verma, A. (2016). *Embedded Value in Bloomberg News & Social Sentiment Data*.

Khurshid, A., Han, J., Hutson, E., Kearney, C., & Liu, S. (2016). Media-expressed negative tone and firm-level stock returns. *Journal of Corporate Finance 37,* 152-172.

Womack, K. L. (1996, March). Do Brokerage Analysts' Recommendations Have Investment Value? *Journal of Finance 51(1)*, 137-167.

Figure 1 – Results by calendar year

Figure 2 – Results by size group

About the Bloomberg Terminal.

Since 1981, business and financial professionals have depended on the Bloomberg Terminal® for the real-time data, news and analytics they need to make the decisions that matter. The Terminal provides information on every asset class — from fixed income to equities, foreign exchange to commodities, derivatives to mortgages — all seamlessly integrated with on-demand multimedia content, extensive electronic-trading capabilities and a superior communications network.

Take the next step.

For additional information,
press the <HELP> key twice
on the Bloomberg Terminal®.

Beijing +86 10 6649 7500	**Hong Kong** +852 2977 6000	**New York** +1 212 318 2000	**Singapore** +65 6212 1000
Dubai +971 4 364 1000	**London** +44 20 7330 7500	**San Francisco** +1 415 912 2960	**Sydney** +61 2 9777 8600
Frankfurt +49 69 9204 1210	**Mumbai** +91 22 6120 3600	**São Paulo** +55 11 2395 9000	**Tokyo** +81 3 3201 8900