Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Sentiment and hype of business media topics and stock market returns during the COVID-19 pandemic

Ernest N. Biktimirov ∗, Tatyana Sokolyk, Anteneh Ayanso

Goodman School of Business, Brock University, 1812 Sir Isaac Brock Way, St. Catharines, ON L2S 3A1, Canada

ARTICLE INFO

Article history:
Received 16 June 2021
Accepted 2 July 2021
Available online 5 July 2021

JEL classification:
G12
G40
G41

Keywords:
COVID-19
Media hype
Media sentiment
Stock market
Topic modeling
Text analytics

ABSTRACT

We examine COVID-19 related topics discussed in the printed edition of the Wall Street Journal. Using text analytics and topic modeling algorithms, we discover 15 distinct topics and present differences in their sentiment (polarity) and hype (intensity of coverage) trends throughout 2020. Importantly, the hype of the topic, not the sentiment, relates to stock market returns. In particular, the hype scores for Debt market and Financial markets have the strongest positive relation to the stock market performance.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The emergence of the Coronavirus (COVID-19) pandemic has impacted every aspect of our lives. At its peak in March 2020, the Coronavirus media coverage index showed that 84 percent of all news sources covered stories about COVID-19; and the index remained above the 70 percent mark through December 2020. 1 By March 23, 2020, the S&P 500 index declined by 34% from its all-time high reached only a month earlier.

In this study, we analyze all COVID-19-related news articles published in the printed edition of the Wall Street Journal. We identify 15 distinct topics and document differences in their sentiment (polarity) and hype (intensity of coverage) trends throughout 2020. Importantly, we find that the hype scores of some topics (in particular, Debt market and Financial markets), but not the sentiment scores, significantly relate to the S&P 500 index returns.

This study extends two strands of literature. First, by examining sentiment and hype at the topic level, it contributes to the set of studies documenting the importance of media sentiment/coverage in predicting changes in asset prices and financial markets (see Tetlock, 2007; Bhattacharya et al., 2009; Fang and Peress, 2009; Shiller, 2015; Uhl et al., 2015), especially during times of economic or political uncertainty (Wisniewski and Lambe, 2013; Smales, 2014; Broadstock and Zhang, 2019; Shi and Ho, 2020). Second, the paper contributes to a rapidly growing literature on the market response to the COVID-19 pandemic and, in particular, to the studies utilizing sentiment and news media data (see O’Donnell et al., 2021; Ali et al., 2020; Al-Awadhi et al., 2020; Schell et al., 2020; Akhtaruzzaman et al., 2020; Zhang et al., 2020; Haroon and Rizvi, 2020; Costola et al., 2020; Baig et al., 2021).

2. Data and methodology

To identify major topics related to COVID-19, we collect articles published during 2020 in the printed edition of the Wall Street Journal. Our sample consists of 6552 articles, comprising over five million words in total. By performing topic modeling based on the Latent Semantic Analysis (LSA) (see Albright, 2004; Chakraborty et al., 2014; Alexander et al., 1998; Manning and Schütze, 1999), we identify 15 distinct and meaningful topics.

1 Corresponding author.
E-mail addresses: ebiktimirov@brocku.ca (E.N. Biktimirov), tsokolyk@brocku.ca (T. Sokolyk), aayanso@brocku.ca (A. Ayanso).

https://doi.org/10.1016/j.jbef.2021.100542
2214-6350/© 2021 Elsevier B.V. All rights reserved.
related to COVID-19. We label these topics by examining the top 20 words with the highest weights. We further confirm our selection of the topic labels by manually and independently examining the top three articles for each topic.

To estimate the media sentiment for each topic, we follow Hu and Liu (2004) and compute topic-specific sentiment score as the difference between the sum of positive and the sum of negative words, normalized by the length of the article. The sentiment score reflects the degree of polarity (i.e., negative, neutral, or positive) of a given article. In addition to the sentiment score, we compute the media hype score for each topic based on the

---

\[ \text{Sentiment Score} = \frac{\text{Sum of Positive Words} - \text{Sum of Negative Words}}{\text{Length of the Article}} \]

\[ \text{Hype Score} = \text{Weights and Number of Documents} \]
number of articles belonging to a particular topic and the weight of the topic in each article. Thus, the hype score captures both the breadth and intensity of coverage.

3. Results

Table 1 presents the top 20 words and the percentage of articles (Panel A), and summary statistics on the sentiment and hype scores (Panel B) for each of the 15 discovered topics. The diversity of the topics covered in the main business newsprint, ranging from financial markets to family, education, and sports, reflects the widespread impact of the pandemic that touched all facets of life. The identified topics listed from the most to least frequent are: Family/social connections, World economy and politics, Workplace safety, Retail business, Debt market, Local pandemic, Contact tracing, Treatment, Government funding, Financial markets, Presidential election, Vaccine, Travel, Education, and Sports.

The mean and median sentiment scores for all topics are negative, which is not surprising given the negative tone of COVID-19 pandemic in the media. The sentiment scores for Family/social connections and Presidential election have the highest medians and means, through still negative at around $-0.02$. Local pandemic stands out among the 15 topics both in terms of sentiment and hype scores. Specifically, it shows the lowest mean (median)
Table 1
Topics, sentiment, and hype scores.

| Panel A: Topics | Panel B: Sentiment (Hype) Scores |
|-----------------|----------------------------------|
| **Topic label** | **Top 20 words** | **Percentage of articles** | **Mean** | **Median** | **Standard deviation** | **Maximum** | **Minimum** |
| Family/social connections | home, family, feel, friend, live, room, husband, life, child, thing, love, house, want, daughter, space, wife, mother, work, restaurant, couple | 8.75% | −0.0254 | −0.0199 | 0.1080 | 0.4991 | −0.1691 |
| World economy and politics | chinese, country, world, government, global, foreign, political, economy, economic, european, policy, power, oil, international, trade, war, production, leader, export, epidemic | 8.73% | −0.0809 | −0.0774 | 0.0616 | 0.2338 | 0.2239 |
| Workplace safety | worker, hospital, plant, facility, production, employee, supply, nurse, factory, oil, equipment, company, patient, ventilator, meat, protective, demand, healthcare, shortage, care | 8.29% | −0.1285 | −0.1265 | 0.0815 | 0.0190 | −0.2682 |
| Retail business | sale, quarter, store, retailer, company, consumer, customer, revenue, share, demand, profit, product, company, analyst, brand, e-commerce, online, chain, shop | 8.26% | −0.1096 | −0.0990 | 0.1265 | 0.1018 | −0.7140 |
| Debt market | debt, investor, billion, fund, company, loan, property, york, bankruptcy, bank, firm, lender, asset, deal, court, bond, investment, city, sell, file | 8.03% | −0.0720 | −0.0762 | 0.0565 | 0.1433 | −0.1529 |
| Local pandemic | death, state, case, cuomo, york, city, gov, governor, infection, resident, johns, reopen, hopkins, new case, official, number, hospitalization, test, restriction, blasio | 7.79% | −0.2317 | −0.2248 | 0.2606 | 0.1411 | −1.7118 |
| Contact tracing | test, employee, information, user, infect, symptom, cdc, expert, contact, app., data, person, tech, company, virus, technology, apps, inc, software, google | 7.63% | −0.0615 | −0.0536 | 0.0995 | 0.1881 | −0.5362 |
| Treatment | patient, hospital, drug, treatment, study, doctor, antibody, treat, blood, disease, remdesivir, plasma, therapy, medical, hospitalize, researcher, care, symptom, clinical, physician | 7.03% | −0.1067 | −0.0963 | 0.0879 | 0.0747 | −0.4580 |
| Government funding | unemployment, bill, worker, job, loan, trillion, senate, tax, payment, federal, economist, economy, package, spending, benefit, aid, stimulus, employer, money, democrat | 6.85% | −0.0921 | −0.0861 | 0.0776 | 0.0577 | −0.3394 |
| Financial markets | investor, stock, market, s&p 500, index, bond, yield, oil, share, price, trade, investment, bank, gain, point, rally, composite, economy, recovery, average | 6.26% | −0.1152 | −0.1037 | 0.0983 | 0.0398 | −0.5030 |
| Presidential election | trump, biden, election, voter, campaign, democrat, vote, democratic, president, senate, presidential, republicans, ballot, house, joe, white, poll, win, party, president | 6.09% | −0.0230 | −0.0193 | 0.0742 | 0.1465 | −0.2129 |
| Vaccine | vaccine, dose, trial, shot, pfizer, moderna, drug, clinical, vaccine, biontech, clinical trial, develop, study, fda, authorization, vaccination, pfizer inc, effective, candidate, astrazeneca | 4.87% | −0.0452 | −0.0487 | 0.0793 | 0.1216 | −0.3320 |
| Travel | airline, passenger, flight, travel, carrier, airlines, traveler, airport, fly, plane, air, trip, ship, aircraft, mask, hotel, cruise, crew, international, booking | 4.42% | −0.0738 | −0.0761 | 0.0467 | 0.0195 | −0.1563 |
| Education | student, school, teacher, learning, child, parent, in-person, district, class, classroom, instruction, education, remote, college, campus, remote learning, fall, online, grade, kid | 4.13% | −0.0821 | −0.0704 | 0.1456 | 0.1210 | −0.9342 |
| Sports | player, game, league, season, football, team, sport, play, coach, athlete, baseball, college, test, sports, fan, positive, mlb, protocol, tournament, ten | 2.87% | −0.0716 | −0.0598 | 0.0857 | 0.0119 | −0.5436 |

Notes: The table presents 15 major topics, along with top 20 words and the percentage of articles represented in each topic (Panel A) and descriptive statistics for the sentiment and hype (in parentheses) scores for each topic (Panel B). Topics are generated from 6552 articles published in the printed edition of the Wall Street Journal during January 1 – December 31, 2020 that mention the word “COVID”.
sentiment score of $-0.2317$ ($-0.2248$) and the highest mean (median) hype score of 11.5167 (12.0490).

Fig. 1 shows the time-series patterns in sentiment and hype scores for each of the 15 topics and highlights the diversity in topic time-series patterns. For some topics, such as Debt market, Treatment, and Presidential election, sentiment scores show smooth patterns without significant spikes and declines. For other topics, such as Retail business and Local pandemic, sentiment scores exhibit sharp increases and decreases. The sentiment for Family/social connections displays a downward trend, whereas the sentiment for Financial markets shows a somewhat consistent upward trend.

The patterns of media hype scores are more striking. Specifically, hype scores for all topics, except Family/social connections, show a rapid surge from February to April. After that, hype scores tend to bounce up and down for most topics but show a steady decline for Financial markets and Vaccine. The patterns of hype scores for Family/social connections and Contact tracing are noticeably different from those of other topics. Specifically, a hype score for Family/social connections shows negative values and a slow increase at the start of the pandemic, but a rapid exponential increase starting in August. A hype score for Contact tracing shows a steady growth from February to December.

To offer a distinct illustration of a change in the degree of topic coverage, Fig. 2 presents the proportional distributions of the topics’ hype values in March 2020 – the drastic drop in the stock market due to COVID-19 pandemic – versus December 2020 – the last month of our sample period representing the year-end recovery of the stock market. The height of each block within a stacked horizontal bar chart shows the proportion of a specific topic. The intensity of coverage of the Family/social connections topics increased drastically, going from a meager 3% in March 2020 to a sizeable 22% in December 2020. Among the topics that experienced a decrease in hype, Vaccine and Local pandemic had the largest declines of 5.8% and 4.7%, respectively.

To examine the relation between the sentiment and hype of the 15 topics and the S&P 500 index, we regress the returns of the S&P 500 index on its first lag and the topic sentiment and hype scores. Thus, for each topic $i$, we estimate the following model:

$$S&P500_t = \alpha + \beta_i \text{Sentiment Score}_{i,t} + \gamma_i \text{Hype Score}_{i,t} + \sigma_i S&P500_{t-1} + \epsilon_{i,t},$$

where $t$ represents a week during a 47-week period from February 12 to December 31, 2020.  

Table 2 presents regression results for six topics with significant standardized coefficients. Coefficients for the hype score for World economy and politics, Workplace safety, Debt market, Local pandemic, Government funding, and Financial markets are positively related to the S&P 500 index returns with the level of significance ranging from 5% to 10%. Taken together, the regression results suggest two conclusions: (1) media hype, not sentiment, relates to stock market returns during the pandemic; (2) among different topics, media hype for topics related to business and economy, such as World economy and politics, Debt market, and Financial markets, demonstrates the strongest relation to stock market returns.

4. Summary

This study offers new empirical evidence on the relation between COVID-19 related news and stock market returns. By examining media sentiment and hype at topic level, we find that the hype scores, which reflect both the breadth and intensity of coverage, not the sentiment scores, which reflect polarity, are significantly related to the S&P 500 index returns during 2020. In particular, the hype scores for Debt market and Financial markets have the strongest positive relation to the stock market performance. These results suggest that when the sentiment scores are generally negative, the intensity of the topic coverage is more informative for predicting stock market performance. The results also imply that business media can play a role in building investors’ confidence in the path for economic recovery by increasing the coverage of COVID-19-related topics. As COVID-19 pandemic continues beyond 2020, more research is needed to

---

5 In unreported analysis, we ran the regressions of the topic sentiment and hype scores separately with qualitatively similar results.

6 We note that the number of observations (equal to 47) is relatively small and, therefore, interpret the results with caution.

7 To conserve space, the regression results are omitted for other topics for which the coefficients on sentiment and hype scores are not significant.
Table 2

| Topic label                | Constant | Sentiment score | Hype score | Lagged S&P 500 index | R²    |
|----------------------------|----------|-----------------|------------|----------------------|-------|
| World economy and politics | −0.03    | 0.01            | 0.29*      | −0.11                | 0.087 |
| (0.11)                     | (0.96)   | (0.06)          | (0.46)     | (0.11)               | (0.96) |
| Workplace safety           | −0.04    | −0.09           | 0.27*      | −0.11                | 0.073 |
| (0.14)                     | (0.56)   | (0.08)          | (0.45)     | (0.06)               | (0.28) |
| Debt market                | −0.03    | 0.06            | 0.30**     | −0.12                | 0.097 |
| (0.15)                     | (0.67)   | (0.05)          | (0.42)     | (0.06)               | (0.28) |
| Local pandemic             | −0.05    | −0.17           | 0.26       | −0.06                | 0.074 |
| (0.11)                     | (0.28)   | (0.10)          | (0.71)     | (0.06)               | (0.28) |
| Government funding         | −0.04    | −0.06           | 0.28*      | −0.09                | 0.083 |
| (0.12)                     | (0.69)   | (0.06)          | (0.55)     | (0.06)               | (0.28) |
| Financial markets          | −0.02    | 0.17            | 0.29**     | −0.14                | 0.110 |
| (0.32)                     | (0.25)   | (0.05)          | (0.36)     | (0.05)               | (0.36) |

Notes: The table presents the results of the ordinary least squares regressions examining the relation between topic sentiment and hype scores and the S&P 500 index returns: 

\[ S_{\text{P}500} = \alpha + \beta_{1}\text{Sentiment Score}_{i,t} + \beta_{2}\text{Hype Score}_{i,t} + \sigma S_{\text{P}500,t-1} + \varepsilon_{i,t}, \] 

where \( t \) represents a week during a 47-week period from February 12 to December 31, 2020 and \( i \) represents one of the 15 topics described in Table 1. The number of observations is 47 in each regression. Two-tailed p values are shown in brackets below the estimated coefficients. To conserve space, the regression results are presented only for six topics with significant coefficients on hype score. 

* \( p < 0.1 \).

** \( p < 0.05 \).

confirm the persistence of the reported relation between the hype of topics and stock market returns.

Acknowledgment

Financial support from the CPA Ontario Center for Public Policy and Innovation in Accounting is gratefully acknowledged.

References

Akhtaruzzaman, M., Boubaker, S., Sensoy, A., 2020. Financial contagion during COVID–19 crisis. Finance Res. Lett. 37, 101604.

Al-Awadhi, A.M., Alsaifi, K., Al-Awadhi, A., Alhamadi, S., 2020. Death and contagious infectious diseases: Impact of the COVID-19 virus on stock market returns. J. Behav. Exp. Finance 27, 100326.

Albright, R., 2004. Taming Text with the SVD. SAS Institute Inc., NC, U.S.A.

Alexander, L., Johnson, R., Weiss, J., 1998. Exploring Zipf’s law. Teach. Math. Appl.: Int. J. IMA 17 (4), 155–158.

Ali, M., Alam, N., Rizvi, S.A.R., 2020. Coronavirus (COVID-19)–An epidemic or pandemic for financial markets. J. Behav. Exp. Finance 27, 100341.

Baig, A.S., Butt, H.A., Haroon, O., Rizvi, S.A.R., 2021. Deaths, panic, lockdowns and US equity markets: The case of COVID-19 pandemic. Finance Res. Lett. 38, 101701.

Bhattacharya, U., Galpin, N., Ray, R., Yu, X., 2009. The role of the media in the internet IPO bubble. J. Financ. Quant. Anal. 44 (3), 657–682.

Broadstock, D., Zhang, D., 2019. Social-media and intraday stock returns: The pricing power of sentiment. Finance Res. Lett. 30, 116–123.

Chakraborty, D.G., Pagolu, M., Garla, S., 2014. Text Mining and Analysis: Practical Methods, Examples, and Case Studies using SAS. SAS Institute, NC, U.S.A.

Costola, M., Lacopini, M., Santagjustina, C.R., 2020. Google search volumes and the financial markets during the COVID-19 outbreak. Finance Res. Lett. 101884.