hinese create doctor

ban

expert fund

program

data turn fight percent

facility reach reduce community

network

patient

expert information testing

employee

taff resident

What are People Saying About COVID-19-Related Regulations?

Research conducted by George Washington University's Zhoudan Xie shows how text and data mining can quickly analyze a massive set of news articles to uncover critical and timely trends.

proquest.com

ProQuest®

To talk to the sales department, contact us at
1-800-779-0137 or **sales@proquest.com**.

Zhoudan Xie is a senior policy analyst in the Regulatory Studies Center and a Ph.D. student in economics at George Washington University. Her research interests include economic effects of regulation, bureaucratic behavior in regulatory processes, and public participation in regulatory policy making.

Senior Policy Analyst, Regulatory Studies Center at George Washington University

Executive Summary

Government agencies have taken various regulatory actions in response to the COVID-19 pandemic. What are people saying about these regulatory responses, and which regulatory issues have prompted the most discussion?

To answer these questions, George Washington University researcher Zhoudan Xie analyzed a massive set of ProQuest news content using TDM Studio, ProQuest's new text and data mining solution. Xie applied topic modeling to recent U.S. news articles related to both COVID-19 and regulation.

This analysis identified 16 topics that are present in more than 3,000 articles published between January and April 2020. These topics demonstrate different levels of prevalence and trends over time, reflecting changes in media attention on regulatory policy response to COVID-19.

Challenge

Responses to the COVID-19 pandemic led to many government regulations. In the U.S., government agencies have taken actions that range from restricting travel to stopping the enforcement of certain regulatory requirements.[1] As with any regulations, the reactions to these have been mixed. But which ones have prompted the most discussion and debate?

Using ProQuest TDM Studio to analyze a massive dataset of COVID-19 news, I set out to answer this question with text and data mining (TDM).

Analyzing newspapers can reveal timely information on the public's perception of regulatory policy responses to the pandemic. This information could suggest the regulatory issues that people and businesses are most concerned with during this crisis, and thus help policymakers prioritize their reforms to minimize unnecessary regulatory burdens.

As part of the research effort at the George Washington University Regulatory Studies Center to understand how government regulations are affected by and respond to the pandemic,[2] I applied a topic modeling analysis to the recent U.S. news articles related to both "COVID-19" and "regulation" using TDM Studio. Topic modeling is a widely used text-mining tool for discovering hidden topical patterns in a collection of documents.

The analysis extracts 16 topics that are the focus of more than 3,000 news articles published between January and April 2020. The topics demonstrate different trends since the beginning of the coronavirus outbreak, reflecting changes in media attention on regulatory issues related to COVID-19.

Analysis

The COVID-19 dataset on TDM Studio covers more than a million articles containing terms related to COVID-19. To focus the analysis on news content in the U.S., I refined the dataset to include news-type articles (i.e., excluding commentaries, letters to the editor, etc.) published by eight major U.S. news sources in 2020.[3] That resulted in 45,654 articles published between January 3 and April 17. Among them, I searched for articles that also contained terms related to regulation (including variations such as "regulatory," "regulator" and "deregulatory").[4]

The final dataset I used for the analysis included 3,149 news articles related to COVID-19 and regulation.

"Analyzing newspapers can reveal timely information on the public's perception of regulatory policy responses to the pandemic. This information could...help policymakers prioritize their reforms to minimize unnecessary regulatory burdens."

Figure 1: Weekly News Article Frequency Related to COVID-19 and Regulation (January–April 2020)
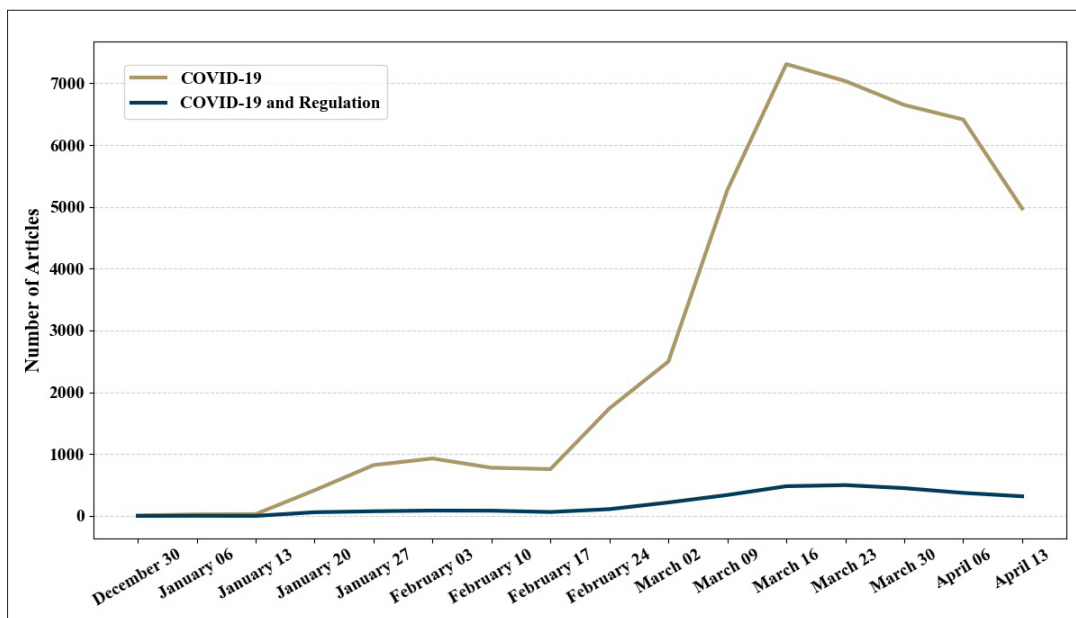
Results

As shown in Figure 1, the weekly number of COVID-19 related articles from the eight news sources has increased substantially since mid-February, coinciding with the increase in reported cases and the occurrence of first death in the U.S. The count reached a peak in the week of March 16, the week after President Donald Trump declared a national emergency regarding the outbreak. The count of articles related to COVID-19 and regulation followed a similar trend, comprising between 5 and 9 percent of all COVID-19 articles since the week of February 17.

Figure 2 provides a word cloud from the news articles about COVID-19 and regulation. It indicates that these articles address various issues with relative emphasis on patients, banks, employees, facilities, communities, residents and quarantines. While the word cloud is helpful for extracting keywords from the articles, topic modeling is needed to look for patterns in the use of words.

Figure 2: Word Cloud of Articles Related to COVID-19 and Regulation

Topic Modeling

I used the Latent Dirichlet Allocation (LDA) method, one of the most popular topic modeling approaches, to discover the topics discussed in the articles related to COVID-19 and regulation.[5] The LDA relies on statistical modeling to identify topics from the collection of words in a document and then the most relevant topics across a corpus of documents. Since topic modeling is an unsupervised machine-learning technique, no prior knowledge about the data is required.

I analyzed the corpus of news articles using the LDA after performing a series of standard preprocessing steps (e.g., lemmatization, bigram formation, customized stopwords removal, etc.). The analysis resulted in 16 coherent topics. As visualized in Figure 3, each topic is represented by a set of probable terms related to the topic, with a larger-sized term indicating a higher probability of attachment to the topic. Terms within a topic are coherent such that they tend to have similar semantic meanings. For example, Topic 10 is composed of terms such as "vote," "election," "sander" (a lemmatized form of "Sanders," as in Bernie Sanders), "campaign" and "candidate," which are all closely linked to election and president. In this way, the title of each topic was inferred from the relevant terms.

Figure 3: Word Cloud by Topic

Notes: This figure was created by the author using data from TDM Studio and the Python WordCloud package. Each word cloud includes the top ten terms related to the topic. Topic titles were inferred from the relevant terms.

Distribution of Topics

Figure 4 shows the distribution of dominant topics across all the articles analyzed in this study.[6] The most prevalent topic discussed in the news articles analyzed relates to quarantine and reopening (Topic 1). Quarantine is perhaps the most salient regulatory response to the pandemic. In addition to the "shelter-in-place" orders placed by state governors, federal regulations govern interstate and foreign quarantine. Unsurprisingly, these issues became a focus of media and public attention. Relatedly, the more recent debate about reopening also likely contributed to the prevalence of this topic in news. For example, a Washington Post article designated with this topic discusses President Trump's authority to reopen the country, saying "various police powers, as well as authority over functions such as zoning and regulation of business, belong to the states because the Constitution does not grant them to the federal government."[7]

Topics related to oil prices (Topic 7) and testing and treatment (Topic 11) are also prevalent in the articles. Regulatory debates around oil prices include potential regulatory restrictions on oil production after a historic price crash. For example, Bloomberg reported on April 14 that "Texas oil regulators are scheduled Tuesday to discuss supply restrictions in response to the price crash."[8] In contrast to the plunged demand for gas, the spread of coronavirus caused an unprecedented demand for testing and treatment equipment which is usually required to go through the Food and Drug Administration (FDA)'s approval process. In response, the FDA has allowed emergency use of many products that were not approved previously to test, treat or prevent COVID-19.[9]

Figure 4: News Articles Distribution by Dominant Topic
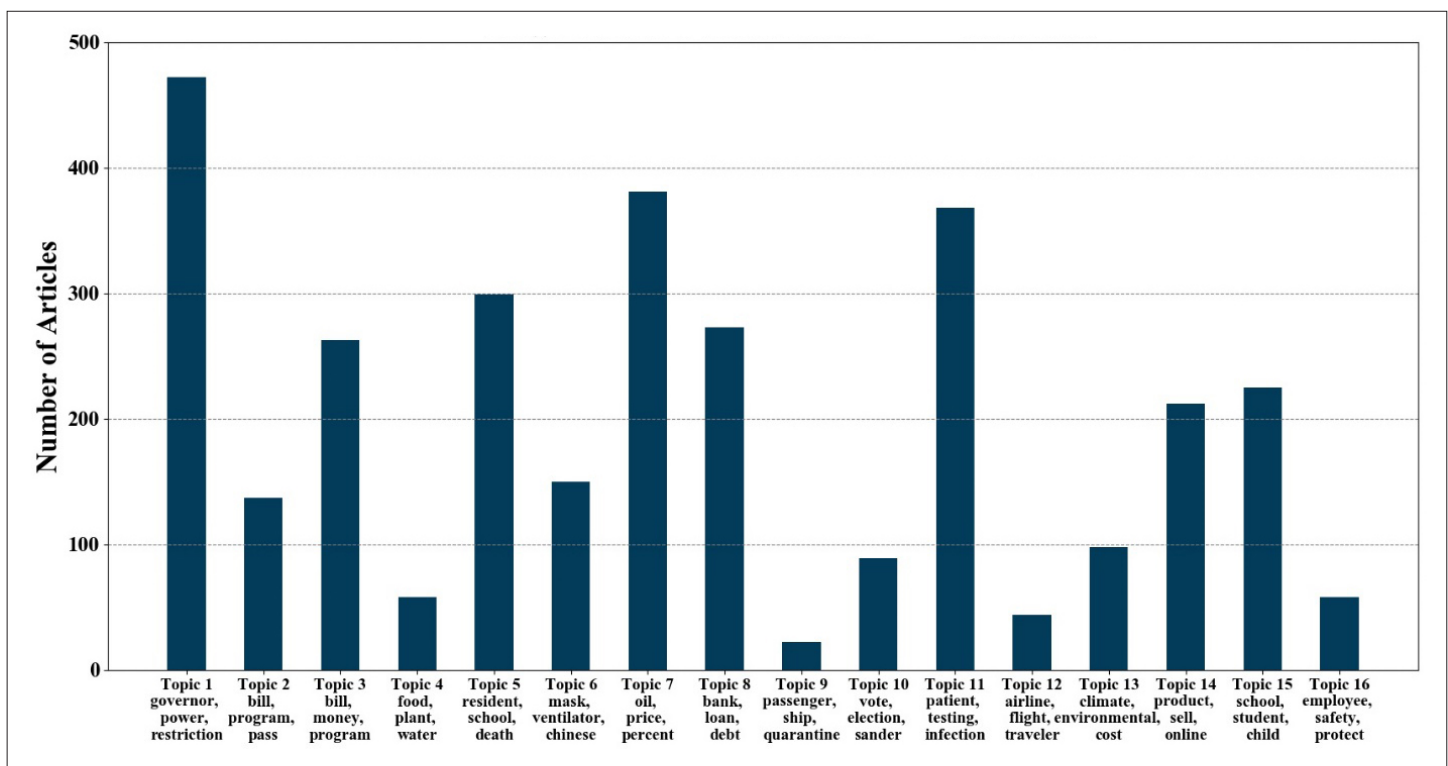
Notes: This figure was created by the author using data from TDM Studio and the Python Matplotlib package. X-axis shows the top three terms associated with each topic.

Other prevalent topics include residency and community (Topic 5), banks and loans (Topic 8), stimulus bills (Topic 3), schools and students (Topic 15), and online shopping and entertainment (Topic 14). Regulations pertaining to these topics are likely affecting people, businesses, and the economy in the responses to the pandemic. The topics and more relevant terms for each topic are also available in the interactive web-based visualization created using pyLDAvis.[10]

## Determining Trends in Topics Over Time

The topics demonstrate different trends over the past months. As shown in Figure 5, the relatively more prevalent topics discussed above largely followed the weekly trend of overall article frequency related to COVID-19 and regulation in Figure 1, presenting a substantial increase starting mid-February and a peak around mid-March. It is noteworthy that the article frequency for Topic 1 on quarantine and reopening rebounded in April after a drop in the second half of March, indicating a possible news focus on reopening after weeks of lockdown in many states.

Other topics, while less prevalent in terms of total article frequencies over the four months, reflect an increasing trend until mid-April. For example, the number of articles related to environmental standards (Topic 13) had its largest increase in April, perhaps influenced by the 50th anniversary of Earth Day, and the federal government's temporary relaxations of many environmental regulations to allow greater flexibility during the pandemic.[11] Another topic that shows an increasing trend since February is food supply (Topic 4), indicating persistent concerns about food shortages. Article frequency for Topic 16 on workplace safety also started to rise in April and is expected to continue rising, as more workers and employers seek government guidance to ensure safety after returning to their workplaces.

Topic modeling allows us to understand the general patterns and trends in regulatory issues being talked about in the news related to COVID-19. With the rich resources provided by TDM Studio, further text mining and analysis can be applied to answer more specific questions in regulatory studies, generating insights that may inform policymakers and help the economy recovers from the pandemic.

Topic modeling allows us to understand the general patterns and trends in regulatory issues being talked about in the news related to COVID-19.

Figure 5: Weekly News Article Frequency by Dominant Topic (January–April 2020)
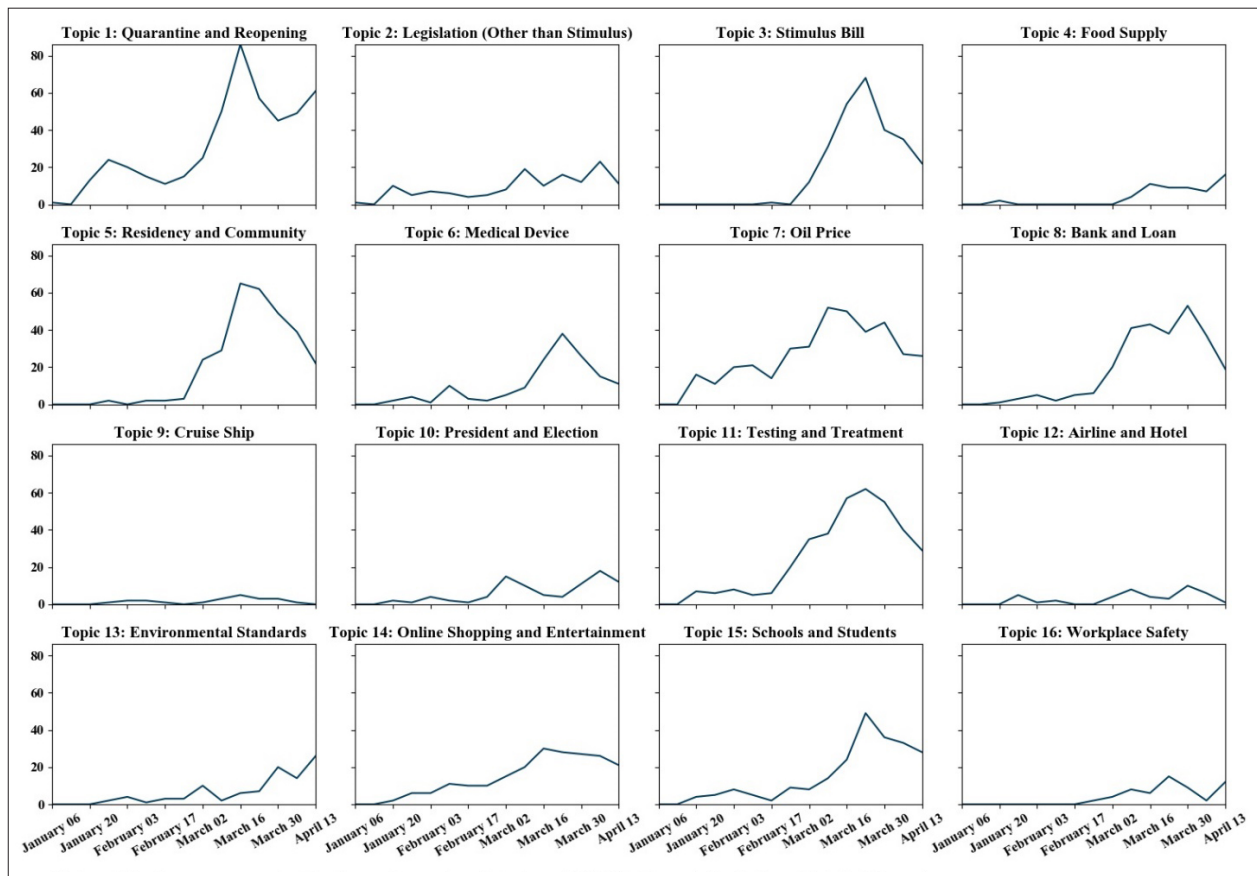
Notes: This figure was created by the author using data from TDM Studio and the Python Matplotlib package. Dates on the x-axis indicate the start date of each week. Data were available until April 17, so the frequency for the week starting April 13 is likely underestimated in the figure. Topic titles were inferred from the relevant terms generated by LDA.

About TDM Studio

ProQuest's workflow solution for text and data mining is designed for research, teaching and learning. TDM Studio provides access to sought-after content including current and historical newspapers, primary sources, scholarly journals, and dissertations and theses. It empowers researchers, students and faculty to analyze documents by uncovering connections and patterns that lead to career-defining discoveries.

Learn more at www.proquest.com/go/tdm-studio
or contact your ProQuest representative today.

1 For example, the Centers for Disease Control and Prevention issued regulations "to suspend the introduction of persons from designated countries or places" (https://www.federalregister.gov/documents/2020/03/24/2020-06238/control-of-communicable-diseases-foreign-quarantine-suspension-of-introduction-of-persons-into). The Institute for Policy Integrity tracks waivers of enforcement of environmental laws and regulations at: https://policyintegrity.org/covid-19-concession-tracker.
2 See the GW Regulatory Studies Center's COVID-19 and Regulation page at: https://regulatorystudies.columbian.gwu.edu/covid-19-and-regulation.
3 The news publishers covered in this analysis include New York Times, The Washington Post, Bloomberg Wire Service, CNN Wire Service, The Examiner, USA Today, Los Angeles Times, and Daily Herald.
4 A caveat is that many sports articles also contain the term "regulation," which is used as a different meaning from government regulation. Therefore, news articles that were published in sport-related sections were removed from the analysis.
5 For details about LDA, see: http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf
6 The LDA allows multiple topics for a document and assigns a probability to each topic for the document. A dominant topic for a document is the topic with the highest probability.
7 The article is available at: https://search.proquest.com/docview/2389506215?accountid=131239
8 The article is available at: https://search.proquest.com/docview/2389060706?accountid=131239
9 See FDA's Emergency Use Authorizations at: https://www.fda.gov/medical-devices/emergency-situations-medical-devices/emergency-use-authorizations#coronavirus2019.
10 The interactive visualization is available at: https://zhoudanxie.github.io/documents/topicvisual_covid19reg.html. For details about PyLDAvis, see: https://pyldavis.readthedocs.io/en/latest/readme.html.
11 See the Institute for Policy Integrity's tracker of waivers at: https://policyintegrity.org/covid-19-concession-tracker.

proquest.com

To talk to the sales department, contact us at
1-800-779-0137 or **sales@proquest.com**.

ProQuest

9122-LTR/SB-AR-CG/07-20