

**WORKING PAPER**

# Economic Impact Analysis of the Coronavirus, An Alternative Data Perspective

Xiyuan Zhang<sup>1</sup> | Qingquan Zhang<sup>2</sup> | Jiayue Wang<sup>3</sup> | Yanjun Li<sup>4</sup> | Beibei Li<sup>5</sup> | Danxia Xie<sup>1</sup><sup>1</sup>Tsinghua University, Beijing, China<sup>2</sup>University of Illinois Urbana-Champaign, Illinois, United States<sup>3</sup>Beijing International Studies University, China<sup>4</sup>Crosspoint Academy, China<sup>5</sup>Carnegie Mellon University, The United States**Abstract**

Alternative data is transforming the investment management process for financial industry, hedge funds, mutual funds, foundations, and pension funds. This paper describes the use of alternative data on the field of finance, particularly illustrating the complex forces driving the stock markets in China and the U.S., through the exploration of the alternative data through the outbreak of COVID-19. The analytic results demonstrate that the subway traffic and commercial housing deal area are positively correlated to the Chinese market in a rather weak level, and that search trends of coronavirus is negatively correlated to the American market with high reliability. The alternative data, which are closely connected with the situation of the coronavirus outbreak, have the ability of predicting where the markets are heading.

**KEYWORDS:**

Alternative Data; Coronavirus; COVID-19; Economic Impact; Linear Regression

## 1 | INTRODUCTION

The worldwide outbreak of COVID-19 has made a hard hit on the global economy. As a result, major stock markets witnessed unprecedented volatility since the great depression. The rapid changing situation among the government and central banks' response, the duration of the economy lock down, and availability of therapies and vaccines requires a fast reaction from the investors, both retailer individuals and institutions. It is getting critical to analyze how much COVID-19 could impact the US economy and stock markets. However, most of the traditional factors or economic indicators will lag the market movement. Therefore, alternative datasets other than the financial data show their explanation power to provide insights into the pandemic. In this article, We want to study the pattern of the market fluctuation from the perspective of alternative data.

### 1.1 | The Outbreak of Coronavirus

Novel coronavirus is the name of an epidemic which is highly infectious and has a long incubation period which could be as long as 14 days. On January 11, 2020, the Wuhan Municipal Health Commission announced the first death caused by COVID-19. It is the first time that the coronavirus starts to draw public attention. From then on, the numbers of suspected cases and confirmed cases in China increased dramatically. Later, on January 23, the Ministry of Transport of the People's Republic of China announced that Wuhan was on lockdown. The responses of Chinese government were quick and decisive. The government took a series of actions to put the spread of the virus under control, for instance, building makeshift hospitals in Wuhan, suggesting all citizens not going out unless absolutely necessary and suggesting colleges and universities reopen on a staggered

schedule. Those orders successfully curbed the spread of the disease but also caused a nationwide shutdown. As a result, shops, businesses and industries remained closed after China's New Year. The economy has taken a hit.

In the following weeks, with the movements of population around the world, more and more people in different countries have been tested positive of COVID-19. At the beginning of March, Florida, Kentucky, New York and many other states in the United States declared states of emergency. Corporations like Facebook and Amazon closed their offices. What is worse is that the outbreak of the coronavirus happened not only in the United States but also in European countries, such as Italy, Spain and Germany.

The catastrophic outbreak has caused world-range panic about public health as well as global economy. For example, many corporations in the manufacturing industry were halted to prevent the spread of the virus. For the same reason, the CEO of Apple announced that Apple decided to temporarily close all stores outside of mainland China for more than ten days. Worse still, all stock markets have experienced unexpected fluctuations during recent weeks. Take the U.S. stock market as an example. On March 9, it triggered mechanism as stock plunged more than 7%, which happened three times more during the following days. 9 days later, on March 18, the New York Stock Exchange declared to temporarily close floor trading and move to electronic trading after two employees tested positive for COVID-19. The unprecedented situations in the U.S. stock market have become the concerns of the multitude. A seemingly unavoidable disaster is yet to come.

## 1.2 | Background of Alternative Data

Alternative data are drawn from non-traditional data sources, which is why the analysis to the data often yield additional insights that complement the information received from traditional sources. Alternative data are presented in different forms like text, images, voices and videos. They can also be generated by digital footprints of large crowds. Besides, the Internet-of-Things (IoT) has become a dominant source of alternative data. A more detailed introduction of the alternative data sources is as follows.

Alternative data in the form of text are probably the most saliently used in finance and business economics since the beginning of the century. They have abundant sources, such as news articles, conference call transcripts, analyst reports etc. Gentzkow, Kelly, and Taddy 2017 discuss the features that make text different from other kinds of data, offering a practical overview of relevant statistical methods as an input to economic research.

Machine learning and the natural language processing (NLP) are two typical tools which are capable of dealing with large-scale, complex data and in the meantime, preserving its syntactic and semantic structure. Hanley and Hoberg 2019 combine LDA and word2vec to predict emerging risks in the financial sector.

Besides texts, many innovative usages of satellite imagery, voices and videos have been emerging, providing information from an unusual aspect. OrbitalInsight n.d. is a private geospatial data platform offering satellite data with broad applications. Rsmetrics n.d. puts satellite images to analysis of commodities and industrial applications. Facial features are also used as an indicator of testosterone level in the area of economic decision making. Jia, Lent, and Zeng 2014 document a positive association between CEO facial masculinity and various misreporting proxies.

The high penetration of mobile devices and the Internet has made it possible to acquire fine-grained user-behavior data (e.g., shopping preference, web browsing and interests). Those data are termed as 'digital footprints' of users and have a broad application in the industry. Data generated from digital footprints are often unbiased and comprehensive. Therefore, more surplus value could be found and obtained.

Advances in the Internet of Things have empowered almost every industry to become more intelligent and efficient by providing with a much more personalized service. With the information from the Internet of Things, the essential problem that who the consumers at the other end of the product are is perfectly solved. Inspirit IoT, an IoT startup in Illinois, develops an AI-based algorithm on a monitoring system to detect safety concerns of environment. Grewal, Roggeveen, and Nordfält 2017 talk about the applications of big data to provide customers friction-less shopping experience.

The sources of alternative data naturally bring them some features. First, they are unusual and non-standard with huge volumes. Second, they represent collective behaviors and are unlikely to swing just because of a change in a single person's behavior. Third, alternative data are various, which means they have abundant sources and have the potential of playing a cutting-edge role in the industry. With all the advantages of alternative data combined, it is promising that alternative data could reflect or even predict the trends of the pandemic.

## 2 | DETAILED ANALYSIS INTO HOW THE ALTERNATIVE DATA REFLECTED THE IMPACT CAUSED BY THE CORONAVIRUS PANDEMIC

### 2.1 | The Impact on the Stock Markets in Both China and the U.S.

The outbreak of the coronavirus has caused a pandemic of respiratory disease worldwide. And since no vaccines or specific treatments for the virus are available now, home isolation is probably the most appropriate response which most people are applying right now. During the prevalence of the coronavirus, almost all factories, schools, shopping malls have been shut down. People have to work from home or worse, lose their job. Both the demand side and the supply side have been tremendously affected, and it has been a hard hit in consumer confidence and investor confidence. A stock market crash is inevitable, followed by seemingly unpredictable fluctuation. However, since the big picture of the economy consists of billions of individual activities, the economy trends may be reflected, or even foreseen by the trends of their activity data. Two sets of alternative data are selected for each of the two countries, China and the U.S., i.e., daily subway traffic index in Shanghai, weekly commercial housing deal area in ten big Chinese cities for the former and weekly initial jobless claims, search trends (search interest relative to the highest point on the chart for the given region and time) of coronavirus for the latter. CSI 300 and S&P 500 are chosen to represent the economy of each country.

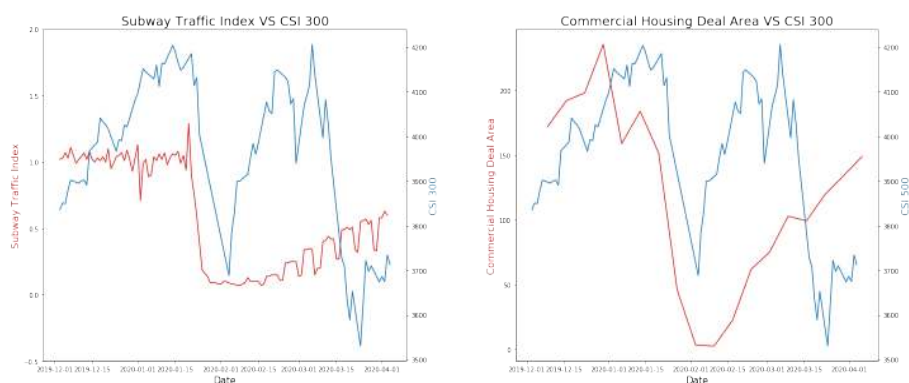


FIGURE 1 China Alternative Data VS CSI 300

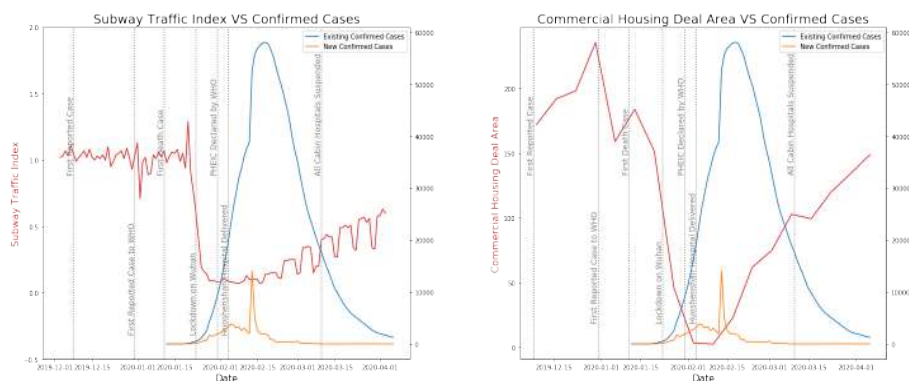
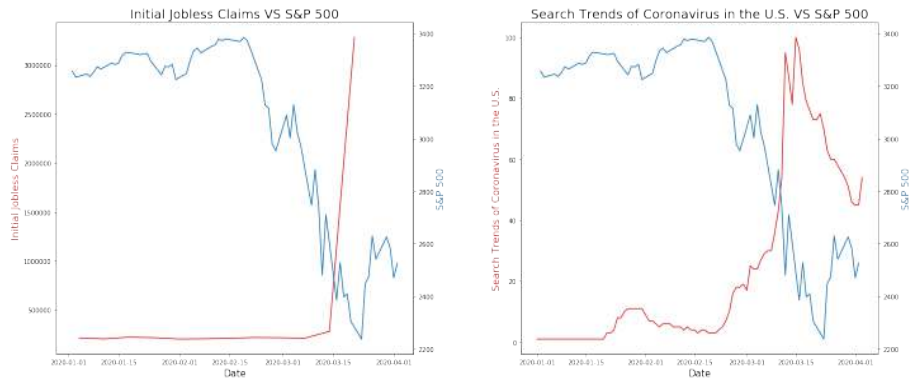


FIGURE 2 China Alternative Data VS Confirmed Cases

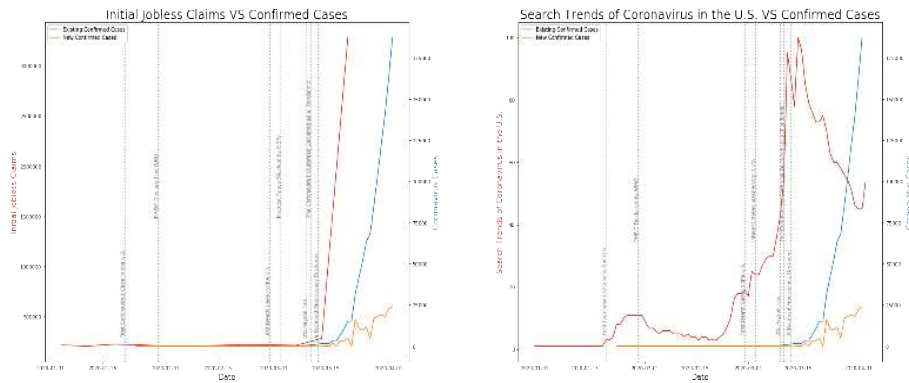
China's New Year Eve is on January 24, 2020 and the holiday was extended to two weeks due to the outbreak of the coronavirus. As can be seen in Figure 1, there are corresponding drops in both lines during that time. Usually the numbers would return to the normal level after a week, but not this time. The numbers are increasing at a slow speed but still haven't reached the normal

level, which means that things are getting better and that the economy is recovering gradually. That also matches the recent trends of confirmed cases in Figure 2: the number of existing confirmed cases is decreasing sharply and the number of new confirmed cases has been around 0 for a while. CSI 300 has a huge drop around March 13. The reason is that the worldwide outbreak of COVID-19 caused panic in Chinese market and the market tends to overreact.

The America's market has inseparable connections with the other parts of the world. When other countries showed a sign of an outbreak of COVID-19 around February 18, the America's market responded immediately. There is a time gap between the outbreak of the disease in America and in other countries, which is the reason why the two sets of alternative data, which mainly reflect the internal severity of the pandemic, are behind the market.



**FIGURE 3** The U.S. Alternative Data VS S&P 500



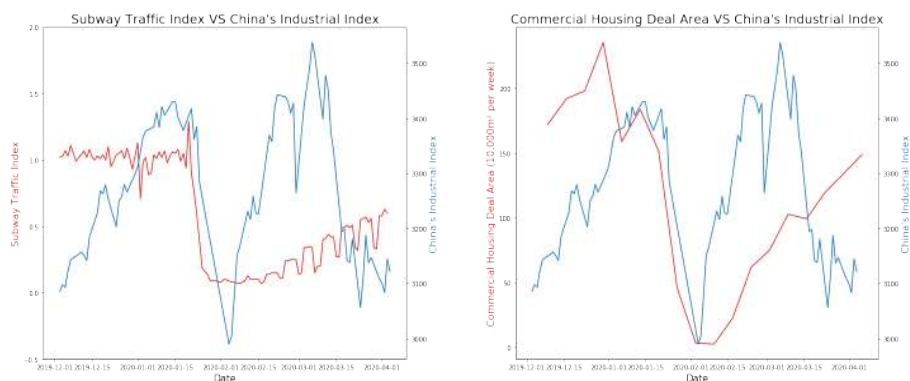
**FIGURE 4** The U.S. Alternative Data VS Confirmed Cases

## 2.2 | The Impact on Three Industries

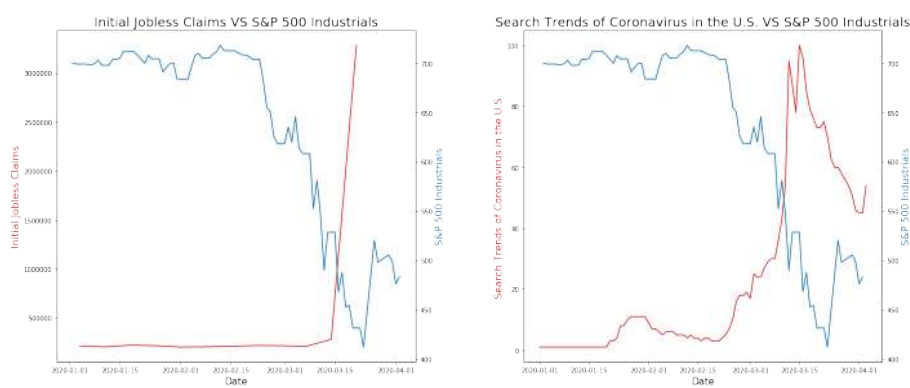
In this section, we are going to discuss how the outbreak of COVID-19 has impacted industries in China and the U.S. from the perspective of alternative data. Three indices and two sets of alternative data are selected to illustrate their relations.

### 2.2.1 | Industrial Index

Both two industrial indices have similar trends with the entire markets. As can be seen in Figure 5, China's Industrial Index fluctuates roughly in sync with the two sets of alternative data, indicating the inner relations between the alternative data and China's Industrial Index. In Figure 6, we can see that weekly initial jobless claims moves a few days after S&P 500 Industrial



**FIGURE 5** China Alternative Data VS Industrial Index



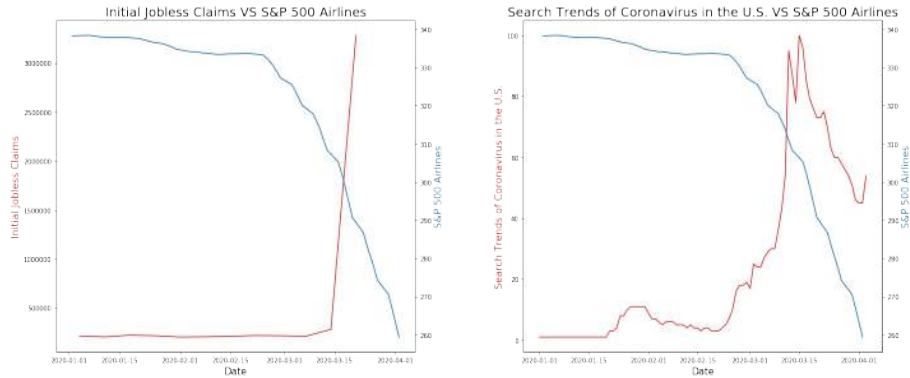
**FIGURE 6** The U.S. Alternative Data VS Industrial Index

does. That is because of the sequence that the economy breaks down first and then people lose their job. So the Initial Jobless Claims is a hysteresis factor of the economy. The search trends of coronavirus, on the other hand, is negatively correlated with the market and moves before S&P 500. After we did linear regression on them, we found a quite strong relation between them. The details are demonstrated in the next section.

## 2.2.2 | Airline Industry



**FIGURE 7** China Alternative Data VS Airline Index

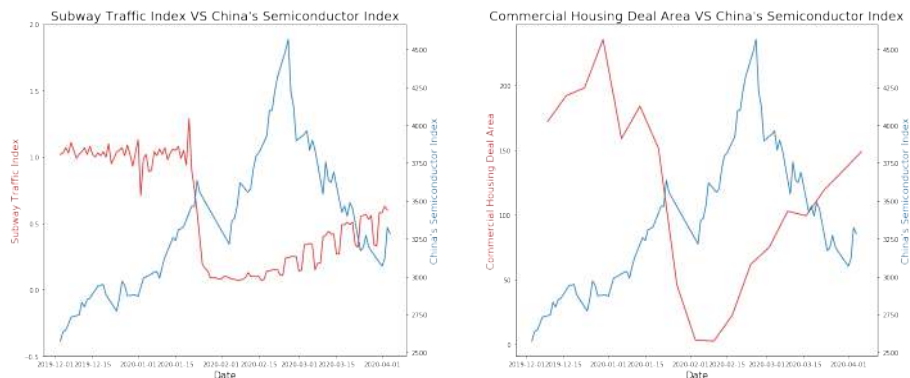


**FIGURE 8** The U.S. Alternative Data VS Airline Index

The Air Transport Index in China is mainly affected by two events: China's Spring Festival and the outbreak of COVID-19. The China's New Year Eve is on January 24. The Air Transport Index reaches its phased peak before that and goes all the way down during the holiday. After the holiday, most people began their home quarantine in case of getting infected, and thousands of airline tickets were canceled. Therefore, the airline index does not recover to the normal level right away.

The airline index in America has not been really influenced until the end of February, when the coronavirus became prevalent worldwide and the spread seemed unstoppable. When people in America realized the severity of the situation, they might avoid outside activities like business trips or travelling. The need for airline tickets has dropped down. And if the plane attendances are not enough to cover the flight' cost, the flights may be canceled by the airline companies. In fact, some American airline companies are on the brink of bankruptcy right now.

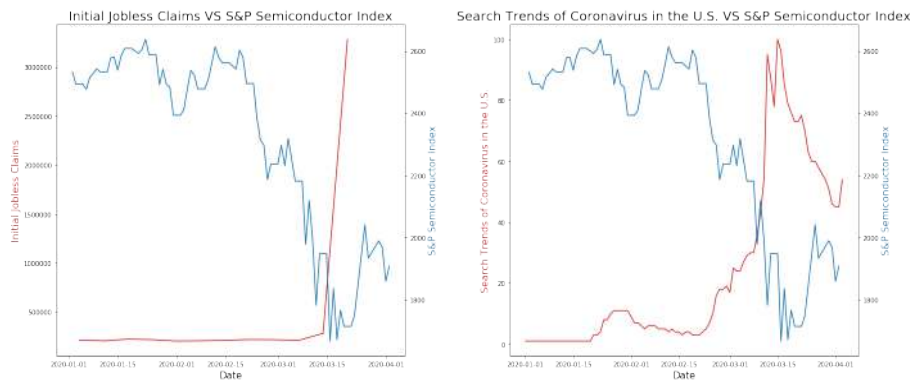
### 2.2.3 | Semiconductor Industry



**FIGURE 9** China Alternative Data VS Semiconductor Index

Due to the coronavirus pandemic, people either work from home or are suspended from job. Students have to take online courses instead of going to school. Households tend to replace their daily social activities with online entertainment, like mobile games and social media. All of those expand the demand for cloud computing services. On the other hand, the pandemic of the coronavirus indirectly caused lots of people jobless. Without salaries, the steady income, people tend to contract their spending. The markets of digital devices like smart phones have shrunk. Those are short-run effects from the pandemic of the coronavirus, while the long-run effects are still unknown.

The U.S. takes up almost half of the worldwide semiconductor market, which means the U.S. is a semiconductor-export country. China, however, takes up merely about 10% of the market, depending strongly on the import from America and other countries. When the prevalent disease forced factories all around the world to be shut down, not to mention the fact that the



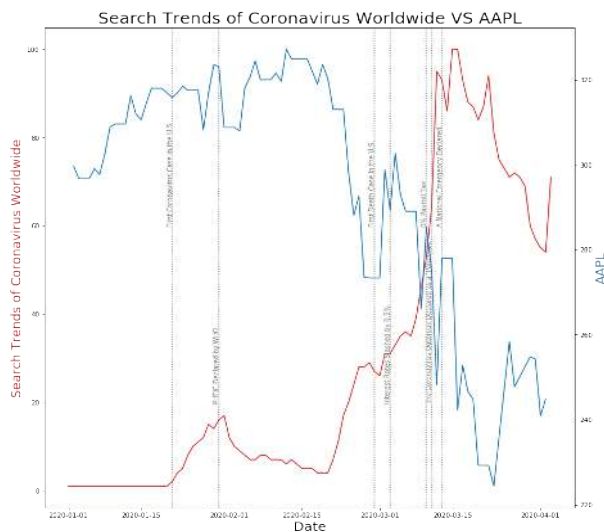
**FIGURE 10** The U.S. Alternative Data VS Semiconductor Index

U.S. government has been suppressing China's semiconductor market by forbidding exports to China and illegalizing Chinese semiconductor companies, it is expected that the S&P 500 Semiconductor Index is dropping all the way down.

For China, as can be seen in Figure 9, there is a sharp rise in the China's Semiconductor Index after the China's New Year, and it slowly falls back when more and more people have returned to work. Since people who work from home have to use online applications and most traditional pastimes are replaced by online entertainment, there is a growing need for cloud computing services. That is a strong push for the semiconductor industry. After people in China are back to work, the short pulse gradually dies down.

### 2.3 | A Deep Look into the Apple Corporation

In order to go into details about the influence of COVID-19 upon the world's economy, we compared the stock of Apple Inc. with the worldwide search trends of coronavirus. It is common knowledge that Apple Inc. has business with companies all over the world. For instance, Samsung Display and LG Display Co., two companies in Korea, are in charge of producing iPhone screens, and many Apple engineers working on cellular modems are based in Munich, Germany. Two of Apple's assembly factories are located in China. Nearly every part of the world has a connection with Apple Inc.



**FIGURE 11** Alternative Data VS AAPL

As can be seen from Figure 11, the trends of the two lines are negatively correlated. The search trends of ‘coronavirus’ reach a small peak around February 1st, nearly at the same time when the stock price of Apple Inc. experienced a minor drop. After that, both lines have a short, slightly peaceful period. When the outbreak of the coronavirus took place in many other countries in the late of February, the search trends go directly up and the stock price of Apple Inc. declines rapidly. What is worse is that after more and more people in America get infected, the search trends surge exponentially. Correspondingly the stock price of Apple Inc., though fluctuating, falls significantly. The fluctuation is because of a series of newly-issued policies trying to reinvigorate the economy such as cutting down tax and interest rates. However, the panic caused by the highly infectious disease is too suffocating to easily dissipate. As long as there are no signs of a turning point of the coronavirus pandemic, it is hardly possible for a multinational corporation like Apple Inc. to fully restore its supply chain and therefore have its stock price go positively. Cautiously speaking, all the impacts for now are short-run. In the long run, the outbreak of the pandemic might be trivial but only if the disease can be under strong control.

### 3 | MATHEMATICAL PROCESS OF THE DATA

We did linear regression analysis of the alternative data (subway traffic index for China and search trends of coronavirus for America) and the stock indices in order to find a preliminary mathematical relation between them. The data concerning China are confined to the range of January 1st to March 3rd to eliminate the influence caused by the China’s New Year and the America’s market. The results are listed in the table.

**TABLE 1** CHina Linear Regression of Alternative Data and Indices

Index	Equation	R-squared	P-value	F-statistics
HS 300	$Y = 3977.79 + 683.20 * T_1 - 501.22 * T_2$	0.481	$1.06 * 10^{-6}$	19.44
Industrials	$Y = 3303.95 + 77.93 * T_1 - 657.12 * T_2$	0.358	$9.05 * 10^{-5}$	11.72
Airline	$Y = 3467.97 + 968.12 * T_1 - 395.75 * T_2$	0.835	$3.67 * 10^{-17}$	106.3
Semiconductor	$Y = 4001.97 + 525.62 * T_1 - 1242.00 * T_2$	0.557	$3.80 * 10^{-8}$	26.37

**TABLE 2** The U.S. Linear Regression of Alternative Data and Indices

Index	Equation	R-squared	P-value	F-statistics
S&P 300	$Y = 3315.62 - 3.7630 * T_1 - 8.0793 * T_2$	0.892	$1.10 * 10^{-29}$	247.0
Industrials	$Y = 706.33 - 1.3480 * T_1 - 1.8371 * T_2$	0.892	$1.45 * 10^{-34}$	289.3
Airline	$Y = 336.29 + 0.3021 * T_1 - 0.9748 * T_2$	0.717	$3.49 * 10^{-17}$	76.09
Semiconductor	$Y = 2535.99 - 6.3945 * T_1 - 3.523 * T_2$	0.878	$1.13 * 10^{-32}$	251.3

In general, the linear regression models of the data related to China give slightly vague linear relations between the Chinese alternative data and all the indices, while the models of the data related to the U.S. produce more significant results. R-squared in the models of China are not big enough to prove their strong linear relations except the one of China’s airline index. The main reason is probably that the alternative data reflect only the situation in China while Chinese market has a strong and inseparable connection with the rest of the world. The pandemic was in control in the mainland of China in the late February but the coronavirus was just about to ravage the world at that time. The number of total confirmed cases in China was overtaken by the sum of total confirmed cases in the other countries, meaning that the severe international situation might cause more panic than the situation in China. As for the R-squared of airline-index model, both airline index and subway traffic index reflect the willingness of travel. Therefore, the R-squared of airline-index model is relatively higher than others.



The linear regression models of the data related to the U.S. produce better results than the models of China, as seen in Table 2. Generally speaking, the search trends of coronavirus have a negative connection with the indices, which is consistent with the previous analysis. The correlation coefficients are at a higher level. The main reason why the results of the models of the U.S. are better is that the coronavirus pandemic in the U.S., though taking place two months later than in China, is expanding exponentially. While the industries in China are gradually re-opening, the economy of the U.S. is falling down. Millions of people have lost their job during the pandemic. From all the numbers and reports, it is obvious that the situation in the U.S. is much worse than that in China. Because there is a more severe breakdown in the U.S., it is reasonable that the coronavirus pandemic in the U.S. has caused a more destructive consequence in the domestic market.

The linear regression model of AAPL and worldwide search trends of coronavirus also produces a significant result, proving the outbreak of the disease is the main reason why AAPL is fluctuating these days.

For all the models, the confidence intervals are set between 5% to 95% and the P-values are low enough to promise the significance of the models. The F-statistics in China's models are not as big as those in the U.S.'s models, which corresponds with the R-squared.

No one knows for sure if all the indices and stock prices are bottoming out now unless this disaster has really passed. But one thing can be sure that the social activities no matter where, can only restore to its normal level on the condition that COVID-19 is no longer a threat to human life.

## 4 | CONCLUSION

Alternative data provide an unbiased and novel perspective into the ongoing pandemic worldwide. Combined with appropriate math methods, alternative data can generate predictive information on the stock markets. It is reasonable to conclude that they have the potential for building up a fine model to predict the trends or patterns of the economy.

## 5 | BIBLIOGRAPHY

### References

- Jia, Yuping, Laurence Van Lent, and Yachang Zeng (Dec. 2014). "Masculinity, Testosterone, and Financial Misreporting". In: *Journal of Accounting Research* 52.5, pp. 1195–1246. DOI: 10.1111/1475-679X.12065. URL: <https://ideas.repec.org/a/bla/joares/v52y2014i5p1195-1246.html>.
- Gentzkow, Matthew, Bryan T Kelly, and Matt Taddy (Mar. 2017). *Text as Data*. Working Paper 23276. National Bureau of Economic Research. DOI: 10.3386/w23276. URL: <http://www.nber.org/papers/w23276>.
- Grewal, Dhruv, Anne L. Roggeveen, and Jens Nordfält (2017). "The Future of Retailing". In: *Journal of Retailing* 93.1. The Future of Retailing, pp. 1–6. ISSN: 0022-4359. DOI: <https://doi.org/10.1016/j.jretai.2016.12.008>. URL: <http://www.sciencedirect.com/science/article/pii/S0022435916300872>.
- Hanley, Kathleen Weiss and Gerard Hoberg (2019). "Dynamic Interpretation of Emerging Risks in the Financial Sector". In: *Review of Financial Studies* 32.12, pp. 4543–4603. URL: <https://ideas.repec.org/a/oup/rfinst/v32y2019i12p4543-4603.html>.
- OrbitalInsight (n.d.). *Orbital Insight Website*. <http://www.orbitalinsight.com>. Accessed 2020.
- Rsmetrics (n.d.). *Rsmetrics Website*. <https://www.rsmetrics.com>. Accessed 2020.

