

Data-driven Analysis of Remote Work in China during the COVID-19 Pandemic

<p>Yufei Wu Information Technology and Web Science (ITWS), Rensselaer Polytechnic Institute littelsilverspoon@gmail.com</p>	<p>Thilanka Munasinghe Information Technology and Web Science (ITWS), Rensselaer Polytechnic Institute munast@rpi.edu</p>	<p>Lydia Manikonda Lally School of Management, Rensselaer Polytechnic Institute manikl@rpi.edu</p>	<p>Oshani Seneviratne Institute for Data Exploration and Applications (IDEA), Rensselaer Polytechnic Institute senevo@rpi.edu</p>
--	---	--	---

Abstract

This paper leverages online content to investigate the biggest impact of COVID-19 – remote work, by using China as a primary case study. Telecommuting has become popular since February 2020 primarily due to the pandemic, and people have been slowly returning to their office from May 2020. This study focuses on two time windows in the year 2020 to calculate the growth of different job sectors. Our results indicate the negative impact of teleworking in manufacturing industry, but shows that information technology-related industries are less affected by working from home. This paper also investigates the impact of COVID-19 on the stock market and discussed what plan of action the policy makers should take to provide a good economic environment for the country. In addition to the overall economic situation, we observed how the psychological situation of employees could affect their job performance, indirectly affecting the development of certain industry sectors. Therefore, misinformation in certain Chinese social media channels was also studied in this paper specifically examining the rumors and their latent topics. We believe that our work will initiate a dialogue between scientists, policy makers and government officials to consider the observations highlighted in this paper.

1. Introduction

Due to the rampant spread of the COVID-19 pandemic, social distancing was put in to effect to help mitigate the spread of the virus through workplace interactions. These measures led to an unexpected shift from in-person to online teleworking or remote working to minimize the spread of the disease [1] and to comply with various health guidelines imposed by governments and industries [2]. In fact, COVID-19

forced employers to allow telework on a massive scale, which had a certain impact on the overall global as well as local economies. However, the impact of teleworking appear to be different for various industries [3], so it needs further analysis. At the same time, providing a good environment for remote work is not only the responsibility of enterprises, but also of the government in terms of formulating relevant policies to manage the prevailing pandemic situation well, and mitigate any economic impacts resulting from the pandemic. The pandemic disrupted people's daily lives that could easily reduce work efficiency, which is why effective policies at various levels, i.e., corporate and governmental, are needed to mitigate such disruptions while providing appropriate support structures. Stakeholders involved should also pay special attention to the public opinion dynamics and combat misinformation to dispel harmful rumors in a timely manner [4]. This paper investigates the impact of remote work and also leverages different topics present in the social media to highlight the various implications on stock market.

Through data-driven analysis using data obtained from online posts (made on a mobile app called *dingtalk*) [5] as well as stock market prices, we attempted to determine the industries that are vulnerable to teleworking. Our analysis reflects that industries that are materials-based, and hospitality industry such as hotels, tourism are vulnerable to teleworking corroborating the existing research [6, 7]. However, telecommunication and information industries appear to be less affected by teleworking. Therefore, there is a need for better preparation and investment to protect jobs in the former industries, and the latter industries can consider moving some of their work online as appropriate for cost-cutting measures.

We further analyzed how the types of public opinions can help the teleworking process. We chose the Shanghai Stock Exchange (SSE) Composite Index and Primary Sector indices to study the markets and different

fields from an economic perspective. Based on our data-driven analysis, we provide some rationale for how the initial epidemic in China and the global pandemic affected the Chinese economy, and the extent of their influence, and the reasons behind their differences. In terms of public opinion, we use the Latent Dirichlet Allocation (LDA) model [8] to investigate rumors that are in the form of text, and determine the corresponding measures to mitigate their spread.

2. Methodology

Data Used in the Analysis: For analyzing the performance of remote work in the Chinese economy, we used stock market and rumor datasets. The datasets used for stock market analysis are listed in Table 1. For misinformation analysis we utilized the *rumor dataset* [9].

2.1. Determining the Remote Work Time Period

Before analyzing remote work due to COVID-19, we need to characterize the points at which remote work began and transitioned back to in-person office work. However, it is difficult to determine the exact date when all the companies went into a remote work setting and return to the regular in-person work setting. Therefore we chose a heuristic using a popular mobile platform used in China. DingTalk [5] is a leading intelligent mobile office platform in China, developed by Alibaba group and provided to all Chinese enterprises free of charge for business communication and work collaboration. According to a third-party data from quest mobile [13], DingTalk has the highest market share in an intelligent mobile office, with the number of active users ranking the first. The dominance of DingTalk in the mobile office platform is the reason why DingTalk data was chosen as the research object of time segmentation in this study. The ranking data of DingTalk comes from *kuchuan* [14], the largest third-party platform for app publishing and app monitoring in China. We obtained DingTalk’s ranking in all free apps in 2020. The plotted graph of the ranking can be seen in Figure 1. The trend of two lines (blue line representing the “ranking of all android and ios applications” and the orange line representing “ranking in all iOS free applications) in the graph shown in the Figure 1 are quite similar, which demonstrates that the trend is invariant to the underlying platform. The orange line shows the rank considering all applications stores in Android and iOS while the blue line shows the rank considering the iOS store. The X-axis represents the specific time of data acquisition, and the Y-axis

represents DingTalk’s ranking in the app store.

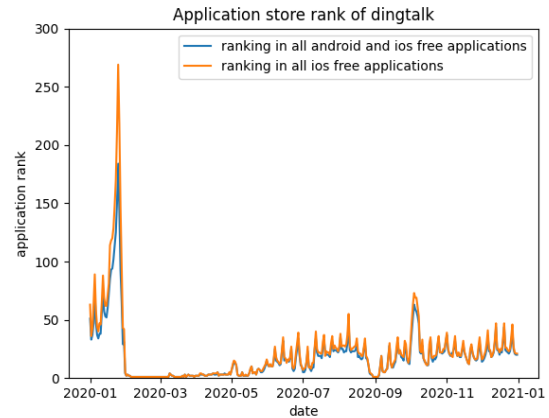


Figure 1. Application Store Rank of DingTalk

The vertical axis shown in Figure 1 shows the ranking – lower the numerical value of ranking, higher the popularity of DingTalk. As seen from Figure 1, DingTalk ranked very low in January. This is reasonable because, before the pandemic started, people mainly worked in the office and had less demand for mobile work platforms. But, as of the rising pandemic, many employees began to work remotely. In order to manage and communicate, they needed to use mobile office platforms to help internal communication and business communication with clients. During the peak period of the pandemic, DingTalk ranked very high in the app store, and even held the rank first for a long time. In the summer of 2020, many people ended their remote work and slowly began to return to the office [15], and their DingTalk usage was reduced. Based on this data Figure 1, we used January 2020 to May 2020 as the target time period to analyze remote work and May 2020 to December 2020 for the return from remote work.

2.2. Analysis of the Growth in Remote Work

According to the time segment determined above, we divided the stock market dataset into two parts – 1) data from Jan’20–May’20 to analyze the impact of COVID-19 and remote work on various areas and 2) data from May’20–Dec’20 to analyze the impact of returning from remote work to in-person work. We first analyzed the overall situation of the stock market for this entire period by calculating the growth rate as:

$$p = (v_{end.date} - v_{start.date}) / v_{start.date} \quad (1)$$

In the above equation, p is the growth value of a certain period, $v_{end.date}$ is the stock index value of end date, and $v_{start.date}$ is the stock index value of start date.

Table 1. Different datasets and a short description

Dataset	Dataset Description	Dataset Usage
COVID-19 Cases [10]	The current confirmed, dead, and recovered instances in different regions in the world	The impact of COVID-19 on economy
The SSE Composite index [11]	Stock market index of all stocks traded at the SSE	The impact of COVID-19 on economy
Primary Sector index [11]	Stock index of different sectors	The impact of COVID-19 on economy and The impact of remote work on different sectors
Rumor Refutation Data Set [12]	The content of rumors and the refutation of it	Study the main content of the rumors during the initial epidemic in China and what the policy makers can do for the public opinion to provide a good environment for remote work.

As seen in Table 2, the growth rate in the first stage appears to be negative, and in the second stage it is positive. Unsurprisingly, as known quite widely, the COVID-19 pandemic significantly impacted the economies around the world. The outbreak was so sudden that organizations and the society at large were not well equipped to handle the significant disruptions caused by the pandemic. For example, software/hardware facilities, company management, and the psychological state of the employees were all needed to be considered in such a short time. Teleworking further reduced the efficiency of employees, which also affected the economy to a certain extent[16].

Table 2. Growth Rate in Two Stages

Jan '20–May '20	May '20–Dec '20
-6.87%	18.58%

To further understand the impact in different industries, we calculated two indicators based on stock market data, including the growth rates of different industries from Jan'20–May'20 and May'20–Dec'20. The average growth rates of the two periods were calculated respectively, and then for each period, all industries were divided into two categories based on the value of the mean, resulting in four different categories in total. We also used the relative value based on the average value to make a better comparison. The average growth value from Jan–May was negative. We observed that for the first period (Jan–May), the higher the relative growth rate of the stock, the more it declines. To compute the relative growth (rg) in both stages, we use two formulae as shown below. Since the average in the first stage is positive while that of the second stage is negative, to better show the extent of influence, we reverse the equation in the second stage. Therefore, the higher the relative growth rate is, the higher the degree

of influence would be.

$$rg_{first_stage} = mean - ag \tag{2}$$

$$rg_{second_stage} = ag - mean \tag{3}$$

In these formulae, rg_{first_stage} is relative growth of the first stage (Jan'20 to May'20), rg_{second_stage} is relative growth of second stage (May'20 to Dec'20) and ag is absolute growth rate of sector.

Please refer to the quadrant diagram shown in Figure 2 that distinguishes the four categories of market sectors. In this diagram, we used different colors to represent sector names starting with SSE and 380¹.

In the **first** quadrant of Figure 2, we see industries whose growth rates in the first and second stages that are higher than the average value are included. These industries are significantly affected in the process of remote work and returning to normal. In the first quadrant, sectors are industries, materials, and optional stocks covering a wide range of sectors, including many industries that rely on employees to work in specific workplaces, such as the hotel and manufacturing industries. All these sectors need employees to operate machines on-site and highly rely on on-site work. The **second** quadrant includes the less affected industries in the first stage but has a significant growth in the second stage. They include service-related indices such as medicine. Remote work will not bring many obstacles to these sectors, but when going back to normal, those sectors would have had a simulation for further development with remote work experience. The indices in the **third** quadrant are less affected in the first and second stages. Generally speaking, they are less affected by the shift in the work location. The relative

¹Sector names starting with **380** means that that it represents the overall performance of emerging blue chip companies of Shanghai Stock Exchange.

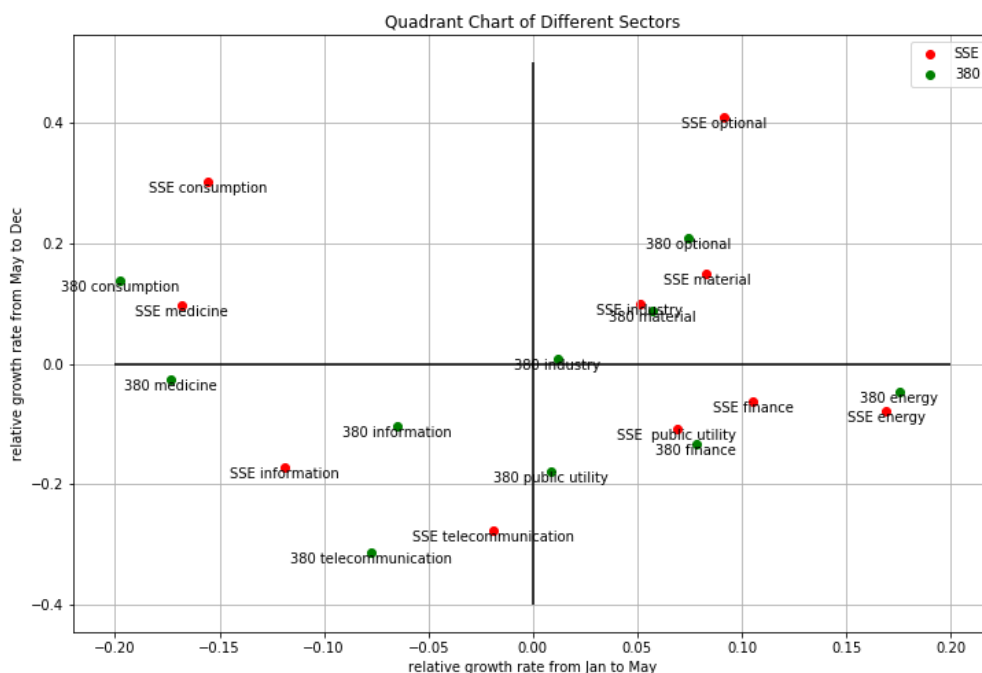


Figure 2. Quadrant Chart of Different Sectors

growth rate of 380 medicine follows the same trend as that of SSE medicine. Other areas are mainly about telecommuting and information which heavily rely on the electronic equipment rather than machines needed to operate on site. In the **fourth** quadrant, in the first stage, the impact of remote work on these areas is more significant, but in the second stage, the growth is less. The two relative growth rate shows that remote working and COVID-19 negatively influence finance, energy, and public utility, and they were not able to recover quickly. In the early stage, they were influenced by international industries, but in the later stage, they did not have enough stimulation to recover and develop.

2.3. Analyzing Misinformation during the COVID-19 Pandemic

During the pandemic, remote work needed the support of software and hardware and more attention to the psychological state of employees. Rumors during the pandemic created initial panic and anxiety in certain populations [17], which was not conducive to improving the efficiency of remote work [18]. As such, the second part of this research considered how false information may have affected people during

the pandemic and any implications for remote work practices. To study misinformation associated with COVID-19 we leveraged a publicly available *rumor dataset* [9]. Since we were focusing on the data about Chinese industries, we utilized the rumors that were in the Chinese language. To handle this, we used Chinese word segmentation and computing TF-IDF method [19], a statistical method to evaluate the importance of a single word in a document set or corpus. The value of TF-IDF will increase with the number of words appearing in the document and decrease with the number of words appearing in the corpus. When processing the Chinese natural language, it is usually necessary to segment words first due to its linguistic properties. We chose the Jieba module [20] to perform this task.

3. Results

This section illustrates the results from the analysis on the various data points collected during the pandemic.

3.1. Economic Sector Analysis

There is no denying that the situation in every sector is closely related to the development of the whole economy. Due to the pandemic, all the major economies were hit, and GDPs have declined [21]. There was also a significant impact on the stock market [22, 23]. Therefore, policymakers must formulate relevant policies to minimize such impacts in the future [24]. Their actions will help in building an efficient remote work environments. However, to formulate more appropriate policies, we need to have a deeper understanding of the state of economy in this situation. Based on the DingTalk data we analyzed (Section 2.1), we determined the period of remote work during the pandemic in China to be from Jan 24, 2020 to May 15, 2020.

There are many attributes in the COVID-19 cases dataset. We chose Multivariate Analysis Of Variance (MANOVA) [25] and Variance Inflation Factor (VIF) [26] to select the suitable independent variables for our model, which enabled the selection of a high correlation and reduced the multi-collinearity of all independent variables. After getting all the elements whose p -value is less than 0.05, we tested the combination of factors to ensure that the VIF value is less than 5. Finally, we chose the independent variables to be daily confirmed cases in the world, and recovered ratio in the world. The recovered ratio is equal to the ratio of cured people to diagnosed people from COVID-19. The result of VIF can be seen in Table 3.

Table 3. The VIF value of the combination of confirmed cases in the world, and recovered ratio in the world.

Attribute	VIF
Confirmed cases (World)	1.986104
Recovered ratio (World)	1.930471

After obtaining the combination of independent variables that will influence the stock market, we aimed to determine to what degree the pandemic affected the index and volume of the stock market. Therefore, we chose linear regression to evaluate those factors. The dependent variable is the index of the SSE composite index, the volume of stocks trading on the Shanghai Stock Exchange, and primary sector indices.

3.2. SSE Composite Index Analysis

The SSE composite index is a statistical index reflecting the general trend of listed stocks in the Shanghai Stock Exchange. The total market value, circulation market value, quantity proportion, and

transaction amount proportion of stocks in the SSE are quite considerable in the Chinese stock market. We chose two attributes in this dataset, volume, and index as dependent variable. We chose confirmed cases in the world, and recovered ratio of the world that comes from the result of VIF as independent variable to evaluate the impact of COVID-19 on the Chinese stock market from a macro perspective. The coefficient of linear regression about the SSE composite index and the volume of the SSE composite index can be seen in table 4.

Table 4. The coefficient of linear regression on the index.

Attribute	Coefficient of Index	Coefficient of Volume
Confirmed cases (World)	-0.19899411	-0.27922668
Recovered ratio (World)	0.32447878	0.32447878

When building the model, we divided the data into two parts at a ratio of 9:1 for the training set and testing set. The reason for splitting data like this, is because the period considered was not a long time before the pandemic began, so our data volume is relatively small. The Root Mean Square Error (RMSE) [27] was used to validate the model's accuracy. RMSE is around 0.19 in the model with the index as the dependent variable, so there is a forecast error. RMSE is 0.07 in the model with volume as the dependent variable, indicating that the model's overall accuracy is relatively high. These errors are considered in the subsequent analysis.

According to the result above, even though there is a difference between index and volume, all factors seem to have a particular impact on the Chinese stock market. Among them, the world's daily recovery ratio positively affect the stock market, while the world's confirmed cases seem to hurt the stock market the most. Furthermore, in order to support more enterprises to return to work and production, the local government issued a series of supportive policies and implemented them, such as providing tax relief and credit support for enterprises, especially for small and medium-sized enterprises [28]. At the same time, literature states that the government offered more jobs for infrastructure construction and promoted infrastructure-related industries [29]. Due to the later stage of the pandemic, the world's pandemic center was transferred, and foreign commercial capital returned. The influx of money further promoted the development of related industries.

Meanwhile, the global situation of COVID-19 also had an impact on China's stock market [30]. It is

undeniable that China's economy is closely connected with the world in the context of economic globalization. In the latter stage of COVID-19 considered in our study, we observe that the economy was dramatically impacted in many regions of the world [31], which, in turn, influenced the Chinese markets to some degree. The close relationship between China and the global economy might have played a positive impact on China's stock market [32].

3.3. Primary Sector Index Analysis

Now we turn our attention to different areas of the stock market. We used the linear regression method to analyze various industries globally, changing the SSE Composite Indices data to Primary Sector Indices. The coefficient of linear regression was used to compare the impact of the pandemic on different industries. The results of linear regression are shown in Table 5. According to this table, SSE optional and SSE finance are the most negatively affected by confirmed cases worldwide. Moreover, some indices are positively correlated with confirmed cases worldwide, for example, SSE energy, and SSE consumption. Although the SSE consumption and SSE optional look similar, they are indeed about different areas of consumption. The former is the most basic and necessary consumer goods in daily life, including agricultural, animal husbandry, fishery products, food, personal household products, etc. The latter is the consumption and the necessary expenditure, including cars, clothing, media, etc.

Based on Table 5, SSE telecommunication and the SSE industry were the most positively affected. SSE medicine index was the only one that decreased with recovery increase. The impact of the recovery ratio of the global pandemic on various industries is also positive, especially for the SSE Information and SSE medicine. That is because, in the middle and late stages of the pandemic, people chose to study and work at home, and the related information technology industry developed. At the same time, people needed to go out to work with the awareness of self-protection in the pandemic, so the demand for Personal Protection Equipment (PPE) increased, which promoted the development of healthcare equipment-related industries. Also, the need for vaccine development may be another reason.

3.4. COVID 19 Misinformation in China

Absorbing a lot of misinformation may cause more anxiety to employees and can drastically damage public trust [33]. Public trust in science and providing evidence

Table 5. The coefficient of linear regression on the primary sector indices.

Global confirmed cases	Global recovered ratio	Sector name
0.578599	0.166588	SSE energy
-0.133817	0.380406	SSE material
0.11084	0.202579	SSE industry
-0.390236	0.381991	SSE optional
0.56165	0.273332	SSE consumption
0.39067	0.407933	SSE medicine
-0.35649	0.381908	SSE finance
0.004645	0.485089	SSE information
0.109616	0.036243	SSE telecommunication
-0.081437	0.206555	SSE public utility

is essential for overcoming a pandemic like COVID-19. In this next analysis, we use Chinese social media as an example to see what type of misinformation appeared during the COVID-19 pandemic.

Table 6 presents the keywords with TF-IDF value above or equal to 0.03 because of its length. We can see that the topics identified are mainly about reducing the spread of the virus to prevent infection. For example, people are concerned about what kind of masks are suitable for virus protection and the duration of wearing different masks. However, by reading the original text [12], we can also see that "heat," "alcohol," "hairdryer" are also in the same category. People are concerned about whether the application of heat and alcohol can effectively eradicate the virus. Also, there are terms such as "coat" and "fur collar" that shows that people are worried about how long the virus will last in different clothes. As rumors often occur in the areas of greatest concern [34] these topics identified could signal some of the rumors circulating in social media.

We could also see keyword "Wuhan" and "Hubei." Considering the timing of data acquisition, this reflects the predominant concentration of cases. Different control policies were implemented for different areas of severity [35]. It is likely that people were more concerned about the COVID-19 situation in

Table 6. The keywords and their corresponding TF-IDF values.

Keyword	TF-IDF value	Keyword	TF-IDF value
Cases	0.41	New	0.06
Confirmed case	0.28	Infections	0.05
Pneumonia new added?	0.23	Input	0.05
epidemic	0.17	Close	0.04
Coronavirus	0.1	Contact	0.04
Discharge	0.1	Time	0.04
death	0.09	Isolation	0.04
Healing	0.09	Overseas	0.04
Suspected case	0.08	Prevention and control	0.04
Medical observation	0.07	Severe patient	0.03
Report	0.06	Testing	0.03
		Ministry Of Health	0.03

geographical areas with high case numbers, and misinformation about COVID-19 would be more rampant in social media posts originating from those regions.

To better understand the keywords in the rumors, we applied LDA, a Bayesian probability model with three levels of structure, including words, topics, and documents. We chose the attribute named *body* as a text source, because it contains rumors and inaccuracies. For the LDA model, we picked a value k to determine how many topics we could get. To get the value k , we computed the perplexity [8] based on different k values to choose the optimal one. The graph of perplexity value based on different k can be seen as the Figure 3.

According to Figure 3, it shows the value of perplexity from $k = 1$ to $k = 50$. Perplexity is a metric from information theory that measures the goodness of a language model where, lower the perplexity is better [8]. The formula for calculating the perplexity is given as:

$$\text{perplexity}(\text{test set } w) = \exp \left\{ -\frac{\mathcal{L}(w)}{\text{count of tokens}} \right\} \quad (4)$$

$\mathcal{L}(w)$ is the log-likelihood of a set of unseen documents. It is the evaluation of the LDA model and the judgment of improved parameters. We selected the value of k at the elbow of the line to determine the optimal k of our LDA topics. Looking at the trend in the image, we could choose a value of k from 5 to

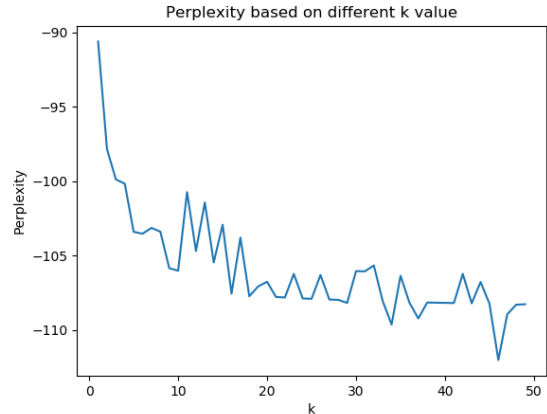


Figure 3. The value of perplexity based on different k-value

10. Furthermore, since the amount of data is not large, we ultimately chose 6 as the k value. After training the models and translating the keywords in the different topics into English, the result can be seen in Table 7, where for each cluster, the five most representative keywords are shown.

Table 7. The keyword topics.

Topic Number	Keywords
1	[personal protective equipment] mask, country, over, immunization, medical
2	[health care system and pets] health, Center, Dog, Disease Control, Diagnosis
3	[the symptoms and the situation in Wuhan] Wuhan, medical, epidemic, symptoms, pneumonia
4	[virus information and virus treatment] effect, report, rumors, isolation heat
5	[virus transmission]respiratory, respiratory tract, droplets, spread, coronavirus

According to Table 7, there are five topics. For the **first** topic, the keyword “mask” has a much higher value than the other words, which includes the usage and effect of the mask and how to identify the most effective mask that people care about. Under this circumstance, it is of considerable significance to guide people on choosing and using masks appropriately and increase the production of pandemic prevention equipment, not only for the prevention and control of the pandemic, but also

to mitigate any public anxiety about masks.

The **second** topic is mainly focused on the healthcare system and pets. As for the healthcare system, it is not the source of rumors but the subject of refutation. After reading the original file of the rumors [12], we realized it is mainly focus on statements of experts from hospitals or health systems. As experts in infectious diseases, their statements are more persuasive and credible than those of ordinary media and can achieve the effect of refuting rumors. For pets, people are concerned about whether diseases can spread between people and pets.

The **third** topic was mainly about the symptoms and the situation in Wuhan. People were worried about what kind of symptoms can be detected in the hospital and the research of asymptomatic patients. The **fourth** topic was mainly about virus information and virus treatment. Because there were many rumors about this specific topic, some professional medical knowledge was needed to refute the rumors. This is crucial as some rumor-mongers dumped drugs or health care products and spread rumors that related drugs are helpful for the prevention and cure of COVID-19. The **fifth** topic is mainly about virus transmission, specifically about the transmission medium. For this topic too professional medical experts are needed to dissuade rumors related to this topic.

4. Discussion

Combined with the analysis of the economic indicators and rumors, the following suggestions are put forward.

First, policymakers should actively understand people's needs using data-driven approaches as described in this paper. A sudden unexpected disaster, such as the COVID-19 pandemic, will seriously affect people's living conditions and mental health, and the workforce is among the most critical factors affecting a country's economy. From the perspective of the well-being of the society as a whole, the ability to understand the plight of the public through data-driven analysis can help effective decision-making. Second, there is a need to set up more production lines to improve the efficiency of products people need, meet people's needs, and control the price of products. Third, based on the initial volume of the rumor dissemination, we conclude that it is better for authoritative medical organizations to promptly provide adequate information about the pandemic on multiple platforms, with a heavy emphasis on diverse social media platforms. This is because rumors often spread because people cannot get reliable and fact-checked information from experts in a time-sensitive manner. The timely release

of information is an effective tool to help people ease their anxiety and tackle the spread of misinformation. At the same time, it is beneficial for policymakers to consider the diversity of access to information. Fourth, for the areas with more COVID-19 cases, policymakers or medical organizations should disseminate official scientific reports on the disease cases and be transparent as possible to have direct impact on curb the spread of rumors. Fifth, policymakers should consider the impact of the pandemic on the economy and adopt relevant macro-control policies to provide a suitable environment for the economy, such as supporting small and medium-sized enterprises, increasing infrastructure to solve employment problems, and other measures. Also, it is necessary to pay attention to the protection of economic security during the pandemic because the global pandemic hit the industries of various countries and caused financial turbulence. Finally, different policies are required to be formulated for different industries. According to our analysis, it can be seen that different industries have different acceptance of teleworking. Therefore, to improve the work efficiency of the industry, the policies should start from the characteristics of the industry to find out the problems of teleworking in a pandemic.

5. Conclusion

We have presented an analysis of the remote work situation in China based on the stock market and social media data collected in 2020. We first divided the time for the analysis based on DingTalk's ranking in the app store between February 2020 to May 2020, when remote work was predominant in China, and June to December 2020, when people went back to their offices. We then used the stock index as an essential indicator of economic development to analyze the development of different sectors during these two phases. Our analysis showed that sectors such as energy, and materials, were more affected by teleworking, and industries such as medicine were less affected. Even though authors understand the importance of evaluating the adverse effects on the supply chain in various sectors, the current scope of this study is not focused on the impact caused on the supply chain due to COVID-19. We also selected confirmed cases in the world, recovered ratio of the world as independent variables using VIF to study their overall impact on the stock and its impact on different sectors. From the aspects of public opinions, we analyzed the social media datasets to determine what concerned people the most during the pandemic. Analyses such as these can elicit information about what people care about during

uncertain and troubling times could help professionals to provide valuable support. Code repository of all the analysis below is available at: https://github.com/lawiet019/DA_FinalProject.

In conclusion, since all observations and conjectures are based on short-term analysis considering a specific region, more data accumulation is needed for longer-term analysis that could further support the insights we observed through this analysis. Furthermore, a horizontal comparison with data from other countries or other pandemic periods might also provide us with a deeper understanding of the stock market trends associated with work conditions and how rumors and misinformation can affect the performance of workers.

References

- [1] A. Belzunegui-Eraso and A. Erro-Garcés, “Teleworking in the context of the covid-19 crisis,” *Sustainability*, vol. 12, no. 9, p. 3662, 2020.
- [2] L.-P. Béland, A. Brodeur, and T. Wright, “The short-term economic consequences of covid-19: exposure to disease, remote work and government response,” 2020.
- [3] G. Gallacher and I. Hossain, “Remote work and employment dynamics under covid-19: Evidence from canada,” *Canadian public policy*, vol. 46, no. S1, pp. S44–S54, 2020.
- [4] C. Biriowu and T. Chris-Biriowu, “Problem-reaction-solution thesis and covid-19: the impact of public opinion formation on global power, economic and employment relations,” *IMPACT: International Journal of Research in Humanities, Arts and Literature (IMPACT: IJRHAL)*, vol. 8, no. 5, pp. 41–48, 2020.
- [5] A. G. H. Limited, “DingTalk.” <https://www.dingtalk.com>, 2021.
- [6] M. Sigala, “Tourism and covid-19: Impacts and implications for advancing and resetting industry and research,” *Journal of business research*, vol. 117, pp. 312–321, 2020.
- [7] Y. Sun, M. Wu, X. Zeng, and Z. Peng, “The impact of covid-19 on the chinese stock market: Sentimental or substantial?,” *Finance Research Letters*, vol. 38, p. 101838, 2021.
- [8] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.
- [9] B. Lin, “DXY-COVID-19-Data: COVID-19/2019-nCoV time series infection data warehouse,” 2020.
- [10] C. f. S. S. Johns Hopkins University and Engineering, “Coronavirus Resource Center - Time Series Summary (csse_covid_19_time_series),” 2020.
- [11] Baostock, “A tool for obtaining historical data of China stock market,” 2019.
- [12] Dingxiangyuan, “Refutation Rumor dataset,” 2020.
- [13] R. Liao, “China Roundup: Amid Coronavirus, tech firms offer ways to maintain China’s lifeblood,” 2020.
- [14] “Kuchuan.” <https://www.kuchuan.com>, 2021.
- [15] J. P. McLaurin, “Lessons From China: What Hybrid Work Really Looks Like in Practice,” 2021.
- [16] N. P. Monteiro, O. R. Straume, and M. Valente, “Does remote work improve or impair firm labour productivity? longitudinal evidence from portugal,” 2019.
- [17] S. Tasnim, M. M. Hossain, and H. Mazumder, “Impact of rumors and misinformation on covid-19 in social media,” *Journal of preventive medicine and public health*, vol. 53, no. 3, pp. 171–174, 2020.
- [18] A. Abdoli, “Gossip, rumors, and the covid-19 crisis,” *Disaster Medicine and Public Health Preparedness*, pp. 1–2, 2020.
- [19] R. Krestel, P. Fankhauser, and W. Nejdl, “Latent dirichlet allocation for tag recommendation,” in *Proceedings of the third ACM conference on Recommender systems*, pp. 61–68, 2009.
- [20] “Jieba : Chinese Words Segmentation Utilities Python Package.” <https://pypi.org/project/jieba>, 2021.
- [21] N. Fernandes, “Economic effects of coronavirus outbreak (covid-19) on the world economy. 2020,” Available at SSRN.
- [22] S. Ramelli and A. Wagner, “What the stock market tells us about the consequences of covid-19,” *Mitigating the COVID Economic Crisis: Act Fast and Do Whatever*, p. 63, 2020.
- [23] H. Liu, Y. Wang, D. He, and C. Wang, “Short term response of chinese stock markets to the outbreak of covid-19,” *Applied Economics*, vol. 52, no. 53, pp. 5859–5872, 2020.
- [24] A. Zarembo, D. Y. Aharon, E. Demir, R. Kizys, and D. Zawadka, “Covid-19, government policy responses, and stock market liquidity around the world: A note,” *Research in International Business and Finance*, vol. 56, p. 101359, 2021.
- [25] G. Carey, “Multivariate analysis of variance (manova): I. theory,” *Retrieved May*, vol. 14, p. 2011, 1998.
- [26] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*, vol. 112. Springer, 2013.
- [27] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
- [28] K. Liu, “Covid-19 and the chinese economy: impacts, policy responses and implications,” *International Review of Applied Economics*, vol. 35, no. 2, pp. 308–330, 2021.
- [29] W. Tian, “How china managed the covid-19 pandemic,” *Asian Economic Papers*, vol. 20, no. 1, pp. 75–101, 2021.
- [30] W. Wu, C.-C. Lee, W. Xing, and S.-J. Ho, “The impact of the covid-19 outbreak on chinese-listed tourism stocks,” *Financial Innovation*, vol. 7, no. 1, pp. 1–18, 2021.
- [31] G. Salton and C. T. Yu, “On the construction of effective vocabularies for information retrieval,” *Acm Sigplan Notices*, vol. 10, no. 1, pp. 48–60, 1973.
- [32] M. Vasiev, K. Bi, A. Denisov, and V. Bocharnikov, “How covid-19 pandemics influences chinese economic sustainability,” vol. 14, no. 2 (eng), pp. 7–22, 2020.
- [33] A. N. Khan, “A diary study of psychological effects of misinformation and covid-19 threat on work engagement of working from home employees,” *Technological Forecasting and Social Change*, p. 120968, 2021.

- [34] W. A. Peterson and N. P. Gist, "Rumor and public opinion," *American Journal of Sociology*, vol. 57, no. 2, pp. 159–167, 1951.
- [35] Y. Wei, Z. Ye, M. Cui, and X. Wei, "Covid-19 prevention and control in china: grid governance," *Journal of Public Health*, vol. 43, no. 1, pp. 76–81, 2021.