

An Investigation of Containment Measures Against the COVID-19 Pandemic in Mainland China

Ji Liu^{1†}, Xiakai Wang^{1†}, Haoyi Xiong¹, Jizhou Huang¹, Siyu Huang¹, Haozhe An¹, Dejing Dou^{1*}, Haifeng Wang¹

[†] These authors contributed equally to this work.

* doudejing@baidu.com

¹ Baidu Inc., Beijing, 100085, China

Abstract—As the recent COVID-19 outbreak rapidly expands all over the world, various containment measures have been carried out to fight against the COVID-19 pandemic. In Mainland China, the containment measures consist of three types, i.e., Wuhan travel ban, intra-city quarantine and isolation, and inter-city travel restriction. In order to carry out the measures, local economy and information acquisition play an important role. In this paper, we investigate the correlation of local economy and the information acquisition on the execution of containment measures to fight against the COVID-19 pandemic in Mainland China. First, we use a parsimonious model, i.e., SIR-X model to estimate the parameters, which represent the execution of intra-city quarantine and isolation in major cities of Mainland China. In order to understand the execution of intra-city quarantine and isolation, we analyze the correlation between the representative parameters including local economy, mobility, and information acquisition. To this end, we collect the data of Gross Domestic Product (GDP), the inflows from Wuhan and outflows, and the COVID-19 related search frequency from a widely-used Web mapping service, i.e., Baidu Maps, and Web search engine, i.e., Baidu Search Engine, in Mainland China. Based on the analysis, we confirm the strong correlation between the local economy and the execution of information acquisition in major cities of Mainland China. We further evidence that, although the cities with high GDP per capita attract more inflows from Wuhan, people are more likely to conduct the quarantine measure and to reduce travelling to other cities. Finally, the correlation analysis using search data shows that well-informed individuals are likely to carry out containment measures.

Index Terms—Covid-19, SIR model, containment, pearson correlation, MCMC

I. INTRODUCTION

Since December 2019, novel coronavirus COVID-19 has been identified and the outbreak has expanded rapidly throughout various countries, e.g., China [1], United States [2], European countries [3], etc. In China, the number of confirmed cases increased from 571 on January 23, 2020 to 84,388 on May 1, 2020 and saturated around 84.5 thousand. The COVID-19 has become a global emergency and is currently spreading throughout the whole world [4], [5]. In order to deal with the rapid outbreak of the COVID-19 pandemic in Mainland China, a range of containment measures have been put in place by Chinese authorities [6]–[8]. Similar containment measures have been adopted in major countries all over the world [9], [10].

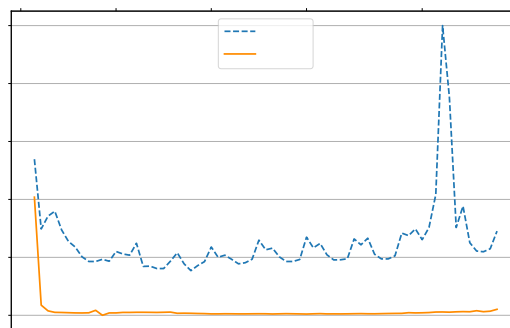


Fig. 1: The comparisons of normalized outflows from Wuhan between 2020 and 2019 in Mainland China. We believe the peak on April 4, 2019 is due to the vacation of Qingming Festival.

In Mainland China, the containment measures consist of intra-city and inter-city measures. For intra-city measures, suspected and confirmed cases have been quarantined in hospitals or monitored self-quarantine at home [11], which is denoted the “quarantine” measure in the paper. The authorities also encouraged citizens to stay-at-home, discouraged mass gatherings, and closed schools [12]. In addition, Wuhan city travel ban was adopted, i.e., all transport was prohibited in and out of Wuhan city from 10:00 a.m. on 23 January 2020, which incurred a significant reduction of the outflows from Wuhan as shown in 1. Inflows and outflows represent counts of the number of people entering and leaving a city. As shown in Figure 2, the national spring vacation has been prolonged and inter-city travel has been discouraged to reduce massive human migration across cities in order to reduce infection.

Mobile applications, e.g., Baidu Migration¹ and search engines, e.g., Baidu², can be easily used to achieve information acquisition for citizens to keep informed during the outbreak of COVID-19. There are a great number of studies [1], [13]–[19] that demonstrated the feasibility to leverage mobile applications for information acquisition. As a result, the history search records can reflect the information acquisition status and the

¹Baidu Migration - <http://qianxi.baidu.com/>

²Baidu - <https://www.baidu.com/>

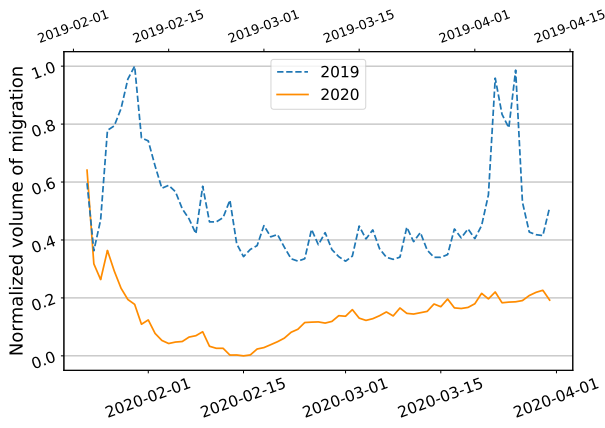


Fig. 2: The comparisons of normalized volumes of the volume of migration between 2020 and 2019 in Mainland China. We believe the peaks between April 4, 2019 and April 7, 2019 are due to the vacation of Qingming Festival.

history statistical migration data can be used to analyze the status of the COVID-19 outbreak. As well-informed individuals are likely to travel less when there is COVID-19 [20], it is interesting to analyze the correlation between information acquisition and the execution of the containment measures. In addition, the correlation between the local economy and the information acquisition can be used to reveal how people in different economic situations react to the COVID-19 pandemic.

In this work, we aim at using a parsimonious model, i.e., SIR-X model, and Markov Chain Monte Carlo (MCMC) [21] methods to estimate the parameters of the execution of intra-city containment measures in major cities of Mainland China. Then, we analyze the correlation among different random variables, i.e., information acquisition status (COVID-19-related search frequency), local economic situation (GDP per capita) and the parameters in the SIR-X model, in order to understand the relationship among economy, information acquisition and the execution of containment measures. More specifically, we would like to investigate following problems:

- *How to construct a model to fit the number of confirmed cases?* This research issue has been studied [7], [11] at province scale using the migration scale index released by Baidu Migration Open Data, where the index is calculated based on the past historical statistical records from a widely-used Web mapping service, i.e., Baidu Maps³. We propose to construct an SIR-X model based on the confirmed cases at city scale using a Markov Chain Monte Carlo (MCMC) method. We provide solid results using exact figures for major Chinese cities.
- *To what degree does the local economy affect the pandemic outbreaks of COVID-19 and the execution of containment measures in major cities of China?* The impact of the COVID-19 pandemic on economy has been studied [22], [23] while the correlations between local economy and the outbreaks of the COVID-19 pandemic or the execution of

containment measures are not analyzed. We hypothesize that the outflows from Wuhan tend to go to the cities where GDP per capital is high. We further hypothesized that there would be more initial confirmed cases as more infected people arrived at these cities. In addition, we hypothesized that the people in the cities where GDP per capital is high tend to perform more information acquisition activities through voluntary COVID-19-related search in order to be well informed on the situation of the COVID-19 pandemic. We analyze the correlations based on the estimated parameters of SIR-X and the statistical data from Baidu Maps and provide solid results for major Chinese cities.

- *To what degree does the information acquisition affect the execution of containment measures in major cities of China?* Strong positive correlation between the pandemic outbreaks of COVID-19 and the information acquisition has been reported in [20] while the correlation between the information acquisition and the execution of containment measures are not analyzed. As reported in [20] and was seen in the collective responses to the emergencies [24], [25] and panics [13], people voluntarily acquire information more frequently when the pandemic situations become worse in their cities. We hypothesized the well-informed people tend to apply the containment measures more strictly in order to avoid the risk to be infected and the risk to make the situation worse. We analyze this correlation based on the estimated parameters of SIR-X and the statistical data from Baidu Maps and Baidu Search Engine, and provide solid results for major Chinese cities as well.

Different from existing research [1], [6], [7], [11], we particularly analyze the correlation between local economy strength and the COVID-19-related search frequency with the city population size (a controlling variable) removed in order to avoid the impact of the scale of city. Compared to the existing work [20], [22], we analyze the correlations not only based on the data from Baidu Maps and Baidu Search Engine but also based on the combination of an SIR-X model and MCMC methods.

II. MODELING EPIDEMIC SPREAD WITH CONTAINMENT MEASURES

In this section, we first present the existing models to capture COVID-19. Then, we propose using SIR-X and MCMC to construct the model. Afterwards, we present the comparison between the official number and the model fitting number of accumulated confirmed cases.

Susceptible Infectious Recovered (SIR) model [26], [27] and Susceptible Exposed Infectious Recovered (SEIR) model [28]–[30] are largely adopted to characterize the outbreak of COVID-19 epidemic. However, the containment measures cannot be described in the standard SIR or SEIR model. A modified SEIR model [31] is proposed with the consideration of mobility while it is still not able to infer the execution of containment measures. A Long-Short-Term-Memory (LSTM)

³Baidu Maps - <https://map.baidu.com/>

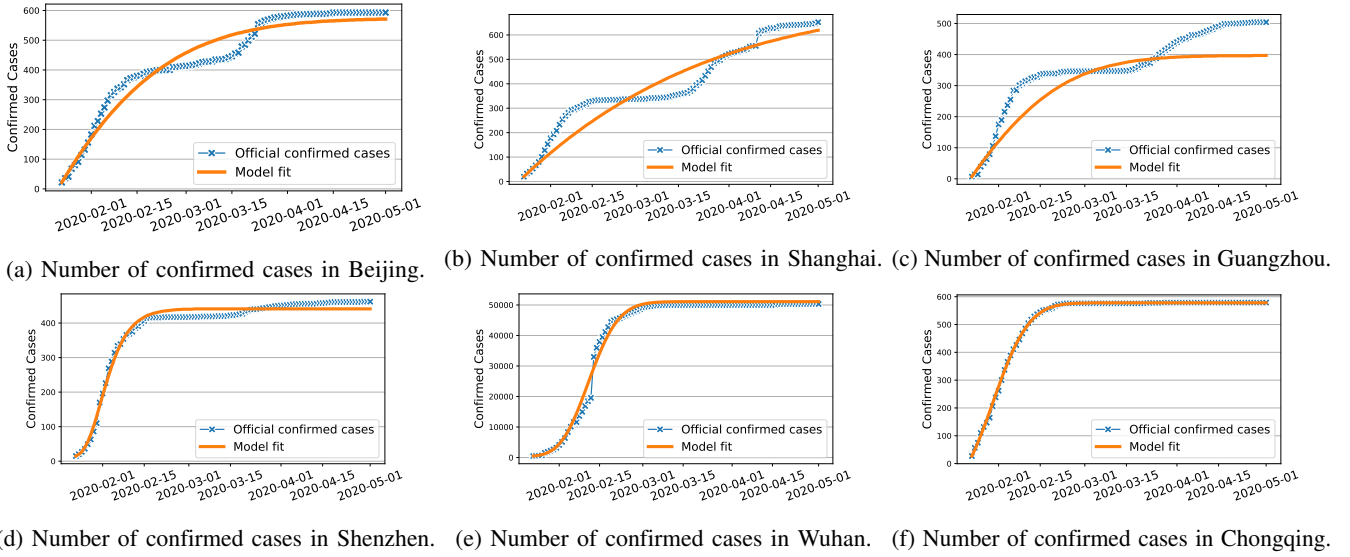


Fig. 3: Comparison between official number of confirmed cases and fitting number from the SIR-X model on May 1, 2020 in Beijing, Shanghai, Guangzhou, Shenzhen, Wuhan and Chongqing.

[31] model is proposed to project the number of accumulated confirmed cases, which is not able to describe the execution of containment measures either.

In order to characterize the outbreak of COVID-19 epidemic with containment measures at city level, we exploit the SIR-X model [11]. The SIR-X model is a modified SIR model, which takes the containment measures into consideration. We have the same assumptions and use the same representative parameters as those in [11]. We assume that there are public containment efforts, e.g., stay-at-home, reduced interaction with other people, which is referred as ‘containment’ and represented by a variable κ_0 . In addition, we assume that infected individuals are quarantined, which is referred as ‘quarantine’ and represented by a variable κ . We use α to represent the infection speed of an infected individual and β^{-1} to represent the average time an infected individual remains infectious before recovery or removal. Then, the SIR-X model is expressed by the following differential equations:

$$\begin{aligned} \partial_t S &= -\alpha SI - \kappa_0 S \\ \partial_t I &= \alpha SI - \beta I - \kappa_0 I - \kappa I \\ \partial_t R &= \beta I + \kappa_0 S \\ \partial_t X &= (\kappa + \kappa_0) I \end{aligned} \quad (1)$$

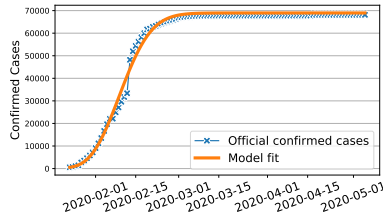
Instead of fixing the same parameters (α and β) for each province in [11], we estimated the parameters using a MCMC [21] method, inspired by [32]. In the model, I_0 represents the number of initial infected individuals. The basic reproduction number R_0 represents the average number of secondary infections an infected will cause before he or she recovers or is removed [11]. The reproduction number can be calculated as: $R_0 = \frac{\alpha}{\beta + \kappa + \kappa_0}$. We use $R_{0,free}$ to represent the reproduction number without containment or quarantine measures. As high temperature and high humidity significantly reduce the transmission of COVID-19 [33], $R_{0,free}$ and β may

be different for different cities because of the diversity of local environments [34]. Thus, we use the MCMC [21] method to estimate the distribution of the parameters, i.e., α , β , κ , κ_0 , I_0 , while the other parameters are fixed (S_0 is the population in the city, R'_0 is fixed as 0 and X_0 is the number of initial confirmed cases) at the beginning (January 23, 2020) with S_0 , R'_0 , X_0 and I_0 representing the initial values of S , I , R and X .

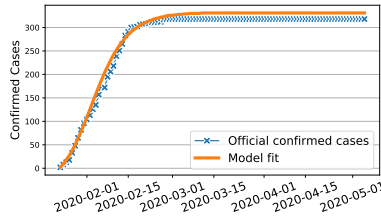
Specifically, we use the uniform distribution as the parameters’ prior distribution. And with the consideration of the nonlinearity of SIR-X model, we adopt the Sequential Monte Carlo sampler to achieve the posterior distribution of model’s parameters including α and β . Finally, we take the expected value of each parameter to construct the model.

In order to have stable results from the MCMC method, we use a priori conditions, i.e., $R_0 < R'_{0,free}$ and $\kappa_0 < \kappa$. The results of MCMC methods may not be stable, i.e., the results of each execution may be different without a priori conditions. Thus, we introduce a priori conditions, i.e., $R_0 < R'_{0,free}$, $\kappa_0 < \kappa$, and the model fit number of accumulated confirmed cases should be equal or bigger than the official number of confirmed cases. $R'_{0,free}$ represents a maximum value of R_0 . We set $R'_{0,free}$ as 6.2, which is in accordance with the result from [11] that the R_0 should be between 1.4 and 3.3. During the fitting process, if the a priori conditions are not met, the fitting process will be executed again until reaching a limit, e.g., 20 times of execution, in order to avoid infinite execution. We assume that the quarantine measure is applied more strictly on the infected individuals than other public citizens, i.e., $\kappa_0 < \kappa$.

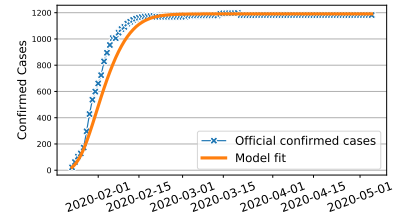
In order to use the SIR-X model, we need to assume that few travelers and symptomatic infected individuals travel into or from a city. As reported in [20], there is strong correlation between the inflows from Wuhan and the confirmed cases in a city. We assume that few infected individuals travelled into major cities after January 23, 2020 as Wuhan travel ban



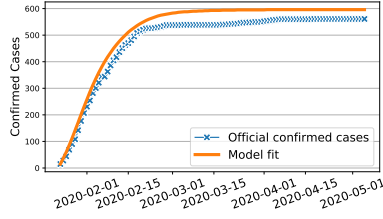
(a) Aggregated number of confirmed cases in Hubei.



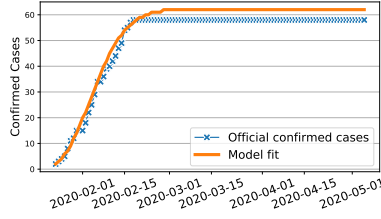
(b) Aggregated number of confirmed cases in Hebei.



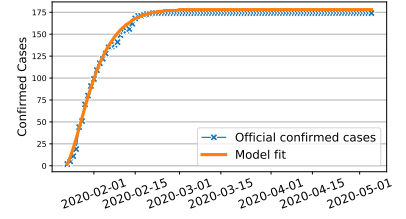
(c) Aggregated number of confirmed cases in Zhejiang.



(d) Aggregated number of confirmed cases in Sichuan.



(e) Aggregated number of confirmed cases in Xingjiang.



(f) Aggregated number of confirmed cases in Yunnan.

Fig. 4: Comparison between official number of confirmed cases and fitting number from the SIR-X model on May 1, 2020 in Hubei, Hebei, Zhejiang, Sichuan, Xingjiang and Yunnan.

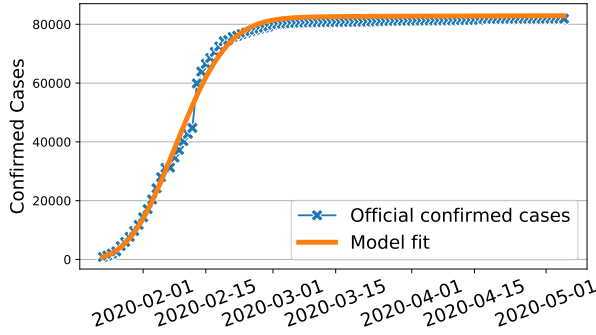


Fig. 5: Comparison between official number of confirmed cases and fitting number from the SIR-X model on May 1, 2020 in China.

has been put in place since January 23, 2020 and few people went to other cities from Wuhan as shown in Figure 1. In addition, we assume that the number of infected individuals in the inflows from other cities can be ignored compared to the number of infected individuals among the local citizens in a city. With these two assumptions, we can use the SIR-X and MCMC to estimate parameters for each major cities in Mainland China based on the number of confirmed cases⁴ from January 23, 2020 to May 1, 2020.

Figures 3a - 3f illustrate the confirmed cases in several major cities of Mainland China. From figures, we can see that the combination of SIR-X and MCMC captures the number of confirmed cases in different cities very well, e.g., Beijing, Shanghai, Shenzhen, Wuhan and Chongqing. However, the model does not well characterize the number of confirmed cases in Guangzhou as there are many (127⁵) infected individuals

⁴COVID-19 statistics - <https://github.com/canghailan/Wuhan-2019-nCoV>

⁵Confirmed cases in Guangzhou from National Health Commission - http://wjw.gz.gov.cn/ztzl/xxfyyqfk/yqtb/content/post_5815637.html

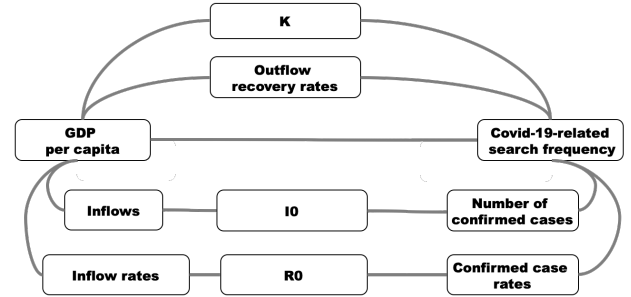


Fig. 6: Significant correlation among different factors.

from other countries, which cannot be captured by the SIR-X model. In addition, we believe that the errors between the confirmed cases and fitted data are mainly due to the travelers from other countries in Beijing (174 confirmed cases from other countries) and Shanghai (326 confirmed cases from other countries). Then, we calculate the confirmed cases of different provinces by adding the number of confirmed cases in each affiliated city. Figures 4a - 4f shows the number of confirmed cases in several provinces of Mainland China. We can see that the SIR-X well captures the aggregated cases at province scale. Furthermore, we use the same method to calculate the number of confirmed cases in Mainland China as shown in Figure 5.

III. CORRELATION STUDY

Besides the number of confirmed cases (May 1, 2020), we collected three datasets, i.e., GDP, mobility and COVID-19-related search frequency for major cities in Mainland China. The GDP dataset was collected from [35], which characterized local economic development in 2019. The mobility data was captured from Baidu Maps and the COVID-19-related search frequency (the ratio between COVID-19-related search volume from January to March 2020 and population in each city) data was gathered from a widely-used Web search engine.

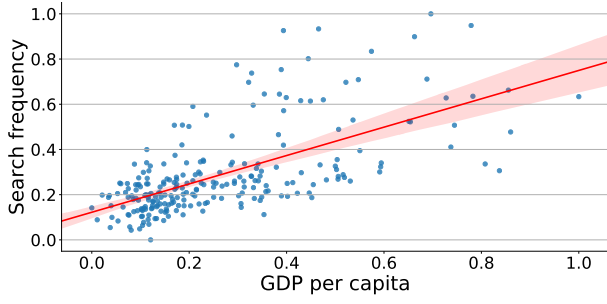


Fig. 7: Significant positive correlation has been found between GDP per capita and COVID-19-related search frequency for major cities in Mainland China. Shaded area represents the 95% Confidence Interval (CI).

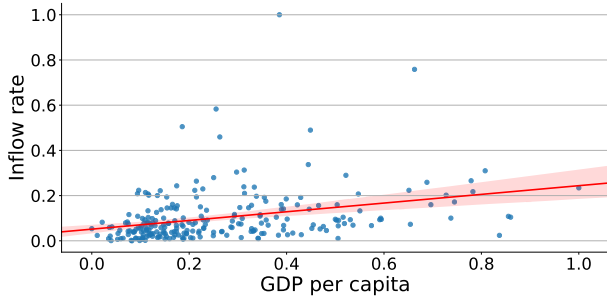


Fig. 8: Significant positive correlations have been found between GDP per capita and inflow rate from Wuhan. Shaded area represents the 95% CI.

We analyze the correlation among local economy, mobility, search behaviors and the parameters estimated based on the combination of the SIR-X model and the MCMC method as presented in Section II for 238 cities in Mainland China (excluding Wuhan). We normalize inflow, outflow, search volume by the following formula:

$$\text{Normalize}(\text{data}) = \frac{\text{data} - \text{data}_{\min}}{\text{data}_{\max} - \text{data}_{\min}}. \quad (2)$$

In this way, the data in the study are curved into the range from 0 to 1 proportionally. The results of our data-driven analysis are summarized as shown in Figure 6. In this section, we present the observations obtained from the analysis.

A. Significant positive correlations have been evidenced between local economy and COVID-19-related search frequency

We have evidenced the significant positive correlations between local economy and COVID-19-related search frequency for major Chinese cities. In order to analyze the correlation between two random variables, we calculated the Pearson correlation coefficients [36] and conducted the Student's T-test (two tails) to verify the significance test (the same for the following analysis in the paper). The Pearson correlation between the local GDP per capita and the total COVID-19-related search volume (between January and March 2020) is $R^{***} = 52.5\%$ ($N = 238$ and $p\text{-value} = 3.06 \times 10^{-18} < 0.0001$) for each city.

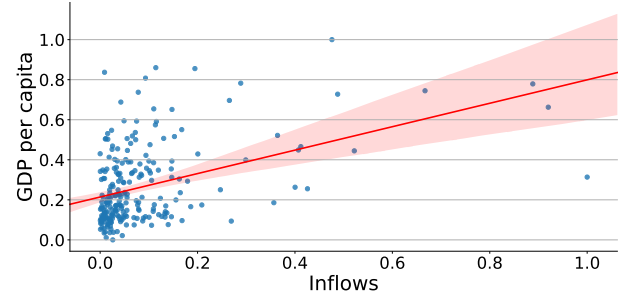


Fig. 9: Significant positive correlations have been found between GDP per capita and inflows from Wuhan. Shaded area represents the 95% CI.

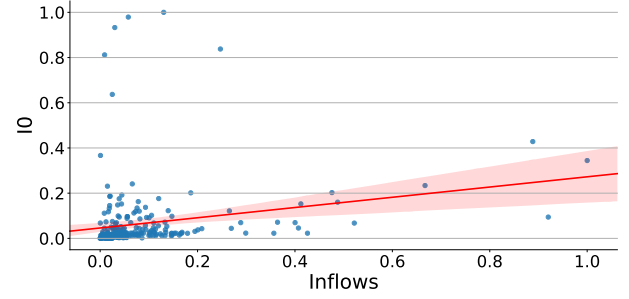


Fig. 10: Significant positive correlations have been found between the inflows from Wuhan and I_0 . Shaded area represents the 95% CI.

However, we considered that this observation can be incurred by the scale of the city, as a larger city would have larger population, and would correspond to bigger GDP and high total COVID-19-related search volume.

We therefore tested the significance of the correlations between GDP per capita and COVID-19-related search frequency, where we evidenced the significance in the correlations as $R^{***} = 63.5\%$ ($N = 238$ and $p\text{-value} = 2.91 \times 10^{-28} < 0.0001$). In addition, in order to obviate the impact of the scale of the city, i.e., the impact of city population, we conducted partial correlation analysis [37], [38] between GDP per capita and the COVID-19-related search frequency with the effects of the city population size (a controlling variable) removed. In order to estimate the partial correlation of random variables X and Y with the random variable Z removed, we expressed the partial correlation coefficient in terms of the Pearson correlation coefficients as

$$\rho(X, Y|Z) \equiv \frac{\rho(X, Y) - \rho(X, Z)\rho(Y, Z)}{\sqrt{(1 - \rho^2(X, Z))(1 - \rho^2(Y, Z))}}. \quad (3)$$

We find a strong correlation with significance as well, such that $R^{***} = 57.0\%$ ($N = 238$ and $p\text{-value} = 7.15 \times 10^{-22} < 0.0001$). Thus, we can conclude that no matter whether the scale of the city is big or small, GDP per capita has a significant positive correlation with the COVID-19-related search frequency. Please see also Figure 7 for the visualization of the correlations.

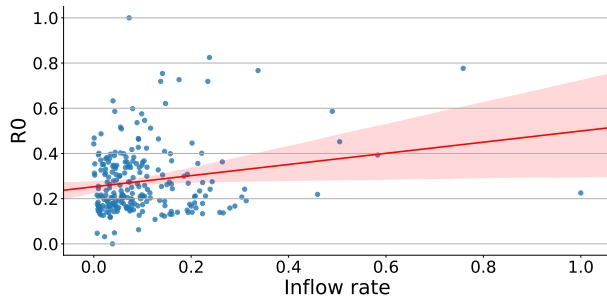


Fig. 11: Significant positive correlations have been found between the inflow rate from Wuhan and R_0 . Shaded area represents the 95% CI.

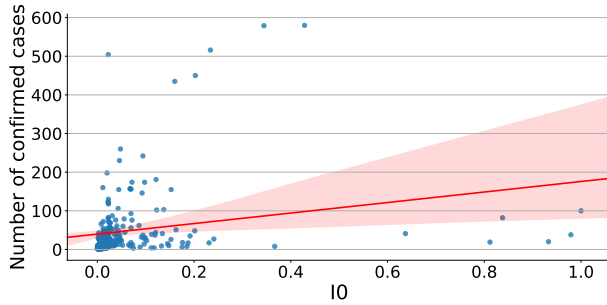


Fig. 12: Significant positive correlations have been found between I_0 and the number of confirmed cases. Shaded area represents the 95% CI.

B. Correlation analysis for the spread of the COVID-19 pandemic

We have evidenced the significant positive correlation between GDP per capita and inflows from Wuhan. We hypothesized that cities with higher GDP per capita would attract larger inflows from Wuhan. Therefore, for every city in the study, we correlated GDP per capita and the inflows from Wuhan, where we obtained Pearson correlation coefficients of $R^{***} = 42.3\%$ ($N = 238$ and $p\text{-value} = 9.88 \times 10^{-12} < 0.0001$). In addition, in order to obviate the impact of the scale of the city, we correlated GDP per capita and the inflows rate from Wuhan, i.e., the ratio between inflows from Wuhan and the population. We found a strong positive correlation with significance as well, such that $R^{***} = 32.6\%$ ($N = 238$ and $p\text{-value} = 2.67 \times 10^{-7} < 0.0001$). Please see also Figures 8 and 9 for the visualization of the correlations.

We have evidenced the significant positive correlation between I_0 and the inflows and the significant positive correlation between R_0 and the inflow rate. We hypothesized that cities with larger inflows from Wuhan have more initial infected cases, i.e., I_0 in the SIR-X model. Thus, we correlated the inflows from Wuhan and I_0 , where we found a strong positive correlation with significance, such that $R^{**} = 21.6\%$ ($N = 238$ and $p\text{-value} = 8.11 \times 10^{-4} < 0.001$). In addition, we analyzed the correlation between the inflow rate and R_0 , where we found a strong positive correlation with significance, such

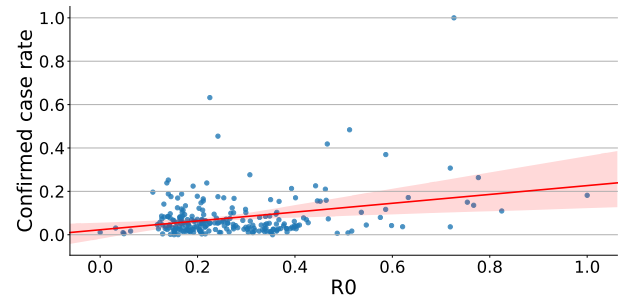


Fig. 13: Significant positive correlations have been found between R_0 and confirmed case rate. Shaded area represents the 95% CI.

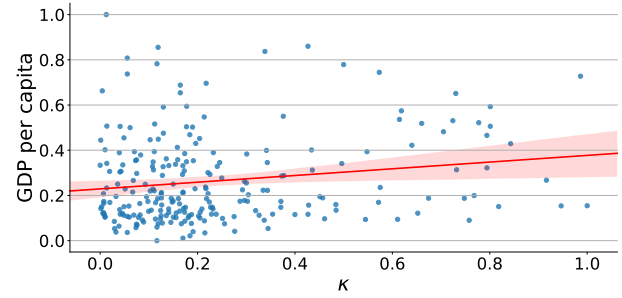


Fig. 14: Significant positive correlations have been found between GDP per capita and κ . Shaded area represents the 95% CI.

that $R^* = 19.2\%$ ($N = 238$ and $p\text{-value} = 2.93 \times 10^{-3} < 0.01$). Please see also Figures 10 and 11 for the visualization of the correlations.

We have evidenced the significance of the positive correlation between the number of initial infected individuals and the number of confirmed cases and the positive correlation between R_0 and the number of confirmed case rate. We hypothesized that cities with bigger I_0 finally have more confirmed cases. To this end, we performed correlation analysis using I_0 and the number of confirmed cases on May 1, 2020. We found a significant positive correlation, such that $R^{**} = 22.5\%$ with $N = 238$ and $p\text{-value} = 4.67 \times 10^{-4} < 0.001$. In addition, we hypothesized that cities with bigger R_0 have more confirmed case rate, i.e., the ratio between the confirmed cases and the population. We performed correlation between R_0 and the confirmed case rate, where we obtained a significant positive correlation, such that $R^{***} = 29.8\%$ with $N = 238$ and $p\text{-value} = 2.82 \times 10^{-6} < 0.0001$. Please see also Figures 12 and 13 for the visualization of the correlations.

We obtained a strong positive correlation with significance between the number of confirmed cases and the COVID-19-related search frequency, such that $R^{***} = 41.5\%$ ($N = 238$ and $p\text{-value} = 2.45 \times 10^{-11} < 0.0001$). Furthermore, we obtained a strong positive correlation with significance between confirmed case rate and the COVID-19-related search frequency, such that $R^{**} = 21.4\%$ ($N = 238$ and $p\text{-value} = 9.01 \times 10^{-4} < 0.001$). Similar results are also reported in [20].

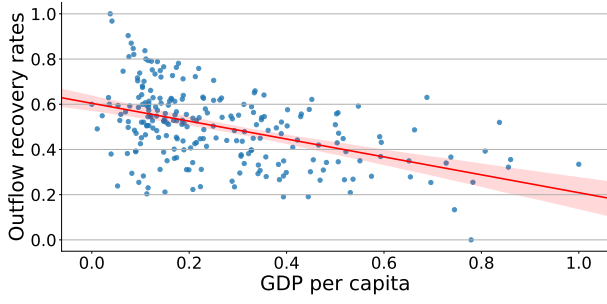


Fig. 15: Significant negative correlations have been found between GDP per capita and the outflow recovery rate. Shaded area represents the 95% CI.

We thus can conclude that for each major city in the study, GDP per capita and the factors that incur infections, e.g., inflows from Wuhan, I_0 , R_0 , have significant positive correlation. This indicates that the rich cities attracted more inflows from Wuhan, which caused infections and in order to fight against COVID-19, the citizens in the rich cities tend to perform more search activities in order to be well-informed.

C. Correlation analysis for the execution of containment measures

In this section, we analyze the correlation among local economy, information acquisition and containment measures. As the quarantine measure (see details in Section II) is directly related to the number of confirmed cases, we analyze the correlation between κ and other factors (local economy and information acquisition). In addition, we are also interested in the realization of inter-city containment measures, i.e., the outflow recovery rate (the ratio between the outflows of 2020 and that of 2019).

We have evidenced the significance of the positive correlation between the execution of the quarantine measure and GDP per capita for major Chinese cities in the study.

Among all 238 cities in the correlation study, we hypothesized that people with high GDP per capita would try harder to implement the quarantine measure for infected individuals. Therefore, we correlated GDP per capita and κ , where Pearson correlation coefficients are $R^* = 17.3\%$ ($N = 238$ and $p\text{-value} = 7.46 \times 10^{-3} < 0.01$). Furthermore, we correlated GDP per capita and the outflow recovery rate, where we obtained Pearson correlation coefficients of $R^{***} = -46.5\%$ ($N = 238$ and $p\text{-value} = 3.82 \times 10^{-14} < 0.0001$). The correlation analysis result suggests that people with higher GDP per capita are more likely to apply the quarantine measure. Please see also Figures 14 and 15 for the visualization of the correlations.

We have evidenced the significance of the positive correlations between the realization of quarantine measure and the COVID-19-related search frequency. We performed correlation using the COVID-19-related search frequency and κ . We found a significant positive correlation, such that $R^* = 17.8\%$ with $N = 238$ and $p\text{-value} = 5.84 \times 10^{-3} < 0.01$. Please see also Figures 16 for the visualization of the correlations. We

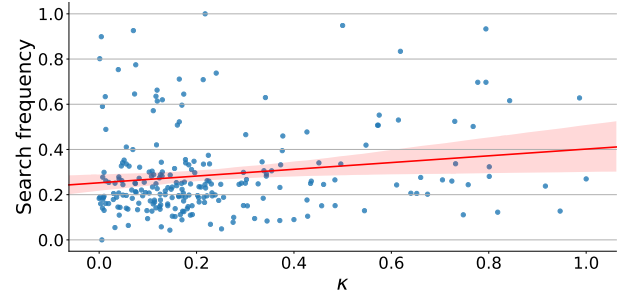


Fig. 16: Significant positive correlations have been found between the COVID-19-related search frequency and κ . Shaded area represents the 95% CI.

found that the negative correlations between the COVID-19-related search frequency and the outflow recovery rate are stronger with $R^{***} = -51.4\%$ ($N = 238$ and $p\text{-value} = 1.88 \times 10^{-17} < 0.0001$) (similar result is also reported in [20]). The correlation analysis result suggests that people with higher per capita COVID-19-related search frequency are more likely to apply the containment measures, i.e., separate the infected individuals and small outflow recovery rate.

We can conclude that for every city in the study GDP per capita and the COVID-19-related search frequency have significant positive correlation to the realization of containment measures. We believe it is due to the will to avoid the risk to be infected and the natural response to the fear and massive panics [13], [20]. In addition, the reason also goes to the fact that the people in the cities of higher GDP per capita tend to have bigger capacity or more tolerance to apply the containment measures.

IV. CONCLUSION

In this work, we first exploit the SIR-X model and MCMC method to estimate the parameters related to the COVID-19 pandemic at the scale of city. Then, we examined the correlation between the local economy and the spread of COVID-19 pandemic and the execution of containment measures in major cities of Mainland China. We conducted correlation analysis based on the mobility data and search data from Baidu Maps and Baidu Search Engine in Mainland China. Our analysis brings novel knowledge of the correlation among different factors related to the COVID-19 pandemic. The cities of higher GDP per capita attracts bigger inflows from Wuhan, which cause more confirmed cases. However, the demands of information from individuals become higher, which incurs the reaction to apply the containment measure. Furthermore, well-informed individuals are more likely to apply the intra- and inter- city containment measures, i.e., quarantine of infected individuals and reducing going to other cities. The implications of these correlations include that, the better the local economy is and the more timely information is attained by residents, the better the containment measures are realized, which help to fight against the COVID-19 pandemic in major cities of Mainland China.

ACKNOWLEDGEMENT

To protect Baidu users' data privacy, all experiments in this paper were carried out using anonymous data and secure data analytics provided by Baidu Data Federation Platform (Baidu FedCube). For data accesses and usages, please contact us via {fedcube, shubang}@baidu.com.

AUTHOR CONTRIBUTIONS STATEMENT

J. Liu formulated the research problems and drafted the manuscript. J. Liu, X. Wang and J. Huang collected data from Baidu, conducted the experiments, and performed data analysis. S. Huang and H. An collected the consensus data and carried out the data visualization. X. Wang, J. Huang, H. Xiong and D. Dou revised the whole paper. D. Dou and H. Wang proposed the research, coordinated the research efforts, and oversaw the whole research process.

REFERENCES

- [1] M. Ahinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun *et al.*, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, no. 6489, pp. 395–400, 2020.
- [2] J. H. University, "Johns Hopkins University Coronavirus Resource Center," <http://https://coronavirus.jhu.edu/>, 2020, [Online; accessed 18-May-2020].
- [3] A. Deslandes, V. Berti, Y. Tandjaoui-Lambotte, C. Alloui, E. Carbonnelle, J. Zahar, S. Brichler, and Y. Cohen, "SARS-CoV-2 was already spreading in france in late december 2019," 2020.
- [4] The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China," *China CDC Weekly*, vol. 41, no. 2, p. 145, 2020.
- [5] C. Sohrabi, Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *International Journal of Surgery*, vol. 76, pp. 71–76, 2020.
- [6] M. U. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage *et al.*, "The effect of human mobility and control measures on the COVID-19 epidemic in China," *Science*, vol. 368, no. 6490, pp. 493–497, 2020.
- [7] H. Tian, Y. Liu, Y. Li, C.-H. Wu, B. Chen, M. U. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang *et al.*, "An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China," 2020.
- [8] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing," 2020.
- [9] S. Engle, J. Stromme, and A. Zhou, "Staying at Home: Mobility Effects of COVID-19," 2020.
- [10] A. Remuzzi and G. Remuzzi, "Covid-19 and italy: what next?" *The Lancet*, vol. 395, no. 10231, pp. 1225 – 1228, 2020.
- [11] B. F. Maier and D. Brockmann, "Effective containment explains subexponential growth in recent confirmed covid-19 cases in china," *Science*, vol. 368, no. 6492, pp. 742–746, 2020.
- [12] S. Chen, J. Yang, W. Yang, C. Wang, and T. Bärnighausen, "COVID-19 control in china during mass population movements at new year," *Lancet*, vol. 395, no. 10226, pp. 764–766, 2020.
- [13] S.-Y. Ren, R.-D. Gao, and Y.-L. Chen, "Fear can be more harmful than the severe acute respiratory syndrome coronavirus 2 in controlling the corona virus disease 2019 epidemic," *World Journal of Clinical Cases*, vol. 8, no. 4, p. 652, 2020.
- [14] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai *et al.*, "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *Journal of Thoracic Disease*, vol. 12, no. 3, p. 165, 2020.
- [15] H. Li, X. Chen, and H. Huang, "The novel coronavirus outbreak: what can be learned from China in public reporting?" *Global Health Research and Policy*, vol. 5, no. 1, pp. 1–3, 2020.
- [16] S. Chen, J. Yang, W. Yang, C. Wang, and T. Bärnighausen, "COVID-19 control in China during mass population movements at new year," *The Lancet*, vol. 395, no. 10226, pp. 764–766, 2020.
- [17] J. Zhang, G. Lin, J. Zeng, J. Lin, J. Tian, and G. Li, "Challenges of SARS-CoV-2 and lessons learnt from SARS in guangdong province, China," *Journal of Clinical Virology*, vol. 126, p. 104341, 2020.
- [18] H. Chen, Y. Chen, B. Sun, P. Wang, L. Wen, Z. Lian, Y. Lu, Y. Qi, S. Zhao, L. Zhang *et al.*, "Correlation between the migration scale index and the number of new confirmed Novel Coronavirus Pneumonia cases in China," *Virus*, vol. 11, p. 3, 2020.
- [19] P. Zhong, S. Guo, and T. Chen, "Correlation between travellers departing from Wuhan before the Spring Festival and subsequent spread of COVID-19 to all provinces in China," *Journal of Travel Medicine*, vol. 27, no. 3, pp. 1–4, 2020.
- [20] H. Xiong, J. Liu, J. Huang, S. Huang, H. An, Q. Kang, Y. Li, D. Dou, and H. Wang, "Understanding the collective responses of populations to the covid-19 pandemic in mainland china," 2020.
- [21] W. R. Gilks, "Markov chain monte carlo," 2005.
- [22] J. Huang, H. Wang, H. Xiong, M. Fan, A. Zhuo, Y. Li, and D. Dou, "Quantifying the economic impact of covid-19 in mainland china using human mobility data," 2020.
- [23] A. Atkeson, "What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios," National Bureau of Economic Research, Working Paper 26867, 2020.
- [24] L. Gao, C. Song, Z. Gao, A.-L. Barabási, J. P. Bagrow, and D. Wang, "Quantifying information flow during emergencies," *Scientific reports*, vol. 4, p. 3997, 2014.
- [25] J. Liu, T. Huang, H. Xiong, J. Huang, J. Zhou, H. Jiang, G. Yang, H. Wang, and D. Dou, "Analysis of collective response reveals that covid-19-related activities start from the end of 2019 in mainland china," *medRxiv*, 2020.
- [26] A. A. Toda, "Susceptible-Infected-Recovered (SIR) Dynamics of COVID-19 and Economic Impact," 2020.
- [27] A. Simha, R. V. Prasad, and S. Narayana, "A simple Stochastic SIR model for COVID 19 Infection Dynamics for Karnataka: Learning from Europe," 2020.
- [28] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of covid-19 in china by dynamical modeling," 2020.
- [29] C. Hou, J. Chen, Y. Zhou, L. Hua, J. Yuan, S. He, Y. Guo, S. Zhang, Q. Jia, C. Zhao, J. Zhang, G. Xu, and E. Jia, "The effectiveness of quarantine of wuhan city against the corona virus disease 2019 (COVID-19): A well-mixed seir model analysis," 2020.
- [30] S. Tang, B. Tang, N. L. Bragazzi, F. Xia, T. Li, S. He, P. Ren, X. Wang, C. Xiang, Z. Peng, J. Wu, and Y. Xiao, "Analysis of covid-19 epidemic traced data and stochastic discrete transmission dynamic model," 2020.
- [31] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai, J. Liang, X. Liu, S. Li, Y. Li, F. Ye, W. Guan, Y. Yang, F. Li, S. Luo, Y. Xie, B. Liu, Z. Wang, S. Zhang, Y. Wang, N. Zhong, and J. He, "Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions," *Journal of Thoracic Disease*, vol. 12, no. 3, pp. 165–174, 2020.
- [32] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: A modelling study," *Lancet*, vol. 395, no. 10225, p. 689–697, 2020.
- [33] J. Wang, K. Tang, K. Feng, and W. Lv, "High Temperature and High Humidity Reduce the Transmission of COVID-19," *Social Science Research Network (SSRN)*, 2020.
- [34] S. Huang, J. Liu, H. Xiong, J. Huang, H. An, and D. Dou, "The weather impacts the outbreak of covid-19 in mainland china," *arXiv preprint arXiv:2006.10376*, 2020.
- [35] N. B. of Statistics of China, "National Bureau of Statistics of China," <http://www.stats.gov.cn/>, 2019, [Online; accessed 21-May-2020].
- [36] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," *Noise Reduction in Speech Processing*, vol. 2, pp. 1–4, 2009.
- [37] K. Baba, R. Shibata, and M. Sibuya, "Partial correlation and conditional correlation as measures of conditional independence," *Australian & New Zealand Journal of Statistics*, vol. 46, no. 4, pp. 657–664, 2004.
- [38] D. Y. Kenett, X. Huang, I. Vodenska, S. Havlin, and H. E. Stanley, "Partial correlation analysis: applications for financial markets," *Quantitative Finance*, vol. 15, no. 4, pp. 569–578, 2015.