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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Emerging study on the transmission of the Novel Coronavirus (COVID-19) from urban perspective: Evidence from China

Lu Liu

School of Economics, Southwestern University of Finance and Economics, China

ARTICLE INFO

Keywords:
COVID-19
The Novel Coronavirus
Transmission
Epicenter
Cities

ABSTRACT

This study presents an in-depth investigation on the transmission of the novel coronavirus (COVID-19) from the urban perspective. It focuses on the “aftermath” of the outbreak and the spread of the infection among cities. Especially, this study provides insights of the fundamentals of the factors that may affect the spread of the infection in cities, where the marginal effects of some most influential factors to the virus transmission are estimated. It reveals that the distance to epicenter is a very strong influential factor, and is negatively linked with the spread of COVID-19. In addition, subway, wastewater and residential garbage are positively connected with the virus transmission. Moreover, both urban area and population density are negatively associated with the spread of COVID-19 at the early stage of the epidemic. Furthermore, this study also provides high precision estimation of the number of COVID-19 infection in Wuhan city, which is the epicenter of the outbreak in China. Based on the real-world data of cities outside Wuhan on March 2, 2020, the estimated number is 56,944.866 (mean value), which is very close to the officially reported number. The methodology and main conclusions shown in this paper are of general interest, and they can be applied to other countries to help understand the local transmission of COVID-19 as well.

1. Introduction

Right before the 2020 Chinese Spring Festival, the outbreak of the novel coronavirus (COVID-19) takes place in China. Fear of SARS which caused large scale of panic in 2003 appears to come back along with the fast spread of the infection in the Wuhan city, other neighboring cities within Hubei province where Wuhan is the capital city, as well as many other cities outside Hubei province across the whole country (Fig. 1).

As early as late December 2019, the novel coronavirus has been laboratory-confirmed in China (Wang et al., 2020; Zhu et al., 2020). Later, as person-to-person transmission of COVID-19 is also confirmed, public concern of the spread of the novel coronavirus increases dramatically (Chan et al., 2020). It has even caused global health concern as well (Wang et al., 2020). On January 27, 2020, World Health Organization (WHO) remains its assessment of risk of the event as very high in China and high level elsewhere.1 However, on January 30, 2020, WHO raises its assessment to a public health emergency of international concern.2

It is really a bad time for the event since the spectacular Spring Transportation in China brings hundreds of millions of people back home for the Chinese Lunar New Year (Liu, 2014). As the Spring Transportation is considered to be the largest scale of temporary human migration on Earth, this tremendous large number of travelers makes it extremely difficult to contain the spread of the infection within the epicenter of outbreak, i.e., Wuhan city. As announced by the local government in the news express on January 26, 2020, 5 million plus people out of 14 million have left Wuhan before the city is officially sealed off.3 Tremendous efforts have been employed to contain the infection and cure the infected patients, and China has taken various measures to combat the novel coronavirus.4 Although the clinical

---

1 Website of World Health Organization (WHO)
https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200127-sitrep-7-2019–ncov.pdf?sfvrsn=98ef79f5_2

2 WHO
https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-nCov)

3 China Daily
http://www.chinadaily.com.cn/a/202001/27/WS5e2dc01a310128217273551.html

4 People’s Daily.
http://en.people.cn/n3/2020/0130/c90000-9652610.html

https://doi.org/10.1016/j.cities.2020.102759
Received 18 March 2020; Received in revised form 12 April 2020; Accepted 25 April 2020
Available online 01 May 2020
0264-2751/ © 2020 Elsevier Ltd. All rights reserved.
features of patients infected with COVID-19 is relatively mild compared with SARS (Huang et al., 2020), its spread is faster. As of April 30, 2020, the accumulated total number of confirmed patients of the novel coronavirus is 84,385 in China, and the number of death cases is 4,643.

While we observe the evolution of the COVID-19 epidemic in China, we notice substantial heterogeneity in the size of the epidemic of cities across the country. Hence, this study tries to answer some essential questions to the transmission of COVID-19 raised by the reality with the evidences from China. First, why do some cities have more infected cases, while some other cities have less? Second, what is the role of the distances of these cities with confirmed cases to the epicenter of the outbreak of an infectious disease with large scale? Third, what are the lessons of the transmission from the urban characteristics of those cities? Fourth, is there a quantitative model which can predict the actual infection scale of the epicenter precisely with the information of the epidemic in other cities?

In the above questions, we keep mentioning “cities”. In fact, people need to pay more attention to the social and especially urban fundamentals of the transmission of COVID-19, not just the biological features of the virus itself. Indeed, the virus is transmitted among persons in the background of the cities. In this particular perspective, the currently ongoing epidemic of COVID-19 can be treated as an urban incident as well, which causes substantial challenges to the urban management or even urban planning. Therefore, how to deal with the current epidemic using urban regulatory methods, and how to prevent it from ever happening in the future, are what we need to learn and discuss from the urban perspective at present.

This study does not discuss how the outbreak happens. Instead, it focuses on the “aftermath” of the outbreak and the spread of the infection among cities. Especially, this study provides insights of the fundamentals to the infection from the urban perspective, where the marginal effects of some most influential factors to the COVID-19 transmission are estimated. This study shows empirical evidence from real world data, and it can provide hints to future scientific research on the novel coronavirus.

In addition, this study provides an alternative way to know the actual size of a large-scale outbreak in the epicenter. Among the four questions mentioned earlier, the last one can be notoriously difficult. Due to the technical difficulty and high cost for the laboratory-confirmation of COVID-19, we are unable to know the exact scale of the infection at the epicenter of the outbreak in its early or even middle stage. Especially in the initial stage of the epidemic in the epicenter, there can be substantial undocumented infection numbers (Li et al., 2020a). Therefore, precise estimation of the real infection scale is badly needed. This is the lesson we really need to take from the experience in Wuhan, China. Similar understanding of the scale of the epidemic needs to be done in other emerging regions with epidemic, before it is too late to contain the epidemic situation. These correct estimates of the size of the epidemic are also crucial to prepare and allocate appropriate medical resources to save lives.

The succeeding part of this paper is structured as follows. After reviewing the related literature, we discuss the data and statistical models. The empirical results are then shown with the focus on the marginal effects of the key variables. In the discussion section later, we compare the estimation results in difference time spots to show the evolution of the COVID-19 epidemic in China, and then we provide our estimates of the infection numbers in the epicenter of China, i.e., Wuhan city, with high precision. At last, we present out concluding remarks.

2. Literature review

Earlier related studies are more commonly seen in regard to the spread of flu or flu-like diseases such as the outbreak of Ebola in west Africa (Kucharski & Edmunds, 2014; Valencia et al., 2017). After the outbreak of COVID-19, the scientific scholars remain active and pay close attention to this emerging epidemic. The evolution of the novel coronavirus can be very complicated (Xu et al., 2020), and there are more and more studies on this virus to come out. By the end of January 2020, some other studies on COVID-19 are already released. Some popular topics commonly seen are listed here but not limited to the following studies such as the use of data (Heymann, 2020), early transmission dynamics (Li et al., 2020b), epidemic in other countries outside China (Holshue et al., 2020), the impact assessment (Munster et al., 2020), the forecasting of the spread (Wu et al., 2020), the genomic characterization (Lu et al., 2020), and the clinical characteristics (Chen et al., 2020).

The most famous epidemic model in mathematical bioscience is credited to the SEIR model, some earlier studies of which are highly cited (Li et al., 1999; Li & Muldowney, 1995). However, after the outbreak of COVID-19, some estimation results using SEIR model are far from the reported numbers shown later (Read et al., 2020; Tang et al., 2020). Perhaps, one big reason for the failure of such biological models is that they ignore the fundamentals of the urban perspective, which in fact provides the real background for the virus transmission. Indeed, the high population density in the metropolitan area as well as the massive level of public transportation in the large cities can play sort of “multiplier effect” for the virus transmission. Therefore, ignoring the urban context in the discussion of the transmission of a severe infectious disease is the mistake we cannot take.

While the researchers in the biological science community are keen to estimate and predict the size of the emerging epidemic of an infectious disease outbreak, scholars in the economics and especially urban study communities keep a much lower profile this time. We have to admit that, not too much related studies can be found in the economics and urban professions. However, now even the biological and medical scholars realize that both challenges and opportunities regarding the current epidemic can be resolved in urban settings (Lee et al., 2020). As noted, besides the economic consequences, cities, especially the large cities, are the most populated area in a region, which are typically travel hubs and employment centers. Therefore, cities are the natural existing soil to the large-scale outbreak of the fast transmitting communicable diseases.

Although the discussion of human health issues in the urban context is popular (Orimoloye et al., 2019; Wang, 2020), the particular studies on the urban infectious disease are rare. Among the limited studies of this strand, typical examples are the discussions on the price change in the neighboring property (Ambrus et al., 2020) and the household-built environment characteristics (Spencer et al., 2020). In addition, the effect of social network isolation on reducing the transmission of flu infection is also discussed (Aiello et al., 2016). However, to our knowledge, studies on the transmission of an infectious disease under the urban setting are barely seen. Once there was the study on a virtual outbreak of a flu epidemic in heavily populated metropolitan area with the simulation of its spread (Lee et al., 2008). Unfortunately, it comes true worldwide this time.

In fact, the use of statistical methods in epidemiology is quite popular and necessary, and many advanced techniques in statistics such as the Bayesian analysis have been employed (Elder et al., 2013). In addition, mathematical models are also commonly used in the studies of virus spreading (Dong et al., 2015). Therefore, in the technical sense, there should be no barriers for the economists and urban scholars to cross the border of such different disciplines.

As discussed earlier (Lee et al., 2008), many factors of biologic and physiologic, as well as social, economic, and geographic can place influences on the transmission and spread of the infectious disease such as flu. Therefore, to achieve such research goal, the methods to be used must be built-in interdisciplinary (Etard et al., 2017). Unlike those studies from purely scientific perspectives such as cells, genes, etc., this study presents a multidisciplinary analysis of all the cities in mainland China with confirmed cases of the novel coronavirus (COVID-19). We will try to reveal the hidden rules of the transmission of COVID-19...
behind the real-world data with proper models and statistical methods.

3. Research method

3.1. Data

This study uses a joint data set combining with several different sources.

The data of the infection number is obtained from Dingxiang Yuan. The data used in this paper is obtained at 09:08 a.m. Beijing time, March 2, 2020, as well as some earlier time spots such as January 31, 2020 and February 5, 2020. In this data source, we only use the number of confirmed patients for a more focused research purpose. Technically speaking, these confirmed numbers are laboratory-confirmed COVID-19 cases, which is not the same as those suspected cases that can be potential COVID-19 cases but not yet laboratory-confirmed.

As for the distances from the target cities to the epicenter of infection outbreak, i.e., Wuhan city, we measure the distances in the digital map provided by Baidu.com. In addition, the temperature data is obtained from the website of Weather China. At last, the local urban map provided by Baidu.com. In addition, the temperature data is obtained from the website of Weather China. At last, the local urban map provided by Baidu.com.

3.2. The model

Here we show a simple log linear regression model, which considers the number of laboratory-confirmed patients as dependent variable, and uses many other explanatory variables to find the influential factors. The regression equation is shown below.

\[
\ln Y = \beta_0 + \beta_1 \ln \text{Dist} + \beta_2 X + \epsilon
\]  

where, \( Y \) is the number of laboratory-confirmed COVID-19 cases. \( \text{Dist} \) is the distance to Wuhan city, which is the key control variable. The relationship of distance to the number of the confirmed patients of the novel coronavirus is shown in Fig. 2(a) and (b). In addition, \( X \) is a vector of explanatory variables, which are chosen as total length of built urban metro lines, urban area, population density, annual quantity of wastewater discharged, residential garbage connected and transported, per capita public recreational green space in this paper. At last, \( \epsilon \) is the stochastic error term.

Although the above regression model appears to be simple, the real hard part is that since there are so many possible factors that may affect the spread of virus, there can always be missing explanatory variables left in the error term, the result of which can lead serious endogeneity problem. Especially for the extremely complicated issue like this, i.e., the outbreak and spread of the mysterious COVID-19, many potential influential factors are technically yet unknown to us, thus we are unable to include them as the independent variables. As a result, the regression parameters with endogeneity issue are inconsistent, which makes the estimation not valid. In our study, which explanatory variable is the endogenous variable? We cast our doubt on population density, since logically it is closely linked to a possible virus transmission in the local area.

The common way to solve the endogeneity issue is to find appropriate instrumental variable(s), i.e., IV(s), and use 2SLS, LIML, or GMM to redo the regression. However, the tricky part here is that the instrumental variable is not easy to find. According to its definition, it must be uncorrelated with the error term, and correlated with the designated endogenous variable which is included in all the explanatory variables. Our question is therefore that does such instrumental variable exist in real world?

Luckily, after numerous rounds of tests, we finally find that temperature is a perfect candidate. In fact, it is the reality in China that the population density tends to be larger in the warmer places, i.e., the southern part of China. Of course, we do rigorous statistical tests to verify whether the use of temperature as the instrumental variable is appropriate, which is shown later in this paper. In addition, since there are a highest temperature and a lowest temperature in the daily records, after tests for many times we choose the highest temperature as the instrumental variable.

4. Empirical results

Here we show the marginal effects of some most influential factors to the transmission of the novel coronavirus (COVID-19) which are estimated from the urban perspective. By analyzing the data from all the cities in mainland China with laboratory-confirmed cases of the novel coronavirus, here we propose the following important findings which are based on Model (6) in Table 2. As noted, we set the time spot of February 5, 2020 as our baseline models which corresponds to Table 2, and we also compare the results before (i.e., January 31, 2020 for Table 4) and after (March 2, 2020 for Table 5) in later section of this paper.

Distance to Wuhan is found to be a very strong influential factor. Cities which are 1% closer to Wuhan city increase the confirmed infection number by 0.962%. The \( p \) value is almost 0%, which shows that distance really matters in statistical level. As we know that 5 million plus people have left Wuhan city before the city is officially sealed off, although it is really difficult to track where exactly these people go, the distance to Wuhan city is a very good proxy variable to represent this massive level of human movement. Logically, the closer to Wuhan city, the easier for the people from Wuhan to travel to those locations. This fact is in very good consistent with our finding in this paper.

Total length of built urban metro lines does matter to affect the virus infection. Every increase of 1 km of lines increases the confirmed infection number by 0.002%. Although this quantity appears tiny, it is statistically significant. It is easy to understand that as the primary tool
of massive transportation in modern cities, the metro carries a very large number of passengers including the potential patients. Therefore, the spread of the virus within the city can be accelerated and extended by the metro system.

Cities with urban area 1% smaller increase the confirmed infection number by 2.571%. In addition, cities with population density 1% smaller increase the confirmed infection number by 2.603%. The above two findings show very clearly about in which direction that we shall pay more attention to contain the virus spread. At the first glance, these findings are counter-intuitive, since the common sense that we are familiar with suggests that the larger cities with denser population are more easily to spread the virus infection, which is generally correct. However, we must consider the very special background of the virus spread in China.

Fig. 1. Epidemic size of COVID-19 in cities of mainland China without Wuhan city as of March 2, 2020.

Table 1

| Variables          | Explanation                                               | Unit       | Mean   | Std. Dev. | Min     | Max       |
|--------------------|-----------------------------------------------------------|------------|--------|-----------|---------|-----------|
| Num_p              | Number of laboratory-confirmed COVID-19 cases             | Person     | 256.016| 2804.774  | 1.000   | 49,315.000|
| Dist               | Distance to Wuhan                                         | Kilometer  | 1004.576| 625.624   | 12.650  | 3263.100  |
| Subway             | Total length of built urban metro lines                  | Kilometer  | 14.706 | 69.928    | 0.000   | 16,410.000|
| Urban_area         | Urban area                                                | Square kilometer | 502.921| 1153.827  | 5.990   | 16,410.000|
| Population_density | Population density                                        | Person/km²| 3638.705| 2379.552  | 77.000  | 11,602.000|
| Wastewater         | Annual quantity of wastewater discharged                  | 10,000 m³  | 13,694.400| 26,494.490| 284.000 | 229,526.000|
| Garbage            | Residential garbage connected and transported             | 10,000 ton | 57.639  | 102.694   | 1.560   | 924.770   |
| Greenspace         | Per capita public recreational green space                | Square meter | 14.209| 4.936     | 2.450   | 51.660    |
| Temp_h             | The daily highest temperature                             | Celsius degree | 5.554 | 7.461     | −16.000 | 25.000    |
| Capital_city       | Dummy variable (1 = capital city, 0 = otherwise)          | NA         | 9.936% | 0.300     | 0       | 1         |

Note: The mean of the number of laboratory-confirmed COVID-19 cases drops to 98.270 using the subsample without the Wuhan city, and it drops further to 43.193 using the subsample without the Hubei Province where Wuhan is the capital city. Data for the confirmed cases is obtained on March 2, 2020.
areas. As the result of the above two reasons, in this special time period, the transmission risk of the novel coronavirus in small cities or even rural areas may be larger than those in the big cities. Our statistical result shown in this paper in fact confirms this proposition.

In addition, cities with annual quantity of wastewater discharged 1% larger increase the confirmed infection number by 1.397%. The p value is around 1%, which is very significant in statistic sense as well. Also, cities with residential garbage connected and transported 1% larger increase the confirmed infection number by 1.633%. As we can see, the p value is around 2%, which is a very strong signal for its influence on the infection. The above two findings are statistically strong enough to provide new clues for possible direction of the investigation on the spread of COVID-19. Although the coronavirus type of disease is typically considered to be the infection via air-borne droplets, the real mechanism of the spread of COVID-19 is yet unknown. Before the mysterious chain of spread is completely understood, the statistical findings found in this paper can give hints to scientists and policy makers to take precautionary activities. The most importantly, appropriate steps must be implemented before it is too late to contain the spread of the virus.

Is the spread of COVID-19 associated with water, especially discharged wastewater? Is there middle host of the virus in the water? Does the residential garbage, which may also include wastewater, have something to do with the virus spread and infection? We are not absolutely certain about the answers to these questions at the present moment. However, the statistical findings shown in this paper can still contribute to the clues for future study and related regulation policies. For example, as many Chinese residents wear face masks after the outbreak of COVID-19, it is easy to predict that the used masks in the residential garbage can reach a massive level. If the residents just randomly junk their used masks without proper attention, it may cause residential garbage can reach a massive level. If the residents just randomly junk their used masks without proper attention, it may cause secondary spread of the virus. Therefore, our statistical results make sense and show the right direction.

At last, we use the per capita public recreational green space as well as a dummy variable for capital city in the province as additional

![Fig. 2(a): Scatter plots of the subsample without Wuhan City (n = 311)](image)

![Fig. 2(b): Scatter plots of the subsample without Hubei Province (n = 296)](image)

Table 2
Empirical estimation results with dependent variable log(Num_p) on February 5, 2020.

|                          | Model (1)  | Model (2)  | Model (3)  | Model (4)  | Model (5)  | Model (6)  |
|--------------------------|------------|------------|------------|------------|------------|------------|
|                          | OLS (Full sample, n = 305) | OLS (Full sample, n = 305) | OLS (Subsample without the city of Wuhan, n = 304) | OLS (Subsample without the city of Wuhan, n = 304) | OLS (Subsample without Hubei Province, n = 289) | OLS (Subsample without Hubei Province, n = 289) |
| Log(Dist)                | -0.971*** | -1.229***  | -1.261***  | -0.856***  | -0.962***  |           |
|                          | (-14.599) | (-15.793)  | (-14.258)  | (-9.600)   | (-7.652)   |           |
| Subway                   | 0.001     | 0.002      | 0.002      | 0.001      | 0.002      |           |
|                          | (0.820)   | (1.213)    | (1.740)    | (1.653)    | (1.904)    |           |
| Log(Urban_area)          | -0.175    | -0.247     | -1.009     | -0.154     | -2.571     |           |
|                          | (-0.735)  | (-1.087)   | (1.789)    | (1.653)    | (1.904)    |           |
| Log(Population_density)  | -0.121    | -0.186     | -0.979     | -0.083     | -2.603     |           |
|                          | (-0.519)  | (-0.840)   | (-0.693)   | (-0.411)   | (-1.841)   |           |
| Log(Wastewater)          | 0.525***  | 0.513***   | 0.816      | 0.445***   | 1.397***   |           |
|                          | (3.201)   | (3.289)    | (1.467)    | (3.115)    | (2.516)    |           |
| Log(Garbage)             | 0.329     | 0.370      | 0.749      | 0.427**    | 1.633**    |           |
|                          | (1.685)   | (1.993)    | (1.083)    | (2.530)    | (2.339)    |           |
| Log(Greenspace)          | -0.218    | -0.233     | -0.292     | -0.126     | -0.318     |           |
|                          | (-1.184)  | (-1.334)   | (-1.416)   | (-0.792)   | (-1.410)   |           |
| Capital_city             | -0.280    | -1.017     | -0.051     | -0.158     | 0.024      |           |
|                          | (-1.131)  | (-1.450)   | (-0.740)   | (0.083)    |           |           |
| Adjusted R²              | 0.597     | 0.616      | 0.599      | 0.591      | 0.362      | 5.114**   |
| Endogeneity Test (Difference in J-stats) | 4.598*** | 7.278      | 0.338      |           |           |           |

Note: The values of the constant terms are not reported. t statistics in parentheses. The data used in this table is obtained at 10:33 a.m. Beijing time, February 5, 2020.

*** p ≤ .01.
** 0.01 < p < .05.
* 0.05 < p < .1.
explanatory variables. As we can see, the smaller such green space area is, the higher the confirmed infection number is. However, it is not statistically significant in Model (6). It is barely significant in Model (4). This result, though not directly, indicates that cities with larger open space may be less vulnerable to the virus infection. It can be considered to be a general result, not just to the spread of COVID-19. The capital city dummy variable is not significant in any model.

As noted earlier, we use three groups of samples, i.e., the full sample, the subsample without Wuhan city, the subsample without Hubei province where Wuhan is the capital city. Presented in Table 2, we see that the estimation is robust in the sense that all the estimation parameters of corresponding variables have the same signs in all the models. However, which one is our best shot?

To answer this question, we examine two things. First, a glance of Table 2 shows that the most key variables are statistically significant in Model (6), which is a very desirable result. Second, both the Endogeneity Test and Weak Instrument Diagnostics show that Model (6) works the best. In model (6), difference in J-stats is very significant, and Cragg-Donald F-stat is very close to 10, which make model (6) pass the required statistical tests for endogeneity and the validity of the instrumental variable. The technical favor of model (6) also makes sense to us. Since Wuhan city is the outbreak epicenter of COVID-19 in China, the confirmed cases in Wuhan is much larger than anywhere else. Therefore, the inclusion of Wuhan to the data set for transmission analysis can be biased.

In addition, with the help of Model (6), we are able to “purify” the data set by deleting those observations with poor prediction performance, the most of which are “remote” cities with very small number of confirmed COVID-19 cases and thus can be considered to be observations with extreme values. After deleting these cities from our data set, how good would the result be? Hence, we present Table 3.

In Table 3, the OLS results in Model (7) and (9) are very good since all the explanatory variables except capital_city are very statistically significant at nearly 0% level, and the adjusted R² values are increased to above 73% level. However, the 2SLS results in Model (8) and (10) are not good any more. After the data “purification” by Model (6), the endogeneity issue can be reduced to a very large extent; hence the OLS results are already good.

5. Discussion

In fact, to understand the COVID-19 epidemic in China, we must first understand the role of the Spring Festival as well as the Spring Transportation in China. Essentially, this is in the “everybody-rush-home” manner. After the national holiday of the Spring Festival, the peak time of the returning travelers in the Spring Transportation would come. It is the real challenge to contain the epidemic situation since travelers from so many small cities or rural areas return to the big cities where they have their jobs, which may cause a second round of outbreak in the large cities. As this paper shows, small cities during the period of Spring Festival suffer from more rapid transmission of COVID-19. In fact, in the days of early February, the rapid increasing number of infections in Shenzhen, one of the most economically advanced cities in China, and who has the highest ratio of migrants, has already signaled a warning.

Therefore, appropriate precautionary steps must be taken immediately to prevent the second round of outbreak in the big cities from ever happening. In Table 4, we show the estimation results in Model (11) to (16) as a comparison. The data used in Table 4 is obtained on January 31, 2020, the time of which is at the end of the national vacation of Spring Festival. As we can see very clearly, results of urban area and population density are less significant in Table 2, which means that the transmission of COVID-19 increases in big cities after the Spring Festival holiday. As we know, strong regulatory intervention policies have been employed timely, one of which is to postpone the back-to-work dates. In fact, the massive level of returning side of the Spring Transportation does not occur as those regular years.

Luckily, the containment of the COVID-19 epidemic in China is a great success that the epidemic outside Hubei province is almost stable since late February. In Table 5, we show the estimation results using data obtained on March 2, 2020 as a comparison. As we can see, the urban area and population density is no longer significant, though they

### Table 3

Empirical estimation results with dependent variable log(Num_p) on February 5, 2020 with “purified” data set.

|                          | Model (7) OLS (Subsample without Hubei Province, n = 227) | Model (8) 2SLS (Subsample without Hubei Province, n = 227) | Model (9) OLS (Subsample without the city of Wuhan, n = 242) | Model (10) 2SLS (Subsample without the city of Wuhan, n = 242) |
|--------------------------|----------------------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|
| log(Dist)                | -0.86,⁎⁎⁎                                                | -0.86,⁎⁎⁎                                                  | -1.22,⁎⁎                                                 | -1.22,⁎⁎                                                 |
| Subway                   | -0.02,⁎                                                 | -0.03,⁎                                                     | 0.00,⁎                                                         | 0.01,⁎                                                         |
| log(Urban_area)          | -1.09,⁎⁎⁎                                               | -1.13,⁎                                                      | -1.19,⁎                                                 | 0.245                                                      |
| log(Population_density)  | -0.98,⁎⁎                                                 | -1.03,⁎                                                     | -1.11,⁎                                                 | 0.411                                                      |
| log(Wastewater)          | 0.84,⁎⁎                                                 | 0.86,⁎                                                     | 0.93,⁎                                                  | 0.394                                                      |
| log(Garbage)             | 0.83,⁎                                                   | 0.86,⁎                                                      | 0.76,⁎                                                 | 0.014                                                      |
| log(Greenspace)          | -0.49,⁎⁎                                                 | -0.40,⁎                                                     | -0.51,⁎                                                | -0.329                                                      |
| Capital_city             | -0.17,⁎                                                 | -0.21,⁎                                                     | -0.05,⁎                                                 | -0.137                                                      |
| Adjusted R²              | 0.73                                                    | 0.73                                                       | 0.73                                                   | 0.677                                                      |
| Endogeneity Test (Difference in J-stats) | 9.728                                                  | 9.728                                                       | 1.685                                                   | 7.962                                                      |

Note: The values of the constant terms are not reported. t statistics in parentheses. The data used in this table is obtained at 10:33 a.m. Beijing time, February 5, 2020.

⁎⁎⁎ p ≤ .01.
⁎⁎ p ≤ .05.
⁎ p ≤ .1.
Empirical estimation results with dependent variable log(Num_p) on January 31, 2020.

| Model (11) | Model (12) | Model (13) | Model (14) | Model (15) | Model (16) |
|------------|------------|------------|------------|------------|------------|
| OLS (Full sample, n = 282) | 2SLS (Full sample, n = 282) | OLS (Subsample without the city of Wuhan, n = 281) | 2SLS (Subsample without the city of Wuhan, n = 281) | OLS (Subsample without Hubei Province, n = 266) | 2SLS (Subsample without Hubei Province, n = 266) |
| log(Dist) | −0.856*** | −0.897*** | −1.094*** | −1.124*** | −0.696*** |
| | (−13.193) | (−9.566) | (−14.063) | (−12.631) | (−7.967) |
| Subway | 0.002** | 0.003** | 0.002** | 0.003*** | 0.002** |
| | (1.699) | (2.064) | (2.578) | (2.687) | (2.565) |
| log (Urban_area) | −0.385 | −0.420 | −0.422 | −2.150 | −0.317 |
| | (−1.624) | (−2.168) | (−1.860) | (−1.630) | (−1.560) |
| log (Population_density) | −0.384 | −4.191*** | −0.420 | −2.074 | −0.304 |
| | (−1.657) | (−2.162) | (−1.896) | (−1.642) | (−1.535) |
| log (Wastewater) | 0.530*** | 1.906*** | 0.502*** | 1.125*** | 0.424*** |
| | (3.284) | (2.621) | (3.249) | (2.263) | (3.051) |
| log (Garbage) | 0.346 | 2.221*** | 0.413*** | 1.267*** | 0.459*** |
| | (1.768) | (2.265) | (1.883) | (2.734) | (2.894) |
| log (Greenspace) | −0.193 | −0.493*** | −0.232 | −0.370 | −0.116 |
| | (−1.067) | (−1.663) | (−1.335) | (−1.696) | (−0.749) |
| Adjusted R² | 0.548 | 0.101 | 0.559 | 0.460 | 0.531 |
| | 7.894*** | 2.178 | 11.700*** | 7.976 | 11.187 |

**Note:** The values of the constant terms are not reported. t statistics in parentheses. The data used in this table is obtained at 07:26 a.m. Beijing time, January 31, 2020.

Table 4

Empirical estimation results with dependent variable log(Num_p) on March 2, 2020 with "purified" data set.

| Model (17) | Model (18) | Model (19) | Model (20) | Model (21) | Model (22) |
|------------|------------|------------|------------|------------|------------|
| OLS (Subsample without Hubei Province, n = 198) | 2SLS (Subsample without Hubei Province, n = 198) | LGM MCMC Draws: 20,000,000 (Subsample without Hubei Province, n = 198) | OLS (Subsample without the city of Wuhan, n = 213) | 2SLS (Subsample without the city of Wuhan, n = 213) | LGM MCMC Draws: 20,000,000 (Subsample without the city of Wuhan, n = 213) |
| log(Dist) | −0.951*** | −0.963*** | −0.949*** | −1.425*** | −1.396*** |
| | (−19.9410) | (−11.211) | (−14.099) | (−20.937) | (−16.651) |
| Subway | 0.001** | 0.002*** | 0.001*** | 0.002*** | 0.002*** |
| | (2.394) | (1.866) | (2.380) | (2.107) | (2.165) |
| log (Urban_area) | −0.211 | −1.672 | −0.200 | −0.328 | −2.948 |
| | (−1.398) | (−1.98756) | (−1.1336) | (−1.590) | (−1.439) |
| log (Population_density) | −0.087 | −1.637 | −0.075 | −0.284 | −2.136 |
| | (−0.585) | (−0.913) | (−0.510) | (−1.375) | (−1.396) |
| log (Wastewater) | 0.346*** | 0.901 | 0.345*** | 0.421*** | 1.093*** |
| | (3.219) | (1.381) | (3.213) | (2.907) | (1.903) |
| log (Garbage) | 0.522*** | 1.174 | 0.515*** | 0.480*** | 1.271*** |
| | (4.206) | (1.533) | (4.172) | (2.811) | (1.881) |
| log (Greenspace) | 0.077 | 0.160 | 0.079 | −0.419*** | −0.826*** |
| | (0.557) | (0.803646) | (0.570) | (−2.203) | (−2.030) |
| Capital_city | −0.045 | 0.191 | −0.048 | 0.029 | 0.224 |
| | (−0.318) | (0.587) | (−0.341) | (0.1474) | (0.795) |
| Adjusted R² | 0.780 | 0.653 | 0.748 | 0.649 | 0.521 |
| | 2.066 | 2.096 | 7.976 | 11.187 |

**Note:** The values of the constant terms are not reported. t statistics in parentheses. The data used in this table is obtained at 09:08 a.m. Beijing time, March 2, 2020.

still remain the negative signs. Again, after the purification of data, the OLS model works fine. Since the number of sample cities is only around 200, we present the Linear Gaussian Model (LGM) estimated by Markov Chain Monte Carlo (MCMC) with Gibbs sampling to check the robustness of the models. As shown in Model (19) and Model (22), even after 20,000,000 MCMC draws, the parameters and their significances are almost the same as the corresponding OLS models. In addition, the frequency illustration of some key variables in Model (19) is also shown in Fig. 3.

The last question in this paper arises. For Model (7) and (9), as well as Model (17) and (20), which one has the best prediction power? Now, we look at Wuhan city again, which is the epicenter of the COVID-19 pandemic.
outbreak in China.

Since the model prediction result of the number of COVID-19 infection cases in Wuhan city relies heavily on the distance to Wuhan, we must discuss the radius we use here before we can proceed. Since Wuhan is a big city, we cannot just simply consider it to be a point with a radius approaching to zero. It is really hard to depict the radius of activities that the most people do in the city. As reported by Didi, the leading sharing car company in China, >90% residents in a city stay in a particular radius from the center of the city. For Wuhan, such radius is measured as 25.300 km, which covers a large proportion of the metropolitan area in Wuhan. We present Fig. 4 to show the validity of this radius. In Fig. 4, we demonstrate the satellite image of Wuhan city with the 25.3 km radius, which is then compared to the map of the official comprehensive planning of Wuhan (2010–2020). As we can see, the 25.3 km radius is clearly in consistent with the official comprehensive planning for the metropolitan area of Wuhan. Moreover, it covers the majority of the built urban area in Wuhan very well as the satellite image confirms.

In Table 6, we show three levels of such radius. The first one is the 25.300 km mentioned above. The second is 17.890 km, which stands for exactly half of the area of circle shown by the first radius. The third one is the half of this radius, which is 12.650 km. As noted, the second radius feels more like the most representative “center of mass” of Wuhan city. Then how can we evaluate the prediction results shown in Table 6? Since the reported number of confirmed COVID-19 cases in Wuhan corresponding to the prediction results of Model (7) and (9) in Table 6 is 8,351, the mean value of the prediction results of Model (7) tends to underestimate the real epidemic size of COVID-19 in Wuhan. This is due to the lack of information from the surrounding cities in Hubei province where Wuhan is the capital city, which limits the prediction ability of Model (7). On the contrary, Model (9) has the valuable information from those cities in Hubei province and hence can predict better.

When we update our prediction using data of March 2, 2020, the results are much closer to the reality. As shown in the last column of Table 6, the prediction with the radius used for Wuhan city as 17.890 km, shows the outcome of 56,944.866 as the mean value of the prediction. Please note that the model prediction result is in the “self-updated” manner. That is, when we update the estimation parameters with the latest data, the prediction result will be updated as well. However, since the local epidemic situation of COVID-19 is almost stabilized in early March 2020, the latest updated numbers of the epidemic in China is in fact reflecting the imported cases from abroad, which is not appropriate for our research purpose. Therefore, we stop updating the epidemic data for this study in early March 2020. As of April 30, 2020, the number of confirmed COVID-19 cases is 50,333 in Wuhan. Considering that there are possibly omitted infected cases, self-cured cases before confirmation, and other possible reasons, our
Fig. 4. Illustration of the 25.3 km radius in Wuhan city
Source: Google Map and Wuhan Natural Resources and Planning Bureau.

| Radius used for Wuhan city (kilometer) | Prediction results of Model (7) | Prediction results of Model (9) | Prediction results of Model (17) | Prediction results of Model (20) |
|----------------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| 25.300                                 | 3492.848                      | 13,766.347                     | 5604.536                       | 34,753.459                     |
|                                        | (868 to 14,054)               | (2784 to 68,070)               | (1844 to 17,031)               | (7940 to 152,000)              |
| 17.890                                 | 4705.333                      | 21,059.862                     | 7793.026                       | 56,944.866                     |
|                                        | (1145 to 19,334)              | (4200 to 106,000)              | (2513 to 24,164)               | (12,800 to 253,000)            |
| 12.650                                 | 6338.711                      | 32,217.530                     | 10,836.09                      | 93,306.308                     |
|                                        | (1508 to 26,637)              | (6320 to 164,000)              | (3420 to 34,337)               | (20,600 to 422,000)            |

Note: 95% confidence intervals in parentheses.
estimation is very accurate. In fact, as of April 14, 2020, the accumulated cases of asymptomatic infection in mainland China is 6,764, where 588 cases are imported from abroad. Therefore, adding these symptom-free cases to the reported confirmed cases mentioned earlier, our result is more accurate than we expect.9

6. Concluding remarks

As a very multidisciplinary research paper, this study may provide valuable information to a variety of readership.

To biological and medical scholars, this study may provide hints to the possible ways of transmission of COVID-19. Since many biologic and physiologic features of this new disease are still unknown, the statistical findings shown in this paper may present valuable clues. For example, we do not know for sure that how the wastewater and residential garbage may affect the transmission of COVID-19, but since the statistical results are significant, scientists can look into these fields. Perhaps, they may have unexpected breakthrough.

To researchers of mathematical bioscience, this paper gives an alternative way other than the SEIR models to estimate and predict the actual size of a large-scale outbreak of epidemic in the epicenter. The method shown here is born to be practical since it is based on only a few information from regions outside the epicenter of the outbreak, the data of which is more accurate and reliable at the initial or even middle stage of the epidemic. As shown in the paper, the estimated number of infections in the epicenter of China, i.e., Wuhan city, is 56,944.866 (mean value), which is very close to the officially reported number. This result is obtained through rigorous statistical tests and examinations, and such method and its result are proven to be accurate given the example of Wuhan as shown in this study.

To policy makers, knowing the accurate size of epidemic or how bad it would evolve to as early as possible in the initial stage of the outbreak is very crucial to get prepared and allocate scarce medical resources, the result of which is hence that more lives can be saved. In addition, the empirical results shown in this paper have important policy implications as well. For example, if the distance to the epicenter is found to have so much influence on the epidemic spread, then the government should seal off the city as epicenter immediately without hesitation. Furthermore, if the subway is shown to be very statistically significant in the transmission, one should close the subway system promptly. We understand that these are all difficult decisions to make in the urban management. However, since it may take a very long time for the scientific research to reveal the hidden mechanism of the virus spread, we cannot wait for months or even years before the infection numbers go exponential. Therefore, if we can observe some influential factors with statistical significance at the initial stage of the outbreak, we should act quickly to cut off the transmission chain by any possible methods we have.

In addition, this study also provides insights to urban planners. For example, this paper has discussed the role of public green space in the epidemic. Although the corresponding coefficient shown in the paper is not quite significant, its negative sign indicates that proper urban planning with more available open space can be essential to the fight with various disasters and catastrophes. As for the epidemic of COVID-19, two major temporary hospitals with >3000 beds are built promptly in Wuhan city’s open space to treat patients with severe symptoms. Many more temporary hospitals are also prepared in the city for the proper medical treatment of patients with mild symptoms or for quarantine use. Later, when the situation gets much better in China in late March and early April 2020, people restore their recreational activities in open green spaces, but the gathering activities in the indoor recreational facilities are still restricted. The above two scenarios are all good examples of the important uses and roles that the (green) open space in the cities can play.

As the epidemic situation is still going on and evolving quickly around the globe, this study provides a timely analysis of the spread of the infection among cities in mainland China. Now the epidemic in China tends to be stable, but the situations in many other countries are still getting worse. The methodology and main conclusions shown in this paper are of general interest, and they can be applied to other countries to help understand the local transmission of COVID-19 as well. As concerns exist that the worse situation is yet to come, the findings and implications of this research may offer methods of prevention to minimize the damage of the disaster from the urban perspective. In the other word, this study may provide valuable and timely hints and clues to the local fight with the terrible disease.

The above implications also apply to the economists and urban scholars. To allocate scarce resources optimally is what the economists do. Thus, economists can do a lot of contributions in the battle with COVID-19, where urban scholars can add more from the city related perspectives. Undoubtedly, economists and urban scholars can do more to the solutions of the ongoing COVID-19 pandemic worldwide. In this sense, this paper is just a beginning.

Contributors

All by Lu Liu.

Data sharing

All data utilized in the present study are publicly available.

Declaration of competing interest

The author declares no competing interests.

Acknowledgement

The author appreciates the comments and suggestions by the editor(s) and anonymous reviewers. All the remaining mistakes in the paper are the author’s own.

References

Aiello, A., et al. (2016). Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial. *Epidemiology*, 15, 38–55.

Ambrus, A., Field, E., & Gonzalez, R. (2020). Loss in the time of cholera: Long-run impact of a disease epidemic on the urban landscape. *American Economic Review*, 110(2), 475–525.

Chan, A., et al. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet*. https://doi.org/10.1016/S0140-6736(20)30154-9.

Chen, N., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus in Wuhan, China: A descriptive study. *The Lancet*. https://doi.org/10.1016/S0140-6736(20)30211-7.

Dong, F., et al. (2015). Evaluation of Ebola spreading in West Africa and decision of optimal medicine delivery strategies based on mathematical models. *Infection, Genetics and Evolution*, 36, 35–40.

Elderd, B., Dywer, G., & Dukic, V. (2013). Population-level differences in disease transmission: A Bayesian analysis of multiple smallpox epidemics. *Epidemics*, 5(3), 146–156.

Etard, J., et al. (2017). Multidisciplinary assessment of post-Ebola sequelae in Guinea (Postebogui): An observational cohort study. *The Lancet Infectious Disease*, 17(5), 545–552.

Heymann, D. (2020). Data sharing and outbreaks: Best practice exemplified. *The Lancet*. https://doi.org/10.1016/S0140-6736(20)30184-7.

Holtz, M., et al. (2020). First case of 2019 novel coronavirus in the United States. *The New England Journal of Medicine*. https://doi.org/10.1056/NEJMoct203191.

Huang, C., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus: The Lancet*. https://doi.org/10.1016/S0140-6736(20)30183-5.

Kucharski, A., & Edmunds, W. (2014). Case fatality rate for Ebola virus disease in west Africa. *The Lancet*, 384(9950), 1266.

Lee, B., et al. (2008). Virtual epidemic in a virtual city: Simulating the spread of influenza
in a US metropolitan area. Translational Research, 151(6), 275–287.
Lee, V., et al. (2020). Epidemic preparedness in urban settings: New challenges and opportunities. The Lancet Infectious Disease. https://doi.org/10.1016/S1473-3099(20)30249-8.
Li, M., Graef, J., Wang, L., & Karsai, J. (1999). Global dynamics of a SEIR model with varying total population size. Mathematical Biosciences, 160(2), 191–213.
Li, M., & Muldowney, J. (1995). Global stability for the SEIR model in epidemiology. Mathematical Biosciences, 125(2), 155–164.
Li, R., et al. (2020a). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science. https://doi.org/10.1126/science.abb3221.
Li, Q., et al. (2020b). Early transmission dynamics in Wuhan, China, of Novel Coronavirus–infected pneumonia. The New England Journal of Medicine. https://doi.org/10.1056/NEJMoa2001316.
Liu, L (2014). Spring transportation in china: the peak-load problem with psychological factors. Emerging Markets Finance & Trade, 50(suppl.2), 100–113.
Lu, R., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. The Lancet. https://doi.org/10.1016/S0140-6736(20)30251-8.
Munster, V., et al. (2020). A novel coronavirus emerging in China: Key questions for impact assessment. The New England Journal of Medicine. https://doi.org/10.1056/NEJMoa2009929.
Orimoloye, I. R., et al. (2019). Implications of climate variability and change on urban and human health: A review. Cities, 91, 213–223.
Read, L., et al. (2020). Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions. https://doi.org/10.1101/2020.01.23.20018549.
Spencer, J., et al. (2020). Emerging infectious disease, the household built environment characteristics, and urban planning: Evidence on avian influenza in Vietnam. Landscape and Urban Planning, 193, 103681.
Tang, B., et al. (2020). Estimation of the transmission risk of 2019-nCoV and its implication for public health interventions. https://ssrn.com/abstract=3525558.
Valencia, C., et al. (2017). Network visualization for outbreak response: Mapping the Ebola Virus Disease (EVD) chains of transmission in N’Zérékoré, Guinea. Journal of Infection, 74(3), 294–301.
Wang, C., Herby, P., Hoyden, F., & Gao, F. (2020). A novel coronavirus outbreak of global health concern. The Lancet. https://doi.org/10.1016/S0140-6736(20)30185-9.
Wang, K. (2020). Neighborhood foreclosures and health disparities in the U.S. cities. Cities, 97, 102526.
Wu, J., Leung, K., & Leung, G. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. The Lancet. https://doi.org/10.1016/S0140-6736(20)30260-9.
Xu, X., et al. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Science China Life Sciences. https://doi.org/10.1007/s11427-020-1637-5.
Zhu, N., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. The New England Journal of Medicine. https://doi.org/10.1056/NEJMo02901017.