

Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China

Yahua Zhang^a, Anming Zhang^b, Jiaoe Wang^{cd*}

May 2020

^a School of Commerce, University of Southern Queensland, Toowoomba, Queensland, 4350 Australia. Email: Yahua Zhang (shane.zhang@usq.edu.au)

^b Sauder School of Business, University of British Columbia, Vancouver, BC, Canada. Email: Anming Zhang (anming.zhang@sauder.ubc.ca)

^c Key Laboratory of Regional Sustainable Development Modelling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;

^d College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China;

* Corresponding author. Email: wangje@igsnr.ac.cn

Abstract

To understand the roles of different transport modes in the spread of COVID-19 pandemic across Chinese cities, this paper looks at the factors influencing the number of imported cases from Wuhan and the spread speed and pattern of the pandemic. We find that frequencies of air flights and high-speed train (HST) services out of Wuhan are significantly associated with the number of COVID-19 cases in the destination cities. The presence of an airport or HST station at a city is significantly related to the speed of the pandemic spread, but its link with the total number of confirmed cases is weak. The farther the distance from Wuhan, the lower number of cases in a city and the slower the dissemination of the pandemic. The longitude and latitude coordinates do not have a significant relationship with the number of total cases but can increase the speed of the COVID-19 spread. Specifically, cities in the higher longitudinal region tended to record a COVID-19 case earlier than their counterparties in the west. Cities in the north were more likely to report the first case later than those in the south. The pandemic may emerge in large cities earlier than in small cities as GDP is a factor positively associated with the spread speed.

Keywords: COVID-19, High-speed rail, Inter-city bus, Air transport, Spread speed, Spread pattern

1. Introduction

Patients with pneumonia of unknown causes were first detected in Wuhan, the capital city of Hubei Province of China, in December 2019. These cases were reported to the World Health Organisation (WHO) China office on 31 December 2019. The cause was later identified as a novel Coronavirus, genetically related to the Middle East Respiratory Syndrome (MERS) virus and the Severe Acute Respiratory Syndrome (SARS) virus. This new virus was subsequently named as COVID-19 by WHO on 11 February 2020. On 11 March, the COVID-19 outbreak was declared by WHO as a pandemic after it quickly spread to Iran, Italy and other parts of Europe and North America.

The rapid human-to-human transmission nature of COVID-19 was formally recognised and announced in China on 20 January 2020. However, by this point in time, the outbreak and spread of COVID-19 seemed to have lost control in Wuhan. In addition, a large number of people had travelled out of Wuhan to other parts of China amid the massive Chinese New Year migration. To halt the spread of the virus, an emergent city lockdown measure was imposed in Wuhan, the epicentre of COVID-19, on 23 January. The lockdown was later extended to the whole province. The shutdown affected a population of about 60 million, including 11 million in Wuhan. Although this radical move was questioned by many people due to the high economic (and other) costs associated, this approach appeared to be vindicated as the curve for the spread of the disease outside Hubei flattened in mid-February 2020, and the same pattern emerged in Wuhan in late February 2020. On 14 March 2020, only Wuhan was classified as a high-risk area and the rest of the province was medium- or low-risk. In the next few days, the lockdown gradually loosened throughout Hubei Province. On 8 April 2020, the city of Wuhan reopened after more than 10 weeks in lockdown. Air and train services have resumed and the city's recovery is being closely watched by the world.

Wuhan is one of China's most important transport hubs, and is therefore very well connected to other areas in China (Zhu et al., 2019). The present research aims to examine the role that each transport mode played in diffusing the cases of COVID-19. Various studies have

considered the link between the movement of people and the imported cases of COVID-19. Zhao et al. (2020) found that the number of air passengers from Wuhan and local population can be used to explain the number of cases in the infected cities. Ai et al. (2020) reported a significant and positive relationship between population movement and the number of COVID-19 cases. They argued that some cases could be avoided and prevented if the city closure was implemented earlier. However, most of these studies only consider one transport mode or the total movement of people regardless of the transport means.

More specifically, this paper investigates the factors influencing the number of imported cases from Wuhan and the spread speed and pattern of the pandemic. The gravity model is used with a consideration of the factors of the frequencies of high-speed train (HST), coach and air services (flights) between Wuhan and the other domestic cities, the cities' GDP and distance from Wuhan as well as longitude and latitude, and the presence of an airport/HST Station. HST, coach (inter-city bus) and air services were the most important means of transporting the five million people out of Wuhan before the lockdown. We believe that exploring the relationship between transport means and the speed and spread pattern of infectious diseases such as COVID-19 is important to both the general public, policy makers and health professionals. Such relationship can also be used for future prediction and control or mitigation when similar incidences take place.

We find significant links between flight and HST frequencies out of Wuhan and the number of COVID-19 cases in the destination cities. The presence of an airport or HST station at a city is significantly related to the speed of the pandemic spread, but its link with the number of total confirmed cases is weak. The farther the distance from Wuhan, the lower number of cases in a city and the slower the dissemination of the pandemic. The longitude and latitude values do not have a significant relationship with the number of total cases, but can increase the speed of the COVID-19 spread. Cities in the south were more likely to report the first case earlier than those in the north. The pandemic may emerge in large cities earlier than in small cities.

The paper is organised as follows. Section 2 describes background and reviews the relevant studies, and Section 3 discusses our empirical methodology. The results are reported and discussed in Section 4. Finally, Section 5 contains the concluding remarks.

2. Background and relevant studies

2.1 Background

Wuhan is one of China's most important transport hubs. It has three railway stations, one international airport, and a significant inland port that lies at the confluence of the Yangtze and Hanjiang rivers. China's first long-haul HST line, the Wuhan-Guangzhou HST, was launched in 2009 (Zhang et al., 2019). This HST line was later extended to Beijing and Hong Kong, forming the longest north-south trunk line of China's "4+4" HST network (four north-south and four east-west trunk lines; see, e.g., Wang et al., 2017). Wuhan is also the midpoint of the Shanghai-Wuhan-Chengdu HST line, which is one of the four east-west trunk lines. Therefore, China's most important cities such as Beijing, Shanghai, Guangzhou, Chengdu, Shenzhen and Hong Kong can be reached within five hours from Wuhan by HST. Inside Hubei Province, the Wuhan-Shiyan HST commenced services at the end of December 2019, connecting Wuhan to Hubei's key industrial cities such as Xiangyang, and Shiyan. With all these connections, Wuhan has been called the heart of China's HST network and it is ranked the fourth place in China by the weighted degree and closeness centrality, and the second place by weighted betweenness centrality (Jiao et al., 2017). It has been ranked the fourth most important rail-air combined node in China's city network (after Shanghai, Beijing and Guangzhou; see Zhu et al., 2018), and fourth in terms of the total rail passengers handled and first in terms of the number of transfer passengers in recent years.¹ Furthermore, Wuhan Tianhe International Airport is the largest airport in Central China, ranking the top-15 airports in China by all three centrality indexes (Wang et al., 2011). It handled 27 million passengers in 2019, an increase of 10.8% from 2018. Besides, Wuhan plays an important role in China's intercity coach network (Wang et al., 2020). The weekly frequencies of flights, HST and coaches out of Wuhan are illustrated in Figures 1, 2 and 3, respectively.

¹ Wuhan is also ranked among the top in criticality and importance. For example, its closure may reduce the performance of HST network by 10.3% (Li et al., 2020); and it is ranked the second in rail-air network by criticality (Li et al., 2019).

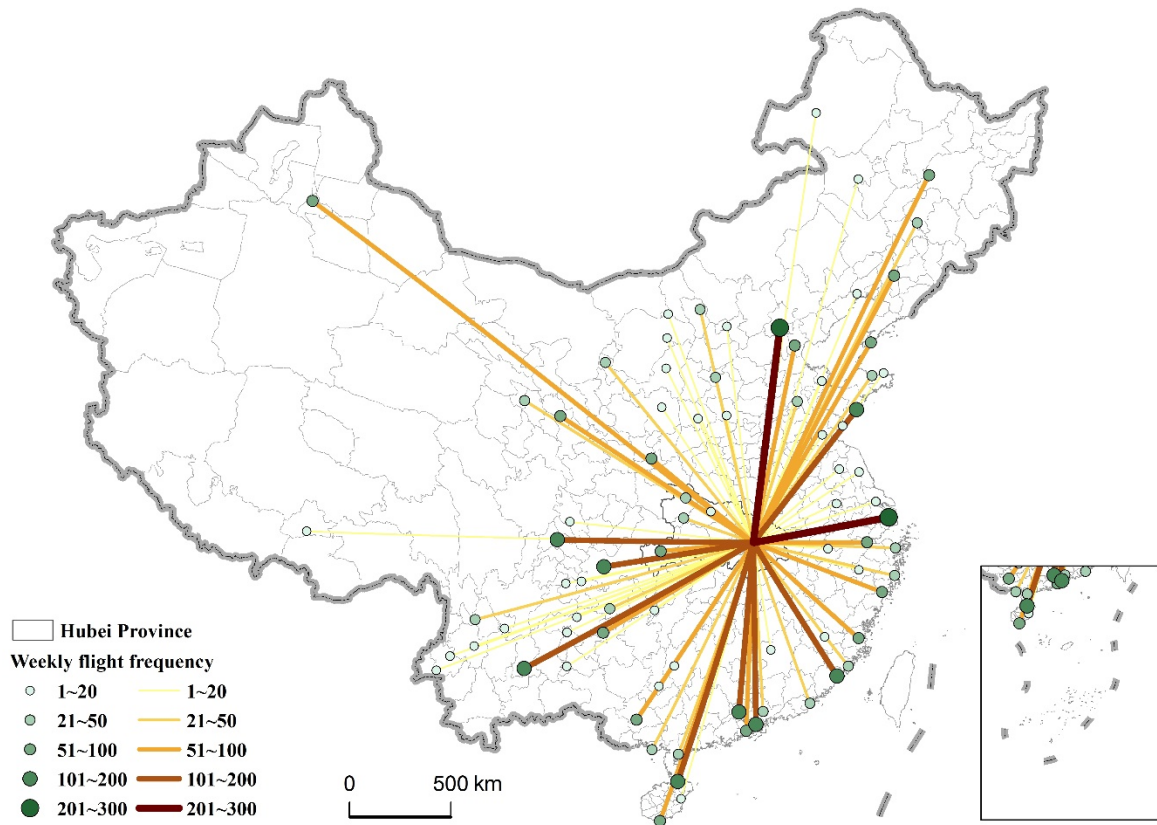


Figure 1: Weekly flight frequency out of Wuhan (2019 data, source: OAG)

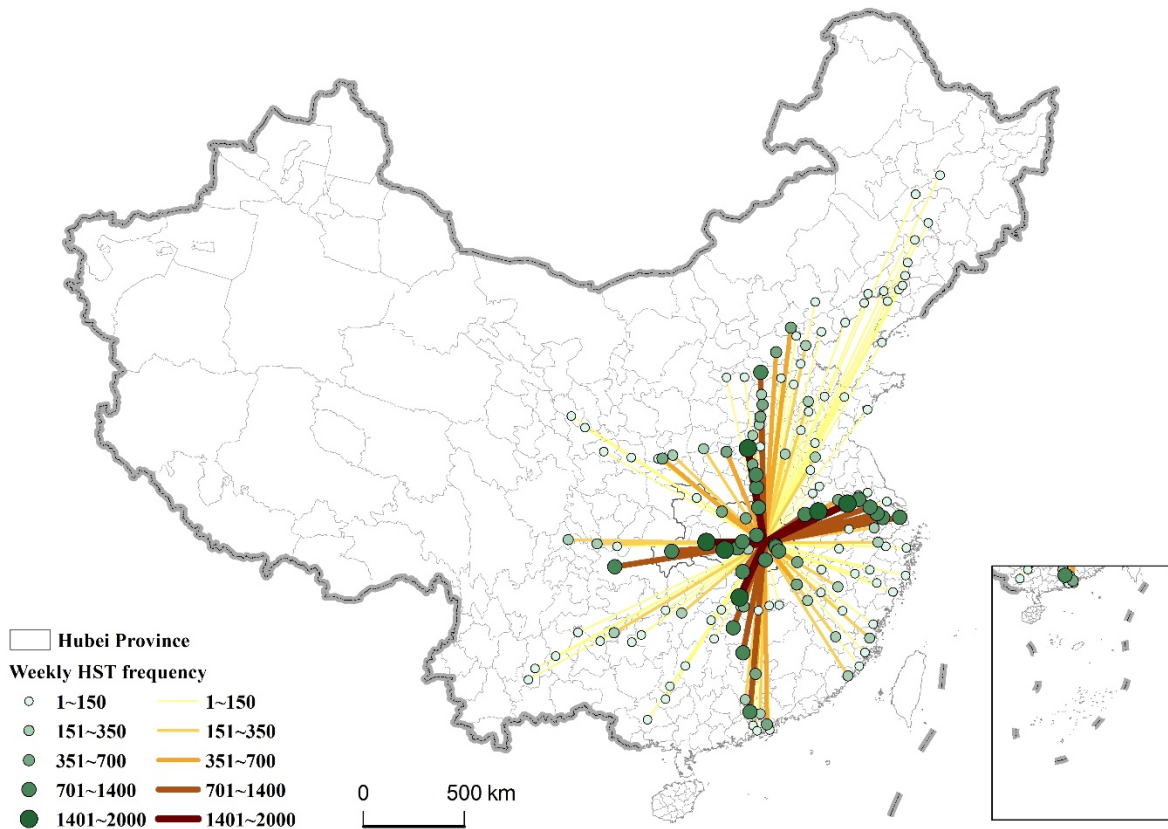


Figure 2: Weekly HST frequency out of Wuhan (2019 data, source: www.12306.cn)

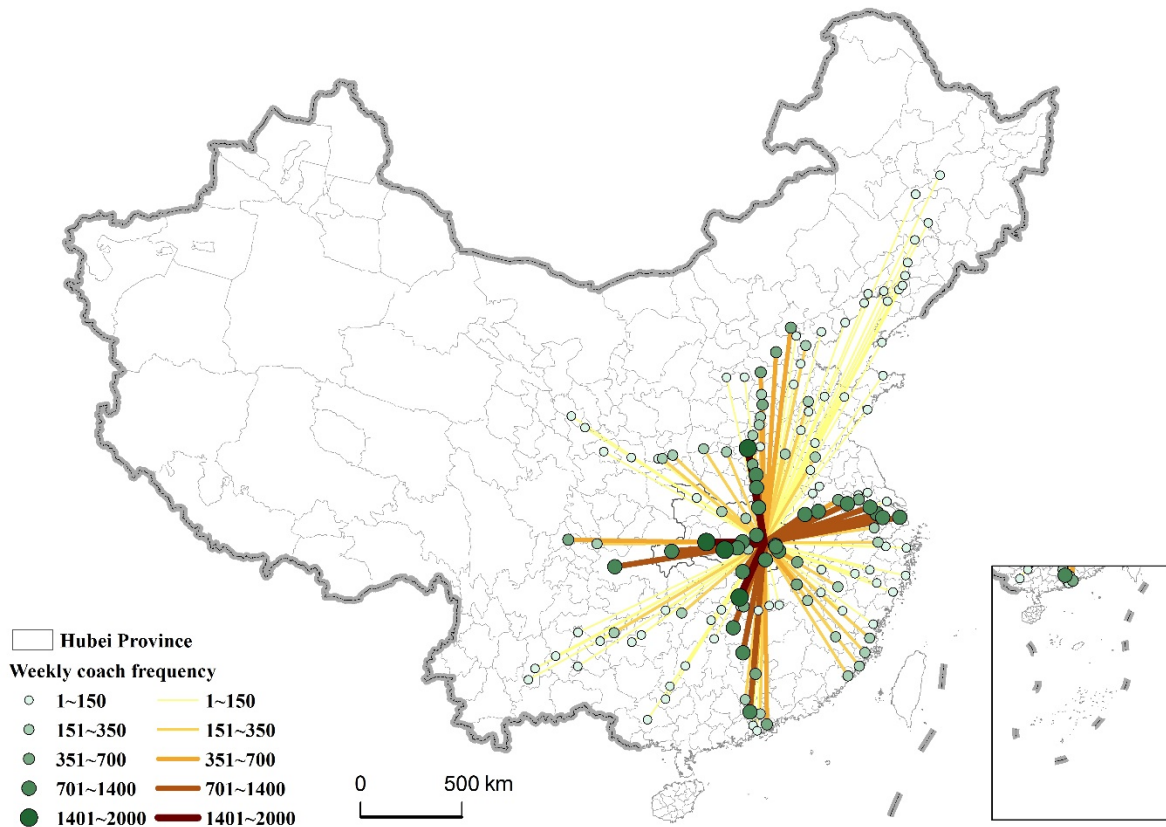


Figure 3: Weekly coach frequency out of Wuhan (2018 data, source: www.cncn.com)

As a well-connected city, Wuhan is the economic and education centre of Central China with dozens of universities and colleges. It has a population of 14 million including 3 million floating population, most of whom are migrant workers from other cities. It also has a large tertiary student population of about 1.3 million. A large outflow of these migrant workers and tertiary students occurred from 14 January 2020 and peaked on 22 January 2020 before the closure of the city. According to The Paper (2020), by 23 January, five million people left Wuhan via different transport methods since 10 January 2020. The top-15 destinations were the cities within Hubei Province, with Huanggang and Xiaogan being the top two. The top-5 provinces were Henan, Hunan, Anhui, Chongqing, and Jiangxi. It is expected that the population that flowed out of Wuhan may have significantly impacted on the outbreaks in other cities. In fact, since the outbreak of COVID-19, tracking passengers from Wuhan and putting them into quarantine were the top priority of almost every city from late January to February 2020, as most of the confirmed cases were imported cases or close contacts of the imported cases.

2.2 Relevant studies

A few centuries ago, infectious diseases could spread only as fast and as far as people could walk, horses could gallop and ships could sail (Tatem et al., 2006). The emergence of modern transport modes has accelerated the spread of new diseases in an unprecedented speed and put more people at risk. In this century, human beings have been hit by a series epidemic diseases including SARS, H1N1 pandemic (in 2009), MERS, and most recently, COVID-19 at a global level, owing largely to the growing availability of affordable air travel. The risk of transmission of respiratory infections on airplanes was a major concern to the public and airline industry. Some studies suggested that air transport is a mode that contributes to the accelerating and amplifying influenza propagation, as such transmission can occur in the flight or at the airports (see a comprehensive survey by Browne et al., 2016). However, other studies on the ventilation systems and patient outcomes showed that the dissemination of pathogens during the flight occurs rarely (see a good survey by Leder and Newman, 2005). For example, quite a few studies suggested that in-flight transmission of SARS was not common. Wilder-Smith et al. (2003) reported that in-flight transmission occurred in one of the three flights with SARS patients on board. The authors note that the risk of in-flight transmission is lower than that reported for influenza, but may increase if super-spreaders are on board. However, while the transmission risk in the flight may be low, flights can carry people with the virus to new places. Bowen Jr and Laroe (2006) attempted to examine the link between air transport accessibility and the speed of SARS diffusion. The accessibility to China was measured by flight frequency and directness of scheduled airlines services. They found that airline network accessibility was a determinant of the speed with which SARS arrived in infected countries, but its influence diminished in later weeks of the outbreak due in part to the public health measures such as health screening at the airports.

Research into the role of ground transport in the propagation of infectious diseases is limited. Among the few studies, Troko et al. (2011) found that the use of public buses or trams is a significant factor for the acquisition of acute respiratory infection. The risk can be greater among occasional bus or tram users. Therefore, it is important to exercise good respiratory hygiene and refrain from making unnecessary travels by public transport during the pandemic outbreak season. Cui (20011) confirmed that trains may have played an important role in the

2009 influenza pandemic as a result of the close contact and long exposure in the train cars. Using the laboratory-confirmed cases of H1N1 for the period from May 2009 to April 2010, Cai (2019) evaluated the effects of airports and railway stations on arrival days and peak days of the virus. The authors found that airports and railways stations in Chinese prefectures significantly advanced the arrival days, but they were not significantly associated with the peak days of the pandemic.

Du et al. (2020) is one of the few studies that estimated the probability of transportation of COVID-19 from Wuhan to 369 other cities in China before the quarantine. They expected that in 130 cities, the COVID-19 risk was greater than 50% and this risk was greater than 99% in the four largest metropolitan areas. However, as with many other studies on COVID-19, they only attributed the spread of COVID-19 to the travel or movement of people without looking at how different transport modes have specifically contributed to the dissemination of the virus. The present research aims to address this issue.

3. Methodology

The gravity model is used to identify the factors, particularly different transport modes, associated with the number of COVID-19 cases in Chinese cities. Our first dependent variable is the cumulative confirmed cases of COVID-19 reported on 15 February 2020, most of which were imported cases with a travel history to Wuhan.² This date was chosen because of the following several reasons. First, there was a shortage of testing kits for COVID-19 before mid-February and patients needed to meet strict criteria to become eligible for a test. Second, patients have an incubation period before various symptoms develop. The WHO estimates that the incubation period ranges from 1 to 14 days, with a median of 5 to 6 days.³ China reported a case with 24 days of incubation. In addition, the quality of the testing kits from the early days may not have been satisfactory, as they reported a high proportion of false-negative results. The definition of a confirmed case set by China's National Health Commission changed over time. From January to March 2020, seven versions of the

² It should be noted that the imported COVID-19 patients could be infected in Wuhan or during the trip from Wuhan to the destination city.

³ <https://www.health.gov.au/sites/default/files/documents/2020/03/coronavirus-covid-19-information-for-clinicians.pdf>.

Novel Coronavirus Pneumonia Diagnosis and Treatment Plan were issued. The fifth edition even recommended that Hubei Province use CT scan results as confirmation of the infections in order to have a quick identification in the presence of high false-negative risk. Although by 15 February, there had been community-transmission cases, the percentage was relatively low and most of these cases were associated with close contact with those who had travel history to Wuhan. It would not distort the results significantly by including them in the imported cases. As a robustness check, we also use the number of cases on 1 February 2020 as the dependent variable, but it should be noted that the number of cases on this date may not accurately reflect the true cases for the abovementioned reasons.

The imported cases are analogous to the trade flows in the gravity model, which has been widely used in cross-country empirical analyses of international trade flows. The gravity model has also been applied to air transport to identify the determinants of bilateral air passenger or air cargo flows (Zhang and Lu, 2013; Zhang and Zhang, 2016). It has been reported that a large portion of the international trade matrix consists of zero trade, either because data are missing or some economies simply do not trade (Helpman et al., 2008). This is also the case in our study where about one tenth of the cities report zero cases during our study period. The problem of zero or missing trade flows when using gravity models has long been ignored in the international trade literature (Zhang et al., 2018). Many researchers simply discard the zero flows from the sample, which results in the loss of information. Some add a small constant value to the zero values to allow the estimation of log-linear equations (van Bergeijk and Brakman, 2010), but this method does not have a theoretical basis.

A more acceptable approach was introduced in Santos Silva and Tenreyro (2006), where Poisson pseudo-maximum likelihood estimator (PPML) technique is proposed. This approach can deal with the zero problem in the dependent variable and is also capable of coping with the heteroscedasticity problem that is common in the gravity data, in which case, the parameters of log-linearised gravity models estimated by ordinary least squares (OLS) will be highly misleading. Other nonlinear transformations such as the nonlinear least squares method and Tobit regressions are also not working if the errors are heteroscedastic. Therefore, a gravity type model with the PPML estimation approach is used in this study.

The empirical gravity model employed is expressed as follows:

$$CASE215 = a_0 + a_1 \ln GDP + a_2 AIR + a_3 COACH + a_4 HST + a_5 HUB + a_6 \ln LAT + a_7 \ln LONG + a_8 \ln DIST + \mu$$

The variables are explained as follows:

- CASE215 is the number of cumulative confirmed COVID-19 cases on 15 February 2020 in each city. We will also replace it with the data of 1 February 2020 (CASE201). The data are from China National Health Commission's daily report.
- GDP is the gross domestic product of each city. The data are obtained from China City Statistics Book 2018 and the 2017 data are used.
- AIR denotes weekly flight frequency from Wuhan to each city in 2019. The data source is Official Aviation Guide (OAG). Bowen Jr. and Laroe(2006) argue that it is important to establish the relationship between the diffusion of the disease and the schedule data as in a potential pandemic crisis, the schedule data are easily obtained while the actual passenger data are only available much later.⁴
- COACH denotes weekly coach (bus) services from Wuhan to each city. The information is extracted from Xinxin Travel (www.cncn.com) using 2018 data.
- HST denotes weekly HST frequency from Wuhan to each city in late 2019. The data are taken from China Railway's booking website (www.12306.cn).
- HUB is a dummy taking the value of one if a city has an airport or HST station, and zero otherwise.
- $\ln LAT$ is the latitude of each city's administration centre in logarithm. $\ln LONG$ is the longitude of each city's administration centre in logarithm. These two variables are included because there have been reports suggesting that the virus is most active in a temperature range between 5 to 11 degrees Celsius (Sajadi et al., 2020). Sajadi et al. (2020) claim that COVID-19 has established a significant community spread pattern in cities and regions along the 30-50 degree north latitude corridor.
- DIST is the great circle distance from Wuhan to each city calculated in ArcGIS. We can also replace this variable with the shortest travel time (TIME)

⁴ We are constrained by the availability of the Spring Festival season data. However, the increase in flight frequency during the Spring Festival season might be a proportional increase across the board for all cities, which implies that distortion caused by using the data of previous periods could be very minimal.

variable from Wuhan to each city via highways. This variable can possibly capture the effect of using private cars. We employ Baidu map API to calculate the shortest travel time by driving in 2018.

We are also interested in the link between the transport modes and the emergence of the first case of COVID-19 in each city, which is to be referred to as the speed of the transmission. Following Brockmann and Helbing (2013) and Cai et al. (2019), we define an arrival day (ARRDAY) variable as the number of days from 10 January 2020 to the date of the first confirmed case in each city until 20 February 2020 (after which the reported new cases in many cities started to decline). In 2020, 10 January was the first day of the 40-day Chinese New Year travel season (or called, in Chinese, the “Chunyun” period). It usually starts 15 days before the Chinese New Year and represents the largest annual human migration in the world. By the time Wuhan was in quarantine, China had been two weeks into the 40-day Spring Festival travel, during which people made millions of trips, which could help spread the COVID-19 pandemic. We adopt a quantile regression model with ARRDAY being the dependent variable and the same set of independent variables as in the gravity model included on the right-hand side of the equation.

The quantile regression was first introduced by Koenker and Bassett (1978) and a comprehensive review of this method can be found in Koenker (2005). It is an extension of the traditional linear regression model used when some conditions are not met, such as in the presence of heteroscedasticity. The quantile regression is also more robust to outliers than traditional regression as it estimates the conditional median (or other quantiles) of the dependent variable instead of conditional mean as the OLS regression does. We will estimate the model at the 0.25, 0.5, and 0.75 quantiles.

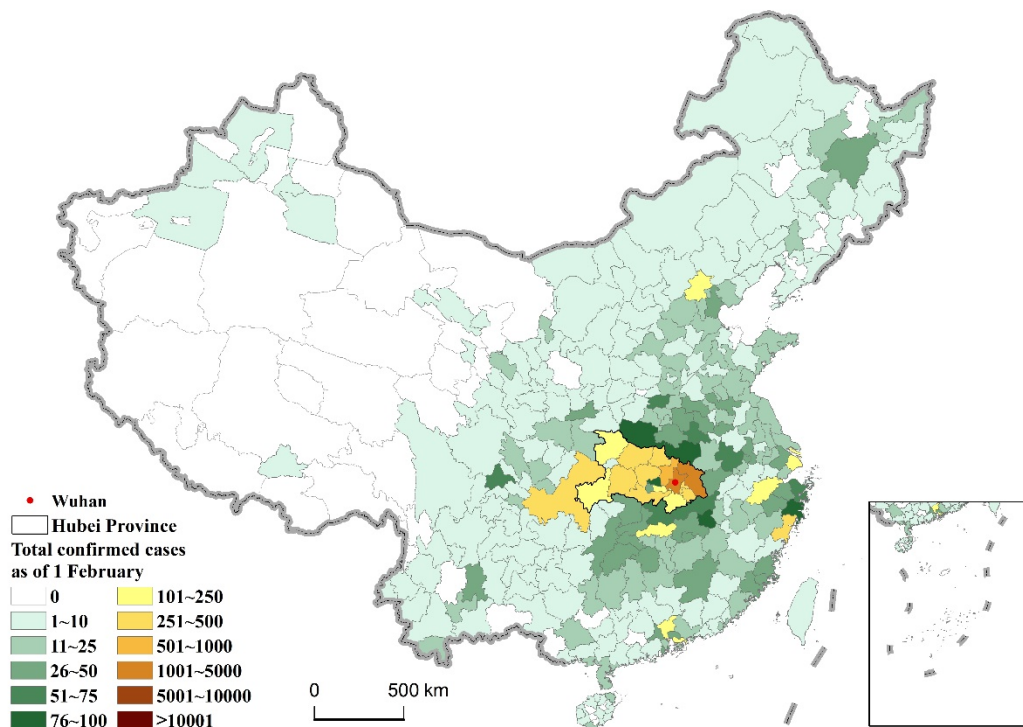
4. Results

Before we run the regressions, we conduct a simple t-test to see whether cities with and without a particular transport hub showed a significant difference in the number of cases using the 15 February data. A city is called a transport hub in this study if it has an airport or an HST station. The Levene test is used to check the equality of variance of the data. The Levene test is not significant at the 5% level but significant at the 10% level. Therefore, the homogeneity of variance assumption can be assumed at 5%. We find that cities without

airports or HST stations recorded an average of 30.7 cases, while those with airports or HST stations reported an average of 87.7. The difference is not statistically significant at a significance level of 5% if equal variances are assumed. Interestingly, if unequal variances are assumed, the difference is statistically significant. This is also the case when the 1 February 2020 data are used. Therefore, it seems that the link between the presence of an airport or HST station and the number of cumulative confirmed cases is not very strong based on the t-test. We should therefore resort to the multiple regression results and include more transport-related variables.

For the arrival day (ARRDAY) variable, we find that cities with airports or HST stations recorded its first case of COVID-19 within a significantly shorter time (17.7 days) calculated from 10 January 2020, than those without such facilities (24.8 days) at the level of 1%. It seems that a transport infrastructure such as airport or HST station can speed up the spread of the virus.

Table 1 reports the descriptive statistics of the main variables, which shows that the number of cases almost tripled from 1 February to 15 February 2020. The average arrival time of the first confirmed cases is 18.6 days. Figure 4 presents the distributions of the cumulative COVID-19 cases on 1 February and 15 February 2020.



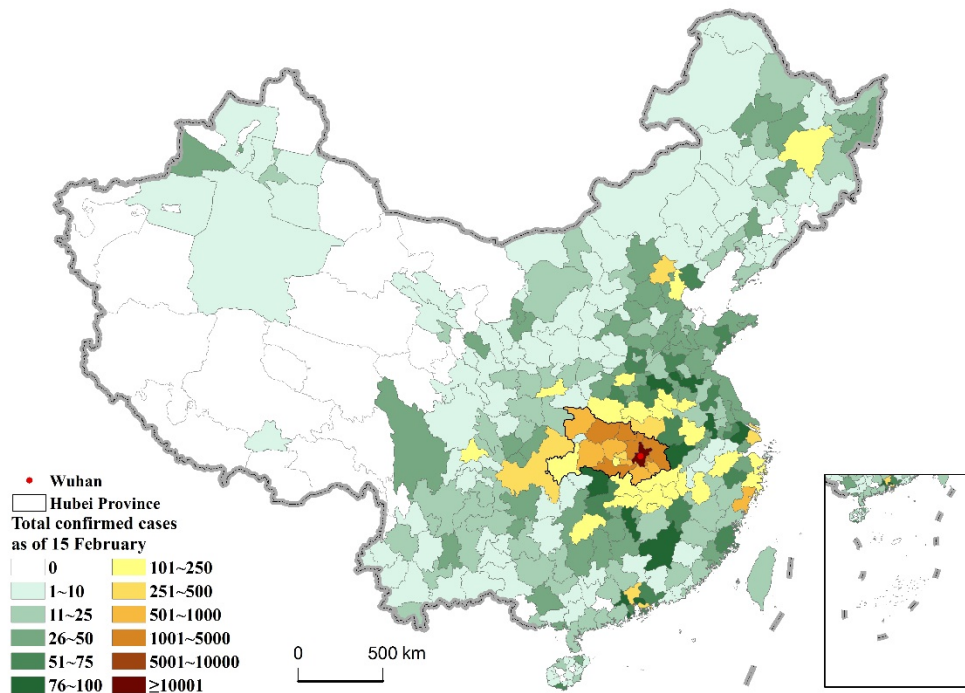


Figure 4: Distributions of COVID-19 cases on 1 and 15 February 2020.

Table 2 reports the results of the gravity model. Specifications (1) to (4) use the number of cases on 15 February as dependent variable while specification (5) uses that on 1 February. Specifications (3) and (4) include the HUB dummy while specifications (1) and (2) do not. The distance variable is included in specifications (1), (3) and (5) while in specifications (2) and (4), this variable is replaced by travel time (TIME). As can be seen, all the specifications have the expected signs, and the levels of significance are largely consistent. However, the presence of an airport or an HST station does not have a significant impact on the number of cases after other variables are controlled for. It is understood that the effect of the HUB dummy can be diluted as it might be correlated with other variables. For example, cities with airport or HST stations tend to be large or at least medium-sized cities, and thus have higher GDP. The correlation matrix of all the independent variables suggest that moderate correlations exist between some variables, but none of them are greater than 0.6. The problem of multicollinearity may arise, but the consequence is not serious particularly because one of the purposes of building the gravity is for future prediction purpose. In fact, multicollinearity does not result in biased estimations. Achen (1982) pointed out that the only effect of multicollinearity is that it makes it hard to get coefficient estimates with small standard error and thus can lead to insignificance of coefficients.

Flight frequency is positively and significantly associated with the number of confirmed cases at the level of 1%. If the weekly frequency increases by one flight, we would expect that the number of the cases increases by 1% according to specification (3) in Table 2. The frequency of HST also has an impact on the number of cases at the 10% level for specifications (1) to (4). An increase in HST frequency by one unit is associated with a much smaller increase in the number of cases compared with air travel. Coach services seem not to have any significant links with the imported cases, which is a bit surprising. One possible reason is that the vast majority of coach services are for short trips. Unlike air travel that requires passengers to report to the airport at least one hour before the departure, and go through several formalities before boarding, Coach passengers normally arrive at the station a short time before the departure and during the trip, they tend to stay in their seats and do not move around as they would do on the HST train. This implies that the risk of coach passengers' exposure to the COVID-19 virus is relatively low.

The distance and travel time are negatively and significantly related to the confirmed cases, which is consistent with the COVID-19 distribution pattern. Considering that Wuhan is located in Central China and it is only a few hours' travel from this city to other major metropolitan areas by HST or air, the wide spread of COVID-19 in China in such a short time is understandable. If the source of infection was located in a city in far north or far west China, it would be a different story. The latitude and longitude coordinates, a proxy of the city location and temperature, are not quite relevant in the diffusion of COVID-19, as none of them is statistically significant once other factors are controlled for.

Table 1: Descriptive statistics (Obs.: No. of observations, Std.Dev.: Standard deviation)

Variable	Obs.	Mean	Std.Dev.	Min	Max
CASE215 (number of cases on 15/2/20)	360	80.14	281.94	0	3201
CASE201 (number of cases on 1/2/20)	360	28.53	86.19	0	1002
ARRDAY (number of days)	360	18.63	8.16	11	41
GDP (Chinse yuan in billions)	360	150.63	338.48	2.6	3063.3

AIR (weekly frequency)	360	9.72	29.66	0	238
HST (weekly frequency)	360	122.98	259.15	0	1547
COACH (weekly frequency)	360	62.50	248.92	0	2170
HUB (presence of airport or HST station)	360	0.87	0.34	0	1
LAT (latitude degree)	360	32.62	7.31	18.25	50.41
LONG (longitude degree)	360	111.18	9.96	75.99	131.15
DIST(km)	360	1062.14	693.85	52.37	3550.72
TIME (hour)	360	16.29	11.44	1.23	73.86

Table 2: The results of the gravity model with PPML estimation

Variable	(1) CASE215	(2) CASE215	(3) CASE215	(4) CASE215	(5) CASE201
lnGDP	0.1691 (0.1101)	0.1364 (0.1164)	0.1801 (0.1127)	0.1422 (0.1177)	0.2923** (0.1141)
AIR	0.0090*** (0.0023)	0.0101*** (0.0024)	0.0090*** (0.0023)	0.0101*** (0.0024)	0.0092*** (0.0023)
HST	0.0006* (0.0003)	0.0005* (0.0003)	0.0006* (0.0003)	0.0005* (0.0003)	0.0004 (0.0003)
COACH	0.0002 (0.0003)	0.0002 (0.0002)	0.0002 (0.0003)	0.0002 (0.0002)	0.0000 (0.0003)
lnLAT	0.3456 (0.4476)	0.0442 (0.4464)	0.2931 (0.4637)	0.0062 (0.4603)	-0.2449 (0.3852)

lnLONG	-0.7624 (1.8825)	-1.2816 (2.2591)	-0.6953 (1.9158)	-1.2156 (2.2851)	-1.7785 (2.0926)
lnDIST	-1.3942*** (0.2473)		-1.4056*** (0.2471)		-1.4732*** (0.2692)
lnTIME		-1.6229*** (0.2704)		- 1.6677*** (0.2701)	
HUB			-0.2924 (0.4712)	-0.1824 (0.4076)	-0.2583 (0.5006)
R-squared	0.74	0.73	0.74	0.74	0.61
Obs.	360	360	360	360	360

*Significant at 10%; ** significant at 5%; ***significant at 1%.

The results of the quantile regressions at the 0.25, 0.5 and 0.75 quantiles are presented in Table 3. The arrival days for the three quantiles are 14, 16 and 19, respectively. We can see that the magnitude and significance of the coefficients can change across quantiles. However, irrespective of the quantile levels, the longitude coordinates are negatively and significantly associated with the arrival days, probably because Wuhan's economic links with cities in the east are closer. For example, Wenzhou, an east coastal city in Zhejiang, reported a large number of imported COVID-19 cases. Wenzhou is famous for its savvy business people scattered across China and the world. About 170,000 business people worked in Wuhan (Ni, 2020), which may have contributed to the quick spread of the COVID-19 virus in Wenzhou when some of them travelled back before the Chinese New Year. The latitude values are statistically significant in the lower quantile range, implying that warmer places may detect the first case of COVID-19 earlier than cities in the north.

The distance variable is positively related to the arrival day variable and statistically significant, implying that the farther the distance from Wuhan, the later the emergence of the confirmed COVID-19 case. GDP has a significantly negative impact on the speed of the spread and its impact is larger for higher quantiles. The air transport frequency could have facilitated the spread of COVID-19 at lower quantiles, but it is not significantly associated with the arrival day variable at the 0.5 and 0.75 quantiles. The frequencies of HST and coach services are not statistically significant. The HUB variable, or the presence of an airport or an

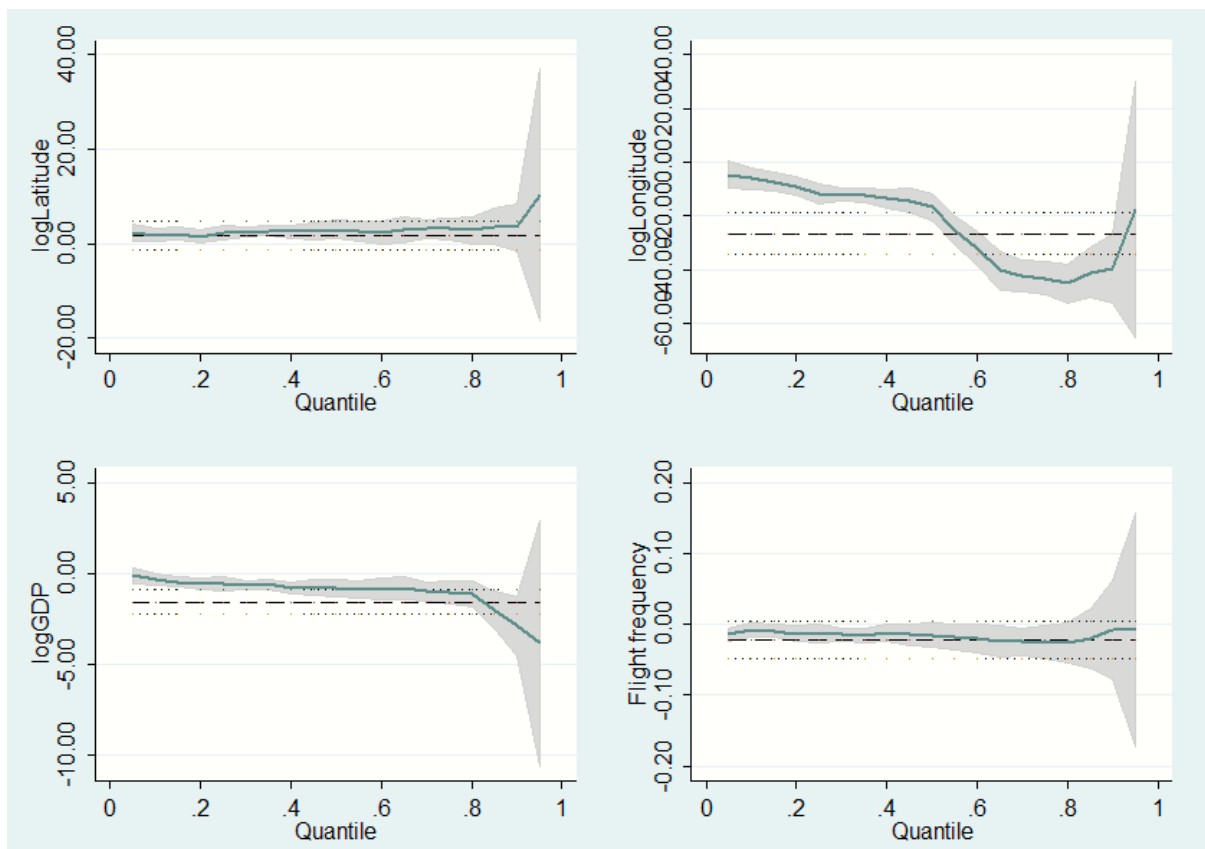
HST station can increase the speed of the COVID-19 transmission as this variable is consistently negative and significant at medium and higher quantiles. This is consistent with Cai et al. (2019) who reported that the presence of airports or high-ranking railway stations in Chinese prefectures significantly advanced arrival days of the H1N1 pandemic. Note that the interpretation of these coefficients in quantile regressions is the same as that for the OLS estimates. For example, at the 0.25 quantile, every one unit increase in flight frequency will lead to a reduction in the arrival days by 0.013.

Table 3: Quantile regression showing the determinants of the 0.25, 0.5, and 0.75 quartiles of arrival days

Variable	0.25 (14 days)	0.5 (16 days)	0.75 (19 days)
lnGDP	-0.5721*** (0.16648)	-0.8046** (0.3678)	-1.0380** (0.5140)
AIR	-0.0131** (0.0064)	-0.0151 (0.0141)	-0.0253 (0.0196)
HST	0.0002 (0.0006)	0.0003 (0.0013)	0.0007 (0.0018)
COACH	0.0007 (0.0007)	0.0006 (0.0017)	0.0025 (0.0024)
lnLAT	2.2785*** (0.7137)	3.0727* (1.5768)	3.1441 (2.2040)
lnLONG	-11.8340*** (1.8597)	-16.7247*** (4.1087)	-43.0951*** (5.7428)
lnDIST	0.8366** (0.3437)	1.4001* (0.7594)	3.5963*** (1.0614)
HUB	-0.4483 (0.4748)	-2.4447** (1.0489)	-6.1464*** (1.4661)
Pseudo R-squared	0.14	0.18	0.34
Obs.	360	360	360

*Significant at 10%; ** significant at 5%; ***significant at 1%.

The effects of the variables of GDP, flight frequency, distance, hub, latitude, and longitude coordinates can be more clearly seen from the graphs in Figure 5. The graphs show the change in quantile coefficients along with the confidence interval of 95% (the shaded area). The quantile process plots allow us to readily identify which predictors are associated with different parts of the response distribution. The black dashed lines are the OLS estimates and their confidence intervals. The coefficient of the log-transformed latitude does not vary much and is no different from its OLS estimate for most of the quantiles less than 0.8. The quantile process plot for the log-transformed longitude variable is quite interesting. The coefficient decreases first but starts to increase after the 0.8 quantile. Its coefficient in the lower quantiles (less than 0.5) well exceeds the OLS estimates. In the flight frequency panel, we can see that the coefficient does not differ much from the OLS estimates and is relatively stable across quantiles. The positive effects of the distance become stronger as the quantile increases. In contrast, the negative effect of the transport hub gradually becomes significant and stronger as the quantile rises.



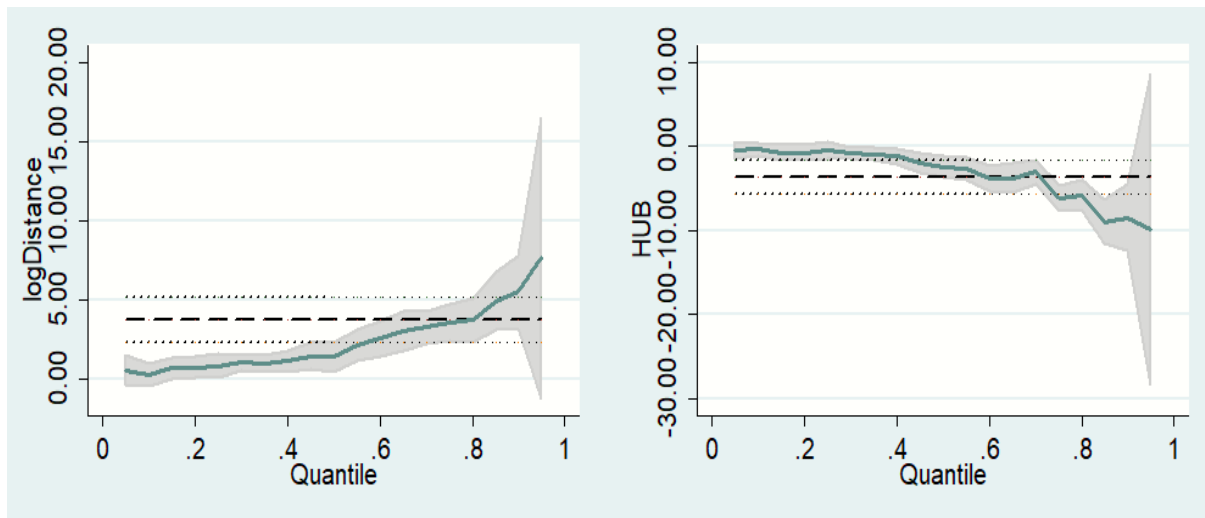


Figure 5: Quantile process regression of the arrival days of COVID-19.

5. Discussion and conclusion

To understand the roles of different transport modes in the spread of COVID-19 pandemic across Chinese cities, we have examined the factors influencing the number of imported cases and the spread speed, and pattern, of the pandemic across Chinese cities. We found that the frequencies of air transport and HST services out of Wuhan are significantly associated with the number of COVID-19 cases in the destination cities, while the frequency of coach services is not quite relevant. The presence of an airport or HST station is significantly related to the spread speed of the pandemic, but its link with the number of confirmed cases is quite weak and actually disappears when other factors are controlled for. These results suggest that reducing air frequency from a pandemic centre might be the first measure that can be taken to slow down travel-related imported infections. This was actually the first response of many countries to the COVID-19 outbreak in China. For example, Australia, Singapore and the US closed borders to Chinese nationals and substantially reduced the number flights from China in late January and early February 2020. By doing so, these countries had successfully contained the number of imported COVID-19 cases from China in February 2020. Unfortunately, Australia did not do the same to the US and European countries in a timely manner, which resulted in a rapid increase in the number of imported cases in early March 2020. In late March, China had successfully controlled the spread of COVID-19 pandemic within its border as a result of the strict quarantine measures; however, imported cases from other countries became a major concern. The government therefore decided to limit the number of international flights to China. Chinese airlines were only

allowed to operate one route to any specific country with one flight each week. Each foreign airline was only allowed to operate one route to China with one flight per week. With these measures, China has quickly brought the imported cases down to a manageable level.

We also found that the magnitude of the HST frequency coefficient is much smaller than that of the flight frequency in the gravity model. This may have implications to the measures used to contain the pandemic. Many countries have adopted a combination of containment and mitigation measures to delay the surges of patients and flatten the curve to avoid overwhelming the health care system. One extreme measure is to lock down a city and cut off all the transport links with other cities. This normally involves the shutdown and halt of most businesses and outdoor activities, which comes at a huge economic cost. Another extreme is to achieve herd immunity. This option allows the infections to rise in a controlled way until 60% of the population gets infected, after which it would be harder for the virus to spread (Chang, 2020). Between the two methods, a discretionary control policy can be used. For the transport links, these measures can be considered: reduce the air transport frequency without cancelling all the flights; cut air services first while keeping most of the HST services, considering the lesser impact of HST in spreading the disease; tighten the restrictions on residents' movements when the risk is high and loosen a bit when the pandemic is under control.

This research reports some interesting spatial distribution patterns of the COVID-19 cases. The farther the distance from Wuhan, the lower number of cases in a city and the slower speed for the pandemic to be disseminated. The longitude and latitude coordinates do not have an impact on the number of cases, but are significantly associated with the speed of the COVID-19 spread. Specifically, cities in the east tended to record a COVID-19 case earlier than their west counterparts. Cities in South China may detect the first case earlier than those in North China. The pandemic might spread in large cities first before they arrive to small cities, as GDP is a factor positively associated with the spread speed. All these results suggest that the outbreak of a pandemic in a large city lying in the centre of a country such as Wuhan can have devastating impact on the whole nation and that the extreme measure to lock down the city might have been a correct move to reduce such impact on other parts of China. This is particularly important to Hubei's neighbouring provinces and the well-developed areas in East and South China such as Shanghai, and Guangdong.

Acknowledgement:

The authors wish to thank Haoran Yang and Delin Du for collecting some of the research data, and Tao Li and Jessica Zhang for helpful comments. This work is financially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant XDA19040402).

References:

- Achen, C. H. (1982). *Interpreting and using regression* (Vol. 29). Sage.
- Ai, S., Zhu, G., Tian, F., Li, H., Gao, Y., Wu, Y., ... & Lin, H. (2020). Population movement, city closure and spatial transmission of the 2019-nCoV infection in China. *medRxiv*.
- Bowen Jr, J. T., & Laroe, C. (2006). Airline networks and the international diffusion of severe acute respiratory syndrome (SARS). *Geographical Journal*, 172(2), 130-144.
- Browne, A., St-Onge Ahmad, S., Beck, C. R., & Nguyen-Van-Tam, J. S. (2016). The roles of transportation and transportation hubs in the propagation of influenza and coronaviruses: a systematic review. *Journal of travel medicine*, 23(1), tav002.
- Brockmann, D. & Helbing, D. (2013). The hidden geometry of complex, *Network-Driven Contagion Phenomena*. *Science*, 342, 1337–1342.
- Cai, J., Xu, B., Chan, K. K. Y., Zhang, X., Zhang, B., Chen, Z., & Xu, B. (2019). Roles of Different Transport Modes in the Spatial Spread of the 2009 Influenza A (H1N1) Pandemic in Mainland China. *International journal of environmental research and public health*, 16(2), 222.
- Chang, C. (2020). The decision Australia needs to make now on how we end the coronavirus epidemic. Available at www.news.com.au.
- Du, Z., Wang, L., Cauchemez, S., Xu, X., Wang, X., Cowling, B. J....Meyers, L. (2020). Risk for Transportation of Coronavirus Disease from Wuhan to Other Cities in China. *Emerging Infectious Diseases*, 26(5), 1049-1052.
- Fan, X. (2020). Tracing the 5 million people leaving Wuhan. Available at <https://www.jiqizhixin.com/>. Accessed 1 March 2020.

- Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The quarterly journal of economics*, 123(2), 441-487.
- Jiao J., & Wang J. (2017). Impacts of high-speed rail lines on the city network in China. *Journal of Transport Geography*, 60, 257-266.
- Koenker. (2005). *Quantile Regression*. Cambridge university press, New Year.
- Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 46(1), 33-50.
- Leder, K., & Newman, D. (2005). Respiratory infections during air travel. *Internal medicine journal*, 35(1), 50-55.
- Li, T., & Rong, L. (2020). A comprehensive method for the robustness assessment of high-speed rail network with operation data: A case in China. *Transportation Research Part A: Policy and Practice*, 132, 666-681.
- Li, T., Rong, L., & Yan, K. (2019). Vulnerability analysis and critical area identification of public transport system: A case of high-speed rail and air transport coupling system in China. *Transportation Research Part A: Policy and Practice*, 127, 55-70.
- Ni, D. (2020) How Wenzhou, 900 km from Wuhan, went into total lockdown. Sixth Tone. Available at <https://www.sixthtone.com/news/1005197/how-wenzhou%2C-900-km-from-wuhan%2C-went-into-total-lockdown>. Accessed 1 April, 2020.
- Sajadi, M. M., Habibzadeh, P., Vintzileos, A., Shokouhi, S., Miralles-Wilhelm, F., & Amoroso, A. (2020). Temperature and latitude analysis to predict potential spread and seasonality for COVID-19. *Available at SSRN 3550308*.
- Tatem, A. J., Rogers, D. J., & Hay, S. I. (2006). Global transport networks and infectious disease spread. *Advances in parasitology*, 62, 293-343.
- The Paper (2020). Five million people left Wuhan: An escape or normal travel? 28 January 2020. Available at https://www.thepaper.cn/newsDetail_forward_5654383. Accessed 1 March 2020.
- Troko, J., Myles, P., Gibson, J., Hashim, A., Enstone, J., Kingdon, S., ... & Van-Tam, J. N. (2011). Is public transport a risk factor for acute respiratory infection?. *BMC Infectious Diseases*, 11(1), 16.

van Bergeijk, P., & Brakman, S. (2010). *The Gravity Model in International Trade: Advances and Applications*, Cambridge University Press, Cambridge, England.

Wang, K., Xia, W., & Zhang, A. (2017). Should China further expand its high-speed rail network? Consider the low-cost carrier factor. *Transportation Research Part A: Policy and Practice*, 100, 105-120.

Wang J., Mo H., Jin F., & Wang F. (2011). Exploring the network structure and nodal centrality of China's air transport network- A complex network approach. *Journal of Transport Geography*, 19(4), 712-721.

Wang J., Du D., & Huang J. (2020). Inter-city connections in China: High-speed train vs. inter-city coach. *Journal of Transport Geography*, 82, DOI:10.1016/j.jtrangeo.2019.102619.

Wilder-Smith, A., Paton, N. I., & Goh, K. T. (2003). Low risk of transmission of severe acute respiratory syndrome on airplanes: the Singapore experience. *Tropical Medicine & International Health*, 8(11), 1035-1037.

Zhang, A., Wan, Y., & Yang, H. (2019). Impacts of high-speed rail on airlines, airports and regional economies: A survey of recent research. *Transport Policy*, 81, A1-A19.

Zhang, Y., Lin, F., & Zhang, A. (2018). Gravity models in air transport research: A survey and an application. In *Handbook of International Trade and Transportation*. Edward Elgar Publishing.

Zhang, Y., & Lu, Z. (2013). Low cost carriers in China and its contribution to passenger traffic flow. *Journal of China Tourism Research*, 9(2), 207-217.

Zhang, Y., & Zhang, A. (2016). Determinants of air passenger flows in China and gravity model: deregulation, LCCs, and high-speed rail. *Journal of Transport Economics and Policy*, 50(3), 287-303.

Zhao, X., Liu, X., & Li, X. (2020). Tracking the spread of novel coronavirus (2019-nCoV) based on big data. *medRxiv*.

Zhu, Z., Zhang, A., & Zhang, Y. (2018). Connectivity of intercity passenger transportation in China: A multi-modal and network approach. *Journal of Transport Geography* 71, 263-276.

Zhu, Z., Zhang, A., & Zhang, Y. (2019). Measuring multi-modal connections and connectivity radiations of transport infrastructure in China. *Transportmetrica A: Transport Science*, 15(2), 1762-1790.