Geospatial Health 2022; volume 17(s1):1060

Impact of socio-economic environment and its interaction on the initial spread of COVID-19 in mainland China

Mao Guo,[1,2] Lin Yang,1 Feixue Shen,[1] Lei Zhang,[1] Anqi Li,[1] Yanyan Cai,[1] Chenghu Zhou,[1-3]

[1]School of Geography and Ocean Science, Nanjing University, Nanjing; [2]Collaborative Innovation Centre of South China Sea Studies, Nanjing University; [3]State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing, China

Abstract

Coronavirus disease 2019 (COVID-19) has strongly impacted society since it was first reported in mainland China in December 2020. Understanding its spread and consequence is crucial to pandemic control, yet difficult to achieve because we deal with a complex context of social environment and variable human behaviour. However, few efforts have been made to comprehensively analyse the socio-economic influences on viral spread and how it promotes the infection numbers in a region. Here we investigated the effect of socio-economic factors and found a strong linear relationship between the gross domestic product (GDP) and the cumulative number of confirmed COVID-19 cases with a high value of R^2 (between 0.57 and 0.88). Structural equation models were constructed to further analyse the social-economic interaction mechanism of the spread of COVID-19. The results show that the total effect of GDP (0.87) on viral spread exceeds that of population influx (0.58) in the central cities of mainland China and that the spread mainly occurred through its interplay with other factors, such as socio-economic development. This evidence can be generalized as socio-economic factors can accelerate the spread of any infectious disease in a megacity environment. Thus, the world is in urgent need of a new plan to prepare for current and future pandemics.

Correspondence: Lin Yang, School of Geography and Ocean Science, Nanjing University, Nanjing, 210023, China. Tel.: 86.18115646683.
E-mail: yanglin@nju.edu.cn

Key words: COVID-19; socio-economic development; coronavirus spread; structural equation models; China.

Acknowledgements: this study is supported by the National Natural Science Foundation of China (Project No. 41971054), and the Leading Funds for the First class Universities (020914912203).

Received for publication: 25 November 2021.
Revision received: 4 March 2022.
Accepted for publication: 5 March 2022.

Geospatial Health 2022; 17(s1):1060
doi:10.4081/gh.2022.1060

Introduction

The coronavirus disease 2019 (COVID-19), first reported in Wuhan, Hubei Province, China in December 2019 and labelled as a pandemic by the World Health Organization (WHO) on 11 March 2020 (Worldometer, 2020), has become a worldwide threat. Two years have passed and we still face the challenge of COVID-19, which has impacted social and economic activities around the world (Bonaccorsi *et al*., 2020; Human development report, 2020; Guan *et al*., 2020). In consequence, global human development may have declined for the first time in the recent 30 years (Human development report, 2020). Source of infection, way of transmission and a susceptible population are the three main elements that decide how infectious diseases spread. Therefore, in addition to the reproduction of the virus itself, socio-economy factors related to these key elements play an important role in the spread of the pandemic. While population movements, population density (PD) and the situation of probable virus-endemic areas obviously impact its spread, the influence of the socio-economic factors remains unclear and so is the mechanism of interaction between these factors. An examination of how the influential factors are related to the spread of COVID-19 would not only be helpful for understanding the mechanism of multiple influential factors on the spatial spread of the pandemic, but would also provide insights into the ways of future prevention against similar infectious diseases (Enserink *et al*., 2020; Qiu *et al*., 2020).

Recent literature on the initial spread of the COVID-19 has highlighted the role of population movement, which is highly related to the level of social and economic development. In modern society, population movement between regions and inside regions is largely due to economic and social life including work, business, family, tourism, *etc*. Studies have found that cities with a higher population influx (PopInflux) from Wuhan usually had more confirmed cases in the early stage (Jia *et al*., 2020; Qiu *et al*., 2020; Zhang *et al*., 2020). Jia *et al*. (2020) developed a spatial-temporal exponential model using PopInflux to estimate the confirmed cases in prefectures, with the gross domestic product (GDP) used as an important index. Zhang *et al*. (2020) used a linear regression model to simulate the relationship between PopInflux and the imported cases suggesting that population movement plays an important role in the spread of COVID-19. These studies verified the importance of population movement in

the spread of the infection. However, the driving force of migration in this context has not been studied.

Studies examining the relationship between social-economic factors and the initial spread of infectious diseases have found that cities with higher GDP per capita have higher COVID-19 transmission rates (Chakraborti *et al*., 2020; Qiu *et al*., 2020; Sun *et al*., 2020). In addition, situation and scale of the PopInflux from Wuhan were contributing to the spread of the disease before Wuhan's shutdown (Wu *et al*., 2021). The expansion of the transportation networks also had a significant influence leading to increased transmission (Adda *et al*., 2020). Besides, since COVID-19 is airborne and supported by close contact, environments such as schools and offices promote transmission (Markowitz *et al*., 2019). However, only few studies have focussed on comprehensive quantification about how social-economic factors influence transmission and the complex interactions among these variables.

Structural equation model (SEM) is an effective causal analysis of the influential mechanism of the social-economic factors model, which can measure the relationship between multiple independent variables and multiple dependent variables (Grace *et al*., 2016). SEM has several advantages for causal analysis. First, it can manage observed variables and the so-called latent variables which cannot be measured directly (Yang *et al*., 2020). Second, SEM can deal with the error level of observed variables so that the estimation of correlations between latent variables are less affected by measurement errors. Third, SEM can analyse direct and indirect effects quantitatively, thereby disentangling complicated variable interactions (Grace *et al*., 2016). The technique has been widely used in soil examination (Angelini *et al*., 2016), environmental science (Hao *et al*., 2020) and ecology (Grace *et al*., 2016) and can therefore be a potential approach when attempting to analyse the influential mechanism(s) of the social-economic factors on virus transmission.

The main objective of this paper was to quantify the impacts of social-economic factors at the early stages of COVID-19 spread, including a study of the interaction of these factors and their direct and indirect effects. We first conducted correlation analysis and regression analysis to recognize the most representative influential factors on the cumulative number of confirmed cases (COVID) and then utilised the SEM models to identify the complex interconnection of socio-economic variables and their effects on transmission. The overall aim was to provide a future epidemiological perspective on COVID-19 leading to a better understanding of its transmission, thereby supporting government decision-making.

Materials and methods

We investigated the effect of social-economic variables on COVID-19 covering the key episodes of the spread of the virus in mainland China. The study involved 306 prefecture-level cities of the 31 provinces from 21 January to 27 February 2020 using SEM. The socio-economic variables, including GDP, PD, road accessibility, PopInflux and distance to the probable epicentre were selected as the influential factors. The samples of confirmed COVID-19 cases used covered the critical periods of virus incubation and transmission

Data

The raw data were collected from the repository of the Centre for Systems Science and Engineering (CSSE) at Johns Hopkins University (Dong *et al*., 2020). Wuhan was the epicentre and the infection number in Hubei Province was the highest among the 31 provinces. Cities in other provinces with high infection numbers included Beijing, Shanghai, Guangzhou, Shenzhen and Chongqing. In general, the infection numbers in southeast China were much higher than those in northwest China. The south-eastern coastal cities are economically advanced areas with a high PD and more developed transportation than other regions and there was, to some extent, a consistency between the severity of the infection and the level of urban development in the geographical distribution (Figure 1A and B).

Confirmed COVID-19 cases and time line

We explored the relationship between the socio-economic fac-

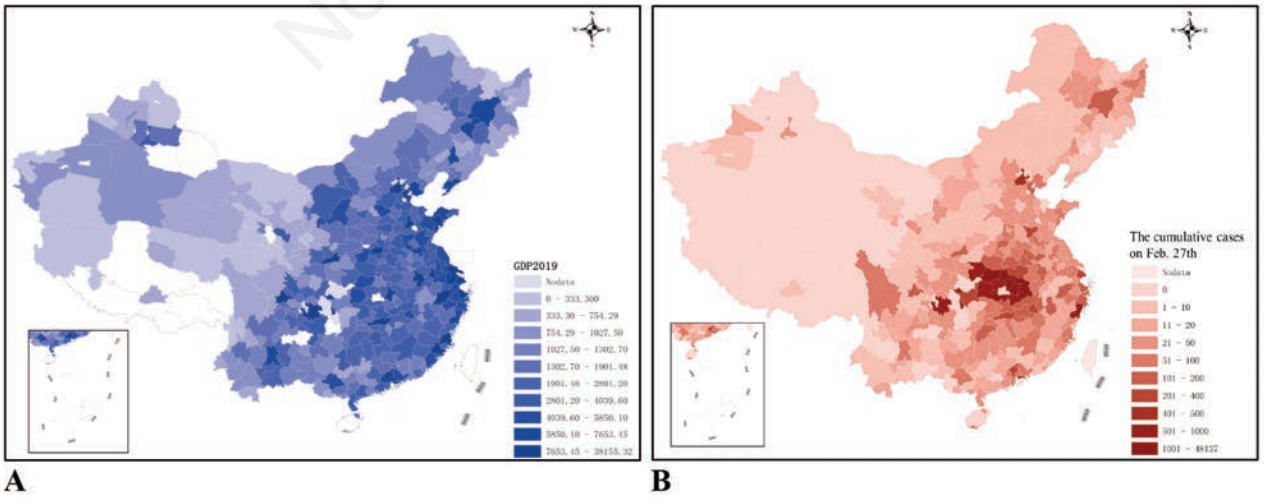

Figure 1. Spatial distribution of gross domestic product (GDP) and COVID-19 in China. A) GDP in 2019; B) the cumulative confirmed number of COVID-19 cases on 27 February 2020.

tors and the number of COVID-19 cases in 288 cities with confirmed cases in mainland China (excluding Hubei Province) and 35 central cities (excluding Wuhan). We also investigated the relationship in local regions, such as Hubei Province, the Yangtze River Delta Area, the Pearl River Delta Area, Jing-Jin-Ji, a big-city cluster in northern China and the Cheng-Yu Economic Circle.

We chose the dates 30 January 2020 (C130) and 6 February 2020 (C206) for SEM modelling based on the COVID-19 incubation period of 14 days (Dong *et al*., 2020; Bendavid *et al*., 2020). We added 13 February 2020 (C213), the 7th day after the lockdown of Wuhan, to have three days with seven days between them (C130, C206 and C213) for linear regression analysis.

Influential variables

Influential variables are the following (Table 1):

- *Gross Domestic Product (GDP)* was used as measure of production activities and economic strength. It is generally recognized as the best indicator to measure the economic situation of a country or a region (Ma *et al*., 2015; Zhang *et al*., 2020) during a certain period calculated according to the national market price (Zhang *et al*., 2020). When the GDP for 2019 was not available, GDP for 2018 was used as alternative (National Bureau of Statistics);
- *Population Density (PD),* the number of people per unit land area (Qiu *et al*., 2020; Wu *et al*., 2021), was downloaded for mainland China from the website www.worldpop.org at a resolution of approximately 1 km^2 at the equator (Lloyd *et al*., 2019) and calculated to city scale by aggregation. This dataset available at intervals of five years, with the currently latest data from 2020.
- *Road Accessibility (RoadAccess),* another reflection of economic development (Jia *et al*., 2020; Qiu *et al*., 2020), was generated from https://download.geofabrik.de. The RoadAccess grid map used was calculated from the main roads extracted from original road vector data for mainland China following Sanderson *et al*. (2002) and Venter *et al*. (2016) using score 8 for distances from the road at <500 m distance and exponentially decaying to 0 at >15 km. To obtain the mean RoadAccessof each city, the grid data were aggregated to city scale (Venter *et al*., 2016; Wu *et al*., 2021);
- *Distance to the epicentre (Dis2Wuhan),* the Euclidean distance from Wuhan to the destination city, was achieved by generating the gravity centre of each city and calculating the distances using ArcGIS 10.2 (ESRI, Redlands, CA, USA);
- *Population influx from Wuhan (PopInflux),* the quantity of population movement, was collected daily between 1 to 23 January 2020 from Wuhan to each city based on Baidu migration big data (http://qianxi.baidu.com/). Baidu uses the change of positioning data during a certain period to get the PopInflux data from the source region to the destination. The total daily PopInflux for each city data were calculated by aggregation (Wu *et al*., 2021).

Statistical approach

Data preprocessing before linear regression and structural equation modelling

The influential variables and COVID-19 on the dates C130, C206 and C213 were normalized by logarithm (base 10) transformation, and then changed into a normal distribution (Ghasemi *et al*., 2012). The PopInflux data obtained from Baidu migration data only made the data of the top 100 cities available, so the daily PopInflux of cities with little population flow from Wuhan could potentially be zero. After summing the data from 1 to 23 January 2020 by city, there were still some zeros. To ensure the stability of the result, we added a very small margin value (0.01) to the Pop Influx data.

Ordinary least square regression

Ordinary least square (OLS) regression (Hutcheson *et al*., 2011; Patton *et al*., 2018; Wu *et al*., 2021) was conducted to investigate the relationship between GDP and the case number. For linear regression of one variable, n groups of observations (X_1, Y_1),(X_2, Y_2), …(X_n, Y_n) were assumed. The principle of the OLS method is to minimize the sum of the squares of the residuals to determine the best fitting curve, which is determined by minimizing the total fitting error, *i.e*. the total residual error. In addition to the convenience of calculation, the estimator obtained this way has excellent characteristics. We also checked the model and the attendant assumptions for adequacy and validity (Ghasemi *et al*., 2012; Patton *et al*., 2018).

Structural equation modelling

This method, a statistical framework combining two or more relational models to obtain multiple relationships, was used to establish, estimate and test causal relationships (Grace *et al*., 2006, 2016; Yang *et al*., 2020). SEM can deal with observed variables and latent variables; it can also have multiple dependent variables

Table 1. Overview of variables investigated.

Variable	Year	Resolution	Sources	Symbol
Gross domestic product (GDP)	2019	City scale	National Bureau of statistics, provincial and Municipal Bureau of Statistics	GDP2019
Population density	2020	1 km grid	Worldpop (https://www.worldpop.org/)	PD2020
Road accessibility	2020	Vector	OpenStreetMap (https://download.geofabrik.de)	RoadAccess
Distance to Wuhan	-	-	Calculated from the administrative division map of mainland China in 2015 downloaded from (http://www.resdc.cn/)	Dis2Wuhan
Population influx from Wuhan	2020/2001	City scale	Aggregated from the daily population influx from Wuhan to each city from 1 to 23 January, 2020 (http://qianxi.baidu.com/)	PopInflux

in one model and is capable of establishing a multivariate relationship, which refers to the sum of direct and indirect interactions between variables (Grace *et al*., 2016). The COVID-19 is a novel type of infection, whose transmission has not yet been fully clarified as it is affected by multiple factors. Here, SEM was utilized to identify the influence of socio-economic factors on COVID-19 transmission. We used the maximum likelihood estimation method, commonly applied in SEM modelling, which iteratively solves the model parameters to obtain the optimal parameter estimation of the fitting model (Grace *et al*., 2016). We established SEM models using the five observed variables described above: GDP, PopInflux, PD, RoadAccess, Dis2Wuhan plus one latent variable that represented COVID-19 on the dates chosen (COVID). Figure 2 shows the potential paths in a hypothesis-oriented SEM model. First we hypothesized that all five socio-economic variables have direct effects on COVID; second that GDP may indirectly affect COVID-19 through its effect on PopInflux, PD and RoadAccess; third that the direct influence of PopInflux on COVID-19 would probably be driven by GDP, PD and RoadAccess. Finally, we thought that RoadAccess may also be affected by GDP and PD and then affect COVID-19. Based on these hypotheses, we fed the data into the model for fitting, adjusted the model by adding the effective paths and removing the non-significant paths to obtain the best-fitted SEM models for different regions (Hu *et al*., 1999; Schumacker and Lomax, 2004). To verify the socio-economic influence when lacking PopInflux data, we also constructed SEM models without PopInflux.

Results

Socio-economic factors and COVID-19

The relationship between the socio-economic factors and COVID-19 (Figure 3) and the correlation among the socio-economic factors (Figure 4) were evaluated by Spearman's correlation. For the 288 cities, PopInflux from Wuhan, GDP in 2019 (GDP2019) and the PD in 2020 (PD2020) were the variables most correlated with COVID-19 (Figure 3A). GDP and PopInflux showed the highest correlations (Figure 4A), *i.e.* the correlation of PopInflux with COVID-19 was higher than that of GDP2019 and PD2020 during this period. All the curves rose first steeply (*i.e*. at the starting dates before 31 January) and then stabilized at a high level. For instance, the correlation of PopInflux was stable at a correlation coefficient of 0.75, while RoadAccess was slightly better correlated with COVID-19 than that of Dis2Wuhan, which was negatively correlated with COVID-19 but at a low level, an indication that the distance to the origin of the virus plays a very limited effect on the development of transport facilities.

As for the 35 central cities, the correlation coefficients of GDP2019, PD2020 and PopInflux increased at the starting dates as before but then also remained stable. In the beginning, the correlations of GDP2019 and PD2020 were higher than that of PopInflux, then the correlation coefficients of the three indices were nearly the same at the 0.84 level. The RoadAccess coefficient achieved a value higher than 0.5 between 23 and 28 January 2020, indicating its effect on the transmission of COVID-19 in the early stage. The negative correlation of Dis2Wuhan with COVID-19 was not as strong as other factors mainly because the big cities with high case numbers, such as Beijing, Guangzhou, *etc*., are situated far from Wuhan. This again shows the effect of transportation conditions on the spread of COVID-19 in modern society. The fluctuations in the correlation coefficients at the starting dates (*i.e*. before 24 January) may be related to the lack of early detection and diagnosis experience of the epidemic in some less developed cities. Compared with the results for the 288 cities, the central cities are usually the most developed, thus having higher socio-economic levels and more convenient transportation. This explains the stronger correlations of GDP2019, PD2020 and RoadAccess in the central cities (Figure 4B).

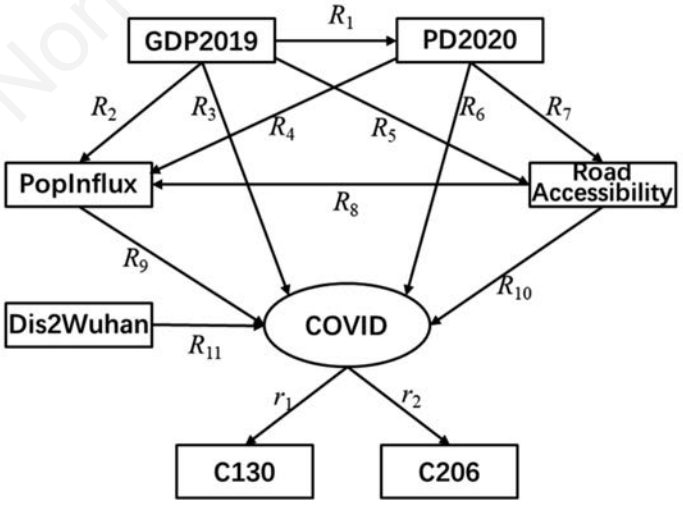

Figure 2. The structural equation model hypothesis showing the potential, influential factors under research. GDP2019, GDP 2019; PD2020, poulation density; PopInflux, population influx; Dis2Wuhan, distance to Wuhan. COVID, the cumulative number of confirmed cases of COVID-19; C130, COVID-19 on 30 January 2020; C206, COVID-19 on 6 February 2020. The paths between variables represent direct influences, with the path coefficients given.

Linear relationship between GDP2019 and COVID-19

The results presented above showed a strong positive correlation between GDP2019 and COVID-19. We then continued by analysing the linear relationship between GDP2019 and COVID-19 at different dates (C130, C206 and C213). Before the analysis, we conducted a normality test on the data and found that they were not normally distributed. Therefore, we performed a logarithmic data transformation before the regression analysis. As shown in Figure 5, COVID-19, in most instances, were strongly, linearly associated with the GDP2019 variable. Thus, the more developed the economy is, the more people are infected. This pattern was consistent, both in the country as a whole and locally in Hubei Province, Jing-Jin-Ji and the Cheng-Yu Economic Circle. In Figure 5A, the highest R^2 of regressions was 0.61 for C130 and C206 in 288 cities, while the R^2 was 0.59 for C213, *i.e.* slightly less than before.

Figure 5B shows the highest R^2 of regressions (0.74) of COVID-19 on the C206 and C213 dates in the 35 central cities and 0.72 for C130. As can be seen in both Figure 5A and B, the slope of the fitting line rises with the date. In Hubei (Figure 5C), GDP and COVID-19 had a positive linear trend. In Jing-Jin-Ji (Figure 5D) and Cheng-Yu Economic Circle (Figure 5E), GDP and COVID-19 had a significant linear relationship and the regressions met the assumptions of adequacy and validity. Due to the limited sample size in these areas, we must be careful when referring to the conclusions of these regression models. In general, GDP and COVID-19 had a clear linear relationship pointing towards the

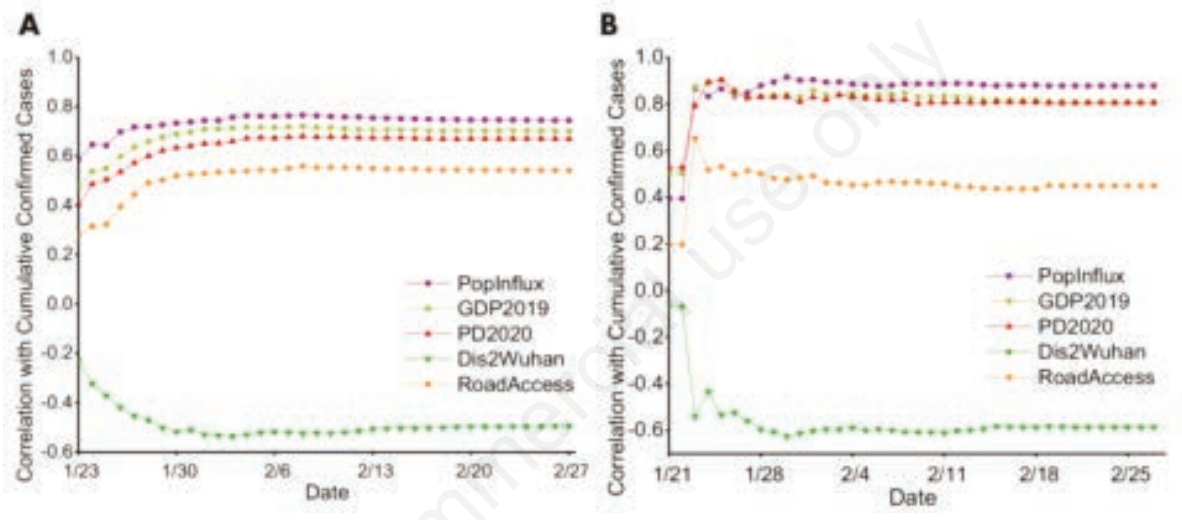

Figure 3. Correlations between the socio-economic variables and the number of cumulative confirmed COVID-19 cases. A) 288 cities in mainland China (Hubei excluded); B) 35 central cities in mainland China (Wuhan excluded).

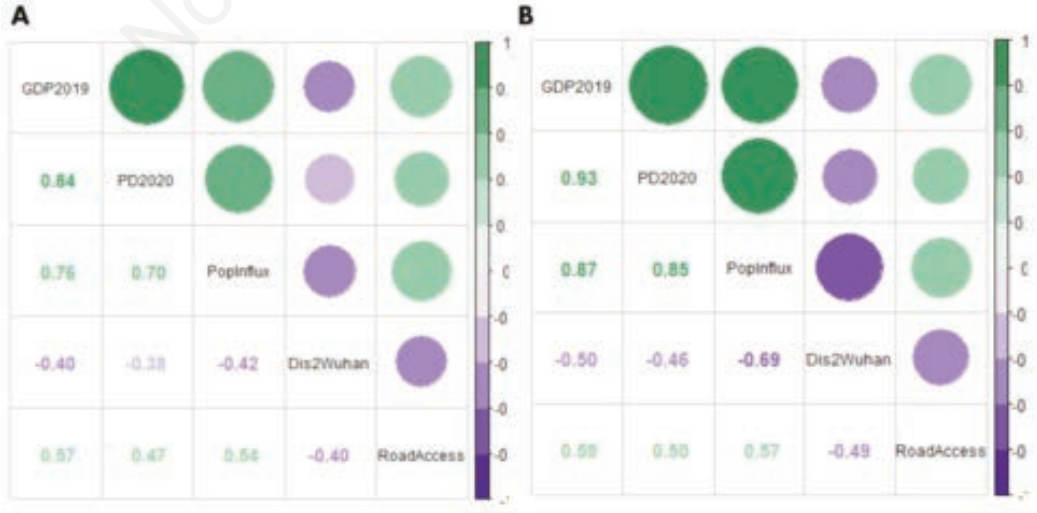

Figure 4. The correlation matrix among the different variables. A) 288 cities in mainland China (Hubei excluded); B) 35 central cities in mainland China (Wuhan excluded). The diameter of the circles represents the degree of correlation.

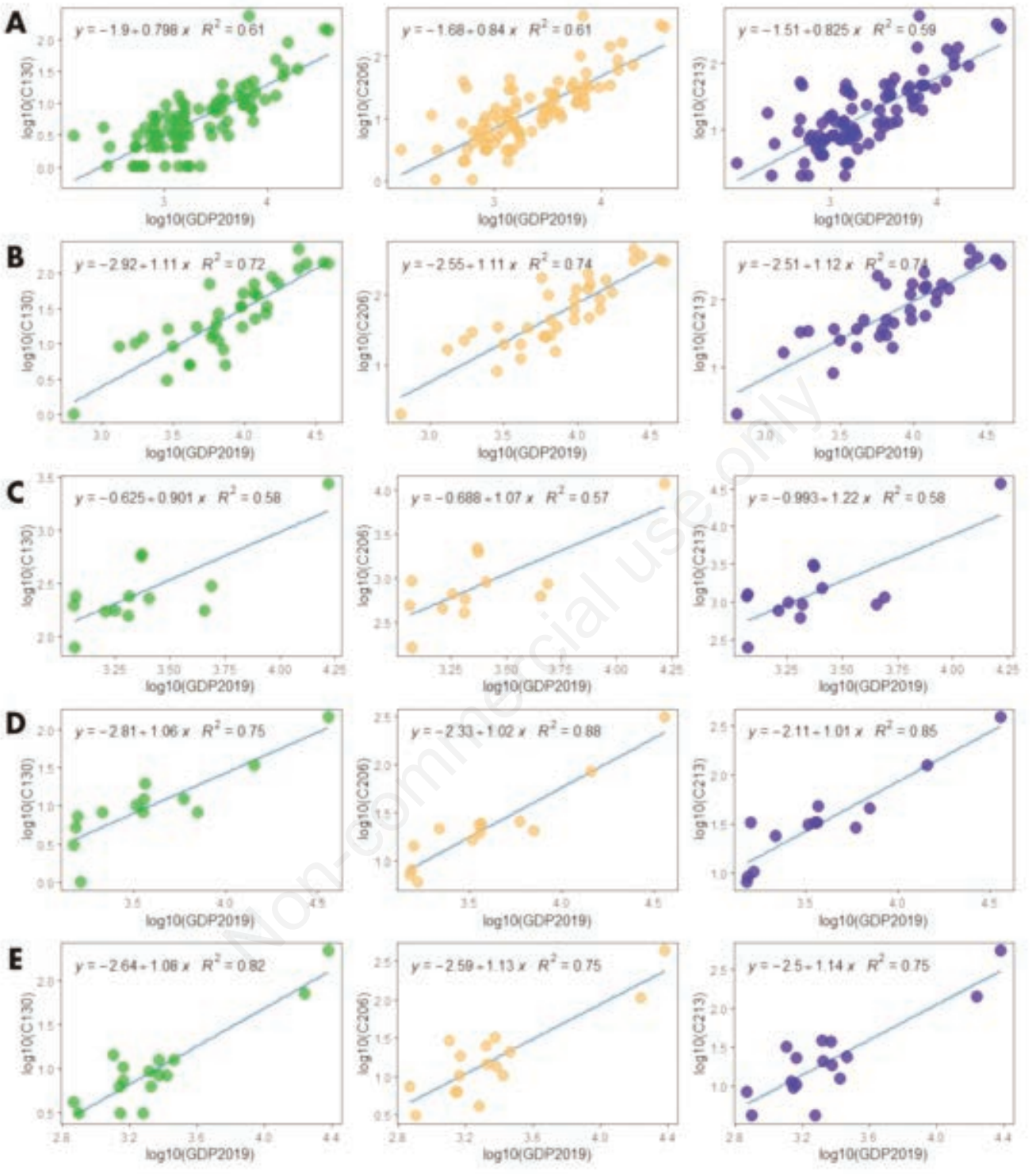

Figure 5. Linear relationships between gross domestic product (GDP) 2019 and the cumulative number of confirmed cases COVID-19 on 30 January, 6 February and 13 February. A) 288 cities (Hubei excluded); B) 35 cities (Wuhan excluded); C) Hubei Province; D) the Jing-Jin-Ji cluster; E) the Cheng-Yu economic circle. Note that there are several points with the same y-ordinate in figure A because in the early days the cumulative number s of COVID-19 cases in some cities were the same.

potential role of economic development in the spread of COVID-19. The $R^2$was between 0.57 and 0.88 for all models. In general, COVID-19's spatial distribution followed this pattern.

Direct and indirect effects of different factors on COVID-19

To further explore the direct and indirect effects of PopInflux, GDP2019, PD2020, RoadAccess, Dis2Wuhan and their interactions, we constructed an SEM model for influential factors and COVID-19 for C130 and C206, which represents the 7th and 14th day of the lockdown in Wuhan.

The SEM models for the 288 cities (excluding Hubei province) and the 35 central cities (excluding Wuhan) explain 72% and 84% of the variances, respectively (Figure 6A and B). In Figure 6A, GDP can be seen to have an indirect impact (0.46) on COVID-19 by strongly affecting PopInflux and PD2020, as well as a small direct effect (0.21) on COVID, while in Figure 6B, GDP has no

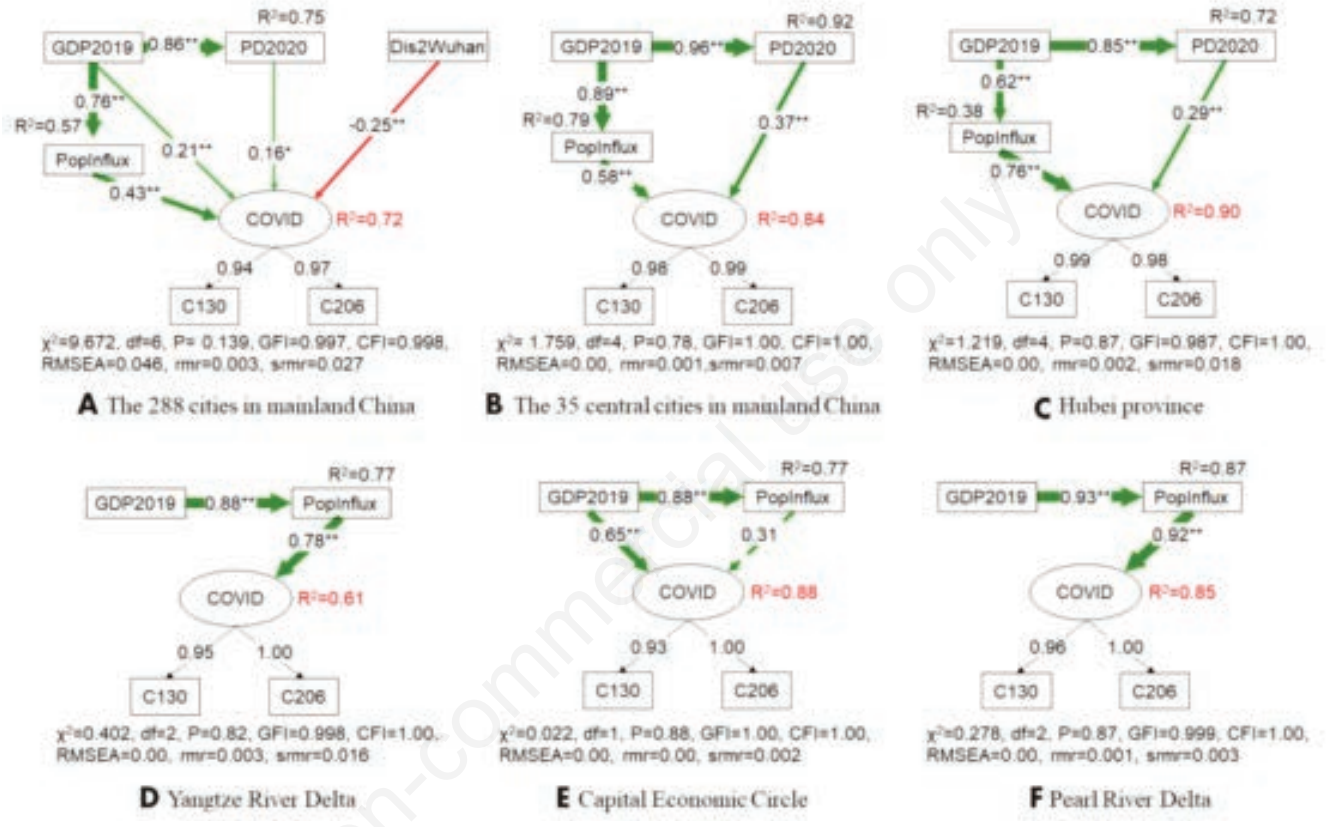

Figure 6. The constructed structural equation models (SEM). The paths between variables represent direct influences, and the numbers next to them are path coefficients. The green paths represent positive effects, while the red ones represent negative effects. The black paths represent the observed variables that make up the latent variables, with the numbers next to them are the coefficients contributed by each observed variable. Solid lines indicate significant paths, while dashed ones indicate lack of insignificance. R^2 represents the explained proportion of the variance. C130 and C206 denote the cumulative confirmed cases on 30 January and 6 February, respectively. The influential factors are gross domestic product (GDP) in 2019 (GDP2019), population influx (PopInflux), population density in 2020 (PD2020) and distance to Wuhan (Dis2Wuhan). The road access path of was removed from the final SEM models as it was not affecting COVID-19 significantly in all models.

Table 2. Effects on COVID-19 of socio-economic factors and distance from Wuhan.

Group	Effect	GDP2019	PopInflux	PD2020	Dis2Wuhan
288 cities in mainland China (excluding Hubei Province)	Total	*0.67*	0.43	0.16	–0.25
	Direct	0.21	*0.43*	0.16	–0.25
		0.009**	0.00**	0.015*	0.00**
	Indirect	*0.46*	0.00	0.00	0.00
35 central Cities in mainland China (excluding Wuhan)	Total	*0.87*	*0.58*	0.37	-
	Direct	*0.00*	*0.58*	0.37	-
			0.00**	0.00**	
	Indirect	*0.87*	0.00	0.00	-

GDP2019, GDP 2019; PopInflux, population influx; PD2020, population density 2020; Dis2Wuhan, distance to Wuhan. **Statistical significance at P≤0.01; *statistical significance at P≤0.05. Significance of the indirect effects and the total effects are not given by structural equation models. Italics indicates the largest effect coefficient in this line.

direct effect on COVID-19 but only indirect effects through PopInflux and PD2020. PopInflux had a strong direct effect on COVID-19 in both models. As shown in Table 2, the total effect of GDP on COVID-19 in the two models was 0.67 and 0.87, respectively, which is higher than that of PopInflux (0.43 and 0.58, respectively). PD2020 had generally a small direct effect (0.16) on COVID-19 (Figure 6A), while its effect for the central cities was more significant with a coefficient of 0.37 (Figure 6B). The relationship between Dis2Wuhan and COVID-19 was not significant in the model for the central cities in mainland China but negatively correlated (–0.25) for the 288 cities.

RoadAccess had no significant effect on COVID-19 and was therefore deleted from the modelling. Overall, GDP had the highest total effect on COVID-19 compared to the other influential factors and PopInflux had the highest direct effect. This indicates that the virus must have broken through the traditional limits of distance to spread by transportation.

The SEM models that were constructed for local regions in China only include GDP2019 and PopInflux (Figure 6AD-F). In Hubei province, PD2020 is also included (Figure 6C). This shows that GDP2019 influenced the spread of COVID-19 mainly by the interaction with the PopInflux in local areas. In Hubei province, the Yangtze River Deltaand the Pearl River Delta (Figure 6AC, D and F), GDP2019 had only an indirect effect on COVID-19 but no significant direct influence, while PopInflux had the highest direct and total effects on COVID-19 (Table 3). In the SEM model of Cheng-Yu Economic Circle (Figure 6E), GDP2019 had a significant, direct effect on COVID-19 with a path coefficient of 0.65, which was greater than that of PopInflux (0.31) as seen in Figure 6E. Importantly,comparted to all other variables, GDP2019 had the highest total effect (0.92) on COVID-19 (Table 3). In these areas (Figure 6AC-F), the SEM models could explain between 61% and 90% of the variance. The overall results indicate that the spread of COVID-19 in local areas is mainly controlled by PopInflux, which is mainly affected by GDP2019, suggesting the potential influence of the economy on the spread of the virus.

The structural equation models with and without PopInflux data

The results of our study showed GDP has a strong total effect and PopInflux a strong direct effect on COVID-19 in mainland China. To verify the socio-economic influence when lacking PopInflux data, we constructed SEM models for cities in mainland China without PopInflux.

For the models of the 288 cities and the 35 central cities, R^2 of the models without PopInflux was 0.65 and 0.82, respectively (Figure 7). R^2 decreased only by 0.07 and 0.03, respectively, compared to the SEM models using all variables. This indicates that the explicatory potential for GDP is strong with respect to spreading the virus. Table 4 shows that GDP had a higher total effect on COVID-19 than PD, Dis2Wuhan and RoadAccess, which had the highest direct effect for the model in 288 cities and the most indirect effect for the model in central cities. This indicates that GDP plays an important role for transmission of the virus.

Using only GDP in the modelling (Table 5), the R^2 is 0.58 (Hubei province), 0.43 (Yangtze River Delta), 0.86 (Cheng-Yu

Table 3. Total, direct and indirect effects on COVID-19 in different geographcal areas.

Effect	Hubei Province			Yangtze River Delta		Cheng-Yu Economic Circle		Pearl River Delta	
	GDP2019	PD2020	PopInflux	GDP2019	PopInflux	GDP2019	PopInflux	GDP2019	PopInflux
Total	0.72	0.29	*0.76*	0.69	*0.78*	*0.92*	0.31	0.86	*0.92*
Direct	0.00	0.29	*0.76*	0.00	*0.78*	*0.65*	0.31	0.00	*0.92*
	-	0.00**	0.00**	-	0.00**	0.00**	0.06	-	0.00**
Indirect	*0.72*	0.00	0.00	*0.69*	0.00	*0.27*	0.00	*0.86*	0.00

GDP2019, GDP 2019; PD2020, population density 2020; PopInflux, population influx. **Statistical significance at P≤0.01; Significance of the indirect effects and the total effects are not given by structural equation models. For each model, italics indicates the largest effect coefficient in this line, i.e., 0.76 (Popflux) is largest than 0.72 (GDP2019) and 0.29 (PD2020) for the total effect in Hubei Province.

Table 4. Effects on COVID-19 of socio-economic factors and distance from Wuhan with exception of PopInflux.

Group	Effect	GDP2019	PD2020	Dis2Wuhan
288 cities in mainland China (excluding Hubei Province)	Total	*0.66*	0.15	–0.30
	Direct	*0.53*	0.15	–0.30
		0.00**	0.047*	0.00**
	Indirect	*0.13*	0.00	0.00
35 central Cities in mainland China (excluding Wuhan)	Total	*0.74*	0.46	–0.25
	Direct	0.30	*0.46*	–0.25
		0.27	0.08	0.00**
	Indirect	*0.44*	0.00	0.00

GDP2019, GDP 2019; PD2020, population density 2020; Dis2Wuhan, distance to Wuhan. **Statistical significance at P≤0.01; *statistical significance at P≤0.05. Significance of the indirect effects and the total effects are not given by structural equation models. Italics indicates the largest effect coefficient in this line.

Table 5. Comparison of R^2 of structural equation models for mainland China and various geographical areas.

Model	The 288 cities in	The 35 central cities	Hubei province	Yangtze River Delta	Cheng-Yu Economic Circle	Pearl River Delta
Models based only on GDP2019	0.57	0.75	0.58	0.43	0.86	0.74
Models without PopInflux	0.65	0.81	-	-	-	-
Models based on all variables (section results)	0.72	0.84	0.90	0.61	0.88	0.85

Economic Circle) and 0.74 (Pearl River Delta), which almost have the same explanatory power as the models using all variables. For the model in Cheng-Yu Economic Circle, the R^2 only declined by 0.03. In Hubei Province, although the R^2 had declined by 0.32, GDP still explains more than 50% of the variance (Table 5). In the Yangtze River Delta and the Pearl River Delta, the R^2 had declined by 0.18 and 0.11. These results show the important influence of GDP on COVID-19.

Discussion

Applicability and limitation

We identified the most influential socio-economic factors on COVID-19 in mainland China. The SEM modelling suggested a strong effect of GDP on COVID-19 through its interactions with other variables, in particular population density and population influx. Although the situation nationwide is complex and might need more explanatory variables, GDP remains one of the most important variables and by explaining more than 50% of the variance. The finding that economy has a positive correlation with infection numbers in a region is supported by the work of Qiu *et al*. (2020) and Zhang *et al*. (2020), and our conclusion that PopInflux has a direct effect on the transmission of COVID-19 is consistent with the studies of Jia *et al*. (2020) and Zhang *et al*. (2020).

The advantage of SEM modelling is that it can deal with observation variables with errors to make the model more reliable. For example, the actual number of infected patients may not be equal to the number of reported confirmed cases on a day (Bendavid *et al*., 2020) because of missing reports, especially in the early stage of viral spread. Further, we used COVID-19 for certain key periods depending on the common14-day incubation period (*i.e*. the 7th and 14th day after the lockdown of Wuhan) as the dependent variables. This choice assisted our attempt at revealing the COVID-19 transmission mechanisms. Finally, the SEM modelling is a partial correlation analysis method, which is more conducive to identifying the direct effects of influential factors by excluding influences from other factors and can therefore disentangle complicated variable interactions.

The strong relationship between GDP and COVID-19 emphasizes the great influence of socio-economic variables on the initial spread of the virus. COVID-19 had a long incubation period in China until its gravity became obvious. Thus, the period before the lockdown allowed an initial, natural spread of COVID-19 without strong interventions, while the later pattern after March 2020 was more complex due to a strict policy based on experience gained (Sun *et al*., 2020). In 2020 and 2021, there were new outbreaks in many places in China, but rapid intervention strategies brought the pandemic under control by lockdowns, forced quarantine at home and travel restrictions. This is consistent with our conclusion that population movement is a key factor affecting the spread of the virus. Beijing has experienced several outbreaks and these newly confirmed cases verify that economically developed cities are at high risk.

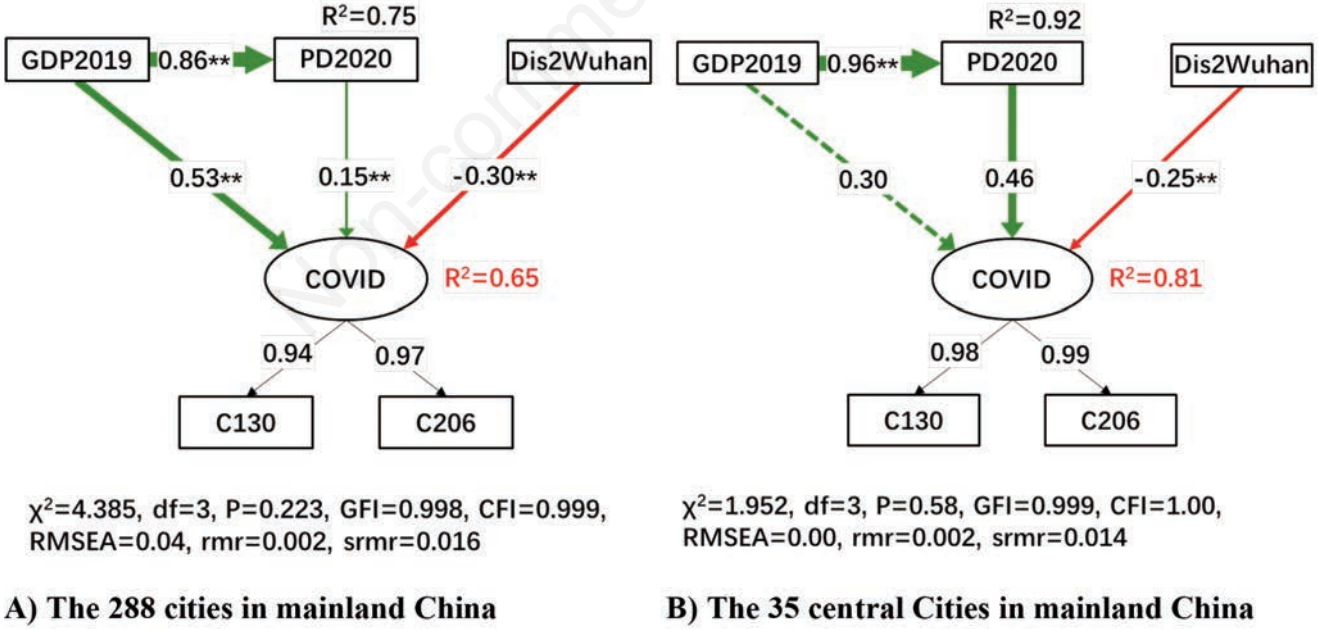

Figure 7. The constructed structural equation models without using PopInflux. The paths between variables represent direct influences and the numbers next to them are path coefficients. The green path represents the positive effect, while the red ones represent negative effects. The black paths represent the observed variables that make up the latent variables, with the numbers next to them the coefficients contributed by each observed variable. Solid lines indicate the path is significant while dashed lines indicate insignificant. R^2 represents the explained proportion of the variance. C130 and C206 denote the cumulative confirmed cases on Jan. 30th and Feb. 6th, respectively. The influential factors are gross domestic product (GDP) in 2019 (GDP2019), population density in 2020 (PD2020), and Dis2Wuhan(Dis2Wuhan).

The possible non-linear relationships between the influential factors and the spread of COVID-19 should be considered by researchers in further studies. Although the sample size of 288 cities across the country met the requirements of SEM modelling, a larger sample size could have been useful, particularly for local area analysis where we had to reduce the number of parameters as much as possible to ensure the reliability of the models. Finally, the fitted SEM model did notfully explain the variances presented by the COVID-19 transmission, possibly because of other factors we did not measured, which requires further investigation. Apart from these limitations, the fitness indices, as well as the R^2 of the fitted models, had a good explicatory capability and provided a valuable reference for better understanding the socio-economic influence on the viral spread.

Although there are many variables representing the socio-economic level of a place, such as city vigour (Wong *et al*., 2002), night-time lights (Ma *et al*., 2015), GDP is one of the most widely used and easy to access (Zhang *et al*., 2020). We verified that its total effects on the spread of COVID-19 exceed that of PopInflux. The population flow data can be applied in the susceptible-exposed-infectious-removed (SEIR) model to dynamically update the specific number of different populations in simulating the spread of infectious diseases (Xia *et al*., 2004; Weinberger *et al*., 2012; Viboud *et al*., 2016). However, PopInflux data are usually unavailable in some regions. Our results show that GDP could be an alternative for PopInflux for modelling when such data are unavailable. In addition, from the available population flow data of some regions, we can infer the unknown data of other regions based on GDP. In epidemic modelling, GDP can also provide references for adjusting the basic reproductive number R_0 for SEIR model (Alexis *et al*., 2003; Chen *et al*., 2003; Lalwani *et al*., 2020) or other related parameters.

Conclusions

Social factors are not only the most important factors for promoting the spread of infectious diseases, but they arte also the key to effectively preventing and eliminating them. For example, when the Chinese Government responded to the outbreak during Spring Festival in 2020, the rapid and strict measures made brought the situation quickly under control (Sun *et al*., 2020). This has also been verified in the latter outbreaks of coronavirus in mainland China. Importantly, our research results provide additional policy suggestions for fighting COVID-19 that could be useful also aginst other epidemic infections.

Travel bans would prevent the wider viral spread, while screening and establishing risk-free areas as quickly as possible is by far the best way to minimize economic damage. Further, high-risk areas such as hub environments and economically developed cities with large population flows requires rapid and strong attention (Delikhoon *et al*., 2021). For long-term respons to epidemics in metropolises, it is suggested to implement a hierarchical re-opening strategy, for example by focusing on less serious areas first to reduce economic loss (Ge *et al*., 2021). Some research shows that the impact of lockdown on mobility is stronger in developed urban areas where income per capita is lower but inequality is higher (Bonaccorsi *et al*., 2020). Interestingly, a recent study across England indicates taht spatial differences in COVID-19 mortality rate are related to socio-economic and environmental factors (Sun *et al*., 2020).

Economic development and health presents a dilemma. Historically, long-run improvements in health have been tied to economic growth through three broad mechanisms: better nutrition, enhancements in public health infrastructure and more effective medical technology (Frakt *et al*., 2018). However, some researchers have found that small particular matter ($PM_{2.5}$), a side effect of economic growth, might increase the COVID-19 mortality rate (Wei *et al*., 2019; Yan *et al*., 2020). Economic development is inevitably accompanied by the increase of the moving populations and the concentration of labour. Developed areas with dense populations, frequent mobility and poor air quality provide conditions for the spread of infectious diseases. Therefore, attention must be paid to balancing economic development and social health in the future.

Although GDP exerts both total and indirect effects on the spread of COVID-19, socio-economic development and COVID-19 of a region have a strong positive correlation. Therefore, the effects of the economy on population influx, population density and road accessibility should not be overlooked. Overall, however, various socio-economic factors are under the control of the economy and have complex interactions with each other. Our results indicate that social and economic factors are closely related to the spread of the virus, both in a direct and an indirect way. Hence, cities with rapid economic development run a greater risk than other areas. For those areas in an outbreak, more stringent intervention and control measures such as a travel bans or home quarantine should be implemented.

References

Adda J, 2016. Economic activity and the spread of viral diseases: evidence from high frequency data. Q J Econ 131:891-941.

Alexis A, 2003. Susceptible-infected-recovered (SIR) dynamics of COVID-19 and economic impact. arXiv:2003.11221.

Angelini M, Heuvelink G, Kempen B, Morrás H, 2016.Mapping the soils of an Argentine Pampas region using structural equation modelling. Geoderma 281:102-18.

Bendavid E, Mulaney B, Sood N, Shah S, Bromley-Dulfano R, Lai C, Weissberg Z, Saavedra-Walker R, Tedrow J, Bogan A, Kupiec T, Eichner D, Gupta R, Ioannidis J, Bhattacharya J, 2021. COVID-19 antibody seroprevalence in Santa Clara County, California. Int J Epidemiol 50:410-9.

Bonaccorsi G, Pierri F, Cinelli M, Flori A, Galeazzi A, Porcelli F, Schmidt, AL, Valensise CM, Scala A, Quattrociocchi W, Pammolli F, 2020. Economic and social consequences of human mobility restrictions under COVID-19. Proc Natl Acad Sci 117:15530-5.

Chakraborti S, Maiti A, Pramanik S, Sannigrahi S, Pilla F, Banerjee A, Das DN, 2021. Evaluating the plausible application of advanced machine learnings in exploring determinant factors of present pandemic: A case for continent specific COVID-19 analysis. Sci Total Environ 765:142723.

Chen Y, Lu P, Chang C, Liu T, 2020. A time-dependent SIR model for COVID-19 with undetectable infected persons. IEEE T Netw Sci Eng 7:3279-94.

COVID-19 and Human Development: Assessing the Crisis, Envisioning the Recovery, 2020. Human development report 2020: Perspectives. Available from: http://hdr.undp.org/en/hdp-covid

Delikhoon M, Guzman MI, Nabizadeh R, NorouzianBaghani A, 2021. Modes of transmission of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) and factors influencing

on the airborne transmission: a review. Int J Environ Res Public Health 18:395.
Dong E, Du H, Gardner L, 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 20:533-4.
Enserink M, Kupferschmidt K, 2020. With COVID-19, modeling takes on life and death importance. Science 367:1414-5.
Frakt A, 2018. How the economy affects health. JAMA 319:1187-8.
Ge Y, Zhang WB, Wang J, Liu M, Ren Z, Zhang X, Zhou C, Tian Z, 2021. Effect of different resumption strategies to flatten the potential COVID-19 outbreaks amid society reopens: a modeling study in China. BMC Public Health 21:604.
Ghasemi A, Zahediasl S, 2012. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab 10:486-9.
Grace J, Keeley J, 2006. A structural equation model analysis of postfire plant diversity in California shrublands. Ecol Appl 16:503-14.
Grace J, Anderson T, Seabloom E, Borer E, Adler P, Harpole W, Hautier Y, Hillebrand H, Lind E, Pärtel M, Bakker J, Buckley Y, Crawley M, Damschen E, Davies K, Fay P, Firn J, Gruner D, Hector A, Smith M, 2016. Integrative modelling reveals mechanisms linking productivity and plant species richness. Nature 529:390-3.
Guan D, Wang D, Hallegatte S, Davis SJ, Huo J, Li S, Bai Y, Lei T, Xue Q, Coffman D, Cheng D, Chen P, Liang X, Xu B, Lu X, Wang S, Hubacek K, Gong P, 2020. Global supply-chain effects of COVID-19 control measures. Nat Hum Behav 4:577-87.
Hao, J, Xu G, Luo L, Zhang Z, Yang H, Li H, 2020. Quantifying the relative contribution of natural and human factors to vegetation coverage variation in coastal wetlands in China. Catena 188:104429.
Hu L, Bentler P, 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equat Model Multidiscipl J 6:1-55.
Hutcheson G, 2011. Ordinary least-squares regression. In: L. Moutinho and G.D. Hutcheson (Eds.), The SAGE dictionary of quantitative management research. SAGE, pp. 224-228.
Jia J, Lu X, Yuan Y, Xu G, Jia J, Christakis N, 2020. Population flow drives spatio-temporal distribution of COVID-19 in China. Nature 582:389-94.
Lalwani S, Sahni G, Mewara B, Kumar R, 2020. Predicting optimal lockdown period with parametric approach using three-phase maturation SIRD model for COVID-19 pandemic. Chaos Solitons Fractals 138:109939.
Li Q, Guan X, Wu P, Wang X, Zhou L, Wt L, 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus infected pneumonia. N Engl J Med 382:1199-207.
Lloyd C, Chamberlain H, Kerr D, Yetman G, Pistolesi L, Stevens F, 2019. Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. Big Earth Data 3:108-39.
Ma T, Zhou Y, Zhou C, Haynie S, Pei T, Xu T, 2015. Night-time light derived estimation of spatio-temporal characteristics of urbanization dynamics using DMSP/OLS satellite data. Remote Sens Environ 158:453-64.
Markowitz S, Nesson E, Robinson J, 2019. The effects of employment on influenza rates. Econ Hum Biol 34:286-95.
Patton N, Lohse K, Godsey S, Crosby B, Seyfried M, 2018. Predicting soil thickness on soil mantled hillslopes. Nat Commun 9:3329.
Qiu Y, Chen X, Shi W, 2020. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. J Popul Econ 9:1-46.
Sanderson E, Jaiteh M, Levy M, Redford K, Wannebo A, Woolmer G, 2002. The Human Footprint and the Last of the Wild. BioSci 52:891-904.
Schumacker R, Lomax G, 2004. A beginner's guide to structural equation modeling. Psychology Press, New York.
Sun Y, Hu X, Xie J, 2021. Spatial inequalities of COVID-19 mortality rate in relation to socioeconomic and environmental factors across England. Sci Total Environ 758:143595.
Sun Z, Zhang H, Yang Y, Wan H, Wang Y, 2020. Impacts of geographic factors and population density on the COVID-19 spreading under the lockdown policies of China. Sci Total Environ 746:141347.
Venter O, Sanderson E, Magrach A, Allan J, Beher J, Jones K, 2016. Global terrestrial Human Footprint maps for 1993 and 2009. Sci Data 3:160067.
Viboud C, Bjørnstad O, Smith D, Simonsen L, Miller M, Grenfell B, 2006. Synchrony, waves, and spatial hierarchies in the spread of influenza. Science 312:447-51.
Wei Y, Wang Y, Di Q, Choirat C, Wang Y, Koutrakis P, Zanobetti A, Dominici F, Schwartz J, 2019. Short term exposure to fine particulate matter and hospital admission risks and costs in the Medicare population: time stratified, case crossover study. BMJ 367:l6258.
Weinberger D, Krause T, Mølbak K, Cliff A, Briem H, Viboud C, Gottfredsson M, 2012. Influenza epidemics in Iceland over 9 decades: changes in timing and synchrony with the United States and Europe. Am J Epidemiol 176:649-55.
Wong C, 2002. Developing indicators to inform local economic development in England. Urban Stud 39:1833-63.
Worldometer, 2020. Available from: https://www.worldometers.info/coronavirus/#repro Accessed: March 26.
Wu X, Yin J, Li C, Xiang H, Lv M, Guo Z, 2021. Natural and human environment interactively drive spread pattern of COVID-19: A city-level modeling study in China. Sci Total Environ 756:143343.
Xia Y, Bjørnstad ON, Grenfell B, 2004. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. Am Nat 164:267-81.
Yan W, Nawaz M, Xu W, Jiang Z, Sun W, Lai J, 2020. Atmospheric pressure and population density as super-factors influencing the transmission of coronavirus disease 2019 (COVID-19). Sci Rep [Epub ahead of print].
Yang L, Shen F, Zhang L, Cai Y, Yi F, Zhou C, 2020. Quantifying influences of natural and anthropogenic factors on vegetation changes using structural equation modeling: A case study in Jiangsu Province, China. J Clean Prod 280. [Epub ahead of print].
Zhang C, Chen C, Shen W, Tang F, Lei H, Xie Y, 2020. Impact of population movement on the spread of 2019-nCoV in China. Emerg Microbes Infect 9:988-90.
ZhangY, Tian H, Zhang Y, Chen Y, 2020. Is the epidemic spread related to GDP? Visualizing the distribution of COVID-19 in Chinese Mainland. arXiv:2004.04387.