A Study on Social and Economic Factors of COVID-19 Prevalence

Junchen Fan

1 Shanghai Pinghe Bilingual School
* wusuv123@sina.com

ABSTRACT
Since Novel Coronavirus Pneumonia (hereinafter referred to as “COVID-19”) was discovered in late December of 2019, it has rampaged worldwide, causing social and financial loss to countries all over the world. There are many factors that affect the spread of COVID-19. This paper will study the economic and social factors that affect the spread of COVID-19 by investigating indicators that represent the economic and social situation of the nation. The dependent variable in this paper measures severity of epidemic while independent variables are the various indicators in each country. The only dynamic indicator is Stringency Index established by Oxford University (hereinafter referred to as “Oxford Stringency Index”). Oxford Stringency Index evaluates the stringency of government responses to COVID-19 epidemic situations, which may be associated with epidemic severity. This paper will first analysis the correlation between Oxford Stringency Index and epidemic severity so as to obtain a preliminary understanding of the Index, followed by regression analysis of other indicators. Regression analysis will uncover the factors that are significantly correlated with epidemic severity in different periods. Ordinary least squares method is adopted to calculate the regression coefficient of the significant indicators. Then, whether the regression coefficients of the significant factors at each time point are positive or negative are discussed to obtain the correlation between each significant factor and the dependent variable, namely whether they are positively or negatively correlated. Meanwhile, this paper also puts forward three policies that are conducive to suppressing epidemic situation.

Keywords: Covid-19 Transmission, social and economic factors, ordinary least squares, epidemic prevention.

1. INTRODUCTION

1.1. Research Background
Reports of Novel Coronavirus surfaced in Wuhan, China, in December 2019. However, due to an incorrect cognition that the virus was not deadly and could not be transmitted from human to human, little attention was paid to the Novel Coronavirus. People were not aware of the severity until the first death occurred in Wuhan, China, in mid-January 2020. Therefore, on January 23, 2020, Wuhan implemented a city lockdown policy. Then, in late January 2020, China implemented traffic controls across the country. With the implementation of all the measures taken to control the epidemic situation, China's epidemic situation has gradually become under control. Lockdown in Wuhan was lifted on April 8th, 2020. In the same month, employees have begun returning to workplace across China.

![United States's Covid-19 trajectory](image)

**Figure 1** Figure of severity of epidemic situation and government response in the US
However, epidemic situation in other countries has not been under effective control yet, since the spread of COVID-19 was faster than expected. To mitigate the spread of COVID-19, it is of necessity to figure out its affecting factors. This paper mainly studies the social and economic factors that influence the spread of COVID-19.

1.1.1. COVID-19 in the most affected countries and government response index

To illustrate the correlation between severity of epidemic and government response, four representative countries, including the US, Brazil, India, and China, are selected. Their figure of correlation between severity of epidemic and government response are listed as follows [1].

1.2. Research Significance

This paper will shed light on the transmission of COVID-19 by revealing the economic and social factors that affects the spread of COVID-19. It is believed that by intervening these factors, the spread of COVID-19 can be prevented or mitigated. COVID-19 transmission risk factors [2] studied the social and economic factors that affect the number of million deaths in a single period while this paper will investigate social and economic factors that affect severity of epidemic over multiple periods.

1.3. Research Problem

This paper mainly studies the economic factors affecting the spread of COVID-19. A country's GDP and unemployment rate are indicators to measure the state of a country's economy. If a country has a high unemployment rate or a low GDP, it indicates that the country is in a poor economic situation. People in a poor economy may suffer from low quality of life and poor health, for example, drug abuse and malnutrition. These are advantageous conditions for outbreaks. In addition, exchange of international personnel, in this case, the number of migrants, is also a vital indicator relevant to the spread of COVID-19. The more people entering a country, the higher the chance that COVID-19 is spread to the country, the higher the chance that COVID-19 is spread to world via the country. Issues studied in this paper are listed as follows. 1) What are the factors that affect the spread of COVID-19? 2) How do these social and economic factors influence the severity of COVID-19 pandemic situation? 3) What implications can the study of these social and economic factors have for future prevention and response to public health emergencies similar to COVID-19?

2. RESEARCH METHOD

2.1. Affecting Factors

The development of the epidemic in different countries depends not only on the transmission characteristics of the virus itself, but also on economic development, social and other factors in the country [3].

2.1.1. Economic Factors

The economic indicators reflect the comprehensive national strength. Countries with a strong economy will have more resources and means to response to epidemic. Unemployment rate is one of the factors affecting the spread of epidemic [4] because being unemployed will lead to lower quality of life [5], thus become more vulnerable to COVID-19. Countries with a strong economy can subsidy those who lost their jobs due to epidemic and quarantine policies, making it feasible for...
the unemployed to stay at home without worrying about the cost of necessities, thus reduce the number of people vulnerable to epidemic. Additionally, they are capable of implementing major quarantine measures, including lockdown of cities, increasing government expenditure on health care, the later was proved to have played an important role in European countries’ response to the spread of the H1N1 in 2009 [6]. These countries are also capable of developing vaccine, eliminating COVID-19 thoroughly. Therefore, severity of epidemic will be significantly reduced. On the contrary, countries with a less developed economy will have less resources at disposal, becoming more vulnerable to epidemic. Therefore, economic factors that measure the level of economic development in each country are likely to influence the spread of the COVID-19 in the country.

The economic factors affecting the spread of COVID-19 are listed as follows:

1) GDP per capita  
2) Total value of GDP  
3) Rural population proportion  
4) Employment rate of citizens elder than 15  
5) Crude Fertility Rate  
6) Number of passengers traveling by air  
7) Number of international arrivals  
8) Cover rate of health insurance  
9) Proportion of citizens having personal access to Internet  
10) Deaths due to lack of clear water and medical facilities (as per 100,000 population)

2.1.2. Social Factors

Indicators in social factors can measure the government response to epidemic and development of a country in a more comprehensive and detailed way. Social factors, which are often linked to economic factors, can analyze the quality of life in a country. For example, Gini coefficient suggests quality of life of citizens by measuring the distribution of wealth. Measured by Oxford Stringency Index, the stringency of government responses to COVID-19 epidemic reflects the importance the society attached to epidemic. Socio-medical factors reflect the health care capacity of a country, which is closely related to mortality rate due to COVID-19, namely the severity of pandemic. Socio-health factors reflect physical fitness of citizens, whose strength may affect the spread of COVID-19 among citizens. In addition, socio-health factors are related to economic factors. If a country has a strong economy, its citizens should have higher levels of physical fitness. Another relevant factor is that citizens who have received information on protecting themselves from epidemic via Internet will stop the spread of COVID-19 consciously through their personal decisions, including self-isolation and reducing going outdoors [7].

2.1.2.1. Social Factors

1. Oxford Stringency Index (A country’s response to the epidemic is measured by the amount of government subsidies for damage caused by the COVID-19, the amount of government investment in healthcare and major decisions made to suppress the epidemic, such as lockdowns of cities and school closures. The index ranges from 0 to 100, with 0 indicating no response and 100 indicating the most positive response. In OLC approach, Oxford Stringency Index is broken down into two indicators: an index of the government response to epidemic combined with health care expenditure (hereinafter referred to as ‘containment health index’), and an index of the government’s economic support.)  
2. Gini coefficient  
3. Number of immigrants (as proportion of total population)  
4. Total population  
5. Population density (person per square kilometers)  

1) Socio-medical Factors  
a) Number of physicians (as per 1,000 citizens)  
b) Number of nurses (as per 1,000 citizens)  
c) Deaths due to infectious diseases and prenatal undernutrition (as the portion of total deaths)  
d) Number of hospital beds (as per 1,000 citizens)  
e) Government expenditure on health care (as per capita annually)  
f) National life expectancy  
2) Socio-health Factors  
a) Population structure (as the proportion of citizens ages under 14 and above 65)  
b) Crude death rate (as per 1,000 citizens)  
c) Diabetes prevalence (population ages between 20-79)  

2.2. Measurement of epidemic severity

In this paper, deaths caused by COVID-19 per 1,000,000 population is adopted to measure epidemic severity. The more deaths caused by COVID-19 per million people in a country, the more severe the epidemic is in the country.
2.2.1. Deaths per million population at different time points

Severity of epidemic can be measured by collecting the number of deaths per million population in each country at different points in time. In this paper, seven points in time including 30, 60, 90, 120, 150, 180, and 210 days after the outbreak are selected. Deaths per million population at these time points are collected and further analyzed against economic and social factors. Namely how and by what extent do economic and social factors affect severity of COVID-19. With 30-day intervals, visible changes in social and economic factors can be observed and documented. It can draw a generalized rule regarding how social and economic factors affect the spread of COVID-19.

2.3. Data Source

Data in this paper was extracted from various authority websites, including official website of World Bank (https://data.worldbank.org/), official website of United Nations (http://data.un.org/), United Nations World Tourism Organization (https://www.unwto.org), and official website of Oxford (https://www.oxfordmartin.ox.ac.uk/global-development)

2.4. Relationship between Oxford Stringency Index and severity of epidemic

Since Oxford Stringency Index evaluates the stringency of government responses to COVID-19 epidemic situations, which may directly influence the epidemic severity, it is decided to analyze the relationship between Oxford Stringency Index and severity of pandemic prior to analyzing other factors.

2.4.1. Relationship between Oxford Stringency Index and severity of pandemic at 30 days after the outbreak

As illustrated in Table 1, Oxford Stringency Index is positively correlated with severity of pandemic. The fit of goodness of the line $R^2 = 0.4224$ is relatively low. $\text{Prob} > F$ is 0.0000 (which is less than 0.05, indicating Oxford Stringency Index has a significant correlation with severity of epidemic). $P > |t|$ is 0.000, which also indicates that Oxford Stringency Index has a significant correlation with severity of epidemic.

![Table 1](image)

2.4.2. Relationship between Oxford Stringency Index and severity of pandemic at 60 days after outbreak

As illustrated in Table 2, Oxford Stringency Index is positively correlated with severity of pandemic. The fit of goodness of the line $R^2 = 0.3089$ is relatively low. $\text{Prob} > F$ is 0.0000 (which is less than 0.05, indicating Oxford Stringency Index has a significant correlation with severity of epidemic). $P > |t|$ is 0.000, which also indicates that Oxford Stringency Index has a significant correlation with severity of epidemic.

![Table 2](image)

2.4.3. Relationship between Oxford Stringency Index and severity of pandemic at 90 days after outbreak

As illustrated in Table 3, Oxford Stringency Index is positively correlated with severity of pandemic. The fit of goodness of the line $R^2 = 0.0809$ is extremely low. $\text{Prob} > F$ is 0.0003 (which is less than 0.05, indicating Oxford Stringency Index has a significant correlation with severity of epidemic). $P > |t|$ is 0.000, which also indicates that Oxford Stringency Index has a significant correlation with severity of epidemic.
Table 3 Regression analysis of Oxford Stringency Index and severity of pandemic at 90 days after the outbreak

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 155</th>
<th>F(1, 153) = 13.46</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>46677.2943</td>
<td>1</td>
<td>46677.2943</td>
<td></td>
<td>0.0003</td>
</tr>
<tr>
<td>Residual</td>
<td>530509.75</td>
<td>153</td>
<td>3467.3845</td>
<td>Adj R-squared = 0.0080</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>577187.053</td>
<td>154</td>
<td>3747.9678</td>
<td>R-squared = 0.7849</td>
<td></td>
</tr>
</tbody>
</table>

2.4.4. Relationship between Oxford Stringency Index and severity of epidemic at 120 days after outbreak

As illustrated in Table 4, Oxford Stringency Index is positively correlated with severity of the epidemic. The fit of goodness of the line $R^2 = 0.0082$ is extremely low. Prob > F is 0.2622 (which is greater than 0.05, indicating Oxford Stringency Index does not have a significant correlation with severity of epidemic.). $P > |t|$ is 0.262, which also indicates that there is no significant correlation between Oxford Stringency Index and severity of epidemic.

Table 4 Regression analysis of Oxford Stringency Index and severity of pandemic at 120 days after the outbreak

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 155</th>
<th>F(1, 153) = 1.27</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>25143.831</td>
<td>1</td>
<td>25143.831</td>
<td></td>
<td>0.2622</td>
</tr>
<tr>
<td>Residual</td>
<td>305775.43</td>
<td>153</td>
<td>19854.579</td>
<td>Adj R-squared = 0.0082</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3632984.58</td>
<td>154</td>
<td>14688.3926</td>
<td>R-squared = 0.0167</td>
<td></td>
</tr>
</tbody>
</table>

2.4.5. Relationship between Oxford Stringency Index and severity of epidemic at 150 days after outbreak

As illustrated in Table 5, Oxford Stringency Index is positively correlated with severity of epidemic. The fit of goodness of the line $R^2 = 0.0101$ is extremely low. Prob > F is 0.7007 (which is greater than 0.05, indicating Oxford Stringency Index does not have a significant correlation with severity of epidemic.). $P > |t|$ is 0.701, which also indicates that there is no significant correlation between Oxford Stringency Index and severity of epidemic.

Table 5 Regression analysis of Oxford Stringency Index and severity of pandemic at 150 days after the outbreak

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 155</th>
<th>F(1, 153) = 0.36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>11281.851</td>
<td>1</td>
<td>11281.851</td>
<td></td>
<td>0.5467</td>
</tr>
<tr>
<td>Residual</td>
<td>464701.59</td>
<td>153</td>
<td>30570.617</td>
<td>Adj R-squared = 0.0124</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>486383.34</td>
<td>154</td>
<td>32495.3607</td>
<td>R-squared = 0.1743</td>
<td></td>
</tr>
</tbody>
</table>

2.4.6. Relationship between Oxford Stringency Index and severity of epidemic at 180 days after outbreak

As illustrated in Table 6, Oxford Stringency Index is positively correlated with severity of epidemic. The fit of goodness of the line $R^2 = 0.0024$ is extremely low. Prob > F is 0.5467 (which is greater than 0.05, indicating Oxford Stringency Index does not have a significant correlation with severity of epidemic.). $P > |t|$ is 0.547, which also indicates that there is no significant correlation between Oxford Stringency Index and severity of epidemic.

Table 6 Regression analysis of Oxford Stringency Index and severity of pandemic at 180 days after the outbreak

As illustrated in Table 7, Oxford Stringency Index is positively correlated with severity of epidemic. The fit of goodness of the line $R^2 = 0.0140$ is extremely low. Prob > F is 0.7007 (which is greater than 0.05, indicating Oxford Stringency Index does not have a significant correlation with severity of epidemic.). $P > |t|$ is 0.701, which also indicates that there is no significant correlation between Oxford Stringency Index and severity of epidemic.

2.4.7. Relationship between Oxford Stringency Index and severity of epidemic at 210 days after outbreak
2.4.8. Reasons for the change in the relationship between Oxford Stringency Index and epidemic severity

In the initial stage of epidemic, countries have not yet been infected with COVID-19 paid little attention while those countries where confirmed cases are reported attached importance to epidemic. Therefore, Oxford Stringency Index is positively correlated with severity of epidemic in the initial stage of epidemic, which is different from common sense. Similarly, in the mid-late stage of epidemic, countries with massive death tightened anti-epidemic policies, scoring high in Oxford Stringency Index. Thus, Oxford Stringency Index is positively correlated with severity of epidemic in the mid-late stage of epidemic. In addition, the correlation between Oxford Stringency Index and epidemic severity is not significant, indicating that there are other factors that affect severity of epidemic.

2.5. Model Setup

2.5.1. Modeling Profile

Mathematical modeling is adopted in this paper to establish a model of severity of epidemic and various factors. Correlation analysis in the regression analysis is used to identify which indicators were strongly associated with severity of epidemic. These indicators are selected and used to further establish a model predicting the relationship between selected indicators and severity of epidemic by adopting the regression equation in the regression analysis. Mathematical models including linear regression and Ordinary least squares are adopted.

2.5.2. Correlation Analysis of association between indicators of factors and severity of epidemic

The correlation between the indicators and severity of epidemic varies. Some indicators may have a strong correlation with epidemic severity, while others may have a weak correlation with epidemic severity. In this paper, correlation analysis is used to analyze the correlation between the indicators and severity of epidemic. According to the established estimation equation, the independent variable X (selected indicators) is used to estimate and predict the value of the dependent variable Y (severity of epidemic). There are three inspection methods, F-test, coefficient of determination R², T test.

Through the SWREG instruction of the data processing software Stata, the regression equation between the indicators and severity of epidemic is established by using the stepwise regression method (forward method), and the values of F-test, R² test and T test of the indicators that are significantly correlated with severity of epidemic are calculated to judge the correlation. The SWREG directive (stepwise regression) automatically removes indicators that are not significantly correlated with severity of epidemic.

1) Stepwise Regression

The idea of stepwise regression is to introduce variables one after another into the model until no variables can be introduced. Variables increase from few to many in the process, which is listed as follows [8].

(a) Establish simple regression models for dependent variable Y and n regressors X₁, X₂,..., Xₙ respectively.

\[ Y = \beta_0 + \beta_1 X_1 + \epsilon, \quad i = 1, \ldots, p \quad (1) \]

Then calculate variable X₀, the corresponding regression coefficient F-test value is denoted as F₁, \ldots, Fₚ. Denote the maximum value as F₁⁺, namely F₁⁺(1) = max \{ F₁(1), \ldots, Fₚ(1) \}. For given significance level a, note the corresponding critical value as F(a), if F₁⁺(1)≥F(a), X₀ will be introduced to model and I₀ will be included into the aggregated of selected indicators.

(b) Build a bivariate regression model of the dependent variable Y with the subset of independent variables \{X₀, X₁, X₂\},..., \{X₀, X₁, X₄\},..., \{X₀, Xₙ\} (i.e. the regressors of this regression model are binary), with a total of p-1. The value of the statistic for the F-test of the regression coefficient of the variable, denoted as F₁⁺(k \equiv I₁), is calculated and the largest of these is chosen and denoted as F₁⁺(2), corresponding to the independent variable foot marked i₂, i.e. F₁⁺(2) = max \{ F₁⁺(2),..., F₁⁺(I₁+1),..., F₁⁺(p) \}. For a given level of significance a, note the corresponding critical value of F(a), if F₁⁺(2)≥F(a), then the variable X₂ is introduced into the regression model. Otherwise, the variable introduction process is terminated.

(c) Consider the regression of the dependent variable on the subset of variables \{X₀, X₃, X₄\} repeating step (b).

This method is repeated, each time selecting one of the independent variables not introduced into the regression model, until it is tested that no variables are introduced. Oppositely, by the backward method, all the independent variables are selected into the regression model beforehand, and any independent variables that contribute less to the sum of squares of the residuals are
removed one after another.

2) Coefficient of Determination $R^2$

Coefficient of determination measures the goodness of fit of the estimated regression equation, which is to examine how closely the sample data is clustered around the sample regression line. In other word, how well the regression equation represents the sample data is thus determined. Apparently, the closer each sample observation (scatter) is to the sample regression line, the greater the value of SSR/SST is and the better the line fits.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum(y - \bar{y})^2}{\sum(y - \bar{y})^2} \tag{3}$$

Usually, the observations fall partially on the regression line, which shows $0 < R^2 < 1$. The closer $R^2$ is to 1, the better the line fits. Oppositely, the closer $R^2$ is to 0, the worse the line fits. If all the observations fall on the regression line, $R^2 = 1$, the model has full explanatory power, the fit is complete and the model has full explanatory power; if the regression line does not explain any of the variance, the total variance in $y$ is all attributable to the sum of squared residuals, i.e., $SST = SSE$, and $R^2 = 0$, indicating that the independent variable $x$ (selected indicators) lacks explanatory power for changes in the dependent variable $y$ (severity of epidemic).

2.5.2.2. Ordinary Least Squares (OLS)

1) Definition of OLS

OLS approach refers to finding a line (or hyperplane) that minimizes the sum of squares of the difference between the predicted and actual values (residuals).

2) Application of OLS

In this paper, the data processing software Stata will be used as an OLS calculation tool to calculate the regression coefficients representing the relationship between each indicator and severity of epidemic.

3) Preparation before establishing a model with OLS

The results of OLS may be inaccurate due to the potential for strong correlation between indicators, as OLS requires that there is no collinearity across indicators. Therefore, in this paper, correlations between indicators were excluded prior to regression analysis. The "corr" command in the data processing software Stata is used to calculate the correlation values ($r$-values) between the indicators, which were then judged according to the criteria ($|r| \geq 0.8$ represents high correlation, $0.5 \leq |r| < 0.8$ represents medium correlation, $0.3 \leq |r| < 0.5$ represents low correlation, $|r| < 0.3$ represents very weak correlation or no correlation). The indicators with high correlation are then excluded. In addition, it is not taken into consideration whether the two indicators, broken down into from Oxford Stringency Index, have a strong correlation with other indicators. This is because Oxford Stringency Index changes over time and it is difficult to determine whether to exclude the indicators. Moreover, Oxford Stringency Index measures the stringency of government responses to epidemic situations and is an important indicator in this paper. Therefore, it is not considered whether Oxford Stringency Index has a strong correlation with other indicators.

1) Correlation analysis between indicators

As illustrated in Figure 2-5-2-2-1, indicators with strong correlation are listed as follows:

1) The proportion of citizens having personal access to Internet, and deaths due to infectious diseases and prenatal undernutrition are strongly correlated with the proportion of citizens ages under 14.

2) The proportion of citizens ages under 14, deaths due to lack of clear water and medical facilities, deaths due to infectious diseases and prenatal undernutrition, and national life expectancy are strongly correlated with the crude fertility rate.

3) The proportion of citizens ages under 14, national life expectancy, crude fertility rate, the proportion of citizens having personal access to Internet, deaths due to lack of clear water and medical facilities, deaths due to infectious diseases, and prenatal undernutrition are strongly correlated with cover rate of health insurance.

4) The proportion of citizens ages under 14, deaths due to lack of clear water and medical facilities, deaths due to infectious diseases, and prenatal undernutrition are strongly correlated with national life expectancy.

5) Deaths due to lack of clear water and medical facilities is strongly correlated with deaths due to infectious diseases and prenatal undernutrition.

6) The proportion of citizens ages under 14 is strongly correlated with the proportion of citizens ages above 65.

7) Government expenditure on health care is strongly correlated with GDP per capita.

2) Correlation between indicators with high correlation
As illustrated in Figure 5, the value of crude fertility rate and the proportion of citizens ages under 14 in each country can be fitted to a straight line. Therefore, there is a strong correlation between the value of crude fertility rate and the proportion of citizens ages under 14 in each country.

As illustrated in Figure 6, national life expectancy and the cover rate of health insurance in each country can be fitted to a straight line. Therefore, there is a strong correlation between national life expectancy and the cover rate of health insurance in each country.

As illustrated in Figure 7, national life expectancy and the proportion of citizens having personal access to Internet in each country can be fitted to a straight line. Therefore, there is a strong correlation between national life expectancy and the proportion of citizens having personal access to Internet in each country.

3) Indicators need to be excluded

The proportion of citizens ages under 14, government expenditure on health care, crude fertility rate, national life expectancy, deaths due to lack of clear water and medical facilities, and the cover rate of health insurance in each country are strongly correlated with various other indicators. Therefore, these indicators will be removed to ensure an accurate result of OLS.

3. EXPERIMENT RESULTS

3.1. Correlation between each indicator and severity of epidemic and the regression coefficients of each indicator

3.1.1. Correlations between indicators and severity of epidemic at 30 days after outbreak and regression coefficients of indicators significantly correlated with severity of epidemic

(1) Value of $R^2$

The correlation between the indicators and severity of epidemic is shown in Table 8. The value of $R^2$ is 0.6668, meaning that the independent variable $X$ (selected indicators) explains about 66.68% of the value of the dependent variable $Y$ (severity of epidemic).

(2) Result of F-test

The value of Prob > F is 0.0000, which is less than 0.05, indicating there is a significant correlation between selected indicators and severity of epidemic.

(3) T-test

According to t-test, the significant indicators are total population, containment health index, and population density of each country. Their value of $P > |t|$ is 0.000, which is less than 0.05.

(4) Regression coefficient

The regression coefficient of total population, population density of each country, and containment health index are respectively $2.38\times10^{-11}$, $-4.50\times10^{-6}$, and 0.0020015.
### 3.1.2. Correlations between indicators and severity of epidemic at 60 days after outbreak and regression coefficients of indicators significantly correlated with severity of epidemic

1. **Value of R²**

The correlation between the indicators and severity of epidemic is shown in Table 9. The value of R² is 0.5044, meaning that the independent variable X (selected indicators) explains about 50.44% of the value of the dependent variable Y (severity of epidemic).

2. **Result of F-test**

The value of Prob > F is 0.0000, which is less than 0.05, indicating that there is a significant correlation between selected indicators and severity of epidemic.

3. **T-test**

According to the t-test, the significant indicators are total population, and containment health index. Their value of P > [t] is 0.000, which is less than 0.05.

4. **Regression coefficient**

The regression coefficient of total population of each country and containment health index are respectively 5.79e-10 and 0.0046872.

### 3.1.3. Correlations between indicators and severity of epidemic at 90 days after outbreak and regression coefficients of indicators significantly correlated with severity of epidemic

Table 9 shows the correlation analysis of selected indicators and severity of epidemic at 60 days after the outbreak.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs =</th>
<th>154</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2.28726337</td>
<td>2</td>
<td>1.14363138</td>
<td>Prob &gt; F =</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>2.47404951</td>
<td>152</td>
<td>0.14681223</td>
<td>R-squared =</td>
<td>0.5044</td>
</tr>
<tr>
<td>Total</td>
<td>4.76131588</td>
<td>154</td>
<td>0.029636501</td>
<td>Root MSE =</td>
<td>0.1219</td>
</tr>
</tbody>
</table>

### 3.1.4. Correlations between indicators and severity of epidemic at 120 days after outbreak and regression coefficients of indicators significantly correlated with severity of epidemic

1. **Value of R²**

The correlation between the indicators and severity of epidemic is shown in Table 10. The value of R² is 0.2965, meaning that the independent variable X (selected indicators) explains about 29.65% of the value of the dependent variable Y (severity of epidemic).

2. **Result of F-test**

The value of Prob > F is 0.0000, which is less than 0.05, indicating there is a significant correlation between selected indicators and severity of epidemic.

3. **T-test**

According to t-test, the significant indicators are the proportion of citizens ages above 65, number of physicians (as per 1,000 citizens), GDP per capita, the proportion of citizens having personal access to Internet, containment health index, and index of the government's economic support of each country. Their value of P > [t] are respectively 0.023, 0.000, 0.015, 0.000, 0.017, and 0.000 which are all less than 0.05.

4. **Regression coefficient**

The regression coefficient of the proportion of citizens ages above 65, number of physicians (as per 1,000 citizens), GDP per capita, the proportion of citizens having personal access to Internet, containment health index, and index of the government's economic support of each country are respectively -2.613378, 19.69898, 0.00767, -1.119718, 0.8158994, and 1.402467.
(1) Value of $R^2$

The correlation between the indicators and severity of epidemic is shown in Table 11. The value of $R^2$ is 0.2345, meaning that the independent variable X (selected indicators) explains about 23.45% of the value of the dependent variable Y (severity of epidemic).

(2) Result of F-test

The value of Prob > F is 0.0000, which is less than 0.05, indicating there is a significant correlation between selected indicators and severity of epidemic.

(3) T-test

According to t-test, the significant indicators are GDP per capita, number of physicians (as per 1,000 citizens), and deaths due to infectious diseases and prenatal undernutrition of each country. Their value of $P > |t|$ are respectively 0.004, 0.000, and 0.014, which are all less than 0.05.

(4) Regression coefficient

The regression coefficient of GDP per capita, number of physicians (as per 1,000 citizens), and deaths due to infectious diseases and prenatal undernutrition of each country are respectively 0.0017123, 40.29059, and 1.670834.

Table 11 Correlation analysis of selected indicators and severity of pandemic at 120 days after the outbreak

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs</th>
<th>F(150)</th>
<th>Adj R-squared</th>
<th>Root MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>710818.567</td>
<td>3</td>
<td>239371.856</td>
<td>154</td>
<td>15.32</td>
<td>0.2345</td>
<td>125</td>
</tr>
<tr>
<td>Residual</td>
<td>234656.251</td>
<td>150</td>
<td>15623.834</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3056701.8</td>
<td>153</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. CONCLUSION

4.1. Reasons for the significant correlation between some indicators and severity of epidemic

The correlation between the explanatory and predicted variables can be judged by the significance of the coefficients of the linear regression model. In the initial stage of epidemic when death cases were only reported in few countries, indicators of these countries were yet very significant. In addition, 150, 180, and 210 days after outbreak, indicators evaluating a country’s medical level, economic strength, and population movement have a significant impact on severity of epidemic.

4.2. Discussion on whether the regression coefficient of total population of the countries is reasonable regarding its positive or negative value

4.2.1. Discussion on whether the regression coefficient of total population of the countries is reasonable regarding its positive or negative value at 30 days after outbreak

The regression coefficient of the total population of each country at 30 days after outbreak is positive, namely severity of epidemic increases with the size of the population, which is in line with common sense. Usually, the greater population a country has, the greater group of citizens who are vulnerable to COVID-19, the more severe the epidemic will become. However, it can be a coincidence since the epidemic outbreak in China. If the epidemic outbreak in a country with a small population, the regression coefficient of the total population of each country at 30 days after outbreak is likely to be negative. The positive or negative value of the regression coefficient of total population of the countries cannot determine or support a positive or negative correlation between total population and severity of epidemic of each country. Therefore, the regression coefficient of the total population of each country at 30 days after outbreak is unreasonable.

4.2.2. Discussion on whether the regression coefficient of total population of the countries is reasonable regarding its positive or negative value at 60 days after outbreak

The regression coefficient of the total population of each country at 60 days after outbreak remains positive, namely severity of epidemic increases with the size of the population. This is because only few countries, including China, France, Italy and Philippine, have reported death cases. It is obvious in the figure that the large population of China has greatly impacted the value of the regression coefficient. Too few countries reporting death cases leads to the fact that the regression coefficient of the total population of each country at 60 days after outbreak is unreasonable.
4.3. Discussion on whether the regression coefficient of index of government’s economic support of the countries is reasonable regarding its positive or negative value

4.3.1. Discussion on whether the regression coefficient of index of government’s economic support of the countries is reasonable regarding its positive or negative value at 90 days after outbreak

Countries with reported death cases have attached their attentions to COVID-19 at 90 days after outbreak. Therefore, those countries score high in regards of index of government’s economic support. The regression coefficient of index of government’s economic support of each country at 90 days after outbreak is positive, namely severity of epidemic increases with government’s economic support, which is different from common sense. Usually, the higher the index of government’s economic support is, the more subsidies government will offer to citizens affected by epidemic. Government aims at encouraging citizens stay at home, thus suppressing the spread of COVID-19. The effect of government economic support to individuals in suppressing the epidemic is likely to be insignificant. Therefore, regression coefficient of index of government’s economic support of each country at 90 days after outbreak is reasonable.

4.4. Discussion on whether the regression coefficient of the number of physicians (as per 1,000 citizens) is reasonable regarding its positive or negative value

4.4.1. Discussion on whether the regression coefficient of the number of physicians (as per 1,000 citizens) is reasonable regarding its positive or negative value at 90 and 120 days after outbreak

The regression coefficient of the number of physicians (as per 1,000 citizens) of each country at 90 and 120 days after outbreak is positive, namely severity of epidemic increases with the number of physicians, which is different from common sense. Usually, the more physician per 1,000 citizens a country has, the higher medical level that country has, the more difficult it is for COVID-19 to spread in the country, the less death cases the country should have. A possible reason for the positive correlation is that the impact of physician on epidemic may take time to take effect. Therefore, regression coefficient of the number of physicians (as per 1,000 citizens) of each country at 90 and 120 days after outbreak is reasonable.

4.5. Discussion on whether the regression coefficient of GDP per capita is reasonable regarding its positive or negative value

4.5.1. Discussion on whether the regression coefficient of GDP per capita is reasonable regarding its positive or negative value at 90 and 120 days after outbreak

The regression coefficient of GDP per capita of each country at 90 and 120 days after outbreak is positive, namely severity of epidemic increases with GDP per capita, which is different from common sense. Usually, the higher GDP per capita is in a country, the stronger a country is by means of economic, the higher medical level that country has [9], the more policies suppressing epidemic can be adopted, the less death cases the country should have. However, GDP per capita is also correlated to population movement. In the initial stage of epidemic, countries having high GDP per capita like France, Italy, Spain, and the US reported more death cases compared with less developed countries. This is because these countries are characterized of significant population movement, making it relatively easy for COVID-19 to spread in these countries. Therefore, regression coefficient of GDP per capita of each country at 90 and 120 days after outbreak is reasonable.

4.5.2. Discussion on whether the regression coefficient of GDP per capita is reasonable regarding its positive or negative value at 150 and 180 days after outbreak

The regression coefficient of GDP per capita of each country at 150 and 180 days after outbreak remains positive, which suggests high GDP per capita cannot suppress epidemic. According to calculation, GDP per capita is not necessarily correlated to number of physicians and containment health index. In addition, impact of physicians and containment health index on suppressing epidemic is insignificant, which aligns with the fact that developed countries, including France, Italy, Spain, Belgium, and the US, experienced relatively severe epidemic. Therefore, regression coefficient of GDP per capita of each country at 150 and 180 days after outbreak is reasonable.

4.6. Discussion on whether the regression coefficient of population density is reasonable regarding its positive or negative value at 30 days after outbreak

The regression coefficient of population density of each country at 90 and 120 days after outbreak is negative, namely severity of epidemic increases when population density decreases. This is because China was the only country with confirmed cases, whose population density...
is not high compared with other countries in the world. Therefore, regression coefficient of population density of each country at 30 days after outbreak is unreasonable.

4.7. Discussion on whether the regression coefficient of population density is reasonable regarding its positive or negative value

4.7.1. Discussion on whether the regression coefficient of containment health index is reasonable regarding its positive or negative value at 30 days after outbreak

Due to the importance attached to epidemic, China scored high in regards of containment health index. Additionally, China was the only country reported death cases in the first 30 days after outbreak. Thus, the regression coefficient of containment health index of each country at 30 days after outbreak is positive, namely severity of epidemic increases with containment health index, which is different from common sense. Usually, the higher containment health index is, the stronger government response to epidemic is, making it more difficult for COVID-19 to spread in the country, thus reducing the death cases. If China had not attached importance to epidemic after death cases were reported, namely China scored low in containment health index, the regression coefficient of containment health index would have remained positive. It can be concluded that the regression coefficient of containment health index cannot determine whether containment health index and severity of epidemic is positively or negatively correlated. Therefore, regression coefficient of containment health index of each country at 30 days after outbreak is unreasonable.

4.7.2. Discussion on whether the regression coefficient of containment health index is reasonable regarding its positive or negative value at 60 days after outbreak

The regression coefficient of containment health index of each country at 60 days after outbreak remains positive. This is because only few countries, including China, France, Italy, and Philippine, have reported death cases. It is obvious that containment health index of China and two other countries have a significant impact on the value of coefficient of regression. Given that only few countries have reported death cases, the value may not make sense. In addition, COVID-19 is highly infectious and its spread cannot be suppressed instantly. For example, measures including lockdowns of cities and school closures may take time to take effect. China, with a high containment health index, reported less death cases than many other countries in the mid-late stage. Therefore, regression coefficient of containment health index of each country at 60 days after outbreak is reasonable.

4.7.3. Discussion on whether the regression coefficient of containment health index is reasonable regarding its positive or negative value at 90 days after outbreak

The regression coefficient of containment health index of each country at 90 days after outbreak remains positive. A period of 90-days is sufficient for measures like lockdowns of cities and school closures to take effect. However, countries with high containment health index reported increasing death cases. This is because effect of government response to epidemic and increased expenditure in the mid stage were too insignificant to suppress the spread of epidemic. Therefore, regression coefficient of containment health index of each country at 90 days after outbreak is reasonable.

4.8. Discussion on whether the regression coefficient of proportion of citizens having personal access to Internet is reasonable regarding its positive or negative value at 90 days after outbreak

The regression coefficient of proportion of citizens having personal access to Internet of each country at 90 days after outbreak is negative, namely severity of outbreak decreases with the rises in proportion of citizens having personal access to Internet. Proportion of citizens having personal access to Internet evaluates a country’s Internet cover rate. The higher Internet cover rate of a country, the greater chance its citizen will have to access mass information. During epidemic, countries provided their citizens with knowledge related to COVID-19, raising citizens’ awareness of its prevention [10]. Therefore, regression coefficient of proportion of citizens having personal access to Internet of each country at 90 days after outbreak is reasonable.

4.9. Discussion on whether the regression coefficient of proportion of citizens ages above 65 is reasonable regarding its positive or negative value at 90 days after outbreak

The regression coefficient of proportion of citizens ages above 65 of each country at 90 days after outbreak is negative, namely severity of outbreak decreases with the rises in proportion of citizens ages above 65. This is because aging population structure usually occurs in developed countries with advanced medical technology and sufficient medical resources. Advanced medical technology and sufficient medical resources suggest greater capability in terms of handling public health emergencies. Therefore, regression coefficient of proportion of citizens ages above 65 of each country at 90 days after outbreak is reasonable.
4.10. Discussion on whether the regression coefficient of deaths due to infectious diseases and prenatal undernutrition is reasonable regarding its positive or negative value at 150 and 180 days after outbreak

The regression coefficient of deaths due to infectious diseases and prenatal undernutrition of each country at 120 days after outbreak is positive, namely severity of epidemic increases with deaths due to infectious diseases and prenatal undernutrition. The more deaths due to infectious diseases and prenatal undernutrition, the worse a country’s health care infrastructures, physical fitness and standard of living of its citizens are, the more vulnerable citizens are to COVID-19. Therefore, regression coefficient of deaths due to infectious diseases and prenatal undernutrition of each country at 150 and 180 days after outbreak is reasonable.

4.11. Summary of regression analysis of the indicators

Regression coefficient of containment health index and which of the number of physicians are positive, which is counter-intuitive. Many developed countries did not pay attention to COVID-19 in the initial stage. It was not until the epidemic had become severe that they started to attach importance to epidemic, which leads to the counter-intuitive result. Therefore, their containment health index started off low, increasing with severity of epidemic. Characterized of significant population movement, developed countries did not realize that significant population movement would lead to potential outbreaks, which is well demonstrated by the positive regression coefficient of the number of international arrivals and the number of passengers traveling by air in each country. Their overlook in the initial stage led to the insignificant effect of their following policies and strategies. Due to strong infectiousness of COVID-19, confirmed cases in the mid-late stage far exceeded the capacity of physicians, making the effect of increased expenditure on health care insignificant. Therefore, it is more of importance to attach attention to epidemic and make relevant preparation in an early stage.

4.12. Suggestions on suppressing the spread of COVID-19

1. Minimize population movement in an early stage

Characterized of highly communicative, COVID-19 has raged in many countries, including those who have implemented major measures in the mid stage. Additionally, countries with significant population movement turned out to be suffering the most from epidemic. Therefore, it is critical to minimize population movement in an early stage.

2. Increase government expenditure on health care

There is a reverse relationship between number of hospital beds (as per 1,000 citizens) and severity of epidemic. An increased government expenditure on health care will sharply increase the number of hospital beds in a short period, making medical treatment available to more COVID-19 patients, thus decreasing the death rate and severity of epidemic. Therefore, an increased government expenditure on health care will contribute to the suppressing of epidemic. Through establishing alternate care sites where extra hospital beds can be set, more COVID-19 patients will receive effective treatment.

3. Publicize knowledge about prevention of COVID-19

There is a reverse relationship between proportion of citizens having personal access to Internet and severity of epidemic. A high cover rate of Internet can be utilized for dissemination of information, raising citizens’ awareness of COVID-19 and addressing its prevention through various channels including television, mobile apps, online news, suppressing the spread of COVID-19, thus decreasing the death rate and severity of epidemic. Therefore, publicizing knowledge related to the prevention of novel coronavirus helps to suppress the spread of epidemic.

4.13. Inspiration from the research and its result

After China’s success in suppressing and controlling epidemic, calls arise, especially in many other countries where COVID-19’s rampage remains, arguing that other countries should reference and copy China’s strategies suppressing epidemic. However, the argument is unreasonable because it can be concluded from the indicators of different countries that the condition and context of each country varies. Therefore, a successful solution in a country will not necessarily work in all the countries’ condition and context. Each country should customize its own epidemic suppressing strategies based on its condition and context rather than copying from other countries. For example, countries with rich and sufficient medical resources can focus on improving Internet coverage so that public awareness of COVID-19 can be increased. Whereas countries with a high Internet cover rate can give top priority to increase medical resources, making medical treatment available to more COVID-19 patients, thus reducing death cases and severity of epidemic.

REFERENCES


