Analysis and prediction of COVID-19 in the US based on the time-varying parameters SIR model

Juan Chen¹,a, Yi Xiong¹, Jiani Tong¹ and Yuanyuan Wang¹*

¹ School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, Zhejiang Province, 310000, China
a hzchj930@zjgsu.edu.cn
*Corresponding author’s e-mail: 1677540432@qq.com

Abstract. With the data from the official website of the World Health Organization, this study calculated the transmission coefficient and other time-varying parameters to analyze the current epidemic situation in the US based on the SIR epidemic model. The reasons for the failure of the first wave of the epidemic were given from the perspectives of protection measures and the actual reproductive number. At the same time, the SIR model was improved based on time-varying parameters-β, and predicted the future development of the epidemic. The results showed that the actual reproductive number of the day fluctuated between 1.5 and 2 in May and June, but the various regions successively implemented the unblocking policies, which led to the crisis of the epidemic again. In this study, a time-varying parameter-β was introduced to predict that the outbreak in the US would turn around in late August. However, the trend of the epidemic is not optimistic, the US should invest more in developing COVID-19 vaccines.

1. Introduction
Novel Corona virus has been rapidly spreading around the world since February 2020. By August 9, 2020, the epidemic in most countries has been effectively controlled. However, the current number of confirmed cases in the US gradually leveled off and then rose rapidly, indicating that the epidemic situation remains very serious.

Since the spread of COVID-19, a number of studies have analyzed and predicted the development of the epidemic based on epidemic models[1]. Based on the SIR model, Zhu Renjie et al. [1] added parameters to represent the change of infection coefficient over time, simulated and predicted the development trend of the epidemic in many countries. Yu Zi et al. [3] used time-varying parameters to predict the development of epidemic situation. However, most of the studies are on the prediction of the epidemic situation in China, and few are in the US. In addition, unlike other countries, the US is experiencing a second outbreak. Therefore, it is necessary to analyze the reasons for the outbreak and predict the future trend for epidemic control research.

This study calculated time-varying parameters to make an analysis of the current epidemic situation in the US. The reasons for the failure of the first wave of the epidemic in the US were given based on the actual reproductive number. At the same time, the SIR epidemic model was improved based on time-varying parameters to predict the future development of the epidemic.

2. SIR epidemic model
In the SIR model, the population consists of three parts: S is the number of susceptible people; I is the number of infected people; R is the number of removed people. If an infected individual is in contact with a susceptible individual, the probability of the susceptible individual being infected is \( \beta \); an individual in I state will change to an R state with probability \( \gamma \) per unit time (The unit time is one day). Therefore, the SIR propagation process can be described by the following three dynamic equations:

\[
\begin{align*}
\frac{dS}{dt} &= -\beta \frac{SI}{N} \\
\frac{dI}{dt} &= \beta \frac{SI}{N} - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{align*}
\]

(1)

Where \( N \) represents the total number of individuals in the system, and \( N = S(t) + I(t) + R(t) \).

It was assumed that the proportion of infected and recovered persons could be ignored. Then, the total number of susceptible persons could always be equal to the total population, that is \( S = N \). Substitute \( S = N \) into formula 2 to obtain the simplified model of SIR:

\[
\begin{align*}
\frac{dI}{dt} &= \beta I - \gamma I \\
\frac{dR}{dt} &= \gamma I 
\end{align*}
\]

(2)

3. Dynamic analysis of COVID-19 outbreak in the US based on time-varying parameters

3.1. Development status of COVID-19 in the US

According to the data from the official website of the WHO, we have collected the COVID-19 epidemic data in the US, and their development trend is shown in Fig. 1.

![Figure 1. U.S. epidemic development trend graph](image)

It can be seen from Fig. 1 that the cumulative and the current existing number of confirmed cases in the US are still rising. The cumulative number of confirmed cases surpassed 5 million for the first time on August 7. The current existing number of confirmed cases continues to rise after a brief decline in early June, making the US the worst-hit country in the world.

3.2. Calculation and Analysis of transmission coefficient and removal coefficient

In the SIR infectious disease model, \( I \) is the number of confirmed cases at time \( T \), which can be obtained by subtracting the cumulative number of confirmed cases from the cumulative number of
deaths and the cumulative number of cured cases. $R$ is the cumulative number of people removed at time $T$, and is the cumulative number of deaths and the cumulative number of cured persons. In addition, according to the actual epidemic prevention work, the data related to the epidemic were updated at a time interval of 1 day. Therefore, the above SIR dynamic equation was discretized in this study to obtain:

\[
\Delta I(t) = (\beta - r)I(t-1) \\
\Delta R(t) = rI(t-1)
\]

Where, $\Delta I(t)$ and $\Delta R(t)$ respectively represent the data difference between the number of confirmed patients $I$ and the number of accumulatively removed people $R$, with an interval of 1 day.

During the spread of the epidemic, the transmission coefficient $\beta$ represents the probability of infection between susceptible people and the source of infection. The removal coefficient $\gamma$ represents the probability that an infected person will be removed from the system by the virus. The US has adopted a series of prevention and control measures, so the spread of the virus ability is affected by the outside world will decline, and the ability to treat diseases also have a certain degree of increase. So the two coefficients $\beta$ and $\gamma$ both are assumed the functions of time. According to the dynamic equations (3) and (4), the expression of the time-varying function, $\beta(t)$ and $\gamma(t)$, are as follows:

\[
\beta(t) = \frac{\Delta I(t) + \Delta R(t)}{I(t-1)} \\
\gamma(t) = \frac{\Delta R(t)}{I(t-1)}
\]

The epidemic data were substituted into the equation (5) and (6) of the transmission coefficient and the removal coefficient. And their day values can be calculated and plotted as follows.

As can be seen from Fig. 2, in the early stage of the epidemic in the US, the transmission coefficient fluctuated sharply and remained high around 0.4; Following the declaration of a COVID-19 state of emergency on March 13, there was a significant linear downward trend; And then the trend slowing goes down exponentially.

According to the observation of Fig. 3 based on the data, on June 5, the transmission coefficient was as low as 0.01674. However, from June 5 to July 10, the transmission coefficient showed a slow upward trend. The change trend of the transmission coefficient first decreased and then increased corresponds to the change trend of the number of confirmed cases in the US over time. Since July 11, there has been a significant downward trend in the coefficient of transmission, and the outbreak in the US has turned a corner.
Similarly, except for the irregular and violent fluctuations at the beginning, the removal coefficient of the US showed a very slight linear upward trend after March 13. The cure coefficient and the death coefficient can be solved similarly with the removal coefficient, and the change trend is not obvious, so the graph analysis will not be done here.

3.3. Calculation and Analysis of the actual reproduction number \( R_0 \)

The actual reproduction number, \( R_0 \), is the average number of infections that can be transmitted by introducing an infected person into a completely susceptible population with an intervention. When \( R_0 < 1 \), the infectious disease will gradually disappear. When \( R_0 > 1 \), the infection spreads exponentially and becomes an epidemic. The larger the number, the more difficult it is to control the epidemic.

In the SIR model, \( \beta S \) and \( \gamma N \) are respectively known as the total number of infected persons infected with susceptible bodies and the total number of patients cured or unfortunately died. Therefore, according to the definition of the actual reproduction number, it's equal to the total number of infected persons infected with the susceptible body divided by the total number of infected persons cured or unfortunately died. And in this study, it is assumed that the total number of vulnerable populations is always equal to the total population. The formula of the actual infectious number is as follows:

\[
R_0 = \frac{\beta}{\gamma} \tag{7}
\]

From the time-varying parameters \( \beta(t) \) and \( \gamma(t) \), the actual reproduction number \( R_0 \) in the US on that day can be calculated, and the figure is shown below.

![Figure 4. the actual reproduction number trend chart](image)

As shown in Fig. 4, in the early stage of the epidemic, due to the lack of timely intervention by the US, the actual reproduction number calculated on that day fluctuated wildly. After March 13, it dropped rapidly and was within 10 in early April and 2 in early May. Therefore, the epidemic was under control in the US and the growth rate of the number of confirmed cases slowed down. However the U.S. government is eager to economic recovery, implement the unlock policy, increase the risk of the spread of the virus, lead to actual number stays near 1.5, still not less than 1, therefore, the existing number of diagnosis in early June after falling short and continue to rise, new champions league second outbreak of epidemic in the US.

4. COVID-19 Forecast IN the US Based on Time-Varying Parameters-SIR Model

4.1. the Establishment of Time-Varying Parameters-SIR Prediction Model

Known from the analysis of the above: Transmission coefficient \( \beta \) is time-varying parameters, it has a significant downward trend, and its variation can be represented by a function of time \( \beta = at + b \). Remove coefficient \( \gamma \), cured coefficient \( \gamma_1 \) and mortality coefficient \( \gamma_2 \) are regarded as an invariant parameter, and they show a very slight linear upward trend. They can be assumed to be constant in the SIR model. Therefore, in this paper, the dynamics function of the SIR infectious disease model was improved as follows:
\[
\frac{dI}{dt} = \beta(t)I - \gamma I \\
\frac{dC}{dt} = \gamma I \\
\frac{dD}{dt} = \gamma I
\]

(8)

(9)

(10)

Solution of differential Formula (8) can be obtained as follows:

\[
I = ce^{(\beta(t)-\gamma)t}
\]

(11)

By substituting \( t = 0 \) into Formula (11), it can be known that, that is \( C = I_0 \), the number of infected persons at the beginning. Take the natural logarithm of the left and right sides of Formula (11) to get:

\[
InI = \left( \beta(t) - \gamma \right) t + InI_0
\]

(12)

\( \beta = at + b \) can be substituted into Formula (12) to obtain SIR Prediction Model of the present number of confirmed patients:

\[
I = e^{(at+b-\gamma)t} + I_0
\]

(13)

Using R, parameters \( a \) and \( b \) can be estimated by quantitative regression.

4.2. Predict the Present Number of Confirmed Patients

4.2.1. Data and Invariant Parameters.

According to the dynamics of the outbreak for the US, transmission coefficient began to show a significant linear downward trend from July 11th till now. Therefore, we collected the COVID-19 data for 30 consecutive days on July 11, solstice, and August 9. And according to the time-varying parameter formula to calculate remove coefficient, cured coefficient and mortality coefficient of the day. They can be substituted for average value within 30 days since all of them are stationary time series with in 30 days, shown in the table below.

| Table 1. Constant parameter concrete value. |
|---------------------------------------------|
| Parameter | Transmission coefficient | Cured coefficient | Mortality coefficient |
| Value     | 0.0204                | 0.0466%            | 1.9651%             |

4.2.2. Quantile regression.

Quantile estimation[7] is the estimation of population percentile, which is an extension of the least square method based on the classical conditional mean model. It uses several quantile functions to estimate the global model[8].

The method were used to estimate the parameters of \( \beta \)-SIR prediction model by R software. Select 0.05, 0.25, 0.50, 0.75 and 0.95 points to fit parameters. Meanwhile, we output the average error of each fitting model. The parameter estimates obtained from the current model for predicting the number of confirmed cases in the United States are shown in Table. When the score of point is 0.5, the average error is the lowest, 2.1259%. Therefore, the present model of predicting the number of confirmed patients in the US based on this point to fit parameter value.

| Table 2. Quantile regression result table. |
|------------------------------------------|
| \( \tau \) | \( b \) | \( b-\gamma \) | average error |
| 0.05 | -0.00023 | 0.01783 | 4.3136% |
| 0.25 | -0.00021 | 0.01749 | 2.4152% |
| 0.50 | -0.00022 | 0.01780 | 2.1259% |
| 0.75 | -0.00022 | 0.01778 | 2.8769% |
4.2.3. Predict the Current Number of Confirmed Patients.
Based on the estimated parameter values, the model was used to predict the current trend in the number of confirmed cases in the US over the next 90 days, as shown in Fig. 6. It can be seen the current number predicted by the model is basically consistent with the actual number of confirmed cases in the previous known stage, which can make a reasonable prediction of the development of the epidemic.

![Figure 5](image)

**Figure 5.** Current projections for the number of confirmed cases in the US over the next 90 days

Combined with the predicted data and observed in Fig. 5, the epidemic in the US will reach an inflection point at the end of August 2020, when the number of confirmed cases in the US will peak at about 2.4 million. But the decline in the number of confirmed cases in the US has been very modest, with half a million people still expected to be diagnosed in November.

4.2.4. Calculate the Actual Transmission Coefficient.
The time-varying parameter-SIR model was used in this study to predict the number of confirmed cases, the cumulative number of deaths and the cumulative number of cured cases. According to the calculation formula of transmission coefficient and remove coefficient, we can calculate transmission coefficient and remove coefficient of the day. And the actual transmission coefficient was further obtained. Some results were listed in the following table.

| time    | Actual Transmission Coefficient |
|---------|---------------------------------|
| 2020-8-20 | 0.98891                         |
| 2020-8-21 | 0.96830                         |
| 2020-9-13 | 0.55111                         |
| 2020-9-14 | 0.53185                         |

As can be seen from the above table, the actual transmission coefficient started to be less than 1 on August 20, which corresponds to the time when the current number of confirmed cases peaks. In the coming years, the number of confirmed cases will decline as quickly as in other countries after reaching the peak if the US can control the movement of people. Meanwhile, the actual transmission coefficient will be less than 1 and continue to decline. In other words, the United States will reach an inflection point and successfully pass the second peak of the epidemic.
5. Conclusion and control suggestions

Based on the above analysis, this study draws the following conclusions:

Firstly, one of the reasons for the second outbreak of the virus in the US: the US government was anxious for economic recovery, implemented easing policies too early, and implemented unsealing in the whole US, which increased the risk of transmission of the virus.

Secondly, the second reason for the second outbreak of the epidemic in the US: Although the actual number of infections on that day during May was much lower than before, its value fluctuated between 1.5 and 2 and was not less than 1, which leading to the crisis of the epidemic again.

Thirdly, projections for future outbreaks in the US: After introducing a time-varying coefficient of transmission, this paper predicts a turnaround in the US by the end of August. Because of the large number of confirmed patients, the trend of the epidemic is still not optimistic.

Based on the above conclusions, this paper proposes the following prevention and control measures:

Firstly, government perspective. Government urgently needs a national policy to control the current outbreak. The central government needs to strengthen coordination and coordination among states, require them to timely and faithfully report and publish daily epidemic information. Meanwhile, they also can organize the base stations for health prevention and control, intensify the research and development of COVID-19 vaccine.

Secondly, media and expert perspective. The major infectious disease experts and media need to strengthen cooperation, on the news on television and the Internet platform publicity of epidemic prevention and control of protective knowledge, experience and popularize knowledge of public health in time and good health habits, calling for the importance of wearing a mask and home quarantine.

Thirdly, the crowd Angle. The effective control of the epidemic in the US still needs the cooperation of the public. People need to understand the government's measures to reduce unnecessary travel to avoid contact with infected people. They should be firm in their confidence and determination to fight against the epidemic and unite as one to overcome the difficulties.

Acknowledgments

First of all, we would like to thank the instructor of the thesis: Professor Juan Chen from the School of Statistics and Mathematics, Zhejiang Gongshang University. In the process of writing the thesis, Mr. Chen gave timely and careful guidance on the difficulties and doubts we encountered. At the same time, I would also like to thank the authors in the references. Through their research articles, we have a good starting point for the research topic. Finally, thank you all for your hard work.

References

[1] Zhu R J, Tang S, Liu T T, et al. (2020) Prediction and Prevention of COVID-19 Based on improved SIR Model on the impact on the development of COVID-19. Journal of Shanxi Normal University (Natural Science edition), 48: 33-38.

[2] Mei W J, Liu Z, Zhu J Y, et al. (2020) Real-time prediction model of COVID-19 limit IR. Journal of University of Electronic Science and Technology of China, 49: 362-368.

[3] Yu Z, Zhang G Q, Liu Q Z, et al. (2020) Evaluation and prediction of COVID-19 epidemic based on time-varying parameter-SIR model. Journal of University of Electronic Science and Technology of China, 49: 357-361.

[4] Li Z X, Gou C Y. (2020) Stability analysis of equilibrium state with diffusion SIR infectious disease model. Journal of Fuzhou University (Natural Science edition), 48: 420-424.

[5] Fan R G, Wang Y B, Luo M, et al. (2020) Transmission model and Inflexion prediction based on SEIR for COVID-19. Journal of University of Electronic Science and Technology of China, 49: 369-374.

[6] Zhi Y C, Chen D M, Yang L S. (2020) Covid-19 situation assessment and development Trend prediction based on improved SEIR model. Journal of Dongguan University of Technology, 27: 11-16.

[7] Pei Y. (2014) Quantile regression and its application [C]. Central China Normal University.
[8] Chen J B, Ding J J. (2008) Review of quantile regression technology. Statistics and Information Forum, 03: 89-96.