Dynamic analysis of Logistic growth rate and change points for the COVID-19 epidemic based on Bootstrap method

Kuo Zhang , Kun-Yu Jiang, Zheng Yang* , Zu-Yao Jia

School of Management and Economics, University Electronic Science and Technology of China, Chengdu 611731

Abstract

The COVID-19 pandemic has disrupted lives the world over for more than a year. Each country is still trying to contain the epidemic by using a variety of measures. This paper proposes a novel method to assess the development of the epidemic. A regression model of logistic function is established and the real-time parameters of the growth rate are estimated for analyzing the effect of control measures. And the growth rate changes are tested to detect the difference in the growth rate at different times. In order to overcome the impact of non-normal distribution of the regression errors, confidence intervals of the growth rate are constructed via Bootstrap algorithm. We apply the new method for estimating and testing the daily confirmed reports of the United States, Italy and Japan. The results show that the development of the epidemic situation in each region is different. This method is suitable for epidemic judge in a variety of situations, which can provide a reference for the current situation of the epidemic and future development trends.

Keywords：COVID-19, growth rate, change point, Logistic function, Bootstrap confidence interval

1 Introduction

Since the outbreak of the novel coronavirus pneumonia (COVID-19), more than 100 million people have been diagnosed globally, and more than 200 countries have been affected to varying degrees. How to evaluate the development of the epidemic, when the epidemic will turn, and how to predict the future development of the epidemic, and other issues have become the focus of global attention.

Recently, some countries have seen a fourth wave of epidemics. Different countries have also introduced corresponding policies, and the EU adjusts the list of permitted countries and regions every half month. According to EU standards, countries and regions with more than 200 new infections in 7 days per 100,000 people are considered high-risk, while countries with severe spread of the new crown variant virus will also be included in the federal disease control agency. List. Some states in the United States

* Corresponding author at：School of Management and Economics, University of Electronic Science and Technology of China, No.2006, Xiyuan Ave, West Hi-Tech Zone, Chengdu, Sichuan, 611731, China.

E-mail address：yangzheng@uestc.edu.cn (Z. Yang)

judge whether the state can be restarted based on the utilization rate of hospital beds, the basic number of new coronavirus infections and the number of consecutive new infections per day. Italy divides the country into red, orange and green regions based on the number of new cases each day. The red zone has the highest infection rate, and all bars, restaurants and most shops, including hairdressers and beauty salons, will be closed. In the orange zone, restaurants and bars will be closed, but hairdressers and beauty salons can continue to operate. The number of vaccines vaccinated globally has exceeded 300 million doses. Developed countries and some middle-income countries have carried out large-scale vaccination. The United States leads all countries with 80.54 million doses of vaccinations, while China and the European Union are not far behind. The support of the international community has eased the vaccine crisis in developing countries to a certain extent. However, the gap is still huge. The above-mentioned indicators for the development of the epidemic are slightly rigid and lack scientific explanations. Inadequate measures have led to repeated outbreaks of the epidemic.

In this paper, the change of the epidemic situation is judged by the change of the growth rate of the logistic population growth model during the epidemic. Looking for changes in the growth rate to reflect the changes in the epidemic in order to provide governments with an intuitive epidemic data that can be referred to. The Logistic equation was put forward by Belgian mathematician and biologist Pierre Francois Verhulst when he was studying the population growth model, which is an improvement of the Malthus population model (Malthus, 1798). Shen compared China's provincial and national levels [1]. Their work examines the applicability of the logical growth model and its implications for the study of the new coronary pneumonia pandemic and other infectious diseases. The results show that the logistic model can fit the epidemic data of 11 countries including China well.

This paper proposes a panel data model with change points for the Logistic growth rate[2-6]. Test whether there is a change point in the growth rate between the seven-day interval and the fourteen-day interval using dummy variables, testing whether the growth rate is significantly different between the seven-day interval and the fourteen-day interval. Based on the above-mentioned data, the extent of changes in the epidemic and the effectiveness of prevention and control are judged.

The basic definition of change point theory is that in a sequence or process, while a certain statistical characteristic changes at a certain point in time under the influence of systemic factors rather than accidental factors, which point in time is called the change point. Dehning et al. combined the established epidemiological model with Bayesian inference to analyze the time dependence of the effective growth rate of new infections[7]. Looking at the spread of COVID-19 in Germany, it is found that the point of change in the effective growth rate is closely related to the timing of publicly announced interventions. Therefore, the effects of interventions can be quantified, and the corresponding change points can be incorporated into the forecast of the number of future programs and cases.

At the same time, due to relatively small sample size, growth rate parameters and change point parameters do not necessarily conform to the standard normal distribution. Therefore, bootstrap method based on residual sampling is used to calculate the confidence interval[8]. Bootstrap is a type of non-parametric Monte Carlo method, which essentially resamples the observation information and then performs statistical inferences on the distribution characteristics of the population. This method makes full use of the given observation information, does not require other assumptions of the model or the addition of new observations, and has the characteristics of robustness and high efficiency. Zhang et al. estimated the R_0 in the early stage of the COVID-19 outbreak by fitting the reported sequence interval with a gamma distribution, simulating the possible cumulative epidemic trajectory and future daily morbidity infectivity[9]. Then the median of 95% confidence interval (CI) of R_0 value is estimated to be 2.28 (2.06-2.52) by using bootstrap resampling method. It is concluded that unless strict infection management and control measures are taken, COVID-19 may cause a larger outbreak.

The results of the study showed that countries in different regions had different timings of the outbreak, with the first outbreak in Asia. The development process of national epidemics in different regions is also not synchronized. Therefore, different countries in different regions should adopt different epidemic prevention measures. Most of the confidence interval of the growth rate does not include the x-axis. Rejecting the null hypothesis indicates that the parameters of the model are significant, which is statistically significant. At the same time, the growth rate and the change point parameters change almost simultaneously, which proves the accuracy and reliability of the model.

In other papers, Cooper studied the effectiveness of modeling the pandemic due to the spread of the new COVID-19 disease, and developed a susceptible removal (SIR) model, which provides a theoretical framework for investigating its spread in the community[10]. Fna et al. proposed a categorized mathematical model for the spread of COVID-19 disease[11]. They calculated the threshold of the basic reproduction number, studied the local stability of the free equilibrium of the disease, studied the basic reproduction number, and investigated the sensitivity of the model to changes in each parameter. Aza et al. have applied these simple algorithms for global prediction of COVID-19 cases, which are recurrent neural network (RNN), long-term short-term memory (LSTM), bidirectional LSTM (BILSTM), gated control Frequent Sending Unit

(GRUS) and Variant Automatic Encoder (VAE) [12]. This study is based on daily confirmed and recurring cases collected from six countries including Italy, Spain, France, China, the United States and Australia. Wang et al. integrated the COVID-19 epidemiological data updated before June 16, 2020 into a logistics model that meets the upper limit of the epidemic trend[13]. Then enter the upper limit value into the FbProphet model, which is a time series prediction model based on machine learning to derive the epidemic curve and predict the epidemic trend. Annas et al. established the SEIR model of COVID-19 and used data from Indonesia to analyze and numerically simulate the stability of the spread of COVID-19 by the SEIR model[14].

The structure of this paper is as follows. The second section derives the regression model of the growth rate parameter according to the logistic function, giving the estimation method of the growth rate and the test method of the change point. The third section introduces the research data. The fourth section presents the estimation results of the growth rate parameters and the test results of the change points. The fifth section is the conclusion.

2 Logistic growth rate regression model and change point test

2.1 Logistic growth rate regression model

When conducting epidemiological research, we usually adopt Logistic model for modeling and research. The differential equation of the infectious disease mechanics model is as follows.

The logistic model proposed by Pierre-François Verhulst to modeling of population growth in ecology. Yu et al. use it to study the new coronavirus epidemic [15]. Cramer proposed that it has become one of the essential tools for bioassays and has been applied in a variety of fields, including finance, statistics, and epidemiology [16].

The differential equation of the logistic function takes the mathematical form

$$\frac{dN_t}{dt} = \beta \cdot \left(1 - \frac{N_t}{K}\right) N_t \qquad (1)$$

where N_t, K and β represent the number of people infected at time t, the maximum population size and the growth rate of the number of infected people, respectively.

The discretized differential equation (1) is as follows:

$$\frac{\Delta N_t}{\Delta t} = \beta \cdot \left(1 - \frac{N_t}{K}\right) N_t \qquad (2)$$

Let $\Delta N_t = N_{t+1} - N_t$ and $\Delta t = 1$, then

$$N_{t+1} - N_t = \beta \cdot \left(1 - \frac{N_t}{K}\right) N_t \qquad (3)$$

Next, equation (3) is transformed as

$$\frac{N_{t+1}-N_t}{K}=\beta\cdot\left(\frac{K-N_t}{K}\right)\frac{N_t}{K} \tag{4}$$

Equation (4) indicates that there are no new confirmed cases on the day $t+1$ while the growth rate β=0. In order to find out the change point of this epidemic, we adopt the regression equation to estimate β and test the hypothesis of $\beta=0$.

Let $y_{t+1}=\frac{N_{t+1}-N_t}{K}, x_t=\left(\frac{K-N_t}{K}\right)\frac{N_t}{K}$. The regression model is

$$y_{t+1}=\alpha+\beta\cdot x_t+\varepsilon_{t+1} \tag{5}$$

where ε_{t+1} is the random error, α is the intercept term, and β is the growth rate. In fact，the parameters of α_t and β_t vary over time as each country changes the control measures. We estimate the dynamic growth rate β_t by using cross-sectional data. The regression model with the panel data is

$$y_{t+1,i}=\alpha_t+\beta_t x_{t,i}+\varepsilon_{t+1,i} \tag{6}$$

where $i=1,\ldots,m$ and m is the number of states of one country. For example, we estimate the dynamic growth rate β_t of the United States using cross-sectional data from 56 states. Moving the time t forward to obtain a series of dynamic growth β_t, $t=1,\ldots,T$.

Equation (6) is a predictive regression model. Based on the estimated parameters, we can get the one-step forecast $\hat{y}_{t+1,,i}$ condition on $x_{t,,i}$. This paper concentrate on the development of the epidemic situation through the dynamic growth rate, not study the issue of prediction. The growth rate in equation (6) can be estimated by the ordinary least square (OLS) method and can be perform the null hypothesis test $H_0:\beta_t=0$ and the alternative hypothesis $H_1:\beta_t\neq 0$ by the t-statistic. For the given significance level, if the t statistic is not greater than its critical value, don't reject H_0, Otherwise, reject H_0.

2.2 Test change points

The detection of change points furtherly analyzes the development of the epidemic. In order to test whether there is a significant change of the growth rate between the s^{th} day and the $s+k^{th}$ day, we set a dummy variable d_t. Let d_t =0 if $t=s$, and d_t =1 if $t=s+k$. The panel regression model with change points is as follows

$$y_{t+1,i}=\alpha_t+\beta_t x_{t,i}+\tau_t x_{t,i}d_t+\varepsilon_{t+1,i}, \tag{7}$$

where i=1,...,m, $t=s,\ s+k$. Estimate the regression model (7) uses OLS again, and test the significance of the change point parameters τ using the t-test method.

The test results of the parameter null hypothesis H_0: $\tau=0$ indicate three situations:

(I) Do not reject H_0, which shows that the epidemic situation has not changed significantly between the s^{th} day and the $s+k^{th}$ day.

(II) Reject H_0 and the sign of $\hat{\tau}$ is positive, which indicate that the epidemic is developing in a worse direction.

(III) Reject H_0 and the sign of $\hat{\tau}$ is negative, which indicate that the epidemic situation is developing in a better direction.

When $k=1$ in equation (7), the significance of $\hat{\tau}$ is to test whether the growth rate of two days has changed significantly. when $k=7$, it is tested whether the growth rate of one week has changed significantly. Different k values can compare the growth rates at different times.

2.3 Boostrap method based on residuals

Let the random error $\varepsilon_{t,i}(i=1,\ldots,m)$ be a sequence of i.i.d. random variables from an unknown probability distribution, F_ε. One can calculate $\hat{\beta}_t$ or $\hat{\tau}_t$ associated with distribution *F*. But, building up the confidence interval estimate of $\hat{\beta}_t$ or $\hat{\tau}_t$ may be difficult because the distribution of $\hat{\beta}_t$ or $\hat{\tau}_t$ is unknown. Bootstrapping provides an alternative estimation method, which is free from assumptions of distribution.

The procedure to establish a bootstrap confidence interval of $\hat{\beta}_t$ is summarized as follows:

Step 1. One can estimate the parameters $\hat{\alpha}_t$ and $\hat{\beta}_t$ in equation (6) and obtain the residuals $\hat{\varepsilon}_{i,t+1} = y_{i,t+1} - \hat{\alpha}_t - \hat{\beta}_t \cdot x_{i,t}$.

Step 2. Sample the centralized residuals $\tilde{\varepsilon}_{i,t+1} = \hat{\varepsilon}_{i,t+1} - \overline{\varepsilon}_{i,t+1}$ where $\overline{\varepsilon}_{i,t+1} = \sum_{i=1}^{m} \hat{\varepsilon}_{i,t+1}$. Let F_ε be estimated by the empirical distribution $\hat{F}_\varepsilon$ putting mass m^{-1} to $\tilde{\varepsilon}_{i,t+1}, i=1,\ldots,m$. The bootstrap data can be generated from the model with F_ε replaced by $\hat{F}_\varepsilon$. One can generate i.i.d. data $\varepsilon^*_{1,t+1},\ldots,\varepsilon^*_{m,t+1}$ from $\hat{F}_\varepsilon$ and define $y^*_{i,t+1} = \hat{\alpha}_t + \hat{\beta}_t \cdot x_{i,t} + \varepsilon^*_{i,t+1}$.

Step 3. Calculate the least squares estimator of $\hat{\beta}^*_t$ based on the data $(y^*_{1,t+1}, x_{1,t}),\ldots,(y^*_{m,t+1}, x_{m,t})$.

Step 4. Repeatedly do steps 2 to 3 until a total of B point estimate values, $\hat{\beta}^*_{t,1},\cdots,\hat{\beta}^*_{t,B}$ are acquired. A percentile bootstrap (PB) confidence interval for $\hat{\beta}_t$ is $[\hat{\beta}^*_{\alpha/2}, \hat{\beta}^*_{1-\alpha/2}]$, where $\hat{\beta}^*_{\alpha/2}$ and $\hat{\beta}^*_{1-\alpha/2}$ are the $\alpha/2$ and $(1-\alpha/2)$ percentiles of the istribution formed by B bootstrap point estimates, respectively. Here α is the significant level. So, we obtain a 95% confidence interval if α=0.05 .

Note that Efron and Tibshirani indicated that a rough minimum of B = 1000 is usually necessary to compute reasonably accurate confidence interval estimates[17]. Similarly, one can produce a confidence interval for $\hat{\tau}_t$ in equation (7) using the bootstrap estimation.

3 Data

As of March 15, 2021, the number of confirmed cases of new coronary pneumonia worldwide is 119,875,420. The epidemic is still raging in some countries. The cumulative number of confirmed cases in the United States is 29,438,775, which is the highest in the world. The cumulative number of confirmed cases in Italy is 3,223,142, while in Japan is 447,326.

The sample data selected in this article contains information on the development of the epidemic from January 21, 2020, the date of the initial epidemic, to February 21, 2021. Data set includes the daily cumulative number of confirmed cases of new coronary pneumonia, the number of newly diagnosed cases, and the total population of each state (or county). The data comes from Wind database [18]. The specific sample data source selection is shown in Table 1. The United States as a whole takes 56 states as a sample, Italy as a whole takes 20 districts, and Japan as a whole takes 47 administrative regions such as prefectures and counties.

Table 1. Sources of sample data

country	date	Include area	Data Sources
United States	2020/1/21-2021/2/21	56 states	Wind database
Italy	2020/1/31-2021/2/21	20 districts	Wind database
Japan	2020/3/7-2020/2/21	47 administrative regions	Wind database

4 Results of estimation

In this section, we estimate the growth rate $\hat{\beta}_7$ of the 7-day panel data and $\hat{\beta}_{14}$ the 14-day panel data by using the regression equation (6), and generate the confidence interval of the growth rate via bootstrap method. Next, the change point parameters $\hat{\tau}_7$ and $\hat{\tau}_{14}$ are estimated from the regression equation (7). The magnitude and significance of the growth rate, indicate the development of the epidemic and the effectiveness of prevention and control. The signs and significance of the change point parameters show the changes in the growth rate at different times. Since the use of one-day cross-sectional data has no significant statistical significance in most cases, the results of using one-day cross-sectional data are not discussed here.

The confidence interval of parameter estimation and test results are shown in Figure 1 to Figure 6. Figures 1 to 6 respectively show the timing diagrams of the parameters , $\hat{\beta}_7$ and $\hat{\beta}_{14}$, $\hat{\tau}_7$ and $\hat{\tau}_{14}$, in the United States, Italy and Japan. A 95% confidence interval is given. If 0 is included in the confidence interval, the null hypothesis is true and the parameter is not significant. Otherwise, the estimated parameter is significant. Each country is broken down into the pre- and post-epidemic period. The stage of pre-epidemic period refers to the period when the epidemic just broke out and the control of the epidemic is relatively lacking, while the stage of pre-epidemic period refers to the period of continuous improvement after government intervention and governance.

What should be noted is that the confidence interval generated by the bootstrap algorithm has a certain degree of randomness. The results may be slightly different each time. At the same time, as the number of sampling increases, the confidence interval distribution will get closer and closer to the true interval distribution.

4.1 United States

(a) The first wave of epidemics: From of February, 2020 to March, 2020

The growth rate rises suddenly and fluctuates violently. The τ value crosses the abscissa back and forth, and most of the time it is positive. What it shows is that the new crown epidemic broke out at this time, the government has taken corresponding measures but the effect is not good, some people do not listen to the government's suggestions, and officials within the government have different opinions. At this time, the epidemic is very unstable and is developing in a worse direction.

(b) The second wave of epidemics: From April, 2020 to June, 2020

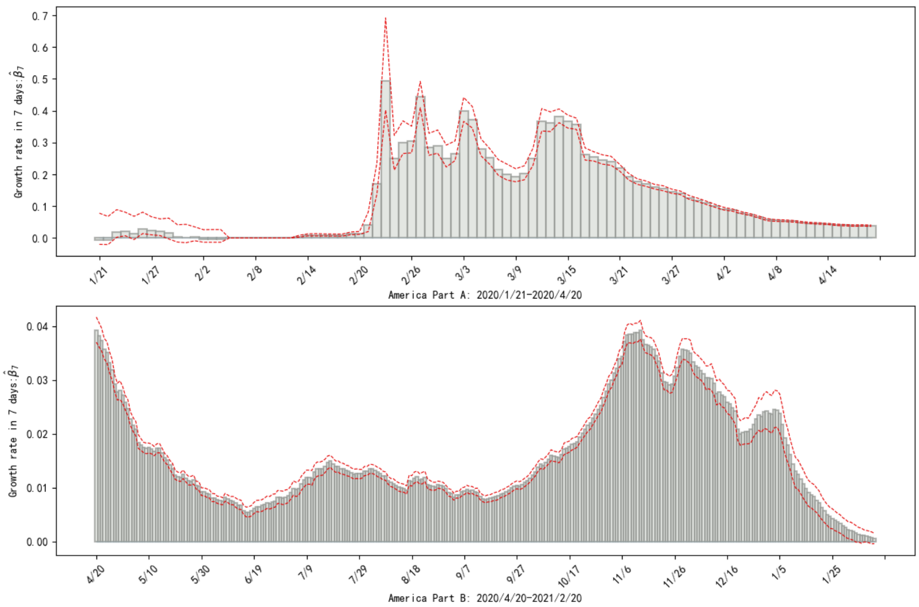
The growth rate gradually decreases. the value of τ is negative mostly, fluctuating on the axis of abscissa, indicating that the growth rate has dropped significantly. What is the reason of it is that as the new crown epidemic has become more serious, the government has increased its control efforts. The growth rate gradually decreases, and the value is mostly negative, fluctuating on the axis of abscissa, indicating that the growth rate has dropped significantly. This is because as the new crown epidemic becomes more serious, the government has increased its control efforts. The nationwide lockdown measures initiated in mid-March have prevented the spread of the new crown virus. Measures such as expanding social distancing, stay-at-home orders, and class suspension orders have achieved results, and the rising curve of the epidemic has gradually flattened. People have gradually realized the harmfulness of the virus. The epidemic has gradually eased.

(c) The third wave of epidemics: From June, 2020 to July, 2020

The growth rate increased significantly, and the value of $\hat{\tau}$ is significantly positive， which means that the growth rate at this stage is increasing and the epidemic situation is getting worse. The main reason is the "sequelae" of riots in many places caused by racial discrimination in the previous period. At the same time, the U.S. government underestimated the epidemic in order to restart the economy for the election. Some people in the U.S. were affected by the government not paying attention to maintaining social distancing and wearing masks, which worsened the epidemic situation.

(d) The fourth wave of epidemics: From October,2020 to January, 2021

The growth rate is significantly positive and rises rapidly, and the value of τ is also significantly positive, indicating that the epidemic has deteriorated significantly, and the third attack of the autumn and winter epidemic has begun, which is caused by the removal of some epidemic prevention measures in order to restore the economy in order to restore the economy in some states of the United States, seeing the situation improving.



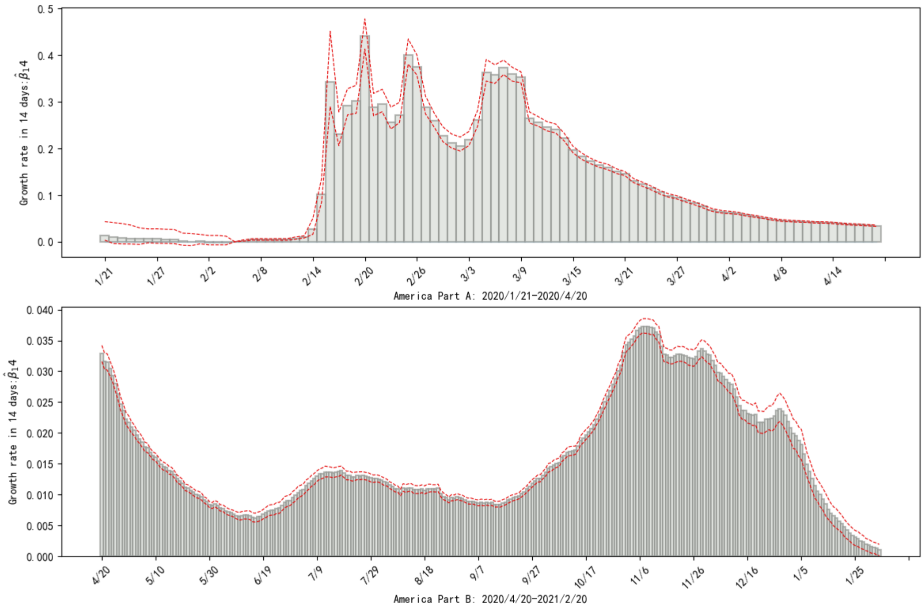
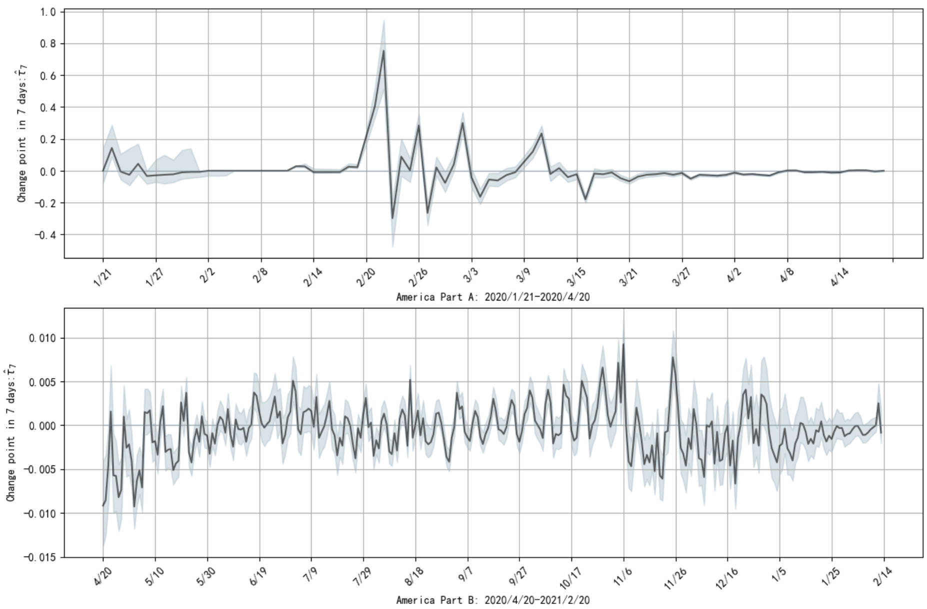


Figure 1. U.S. growth rate: The black bars represent parameter estimates, and the gray shading represents the 95% confidence interval. The abscissa is time (unit: day), the ordinate $\hat{\beta}_7$ is the weekly growth rate, $\hat{\beta}_{14}$ is the bi-weekly growth rate.



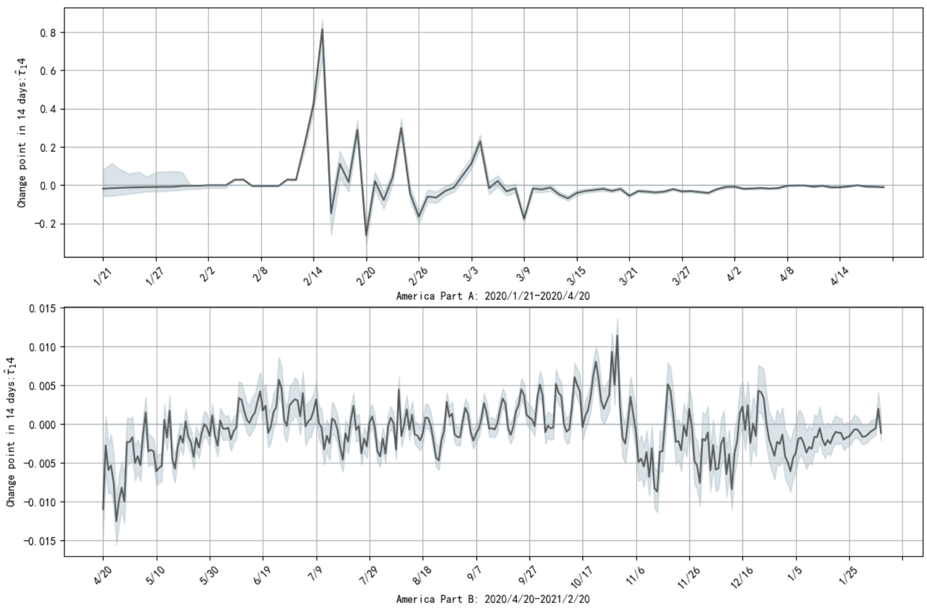


Figure 2 U.S. Change Point Parameters: Black bars indicate estimated values, and gray shading indicates 95% confidence interval. The abscissa is time (unit: day), the ordinate τ_7 is the estimated value at one week interval, and the estimated value τ_{14} is at two-week interval.

4.2 Italy

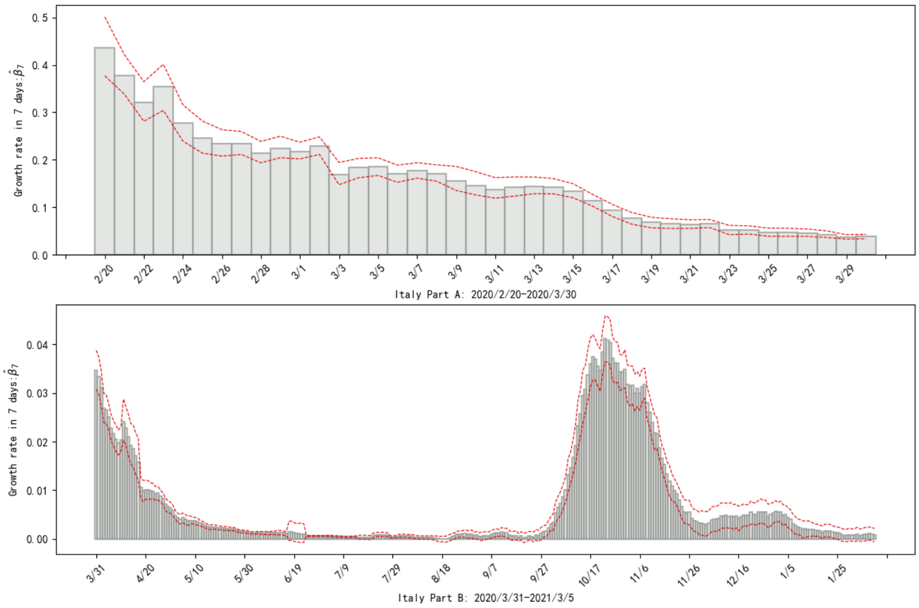
(a) The first wave of epidemics: From February, 2020 to September, 2020:

The growth rate is close to the highest peak and then continues to decline. the value of τ crosses the abscissa back and forth, and is negative most of the time. It shows that the new crown epidemic broke out in February, and then the government intervention measures were appropriate and the results were remarkable, so the epidemic was gradually controlled. This is due to the decree passed by the Italian Council of Ministers on February 23: restricting traffic in 11 northern cities and towns where cases are concentrated, suspending production, public gatherings and teaching activities, and strengthening publicity and disinfection throughout the country, which has a certain degree of restraint on the development of the epidemic. Moreover, the Italian quarantine bill was upgraded to the entire territory of Italy on March 1. The Italian government has successively issued closure orders to close schools, businesses, parks, etc., and suspend flights. Although these measures have not been able to completely control the development of the epidemic, judging from the results of the increase in the infection rate, they have also reduced the rapid change in the growth rate of the epidemic to a certain extent.

In addition, the insignificant growth rate around June 15 can be interpreted as the effectiveness of Italy's epidemic prevention measures. At this time, according to the government's plan, the government will further relax its control from June 15th, reopening children's summer camps, dance halls, theaters and other places under the premise of maintaining social distancing and wearing masks.

(b) The second wave of epidemics: From October,2020 to January, 2021

The growth rate suddenly rises and then declines. Most of the confidence interval for the growth rate does not contain 0, which is significant at the 95% level. The value of τ rises rapidly to positive, and then fluctuates on the axis of abscissas, indicating that the growth rate from October to January increased first and then decreased. This shows that the epidemic rebounded at this time, and the second wave of epidemics broke out in autumn and winter, which is related to the government's relaxation of epidemic prevention measures.



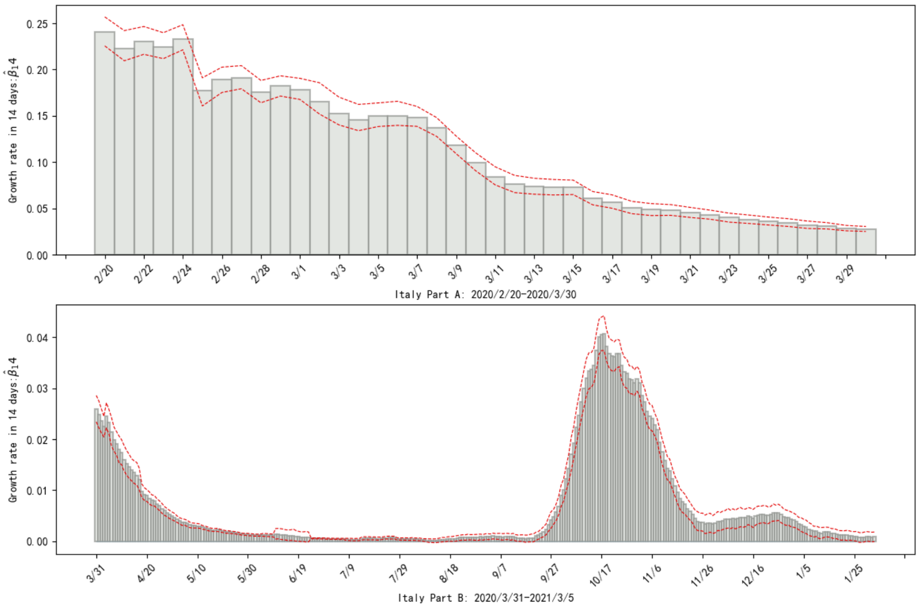
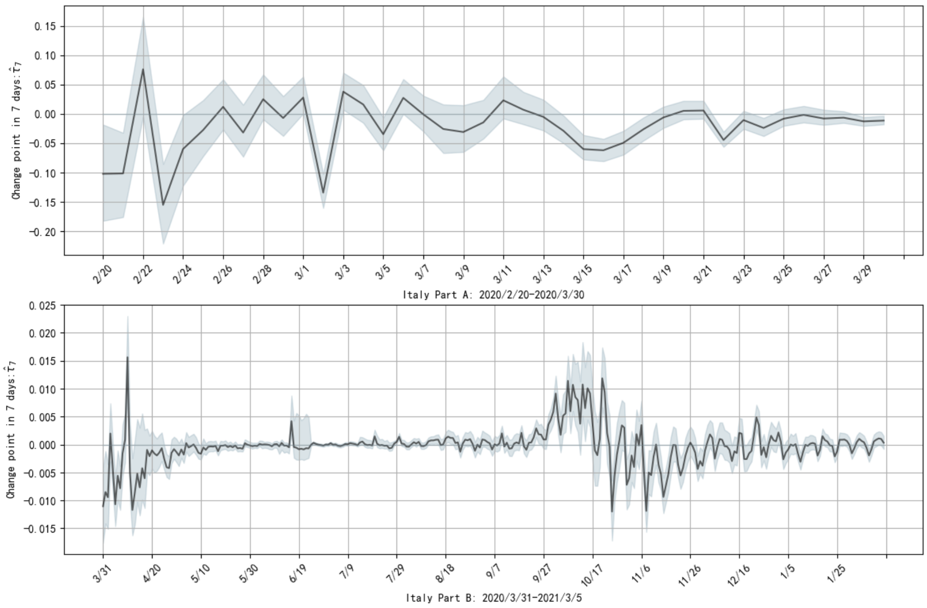


Figure 3 Italian growth rate: The black bars represent parameter estimates, and the gray shading represents the 95% confidence interval. The abscissa is time (unit: day), and the ordinate $\hat{\beta}_7$ is the weekly growth rate, $\hat{\beta}_{14}$ is the bi-weekly growth rate.



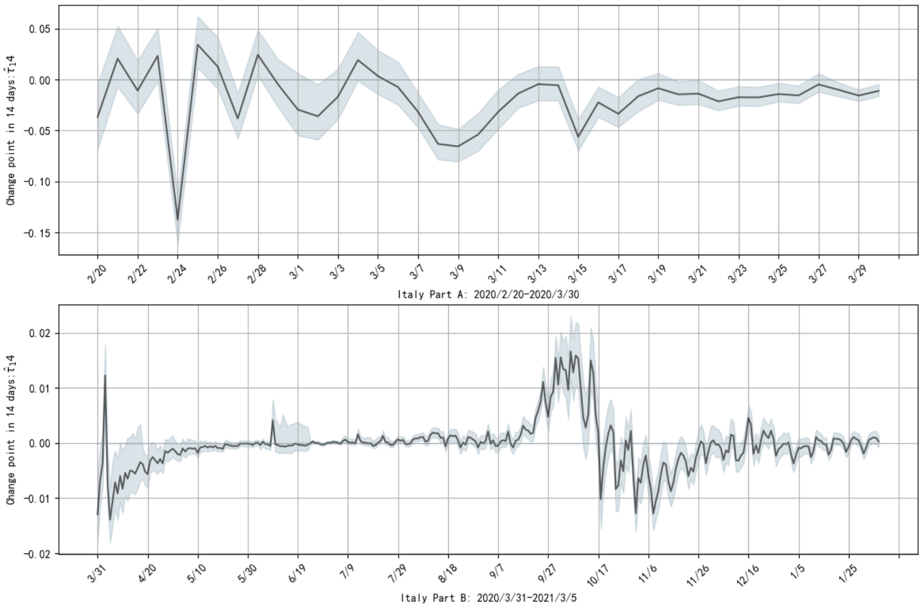


Figure 4. Italian change point parameters: black bars indicate estimated values, and gray shades indicate 95% confidence intervals. The abscissa is time (unit: day), the ordinate τ_7 is the estimated value at a one-week interval, and the estimated value τ_{14} is at a two-week interval.

4.3 Japan

(a) The first wave of epidemics: from March, 2020 to April, 2020

The growth rate in March was significantly positive, and τ the value was significantly positive, indicating that the first wave of epidemics swept through in March. The main reason is that Japan discovered its first new crown case in February, and the Abe's government has been playing down the epidemic in order to protect the Olympics, which has gradually worsened and led to the first major outbreak.

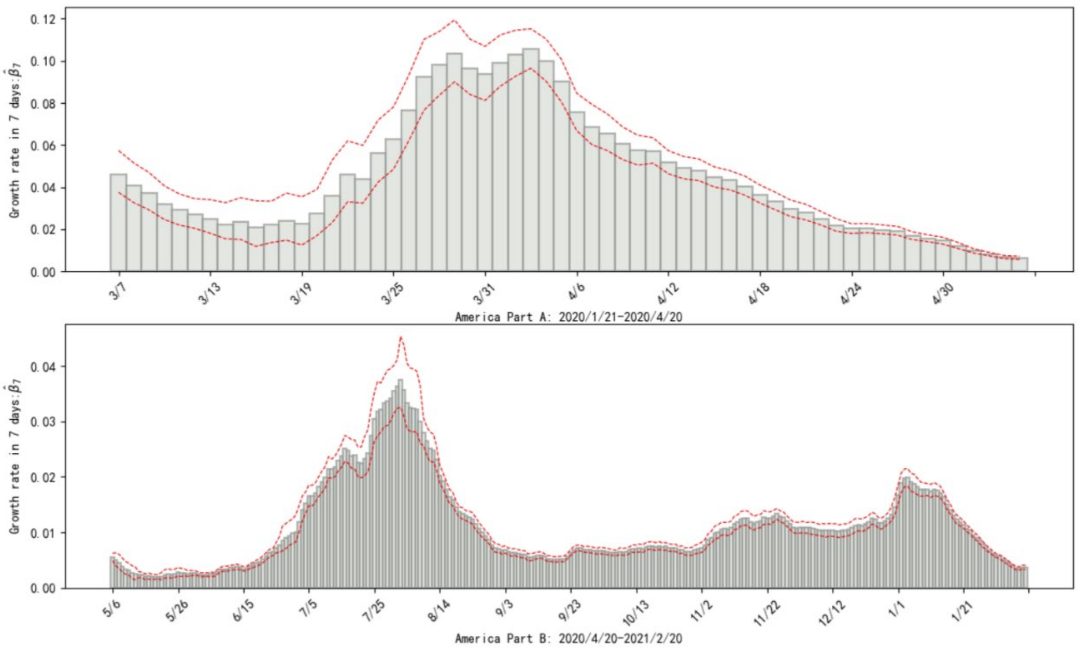
(b) The second wave of epidemics: from June, 2020 to August, 2020

The growth rate in July was significantly positive, and the value of τ was significantly positive, indicating that the first wave of epidemics swept through in July. The main reason is that Tokyo issued the "Tokyo Alert" formulated by the Metropolitan Government on June 2 after the central government lifted the state of emergency. After that, the alert was lifted on June 11. On June 19, night entertainment venues were lifted. Business restrictions have completely lifted the restrictions on business closures, which paved the way for the rebound of the epidemic.

Nightlife venues have become hotbeds for the rebound of the epidemic, especially Tokyo's entertainment venues have formed a transmission chain in the metropolitan area. For example, 31 of the 54 confirmed cases in Tokyo on June 26 were related personnel in the downtown area at night, such as the Cowboy Bar. Among the 54 new cases confirmed on June 30, 15 cases of them were staff and customers at night-time entertainment venues. Since then, the chain of infection has been further extended to nearby areas with the movement of people. The city of Yokohama, which is adjacent to Tokyo, notified 28 new cases on June 30, 26 cases of which came from the same Cowherd nightclub. According to reports, some of them had previously worked in a nightclub in Tokyo where the infection had occurred. This nightclub in Yokohama has now a total of 32 confirmed cases. Another similar incident occurred in Saitama Prefecture, which is adjacent to Tokyo. A staff member of a local nightclub was infected in Tokyo and spread in the nightclub.

(c) The third wave of epidemics: from November, 2020 to January, 2021

At the end of December, the growth rate increased and decreased again, which shows that the epidemic has rebounded slightly. Significantly the positive value of τ indicate a significant increase in growth. What is the reason is that Japan did not face up to and attach importance to the fight against the epidemic. In terms of epidemic prevention and control measures, Japan has not implemented a nationwide blockade. It is eager to open up when the epidemic has not been contained, and even encourages travel to activate the economy. Japanese companies, catering and commerce have basically not closed down. Unrestricted social mobility provides a basis for the spread of the virus and cross-infection.



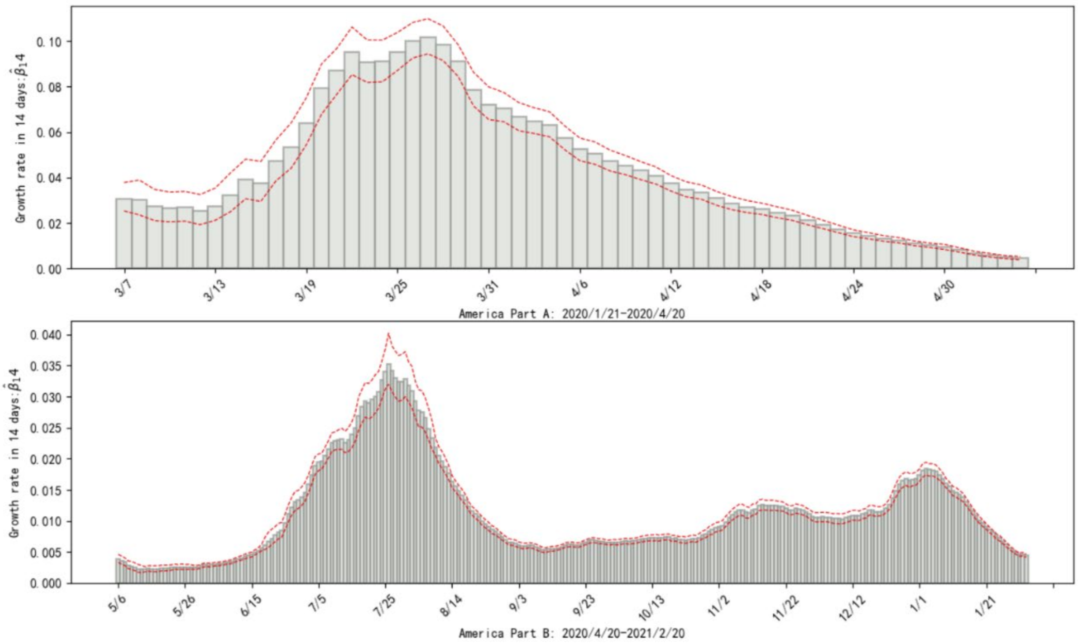
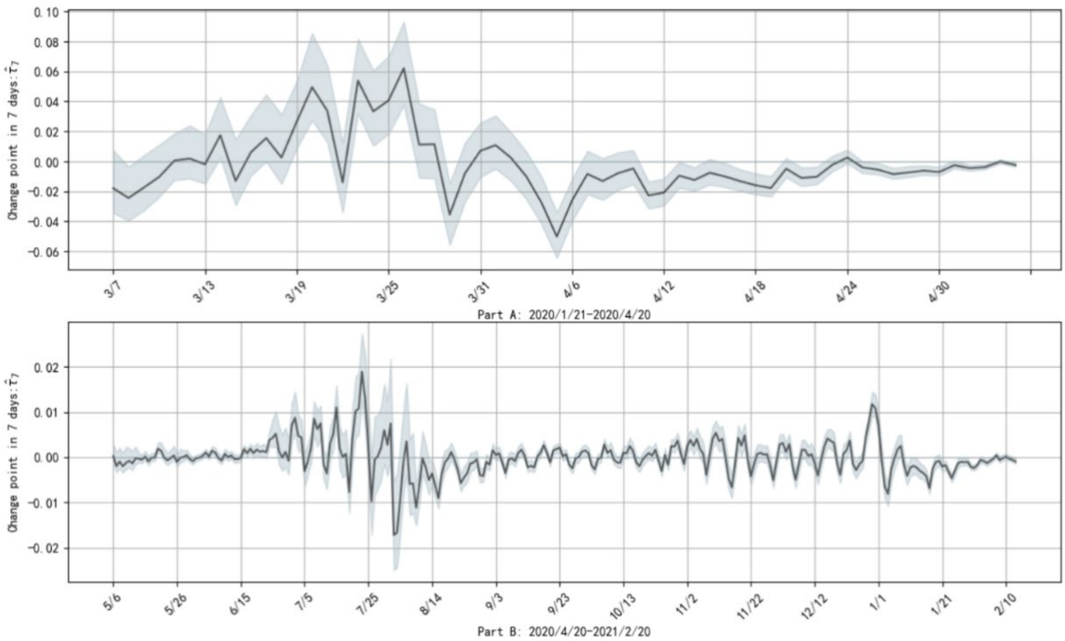


Figure 5 Japanese growth rate: black bars represent parameter estimates, and gray shading represents 95% confidence interval. The abscissa is time (unit: day), and the ordinate $\hat{\beta}_7$ is the weekly growth rate, $\hat{\beta}_{14}$ is the bi-weekly growth rate.



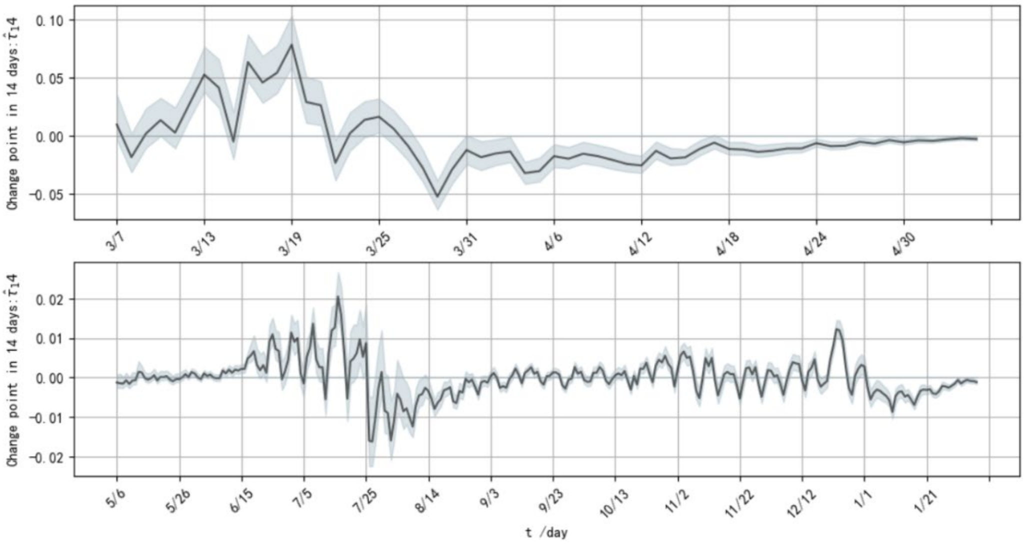


Figure 6: Japanese change point parameters: black bars indicate estimated values, and gray shading indicates 95% confidence interval. The abscissa is time (unit: day), the ordinate τ_7 is the estimated value of one week interval τ_{14} , and the estimated value of two weeks interval.

4.4 Summary

Comparing and summarizing the daily development of the epidemic situation in the three countries, we have the following findings.

First, the time of the outbreak of the epidemic in different regions is different. In this article, Japan is the first to break out, followed by Italy, and then the United States. As of February 21, 2021, their growth rates are finally close to 0, indicating that the final epidemic will eventually be controlled by humans.

Second, in the process of real-time calculation of growth rate estimates and significance testing, even if the growth rate is not significant for a long period of time, this model can quickly identify the point when the epidemic has rebounded significantly again. At the same time, the growth rate change and the change point parameter change are almost synchronized, which shows the accuracy and reliability of the model.

Finally, the development of the epidemic situation in countries in different regions is not synchronized. The epidemic broke out in Japan first, and then properly managed to quickly control the epidemic. Although there were two slight rebounds in the epidemic later, it is generally possible to see that the growth rate is showing a downward trend. After the outbreak of the epidemic in Italy, the growth rate began to fluctuate sharply after a rapid increase. After a period of continuous decline, the epidemic suddenly began to rebound again. The United States has performed the worst in this regard. After the growth rate rose rapidly and slowly fell, it rose rapidly and then fell again.

This shows that different prevention and control management measures should be adopted for different regions. At the same time, the confidence interval of the growth rate in most periods of this article does not contain 0, indicating that the model is supported by the data and is statistically significant.

5 Conclusions

The growth rate estimation and real-time change point test considered in this paper have wide applicability (target objects) and high flexibility. In terms of regions, in addition to estimating the growth rate of the country's administrative regions, it can also be extended to a specific place at the micro level, such as New York, the United States, Wuhan, China, and so on. The same method can be used to estimate and test the significance of the area’s growth rate, using the data on the number of infected people counted by smaller administrative units. In the change point test of the growth rate, in addition to the day and week as the comparison time interval, the test can also be performed at the interval of months. One can consider choosing the time freely, so as to more clearly judge the degree of epidemic change and the effect of prevention and control. At the same time, this paper takes into account the problem of small samples, uses the bootstrap algorithm based on residuals to obtain the confidence interval, and also solves the problem of variables that do not obey the t distribution, making the results more accurate.

Logistic algorithm is the scientific authority, which is suitable as a supplementary standard for epidemic judgment. The national epidemic situation can be judged from the significance and magnitude of the growth rate, which serves as a unified reference index between countries. It can also be used to predict the number of infections and predict the future development of the epidemic.

Since the research results are based on data-driven calculations, which depends on the authenticity of the data, detailed and accurate data is very important. Each epidemic country shall provide scientific advice for assessing the epidemic situation and restarting the economy based on the accurate epidemic data obtained and the estimated growth rate.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Disclosures

There are no disclosures. This study has not been submitted elsewhere for publication.

References

[1]. Shen, C. Y. (2020). Logistic growth modelling of COVID-19 proliferation in China and its international implications. International Journal of Infectious Diseases, 96, 582-589.

[2]. Pelinovsky, E. , Kurkin, A. , Kurkina, O. , Kokoulina, M. , & Epifanova, A. . (2020). Logistic equation and covid-19.

[3]. Ma, M., Bao, J., Fang, J., Li, C., Li, Y., Ge, W., ... & He, L. (2021). Impact of Headache Frequency and Sleep during the COVID-19 Pandemic among Migraine Patients. Current Neurovascular Research.

[4]. Wang, J., Jing, R., Lai, X., Zhang, H., Lyu, Y., Knoll, M. D., & Fang, H. (2020). Acceptance of COVID-19 Vaccination during the COVID-19 Pandemic in China. Vaccines, 8(3), 482.

[5]. Pariente, G. , Broder, O. W. , Sheiner, E. , Battat, T. L. , & Wainstock, T. . (2020). Risk for probable post-partum depression among women during the covid-19 pandemic. Archives of Women's Mental Health(5), 1-7.

[6]. D Tάtrai, & Z Vάrallyay. (2020). Covid-19 epidemic outcome predictions based on logistic fitting and estimation of its reliability.

[7]. Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., & Priesemann, V. (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science, 369(6500).

[8]. Chatterjee, A., & Lahiri, S. (2010). Asymptotic properties of the residual bootstrap for lasso estimators. Proceedings of the American Mathematical Society, 138(12), 4497-4509.

[9]. Zhang, S., Diao, M., Yu, W., Pei, L., Lin, Z., & Chen, D. (2020). Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. International journal of infectious diseases, 93, 201-204.

[10]. Cooper, I., Mondal, A., & Antonopoulos, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. Chaos, Solitons & Fractals, 139, 110057.

[11]. Ndaïrou, F., Area, I., Nieto, J. J., & Torres, D. F. (2020). Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. Chaos, Solitons & Fractals, 135, 109846.

[12]. Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. Chaos, Solitons & Fractals, 140, 110121.

[13]. Wang, P., Zheng, X., Li, J., & Zhu, B. (2020). Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. Chaos, Solitons & Fractals, 139, 110058.

[14]. Annas, S., Pratama, M. I., Rifandi, M., Sanusi, W., & Side, S. (2020). Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia. Chaos, Solitons & Fractals, 139, 110072.

[15]. Yu, Q., Liu, J., Zhang, Y., & Li, J. (2002). Simulation of rice biomass accumulation by an extended logistic model including influence of meteorological factors. International Journal of Biometeorology, 46(4), 185-191.

[16]. Cramer, J. S. (2004). The early origins of the logit model. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 35(4), 613-626.

[17]. Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical science, 54-75.

[18]. Hu, T., Guan, W. W., Zhu, X., Shao, Y., Liu, L., Du, J., ... & Bao, S. (2020). Building an open resources repository for COVID-19 research. Data and Information Management, 4(3), 130-147.. https://doi.org/10.2478/dim-2020-0012.