Tracking Reproductivity of COVID-19 Epidemic in China with Varying Coefficient SIR Model

Haoxuan Sun [1], Yumou Qiu[2], Han Yan[3], Yaxuan Huang[4], Yuru Zhu[5], Jia Gu[5] and Song Xi Chen[6,5]

Abstract

We propose a varying coefficient Susceptible-Infected-Removal (vSIR) model that allows changing infection and removal rates for the latest corona virus (COVID-19) outbreak in China. The vSIR model together with proposed estimation procedures allow one to track the reproductivity of the COVID-19 through time and to assess the effectiveness of the control measures implemented since Jan 23 2020 when the city of Wuhan was lockdown followed by an extremely high level of self-isolation in the population. Our study founds that the reproductibility of COVID-19 has been significantly slowed down in the three weeks from January 27 to February 17th with 96.3% and 95.1% reductions in the effective reproduction numbers R among the 30 provinces and 15 Hubei cities, respectively. Predictions to the ending times and the total numbers of infected are made under three scenarios of the removal rates. The paper provides a timely model and associated estimation and prediction methods which may be applied in other countries to track, assess and predict the epidemic of the COVID-19 or other infectious diseases.

Keywords: Epidemic assessment; Estimation of Basic reproductive number; SIR model; Varying coefficient model;

1: Center for Data Science, Peking University; 2: Department of Statistics, Iowa State University, Joint Corresponding Author; 3: School of Mathematical Sciences, Sichuan University; 4: Yuanpei College, Peking University; 5: Center for Statistical Science, Peking University; 6: Guanghua School of Management, Peking University, Corresponding Author.

1. Introduction

The Corona Virus Disease 2019 (COVID-19) has created a profound public health emergency in China and has spread to 25 countries so far (World Health Organization (WHO), 2020). It has become an epidemic with more than 76,000 confirmed infections and 2,244 reported deaths worldwide as on February 20 2020. The COVID-19 is caused by a new corona viruses that is genetically similar to the viruses causing severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). Despite a relatively lower fatality rate comparing to SARS and MERS, the COVID-19 spreads faster and infects much more people than the SARS-03 outbreak.

The city of Wuhan, the origin of the outbreak, has been locked up to reduce population movement since January 23 in an effort to stop the spread of the epidemic, followed by more than 50 prefecture level cities (as on 8th of February) and countless number of towns and villages in China. A high percentage of the population are exercising self-isolation in their homes. The spring festival holiday period had been extended with all schools and universities closed and all students staying where they are indefinitely. The country is virtually in a stand-still, and the economy and people's livelihood have been severely affected by the epidemic.

There is an urgent need to assess the speed of the disease transmission and to check if the existing containment measures have successfully slowed down the spread of the disease or not. The Susceptible-Infected-Removal (SIR) model (Kermack and McKendrick, 1927) and its generalizations, for instance the Susceptible-Exposed-Infected-Removal (SEIR) model (Hethcote, 2000) with four or more compartments are commonly used to model the dynamics of infectious disease outbreaks. See (Becker, 1977; Becker and Britton, 1999; Yip and Chen, 1998; Ball and Clancy, 1993) for statistical estimation and inference for stochastic versions of the SIR model. SEIR models have been used to produce early results on COVID-19 in (Wu et al., 2020; Read et al., 2020; Tang et al., 2020), which produced the first three estimates of the basic reproduction number R_0: 2.68 by (Wu et al., 2020), 3.81 by (Read et al., 2020) and 6.47 by (Tang et al., 2020). The R_0 is the expected number of infections by one infectious person over his/her infectious period *at the start of the epidemic*, which is closely connected to the effective reproduction number R_t. The latter R_t is the expected number of infections by one infected over infectious period at time t of the epidemic. Both R_0 and R_t are key measures of an epidemic. For fixed coefficient models, if $R_0 < 1$, the epidemic will die down eventually with the speed of the decline depends on the size of R_0; otherwise, the epidemic will explode until it runs out of its course.

The SEIR models that was employed in the above three cited works for the COVID-19 assume constant model coefficients, implying a constant regime of transmission during the course of the epidemic. This is idealistic for modeling COVID-19 as it cannot reflect the intervention measures by the authorities and the citizens, which should have made the infectious rate (β) and the effective reproduction number (R_t) varying with respect to time. Here, the effective reproductive number R_t is the average number of secondary infections made by each infectious case during an epidemic, which contrasts the basic reproductive number R_0 that measures the average number of secondary infections at the beginning of an epidemic.

To reflect the changing dynamic regimes due to the strong government intervention and the self protective reactions by citizens, we propose a varying coefficient SIR (vSIR) model. The vSIR model is easy to be implemented via the locally weighted regression approach (Cleveland and Devlin, 1988) that produces estimates with desired smoothness, and yet is able to capture the changing dynamics of COVID-19's reproduction, with guaranteed statistical consistency and needed standard errors. The consistent estimator and its confidence interval are proposed for estimating the trend of R, assessing the effectiveness of infection control, and predicting the ending time and the final number of infection cases with 95% prediction intervals.

As COVID-19 is quickly spreading outside China, the vSIR model and the associated estimation and prediction methods may be applied to other countries to track, assess and predict the epidemic of the COVID-19 or other infectious diseases.

2. Main Results

By applying the vSIR model, we produce daily estimates of the infectious rate $\beta(t)$ and the effective reproduction number R_t^D (t denotes time) based on three values of infectious duration D: 7, 10.5 and 14 days for 30 provinces and 15 major cities (including Wuhan) in Hubei province from January 21 or a later date between January 24-29 depending on the first confirmed case to February 17.

- Despite the total number of confirmed cases and the death are increasing, the spread of COVID-19 has shown a great slowing down in China within the two weeks from January 27 to February 17 as shown by 96.3% and 95.1% reductions in the effective reproduction number R_t among the 30 provinces and the 15 cities in Hubei, respectively.

- The average R_t^{14} (based on 14-day infectious duration) on January 27th was 6.14 (1.49) and 7.59 (2.38), respectively, for the 27 provinces and the 7 Hubei cities with confirmed cases by January 23rd. The numbers in the parentheses are the standard error. One week later on February 3rd, the R_t^{14} was averaged at 2.18 (0.67) for the 30 provinces and 2.84 (0.59) for the 15 Hubei cities, representing 64.5% and 62.6% reductions, respectively, over the 7 days. On February 10th, the average R_t^{14} dropped further to 0.86 (0.38) for the 30 provinces and 1.23 (0.55) for the 15 Hubei cities, which were either below or close to the critical threshold level 1.

- On February 17th, the average R_t^{14} has reached 0.23(0.15) and 0.37(0.24) for the 30 provinces and the 15 Hubei cities, with 22 provinces' and 8 Hubei cities' R_t^{14} being statistically significantly below 1 for more than 7 days. These indicate a further slowing down in the re-productivity of COVID-19 in China in the week from February 10 to 17.

- The profound slowing down in the reproductivity of COVID-19 can be attributed to a series of containment measures by the government and the public, which include cutting off Wuhan and other cities from January 23, a rapid public awareness of the epidemic and the extensive self protection taken and high level of self isolation at home exercised over a much extended Spring Festival holiday period.

- There are increasing numbers of provinces and cities in Hubei whose 14-day R_t has been statistically below 1, as detailed in Table 1, which would foreshadow the coming of the turning point for containment of the epidemic, if the control measures implemented since January 23 can be continued.

- If the recovery rate can be increased to 0.1 meaning the average recover time is 10 days after diagnosis, the number of infected patients $I(t)$ will be dramatically reduced in March, and the epidemic will end in April for non-Hubei provinces and end in June for Hubei.

3. Time-varying coefficient SIR model

Let $S(t)$, $I(t)$ and $R(t)$ be the counts of susceptible, infected and removed (including dead) persons in a given city or province at time t, respectively. Let N be the total population of the city/province. We propose a varying coefficient Susceptible-Infected-Removed (vSIR) model for the conditional means of the Poisson increments $\Delta I(t)$ and $\Delta R(t)$ given $I(t)$ and $R(t)$. This vSIR-Poisson framework permits estimating the parameters and the effective reproduction number R_t for the dynamics of COVID-19, which are then used for predicting the future spread of the disease.

The SIR model (Kermack and McKendrick, 1927) is a commonly used epidemiology model for the dynamic of susceptible $S(t)$, infected $I(t)$ and recovered $R(t)$ as a system of ordinary differential equations (ODEs). Here we consider a generalized version of the SIR model in that the infectious rate β and the removal rate γ may vary with respect to time so that the deterministic ODEs are

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\beta(t)I(t)\frac{S(t)}{N}, \\
\frac{dI(t)}{dt} &= \beta(t)I(t)\frac{S(t)}{N} - \gamma(t)I(t), \\
\frac{dR(t)}{dt} &= \gamma(t)I(t),
\end{aligned} \tag{1}$$

where $\beta(t)$ and $\gamma(t)$ are unknown infection and the removal rate functions, respectively. Once an individual is removed, including dead, the individual can not return to the susceptible group.

The rationale for using a time-varying $\beta(t)$ function, rather than a constant β, is that $\beta(t)$ is the average rate of contact per unit time multiplied by the probability of disease transmission per contact between a susceptible and an infectious subject. Due to an increasing public awareness of the epidemic and the control measures as mentioned earlier, both the transmission probability and the contact rate have been largely reduced. These favor for a time-varying $\beta(t)$, which are also confirmed by the sharp declined in $R_t^D = \beta(t)D$, where D denote the infectious durations in Figures 1 and 2. The removal rate also changes over time as treatments improve over time as shown in Figure 3. However, Figure 3 shows $\gamma(t)$ is much slowly changing for most of the provinces, which led us to treat $\gamma(t) = \gamma$ at the early stage of the outbreak, whose value gradually increased as the recover rate improved as the time progress and better treatments are available.

The deterministic vSIR model as specified by the ODEs in (1) specifies the conditional means of the Poisson increments $\Delta I(t)$ and $\Delta R(t)$ given $S(t)$, $I(t)$ and $R(t)$ at each discrete time point t. This conditional mean specification leads to a Poisson-vSIR model framework, which can be used to construct conditional likelihood for $(\beta(t), \gamma(t))$ over moving time windows and leads to statistical inference for the effective reproduction ratio estimation and its standard error. The Poisson-vSIR framework is also the basis for the bootstrap re-sampling algorithm that we will propose for generating predictive intervals.

SEIR model is an extension of SIR with an added compartment E for the exposed between S and I. A time-varying SEIR model (vSEIR) satisfies the ODEs

$$\begin{aligned}\frac{dS(t)}{dt} &= -\beta(\beta)I(t)s(t),\\ \frac{dE(t)}{dt} &= \beta(t)I(t)s(t) - \alpha(t)E(t),\\ \frac{dI(t)}{dt} &= \alpha(t)E(t) - \gamma(t)I(t),\\ \frac{dR(t)}{dt} &= \gamma(t)I(t),\end{aligned} \tag{2}$$

where $\alpha(t)$ is the confirmation or diagnosis rate from E to I. The ODEs in (2) specifies the conditional means of the independent Poisson increments. However, like SIR model, the states before I are not infectious.

The basic and the effective reproduction numbers (RN), R_0 and R_t, are important notions in epidemiology as they quantify the reproduction ability of an epidemic at the start (R_0) and during (R_t) an epidemic. For both SIR and SEIR models, $R_0 = R = \beta/\gamma$ (Hethcote, 2000). We are to demonstrate that for the vSIR model, $R_0 = \beta(0)/\gamma(0)$ and the effective RN $R_t = \tilde{\beta}(t)/\gamma(t)$ at time t where $\tilde{\beta}_t = \beta(t)s(t)$ and $s(t) = S(t)/N$. The susceptible rate $s(t)$ is approximately 1 at the start of an epidemic. However, $s(t) < 1$ has to be taken into account as the number of susceptibles declines.

Figure 4 provides the vSIR and vSEIR epidemic progression networks from an initial $I(0)$ infected and initial $I(0)$ and $E(0)$ infected, respectively. The figure also provides the According to the Poisson-vSIR model, at the start of the epidemic, conditioning on $I(0)$, in average $I(1) = (1 - \gamma(0) + \tilde{\beta}(0))I(0) > I(0)$ if and only if $1 - \gamma(t-1) + \tilde{\beta}(t-1) > 1$, which is if and only if $R_0 = \tilde{\beta}(0)/\gamma(0) > 1$. In general, at time t, conditioning on $I(t-1)$, in average $I(t) = (1 - \gamma(t-1) + \tilde{\beta}(t-1))I(t-1) > I(t-1)$ if and only if $1 - \gamma(t-1) + \tilde{\beta}(t-1) > 1$, which is the case iff the effective RN $R_{t-1} > 1$. Thus, indeed, R_t can track the trend of an epidemic being expanding or shrinking. A similar argument can be made under the vSEIR model.

4. Data

Daily records of infected, dead and recovered patients released by National Health Commission of China (NHCC) are obtained from the NHCC website, with the first confirmed record for Wuhan on December 8th, 2019, followed by 30 provinces in mainland China and 15 cities in Hubei province where Wuhan is the capital city. We did not consider data from Tibet due to very small number of cases. Due to severe under-reporting in the first 39 days of the epidemics in Wuhan and Hubei, we consider data from January 16th for Wuhan and Hubei. For other provinces and Hubei cities, the starting dates for data are those of first confirmed case, and the analysis date starts four days after to accommodate the estimation approach for the infectious rate $\beta(t)$. The latest start data for analysis was January 29th for Qinghai province and three cities in Hubei province. The second last analysis starting date was January 28th with two provinces and five Hubei cities. Table A1 in the Supplementary Information (SI) provides the starting dates of the data records and analysis for each province and Hubei city.

To guide for the choice of the infectious duration D used when calculating the reproduction number, we consider two public data sources. The first one is obtained in *Shenzhen Government Online*, which contain datasets released by the Shenzhen Municipal Health Commission from January 19th to February 13th (Shenzhen Municipal Affairs Service Data Administration, 2020). One dataset provides information on the confirmed cases that include the time of onset, time of hospital admission, cause of illness and other information of 391 cases, consisting 188 males and 203 females. The admission time of these cases ranged from January 9th to February 11th. Another Shenzhen dataset reports the discharge times for 94 recovered cases, contianed in the former dataset. The second data source comes from Shaoyang Municipal Health Committee (Shaoyang Municipal Health Commission, 2020) with a dataset of 100 confirmed cases released on February 14 that includes 48 male and 52 female patients with the onset dates ranging from January 12 to February 11.

5. Estimation and Confidence Intervals

The reported numbers of infected $I(t)$ and removed cases $R(t)$ are subject to measurement errors. To reduce the errors, we apply a three point moving average filter on the reported counts to obtain $\bar{I}(t) = 0.3I(t-1) + 0.4I(t) + 0.3I(t+1)$ for $2 \le t \le T-1$ where T is the latest time point of observation. In our analysis, T was February 20 of 2020,. For $t = 1$ or T, we apply two point averaging with $7/10$ weight at $t = 1$ or T, and $3/10$ for $t = 2$ or $T-1$. Apply the same filtering on the recovered process $R(t)$ and obtain $\bar{R}(t)$. To simplify the notation, we denote the filtered data $\bar{I}(t)$ and $\bar{R}(t)$ as $I(t)$ and $R(t)$ respectively, wherever there is no confusion.

Hubei started to report the "clinically diagnosed" cases on February 12th (13th for city Xianning) which created spikes in the newly reported cases. We applied a one-off linear filter that re-distributes the spikes in the Hubei cities and Hubei to the previous 7 days with decreasing weights ranging from 7/28 to 1/28.

Let $N(t) = I(t) + R(t)$ denotes the cumulative number of diagnosed cases and $\Delta N(t) = N(t+1) - N(t)$ denotes the daily change. Let $I(t+1, t+1)$ be the newly infected case at date $t+1$. Then, $\Delta R(t) = R(t+1) - R(t) = I(t) - I(t+1) + I(t+1, t+1)$ and $\Delta N(t) = \Delta I(t) + \Delta R(t) = I(t+1, t+1)$ which is the newly infected cases. Thus, conditional on $I(t)$, $(\Delta N(t), \Delta R(t))$ are conditionally independent Poisson random variables. There is a slight confusion between $N(t)$ and N, as the latter is used to denote the total population size.

We consider the likelihood for the vSIR-Poisson process framework for parameter estimation by treating β and γ as fixed and later we will relax it to allow they vary over a window of time t. Then,

$$\Delta N(t) \sim \text{Poisson}\left\{\beta(t)S(t)I(t)/N\right\} \text{ and } \Delta R(t) \sim \text{Poisson}\left\{\gamma I(t)\right\}.$$

The likelihood function for $(\Delta I(t), \Delta R(t))$ given $I(t)$ and $R(t)$ is

$$L(\beta, \gamma) = f_1(\Delta N(t)|I(t)) \times f_2(\Delta R(t)|I(t)) \tag{3}$$

where f_1 and f_2 are the conditional Poisson density functions. The log likelihood based on the increments at t is

$$\begin{aligned} l(\beta, \alpha, \gamma) \quad &\propto \quad -\beta(t)s(t)I(t) + \Delta N(t)\log\{\beta(t)s(t)I(t)\} - \gamma(t)I(t) \\ &\quad +\Delta R(t)\log\{\gamma(t)I(t)\}. \end{aligned}$$

As the population of each province/city is large and the number of total infected patients is still relatively small, the ratio $S(t)/N$ appeared in (1) is very close to 1. By approximating $S(t)/N = 1$, the likelihood score equations are

$$\frac{\partial l}{\partial \beta} = -I(t) + \frac{\Delta N(t)}{\beta(t)s(t)} \tag{4}$$

$$\frac{\partial l}{\partial \gamma} = -I(t) + \frac{\Delta R(t)}{\gamma(t)} \tag{5}$$

It can be checked that the score functions have (approximate) zero means. The approximation of $S(t)/N = 1$ is just to simplify the expression as everything carries through by using $\tilde{\beta}(t) = \beta(t)s(t)$.

While one can use the above likelihood based inference, an equivalent approach we use in our analysis is based on the (approximate) solution for $I(t)$ via (6)

$$I(t) \approx I(t_1) \exp\{(\beta(t) - \gamma)(t - t_1)\}, \tag{6}$$

for $t_1 = t - w + 1, \cdots, t$ and a window $w > 0$ which satisfies $w \to 0$ and $Tw \to \infty$. Here T is the total number of observational time for the processes. Take logarithm transform on (6), $\log\{I(t)\} \approx \log\{I(t_1)\} + (\beta(t) - \gamma)(t - t_1)$. We propose estimating $\beta(t) - \gamma$ by a local linear regression of $\log\{I(t)\}$ on $t - t_1$. The above log-linear regression may be viewed as a version of the Poisson increment mean model by noting that $\log\{I(t)\} - \log\{I(t_1)\} \approx \frac{I(t) - I(t_1)}{I(t_1)}$ which is approximately $(\beta(t) - \gamma(t))(t - t_1)$ in the mean.

Let $\widehat{\beta(t) - \gamma}$ be the estimated slope from the local linear regression, and $\widehat{\mathrm{Var}}(\beta(t) - \gamma)$ be the estimated variance of $\widehat{\beta(t) - \gamma}$. Their close form expressions are provided in Section S.1 in SI.

Let $\Delta_\delta R_t = R_{t+\delta} - R_t$ for $t = 1, \cdots, T - \delta$. From the second score equation (5), we estimate $\gamma(t)$ by the local least square fitting of $\Delta_\delta R_t$ on $I(t)$ without intercept. Let $\hat{\gamma}(t)$ and $\widehat{\mathrm{Var}}(\hat{\gamma})$ be the estimator of γ and its corresponding estimated variance, respectively. Their expressions are provided in SI. Then, $\hat{\beta}(t) = \widehat{\beta(t) - \gamma} + \hat{\gamma}$ is the estimate for the varying coefficient $\beta(t)$ in (1). The standard error of $\hat{\beta}(t)$ can be obtained as $\mathrm{SE}_\beta(t) = \{\widehat{\mathrm{Var}}(\widehat{\beta(t) - \gamma}) + \widehat{\mathrm{Var}}(\hat{\gamma}) + 2Cov(\widehat{\beta(t) - \gamma}, \hat{\gamma}\}^{1/2}$. The 95% confidence interval for $\beta(t)$ can be constructed as

$$(\hat{\beta}(t) - 1.96\mathrm{SE}_\beta(t), \hat{\beta}(t) + 1.96\mathrm{SE}_\beta(t)). \tag{7}$$

In the implementation, we chose $\delta = 1$ and $w = 5$.

To assess the goodness of fitting, Figure S1 in SI shows the observed infected number $I(t)$ versus the fitted values by the proposed varying coefficient SIR model for 30 provinces in China. It demonstrates the proposed method is well suitable for the dynamics of COVID-19 outbreak. Figure S2 in SI plots the estimated the effective reproductive number R_t^{14}, calculated as $R_t^{14} = \hat{\beta}(t) \times 14$, with its 95% confidence interval for 30 provinces in China.

6. Effective Reproduction Number

The effective reproduction number R_t is the most important parameter in determining the state of an epidemic. It measures the average number of infection made by an infectious person during the course of his/her being infectious. If $R_t < 1 (> 1)$ for t larger than a t_0, then the epidemic will die down eventually (explode). There are two widely adopted definitions of $R(t)$. One is based on the average duration of infection of the disease, and the other is via the removal rate $\gamma(t)$.

At a date t, the effective reproduction number based on an average infectious duration D is $R_t^D = \beta(t)D$ where $\beta(t)$ is the daily infection rate at t. We do not adopt the version involving γ, the removal rate, since its estimation is highly volatile at the early stage of an epidemic. A general version of $R(t)$ may be defined as $\int_{t-D_1}^{t+D_2} \beta(u)du$ where positive D_1 and D_2 represent the infectious durations before and after diagnosis, respectively. The R_t^D given above can be viewed as an approximation by the Mean Value Theorem in calculus with $D = D_1 + D_2$.

Research works (Li et al., 2020; Guan et al., 2020; Chen et al., 2020) so far on COVID-19 have informed a range of duration for incubation, from onset of illness to diagnosis and then to hospitalization. The average incubation period from the three studies ranged from 3.0 to 5.2 days; the median duration from onset to diagnosis was 4 days (Guan et al., 2020); and the mean duration from onset to first medical visit and then to hospitalization were 4.6 and 9.1 days (Li et al., 2020), respectively. Based on a data sample of 391 cases from Shenzhen, the average incubation period was 4.46 (0.26) days and the average duration from onset to hospitalization were 3.9 (0.19) days, respectively, where standard error is reported in the parentheses. Another dataset of 100 confirmed cases in Shaoyang (Hunan Province) revealed the average durations from onset to diagnosis and from diagnosis to discharge were 5.67 (0.39) and 10.12 (0.43) days, respectively. There is a recent revelation (Guan et al., 2020) that asymptomatic patients can be infectious, which would certainly prolong the infectious duration.

There are much variation in the medical capability in timely diagnosis and hospitalization (thus quarantine) of the infected across the country. Thus, the infectious duration D would vary among the provinces and cities, and would change with respect to the stage of the epidemic as well.

Given the diverse range of infectious duration across the provinces and cities, in order to standardize and make the effective reproduction number R_t readily comparable, we calculated the R_t^D based on three levels of D: $7, 10.5$ and 14 days, which represent three scenarios of responsiveness in diagnosing, hospitalization and hence quarantine of the infected. Calculation of the R_t at other duration can be made by inflating or deflating a R_t^D proportionally to reflect a local reality.

7. Reproductivity of COVID-19 in China

By calculating the time-varying infection rate function $\beta(t)$, we present in Figures 1 the time series of estimated R_t^D at the three levels of D for the 30+15 provinces/cities from late January to February 17th. Figure 2 displays four cross sectional R_t^{14} and their confidence intervals on January 27th, February 3rd, 10th and 17th, respectively.

Figure 1 reveals a monotone decreasing trend for almost all the provinces and cities with only exceptions for Hubei, Guizhou, Jinlin, Neimenggu and Qinghai. Even for those exceptional provinces, the recent trend is largely declining. The non-monotone pattern for non-Hubei provinces were largely due to relative small number of infected cases and waves of introduced infections. However, the one for Hubei and Wuhan suggests low data quality and in particularly under reporting and reporting delay. The epidemic statistics from Hubei and the city of Wuhan before January 21th were severely incomplete and with irregular patterns, as millions of people fled from Wuhan before the lockdown.This was the reason we start Hubei's analysis from January 21th.

The average R_t^{14} among the 27 provinces (with confirmed cases on and prior to January 23rd) was 6.14 (1.49), and 7.59 (2.38) for 7 of the 15 Hubei cities on January 27. These levels were comparable to the level of R (6.47) given in (Tang et al., 2020).

One week later on February 3rd, R_t^{14} was averaged at 2.18 (0.67) for the 30 provinces and 2.84 (0.59) for the 15 Hubei cities, indicating that cutting off Wuhan and other cities, and the start of wearing face masks and self isolation at home from January 23th had contributed to 64.5% and 62.6% reduction in the R_t^{14}. In the following week starting from February 4th, the average R_t^{14} came down to 0.86 (0.38) for the 30 provinces and 1.23 (0.55) for the 15 Hubei cities on February 10th, representing further 60.5% and 56.7% reductions, respectively, during the second week. This reflects the beneficial effects of the continued large scale self-isolation within the extended spring festival holiday period.

Table 1 provides the reproduction number R_t^D at the two durations on February 10th. It shows that 5 provinces and 5 Hubei cities' R_t^{14} were significantly above 1 (at 5% significance level). There are 14 provinces and 2 Hubei cities' R_t^{14} were significantly below 1, which were 1 and 1 more than those a day earlier on February 9th, and 9 and 2 more than those on February 8th, respectively. If we use the shorter $D = 10.5$, 22 provinces and 11 Hubei cities had been significantly below 1 for 1-7 consecutive days. These indicated that the reproduction number R_t has showed signs of crossing below the critical threshold 1 in increasing number of provinces and cities in Hubei around February 8-10. An updated Table 1 for February 17th are available in Table 2, which showed further improvement since February 10.

On February 17th, R_t^{14} of all provinces and cities under consideration have all been statistically significantly below 1, among which 22 provinces and 8 Hubei cities had been for at least seven consecutive days.

Given the significant decline in the reproduction numbers, it was time to discuss the turning point for COVID-19 for China. If a province or city's R_t^D started to be below 1 significantly (at 5% level), we would say the province or city have showed signs of the turning point. Given the uncertainty with the data records, especially those large variation in daily infected numbers coming out of Wuhan and Hubei, the turning point of the epidemic would be confirmed if R_t^D have been significantly below 1 for D_1 days, where D_1 is the period of infection before diagnosis, assuming all diagnosed cases can be quarantine immediately. Based on the results in (Li et al., 2020; Guan et al., 2020; Chen et al., 2020), $D_1 = 7$ may be considered. Then, based on this criterion, some of the 30+15 provinces/cities had already reached the turning point on February 17, and more would follow in the coming days according to latest Table 2

8. Prediction for Infection Rate and State Variables

As $R_t^D = \beta(t)D$, predicting $\beta(t)$ is equivalent to predicting R_t^D. From Figure 1 and Figure S2 in SI, we see that the overall trends of $\beta(t)$ is decreasing. But the rate of deceasing gets smaller as time travels. To model such trend, we consider the reciprocal regression

$$\beta(t) = \frac{b}{t^\eta - a} + e_t \tag{8}$$

with error e_t and unknown parameters a, b and η. The parameters a and b are estimated by minimizing the sum-of-squares distance between the estimates $\hat{\beta}(t)$ and their fitted values for a given η, and then the optimal η is chosen to be the one that gives the minimum mean square error over a set of candidate values from 0.5 to 5 with 0.1 increment. Let $\tilde{a}$, $\tilde{b}$ and $\tilde{\eta}$ be the estimated parameters, and $\tilde{\beta}(t) = \tilde{b}/(t^{\tilde{\eta}} - \tilde{a})$ be the fitted function. Figure S3 in SI shows the reciprocal model fits $\hat{\beta}(t)$ quite well for most of the provinces, especially those with large number of infected cases.

With the fitted $\tilde{\beta}(t)$, we project $\{S(t), I(t), R(t)\}$ via the ODEs

$$\begin{aligned}
\frac{d\hat{S}(t)}{dt} &= -\tilde{\beta}(t)\hat{I}(t)\frac{\hat{S}(t)}{N}, \\
\frac{d\hat{I}(t)}{dt} &= \tilde{\beta}(t)\hat{I}(t)\frac{\hat{S}(t)}{N} - \hat{\gamma}_T\hat{I}(t), \\
\frac{d\hat{R}(t)}{dt} &= \hat{\gamma}_T\hat{I}(t).
\end{aligned} \tag{9}$$

where $\hat{\gamma}_{\mathrm{T}}$ is the estimated recovery rate at time T using the last five days' data. With the observed $\{S(T), I(T), R(T)\}$ at the current time T as the initial values, numerical solutions $\{(\hat{S}(t), \hat{I}(t), \hat{R}(t)) : T \le t < \infty\}$ for the system (9) could be obtained using the Euler method. Then, the end time of the epidemic can be predicted as $t_{\text{end}} = \min\{t : \hat{I}(t) < 1\}$, and the estimated final infected number is $\hat{N}_{\text{final}} = \hat{R}(t_{\text{end}}) + \hat{I}(t_{\text{end}})$.

To conduct statistical inference for the epidemic predictions, we use the bootstrap method. In particular, we generate parametric bootstrap resampled processes based on the vSIR model which facilitate the construction of prediction intervals. We regard that the increments of $S(t)$ and $R(t)$ follow the Poisson processes (Bretó et al., 2009) over time as

$$-\Delta S(t) \sim \text{Poisson}\{\beta(t)S(t)I(t)/N\} \text{ and } \Delta R(t) \sim \text{Poisson}\{\gamma I(t)\}$$

where $\Delta S(t) = S(t+1) - S(t)$ and $\Delta R(t) = R(t+1) - R(t)$. With the estimated $\hat{\gamma}$ and $\hat{\beta}(t)$, we generate bootstrap samples $\{(S^{(b)}(t), I^{(b)}(t), R^{(b)}(t))\}_{t=1}^{T}$ of the original process for $b = 1, 2, \ldots, B$.

For each bootstrap resampled $\left\{(S^{(b)}(t), I^{(b)}(t), R^{(b)}(t))\right\}_{t=1}^{T}$, we obtain the estimates $\beta_\star^b(t)$ and $\gamma_\star^b$ for $\beta(t)$ and γ in the same way as for the original sample. Let $\bar{\beta}_\star(t) = \sum_{b=1}^{B} \beta_\star^b(t)/B$ and $\bar{\gamma}_\star = \sum_{b=1}^{B} \gamma_\star^b(t)/B$ be the average of the bootstrap estimates. We employ the bias corrected bootstrap estimates for $\beta(t)$ and γ as

$$\hat{\beta}^b(t) = \hat{\beta}(t) + (\beta_\star^b(t) - \bar{\beta}_\star(t)) \quad \text{and} \quad \hat{\gamma}^b = \hat{\gamma} + (\gamma_\star^b - \bar{\gamma}_\star)$$

for $b = 1, 2, \ldots, B$. We then use the reciprocal model (8) to project the future path of $\hat{\beta}^b(t)$, and use the numerical solution of the vSIR ODEs to predict the end time and the accumulative number of final infected cases as we described in section 3.4. Let the bootstrap estimates for the peak time be $\{t_{\text{end}}^b\}_{b=1}^{B}$. The 95% prediction interval for the peak time is constructed as the 2.5% and 97.5% quantiles of $\{t_{\text{end}}^b\}_{b=1}^{B}$. Similar bootstrap prediction intervals can be constructed for the final accumulative infection number N_{final} of the epidemic.

9. Prediction Results

Based on the estimated $\beta(t)$ over time, we predict COVID-19's future trajectories as solutions to the vSIR model. We used data up to February 19 2020 for the prediction under three scenarios for the recovery rate γ. One uses the empirical estimate based on data to February 19th. As an effective cure for the virus has not been found, the estimated recovery rates are quite low. Among the provinces with more than 100 infections on February 19, Jiangsu had the highest recovery rate 0.08, followed by Jiangxi, Hebei, Shanghai, Shanxi, Chongqing, Henan (0.07). Hubei, the province at the center of the epidemic, is 0.025. The other scenarios was to choose $\gamma = 1/14$ and $\gamma = 0.1$, which mean the average removal time from diagnosis was 14 and 10 days, respectively, representing improvement in the treatment for COVID-19 patients as time progressed.

Tables 3 presents the 95% prediction intervals for the end times of the epidemic and the cumulative number of infected at the ending. The trajectories of $I(t)$ of the proposed vSIR model are presented in Figure 5 under the three scenarios of the recovery rate. The predicted infection number $\hat{I}(t)$ is within

10% deviation from its observed value based on data up to Feb 19th, see Table A2 in SI for the detailed prediction error. From the trajectory of the vSIR model in Figure 5, for the non-Hubei provinces, the number of infected would quickly decease in late February and March with very few cases left in April under all the three scenarios. Some provinces with few number of total infected cases may end as early as March (Qinghai, Jilin, Gansu, Ningxia). For Hubei, with a higher recovery rate of 0.1, the duration of the epidemic would be shorten substantially. The ending time for Hubei is around June 20 2020 with total number of infection in the range 73,857–74,596. This shows that improving the recovery rate is an efficient way to end the COVID-19 infection early given the current decreasing trend of $\beta(t)$, as it leads to the reduction of the infectious duration.

10. Discussion

The implications of China's experience in combating COVID-19 to other countries facing the epidemic are two folds. One is to reduce the person-to-person contact rate by self isolation and curtailing population movement; another is to reduce the transmission probability by wearing protective wears when a contact has to be made.

The eventual control of COVID-19 is rested on if the existing control measures can be continued further for a period of time. The biggest challenges that can jeopardize the great effort from late January are from the impatient populations eager to get out of the self-isolation driven by either economic needs (migrant workers eager to coming back to cities for income) or people trying to escape from the boredness of self isolation while encouraged by the declining infections in the last two weeks. In any case, the vSIR model and its statistical estimation and inference can be used to model and the assess the COVID-19 epidemics in other countries.

References

World Health Organization (WHO), Situation report - 26, Available at `https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200215-sitrep-26-covid-19.pdf`, 2020. [Accessed February 16, 2020].

W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character 115 (1927) 700–721.

H. W. Hethcote, The mathematics of infectious diseases, SIAM review 42 (2000) 599–653.

N. G. Becker, On a general stochastic epidemic model, Theoretical Population Biology 11 (1977) 23–36.

N. G. Becker, T. Britton, Statistical studies of infectious disease incidence, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61 (1999) 287–307.

P. S. Yip, Q. Chen, Statistical inference for a multitype epidemic model, Journal of statistical planning and inference 71 (1998) 229–244.

F. Ball, D. Clancy, The final size and severity of a generalised stochastic multitype epidemic model, Advances in applied probability 25 (1993) 721–736.

J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study, The Lancet (2020).

J. M. Read, J. R. Bridgen, D. A. Cummings, A. Ho, C. P. Jewell, Novel coronavirus 2019-ncov: early estimation of epidemiological parameters and epidemic predictions, medRxiv (2020).

B. Tang, X. Wang, Q. Li, N. L. Bragazzi, S. Tang, Y. Xiao, J. Wu, Estimation of the transmission risk of 2019-ncov and its implication for public health interventions, Available at SSRN 3525558 (2020).

W. S. Cleveland, S. J. Devlin, Locally weighted regression: an approach to regression analysis by local fitting, Journal of the American statistical association 83 (1988) 596–610.

Shenzhen Municipal Affairs Service Data Administration, Latest data, Available at `https://opendata.sz.gov.cn/data/dataSet/toDataSet`, 2020. [Accessed February 14, 2020].

Shaoyang Municipal Health Commission, Dynamic information on prevention and control of new coronavirus-infected pneumonia in shaoyang, Available at https://wjw.shaoyang.gov.cn/wjw/zyxw/202002/c49df53092784c85aaac769149f30265.shtml, 2020. [Accessed February 14, 2020].

Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, et al., Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia, New England Journal of Medicine (2020).

W. Guan, Z. Ni, Y. Hu, W. Liang, C. Ou, J. He, et al., N. Zhong, Clinical characteristics of 2019 novel coronavirus infection in china, medRxiv (2020).

N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study, The Lancet (2020).

C. Bretó, D. He, E. L. Ionides, A. A. King, et al., Time series analysis via mechanistic models, The Annals of Applied Statistics 3 (2009) 319–348.

Table 1: The reproduction number R_t^D at two infectious duration: 10.5 and 14 days, for the 30 mainland provinces and 15 cities in Hubei province on February 10th with extended results on February 17th. The symbols + (−) indicate that the R_t^{14} was significantly above (below) 1 at 5% level of statistical significance, and the numbers inside the square brackets were the consecutive days the R_t^{14} were significantly below 1. The column ΔR gives the percentages of decline in the R_t^{14} from the beginning of analysis to February 10th (the first two weeks of the analysis). The columns $\Delta R(1^{\text{st}})$, $\Delta R(2^{\text{nd}})$ and $\Delta R(3^{\text{rd}})$ are the percentages of decline in the first week (January 27 to February 3rd), the second week (February 3-10), and the third week (February 10-17), respectively.

Province/City	$R_t^{10.5}$	R_t^{14}	ΔR	$\Delta R(1^{\text{st}})$	$\Delta R(2^{\text{nd}})$	$\Delta R(3^{\text{rd}})$
Wuhan	1.99+	2.66+	58.7%	45.9%	23.7%	72.5%
Ezhou	1.64+	2.18+	80%	79.3%	3.6%	67.7%
Hubei	1.48+	1.98+	74.2%	58.2%	38.3%	69.7%
Tianmen	1.33+	1.78+	75%	67.4%	23.4%	52.5%
Guizhou	1.25+	1.67+	62.4%	9.3%	58.5%	91.5%
Xiantao	0.99	1.32	76.9%	46.4%	57%	71.9%
Heilongjiang	0.95	1.27+	81.8%	54.3%	60.3%	62.6%
Hebei	0.94	1.25+	85.4%	82.4%	16.7%	70.5%
Xinjiang	0.9	1.2+	75.6%	60.7%	37.9%	53.1%
Enshizhou	0.86−[5]	1.14+	74.1%	60.3%	34.6%	76%
Jingzhou	0.85−[1]	1.14+	84.8%	50.5%	69.3%	76.9%
Gansu	0.8	1.07	75.1%	47.8%	52.3%	100%
Jingmen	0.79−[1]	1.05	88.3%	77.1%	49.2%	84.1%
Huangshi	0.79−[1]	1.05	78.3%	31.2%	68.4%	83.6%
Anhui	0.74−[1]	0.99	88.3%	71.7%	58.7%	77.6%
Shanxi	0.74−[2]	0.98	86.7%	69.6%	56.1%	87.1%
Ningxia	0.73	0.97	84.9%	75.8%	37.6%	75.7%
Shandong	0.73−[2]	0.97	90.3%	84.5%	37.1%	80.3%
Jiangsu	0.72−[3]	0.96	87.1%	70.5%	56.1%	72.1%
Xianning	0.71−[4]	0.95	70.7%	21.6%	62.6%	33.6%
Shiyan	0.71−[1]	0.94	89.1%	72.4%	60.3%	67.5%
Jilin	0.7	0.93	80.4%	17.6%	76.1%	82.5%
Yichang	0.69−[3]	0.92	87%	56.6%	70%	75.7%
Huanggang	0.69−[3]	0.92	88.9%	60.7%	71.8%	89%
Tianjin	0.69−[4]	0.91	82.9%	51.8%	64.6%	64%
Hainan	0.68−[1]	0.91	80.9%	65.2%	45.2%	96.1%
Guangxi	0.66−[5]	0.88	81.6%	64.9%	47.8%	72.1%
Xiangyang	0.63−[3]	0.84−[1]	88.4%	57.9%	72.4%	80.8%
Sichuan	0.62−[5]	0.83−[2]	89.4%	78.5%	50.7%	47.6%

Continued on next page

Table 1 – continued from previous page

Province/City	$R_t^{10.5}$	R_t^{14}	ΔR	$\Delta R(1^{\text{st}})$	$\Delta R(2^{\text{nd}})$	$\Delta R(3^{\text{rd}})$
Jiangxi	0.61−[2]	0.82	90.8%	70.9%	68.5%	79.6%
Xiaogan	0.6−[1]	0.81	88.9%	61.5%	71.3%	50.1%
Hunan	0.57−[3]	0.76−[2]	91.5%	77%	63.2%	90.4%
Henan	0.56−[2]	0.75−[1]	93.2%	78.8%	67.8%	64.2%
Suizhou	0.52−[2]	0.69−[2]	88.2%	40.9%	80%	65.5%
Chongqing	0.51−[4]	0.68−[3]	90.4%	75%	61.6%	73.9%
Shaanxi	0.51−[3]	0.68−[2]	86.5%	63%	63.6%	72.9%
Neimenggu	0.49−[3]	0.66−[2]	82.4%	43%	69.1%	27.9%
Fujian	0.49−[6]	0.66−[4]	90.5%	76.2%	60%	76.9%
Guangdong	0.45−[3]	0.61−[2]	88.2%	54.4%	74.2%	62%
Liaoning	0.45−[6]	0.6−[2]	89.3%	72.7%	61%	81.7%
Beijing	0.45−[4]	0.6−[2]	90.3%	59.2%	76.2%	65.1%
Shanghai	0.34−[4]	0.46−[2]	92.1%	68%	75.2%	53.4%
Zhejiang	0.31−[4]	0.42−[3]	94.3%	77.9%	74.2%	86.3%
Yunnan	0.28−[7]	0.38−[5]	96.2%	86.8%	71.5%	30.8%
Qinghai	0.02−[4]	0.03–[3]	98.9%	-1.6%	98.9%	100%
Ave(sd)	0.74(0.35)	0.98(0.47)	85.3%	64.2%	59%	71.6%

Table 2: The reproduction number R_t^D at two infectious durations: 10.5 and 14, for the 30 mainland provinces and 15 cities in Hubei province on February 17th. The symbols + (−) indicate that the $R_t^{10.5}(R_t^{14})$ was significantly above (below) 1 at 5% level of statistical significance, and the numbers inside the square brackets were the consecutive days the $R_t^{10.5}(R_t^{14})$ were significantly below 1.

Province/City	$R_t^{10.5}$	R_t^{14}
Wuhan	0.55−[9]	0.73−[8]
Ezhou	0.52−[3]	0.69−[2]
Tianmen	0.63−[9]	0.85
Hubei	0.43−[3]	0.56−[2]
Sichuan	0.33−[19]	0.44−[16]
Xiantao	0.28−[14]	0.37−[13]
Tianjin	0.25−[11]	0.33−[10]
Heilongjiang	0.34−[7]	0.46−[5]
Shiyan	0.23−[15]	0.31−[14]
Neimenggu	0.36−[17]	0.47−[16]

Continued on next page

Table 2 – continued from previous page

Province/City	$R_t^{10.5}$	R_t^{14}
Xiaogan	0.28−[14]	0.37−[13]
Xinjiang	0.42−[12]	0.56−[10]
Beijing	0.13−[11]	0.18−[9]
Jingzhou	0.19−[8]	0.25−[3]
Shaanxi	0.14−[17]	0.18−[16]
Chongqing	0.13−[11]	0.17−[10]
Henan	0.2−[9]	0.26−[8]
Hebei	0.26−[5]	0.35−[3]
Guangxi	0.17−[12]	0.23−[7]
Shanghai	0.16−[18]	0.21−[16]
Jiangsu	0.2−[10]	0.26−[6]
Jilin	0.11−[7]	0.15−[7]
Fujian	0.11−[13]	0.15−[11]
Shanxi	0.09−[16]	0.13−[14]
Yichang	0.17−[17]	0.22−[14]
Yunnan	0.2−[21]	0.26−[19]
Anhui	0.16−[8]	0.21−[7]
Xianning	0.47−[8]	0.63−[8]
Suizhou	0.18−[16]	0.25−[16]
Shandong	0.14−[12]	0.19−[9]
Guangdong	0.17−[10]	0.23−[9]
Xiangyang	0.12−[17]	0.16−[15]
Zhejiang	0.04−[18]	0.06−[17]
Enshizhou	0.2−[12]	0.27−[4]
Huanggang	0.07−[10]	0.09−[7]
Guizhou	0.11−[4]	0.15−[3]
Jingmen	0.12−[8]	0.16−[4]
Gansu	0−[7]	0−[6]
Hainan	0.03−[8]	0.04−[7]
Hunan	0.02−[10]	0.07−[9]
Ningxia	0.18−[9]	0.24−[9]
Huangshi	0.13−[8]	0.17−[7]
Jiangxi	0.12−[9]	0.16−[7]
Liaoning	0.08−[20]	0.11−[16]

Continued on next page

Table 2 – continued from previous page

Province/City	$R_t^{10.5}$	R_t^{14}
Qinghai	0−[4]	0−[4]

Table 3: The 95% prediction intervals for the ending times, and the final accumulative number of infected cases of COVID-19 epidemic in the 30 provinces based on data to Feb 19 2020 with $\gamma = 0.1$. The last column lists the total infected cases ($I(t) + R(t)$) as Feb 19, 2020.

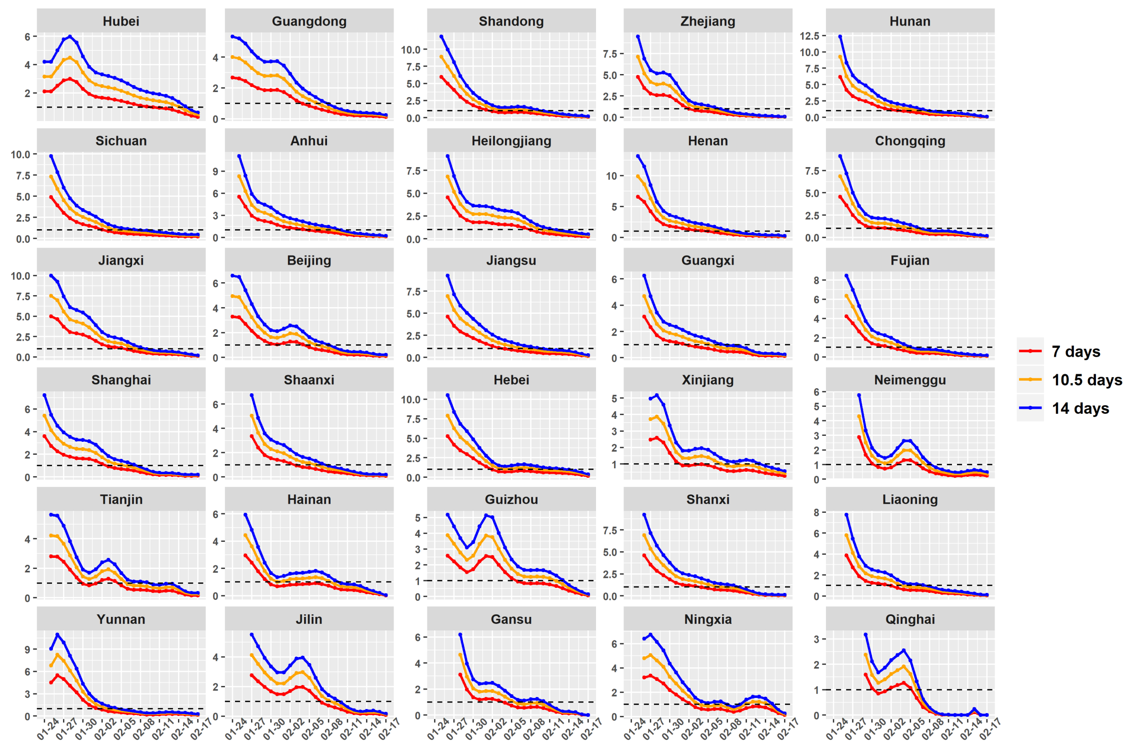
Province	Ending time	$\hat{N}_{\text{final}}$	Current
Hubei	6/20 - 6/21	73857 - 74596	62322
Guangdong	4/27 - 4/29	1368 - 1412	1347
Zhejiang	4/26 - 4/27	1225 - 1245	1195
Beijing	4/17 - 4/20	416 - 436	397
Chongqing	4/18 - 4/21	581 - 600	565
Hunan	4/21 - 4/23	1028 - 1046	1021
Guangxi	4/11 - 4/15	254 - 271	248
Shanghai	4/12 - 4/16	345 - 365	336
Jiangxi	4/23 - 4/25	969 - 994	955
Sichuan	4/25 - 4/28	589 - 619	525
Shandong	4/19 - 4/21	567 - 584	553
Anhui	4/26 - 4/28	1044 - 1068	1006
Fujian	4/13 - 4/16	306 - 320	299
Henan	4/29 - 5/1	1358 - 1387	1283
Jiangsu	4/20 - 4/23	662 - 687	640
Hainan	4/6 - 4/9	174 - 183	168
Tianjin	4/7 - 4/14	141 - 159	132
Yunnan	4/7 - 4/11	174 - 187	174
Shaanxi	4/12 - 4/16	262 - 276	250
Heilongjiang	4/23 - 4/26	519 - 554	479
Liaoning	3/31 - 4/3	120 - 126	122
Guizhou	4/4 - 4/10	150 - 165	147
Jilin	3/30 - 4/4	93 - 101	92
Ningxia	3/18 - 3/26	65 - 73	71
Hebei	4/12 - 4/16	319 - 336	312
Gansu	3/20 - 3/26	90 - 96	92
Xinjiang	3/31 - 4/9	78 - 96	78

Continued on next page

Table 3 – continued from previous page

Province	Ending time	$\hat{N}_{\text{final}}$	Current
Shanxi	4/1 - 4/5	134 - 142	134
Neimenggu	4/2 - 4/11	78 - 98	76
Qinghai	2/23 - 3/6	17 - 20	19

(a) R_t^D for 30 provinces



(b) R_t^D for 15 cities in Hubei

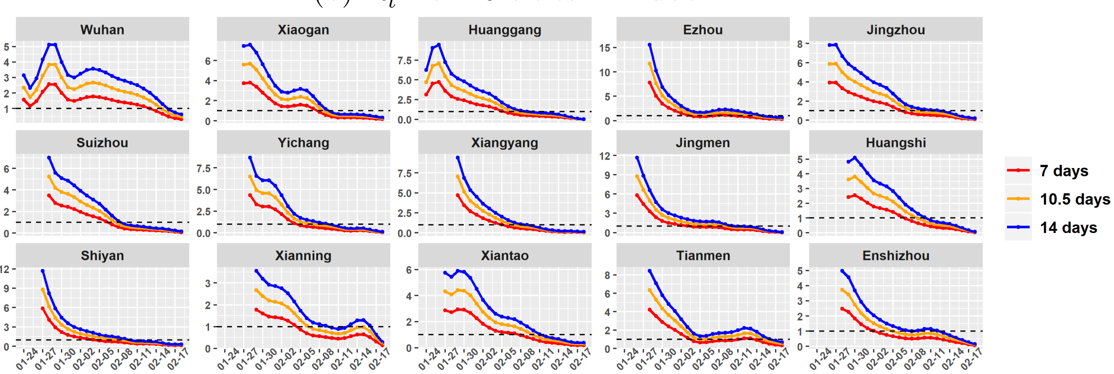
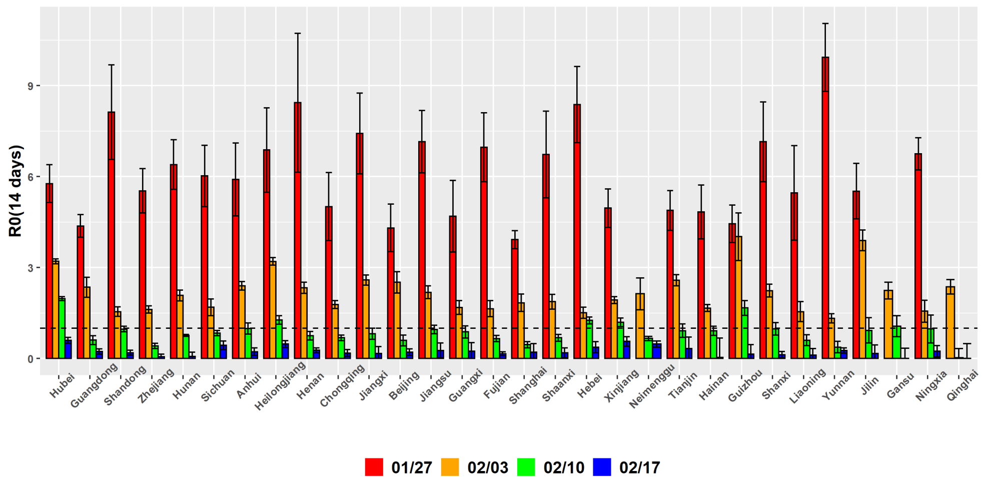


Figure 1: Time series of the reproduction number R_t^D at three infectious durations: $D = 7$ (red), 10.5 (orange), 14 (blue), for the 30 mainland provinces (a) and the 15 cities in Hubei province (b) from Jan 21 to Feb 17 2020. The black horizontal line is the critical threshold level 1.

(a) R_t^{14} for 30 provinces



(b) R_t^{14} for 15 cities in Hubei

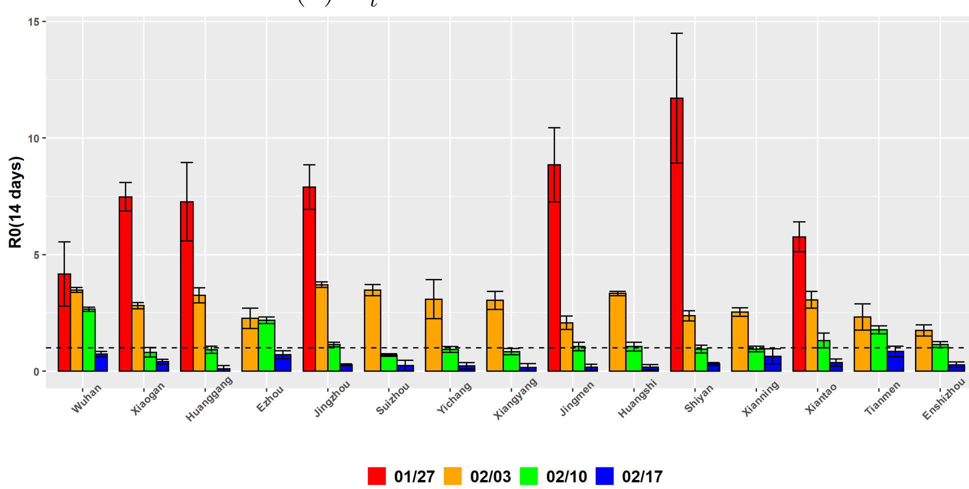


Figure 2: Elevated 95% confidence intervals (black) of the 14-day R_t for the 30 mainland provinces (a) and the 15 Hubei cities (b) on Jan 27 (red), Feb 3 (orange), Feb 10 2020 (green) and Feb 17 (blue). The black horizontal lines mark the critical threshold 1.

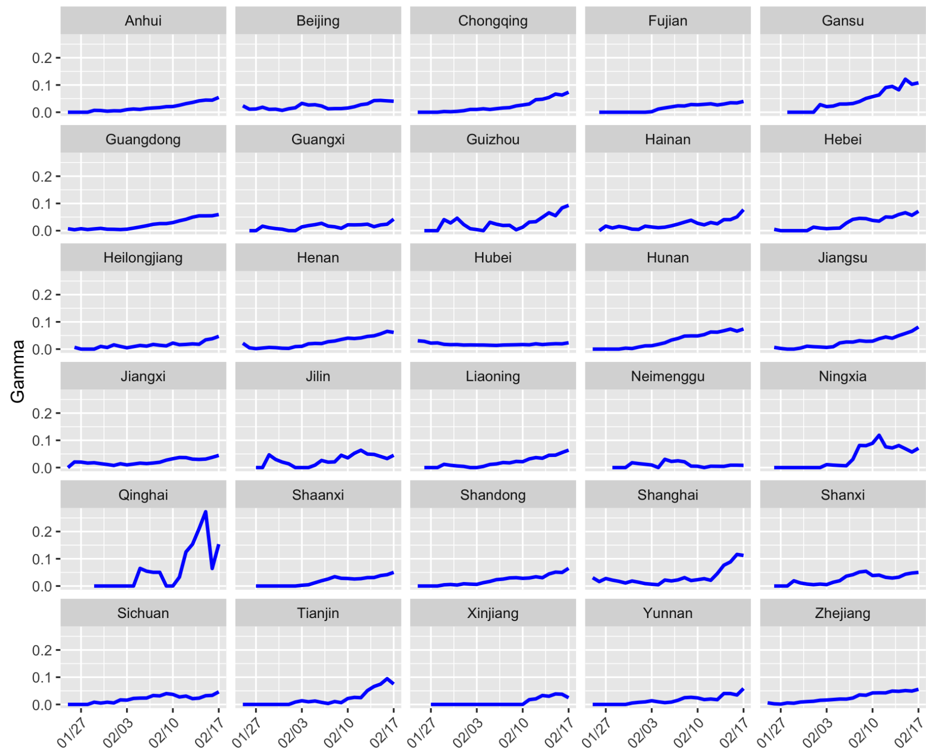


Figure 3: The estimated $\hat{\gamma}(t)$ from the varying coefficient SIR model (1) for the data to Feb 17th 2020 for 30 provinces.

(a) vSIR model

(b) vSEIR model

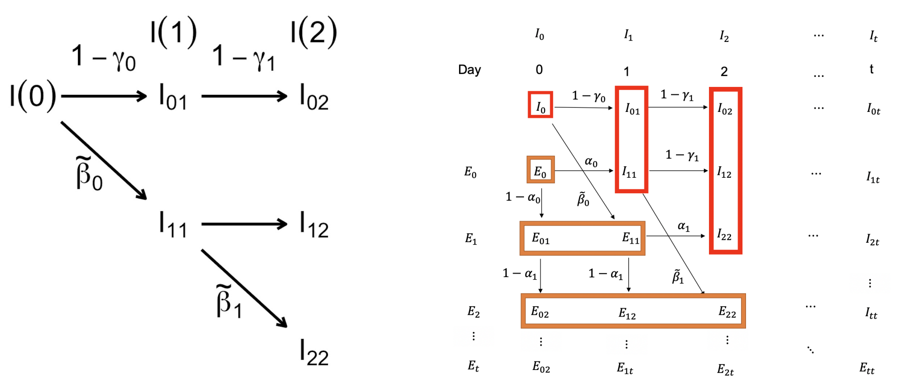


Figure 4: Epidemic progression networks under vSIR and vSEIR models

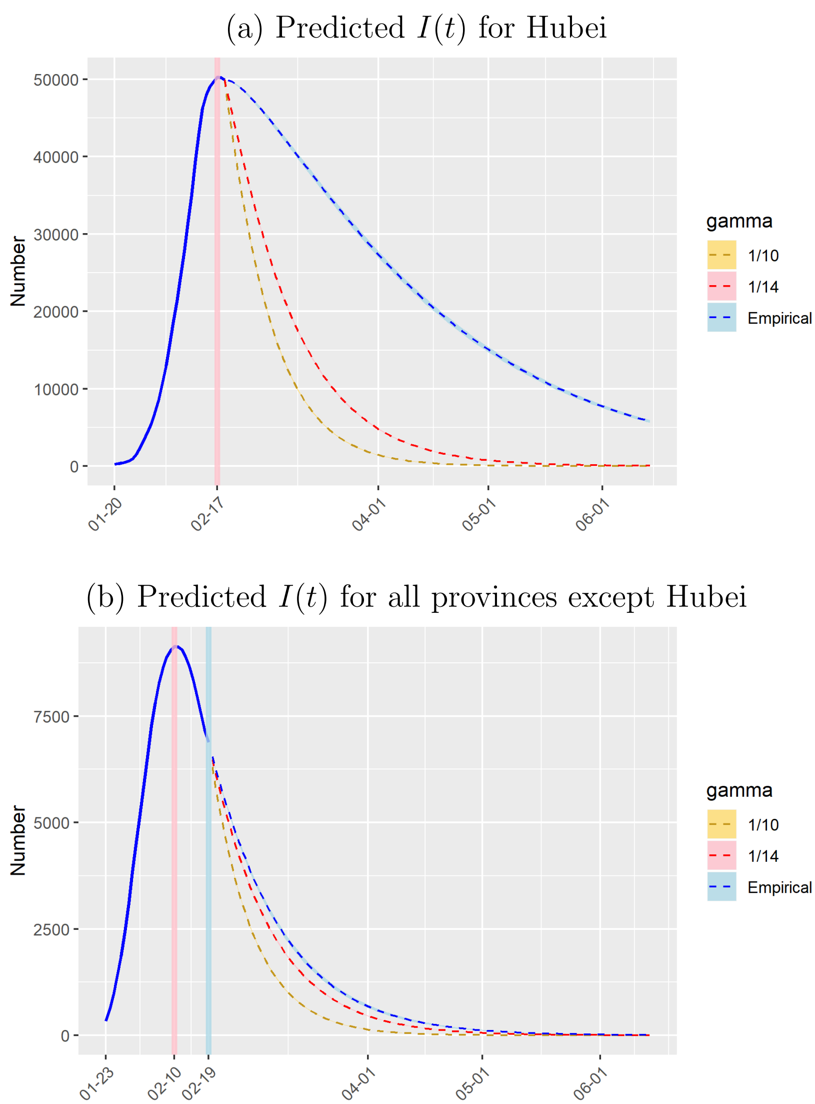


Figure 5: Predicted number of infected cases $I(t)$ with 95% prediction interval for Hubei Province in panel (a) and all other provinces combined except Hubei in panel (b). The grey vertical line indicates the current date of observation; the blue solid line plots the observed $I(t)$ before Feb 19th; the blue dashed line gives the predicted $I(t)$ with 95% prediction interval (blue shaded area) with the estimated $\hat{\gamma}_{\mathrm{T}}$; the pink vertical line indicates the peak date of $I(t)$; the orange and red dashed line gives the predicted $I(t)$ with 95% prediction interval (shaded area) with fixed recovery rate $\gamma = 0.1$ and $\gamma = 1/14$ respectively.