Journal of Data Science 18 (3), 477–479
July 2020
DOI: 10.6339/JDS.202007_18(3).0013

Discussion of "Tracking reproductivity of COVID-19 epidemic in China with varying coefficient SIR model"

Lili Wang[1], Fei Wang[2], Yiwang Zhou[1], and Peter X.K. Song*[1]

[1]*Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA*
[2]*Data Science Team, CarGurus, Cambridge, Massachusetts, USA*

We congratulate the authors on an interesting contribution to the Susceptible-Infection-Removed (SIR) modeling. In this paper the authors have focused on the development of a fast and flexible framework to incorporate a time-varying effective reproduction number ($\mathcal{R}_t$) into the SIR model. Through estimation and extrapolation of such a time-varying effective reproduction number ($\mathcal{R}_t$), this new model can predict time-varying reproduction profiles critical to the understanding of covid-19 disease evolution in China. The use of a reciprocal regression technique enables forecast the number of future infectious cases with a certain confidence level.

Let the numbers of individuals in the compartments of susceptible, infected and removed be, respectively, S_t, I_t and R_t, which are changing over time t. Let $N_t = I_t + R_t$ be the cumulative number of infected cases and $N = S_t + I_t + R_t$ be the target population under investigation, which is fixed constant over time. To reduce noise, the authors applied a local smoothing by the moving average of three data points on the three compartment time series. The authors proposed a local linear fitting approach to estimating the nonlinear transmission rate β_t and removal rate γ_t at each target time t conditional on I_t via the following two local linear models at a target time t_1: $\log(I_t) \sim \log(I_{t_1}) + (\beta_t - \gamma_t)(t - t_1)$, and $R_{t_1+\delta} - R_{t_1} \sim \gamma_t I_{t_1}$, where $\delta > 0$ is set to be 1. Because the number of removed cases, i.e. a total number of deaths and recovered cases, are often collected and reported mostly by hospitals several weeks after virus contract and infection, the observed series R_t is typically delayed and incomplete, and thus there is a concern with the reliability of such data in a small local time window used in the estimation. The data quality issue may result in unstable local estimation of the transmission rate γ_t at a time point. The authors suggested using $\mathcal{R}_t^D = \beta_t D$ to estimate the time-varying effective reproduction number, where D is a prefixed infectious duration varying from 7 to 14 days.

For the prediction purpose, the authors considered a parsimonious nonlinear model that enables extrapolation of the estimated β_t into the future time. The transmission rate is assumed to satisfy the reciprocal regression model

$$\beta_t = \frac{b}{t^\eta - a} + e_t,$$

where a, b, and η are the parameters to be estimated and e_t is the error. In addition, both projected N_t and R_t are obtained following the SIR ordinary differential equations using Euler method. The bias correction and estimating inference were achieved via a bootstrap method. In particular, we notice that the confidence intervals provided in their Figure 5 were very narrow, which does not seem to capture much certainty from the infection dynamic system well. This underestimation of uncertainty might result from maximizing a conditional likelihood (given I_t) in their equation (3) and data pre-processing via an average filtering procedure.

To gain some insights of the above reciprocal regression model for the transmission rate, we implemented this time-varying β_t in an analysis of the COVID19 data of Hubei, China using

*Corresponding author. Email: pxsong@umich.edu.

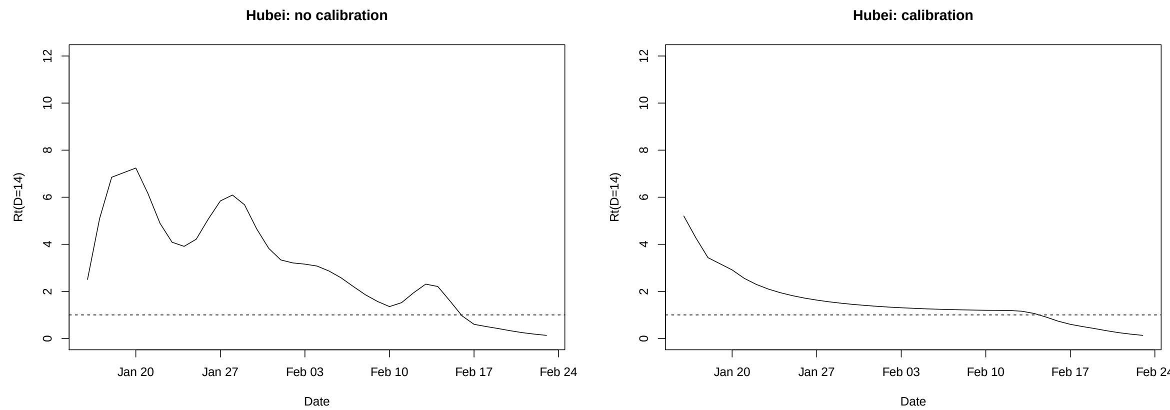

Figure 1: The time-varying effective reproduction number $\mathcal{R}_t^{14}$ with or without data calibration, where the dashed line denotes the event of ending infection with $\mathcal{R}_t^{14} = 1$.

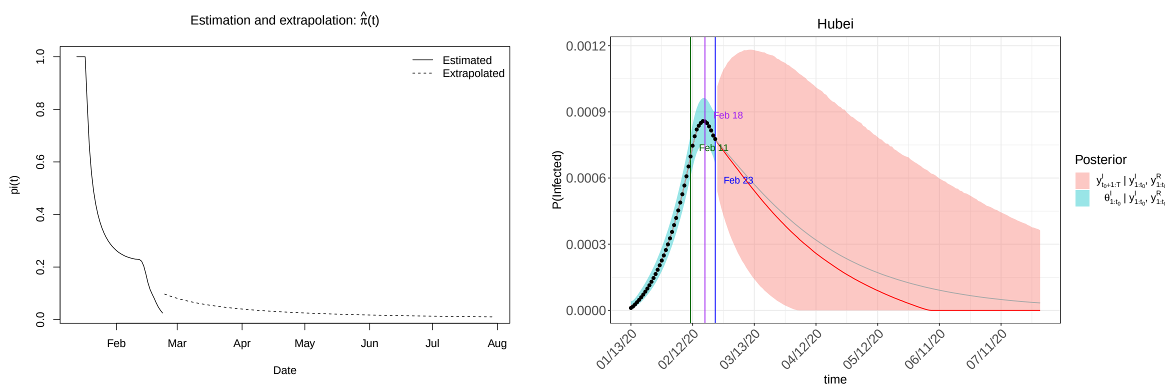

Figure 2: The estimated and predicted mean (grey) and median (red) infection prevalence (right) with transmission rate modifiers $\widehat{\pi}(t)$ (left) fitted by `eSIR` model using data up to February 18, 2020.

our epidemiological forecast SIR model (Wang et al., 2020). We set $D = 14$ (or two weeks or the infectious duration that sums the average incubation period and hospitalization time) and computed $\mathcal{R}_t^{14}$ with or without data calibration used to deal with an abrupt jump of infected cases on February 12, which is mostly due to the issue of under-reporting before February 12. Figure 1 indicates that the estimated $\mathcal{R}_t^{14}$ appears much smoother (the right panel) with no artificial bumps for the one obtained with the correction for under-reporting than the other obtained with no correction for the under-reporting (the left panel). Note that the reported turning date in Hubei is February 18, which is roughly coincides with the estimated $\mathcal{R}_t^{14} = 1$; see the crossing point of the dashed line and the estimated $\mathcal{R}_t^{14}$ in Figure 1.

Given that the proportion of cumulative infected cases is very small in comparison to the size of population, one may assume that $s(t) = S_t/N \approx 1$. In this case, we may rewrite the transmission rate β_t as a multiplicative form $\beta_t = \beta_0\pi(t)$, in which β_0 is the transmission rate and $\pi(t)$ is termed as a transmission rate modifier (Wang et al., 2020). This modifier may be thought of as a consequence of social distancing, self-quarantine, and other preventive interventions issued during the period of the epidemic in Hubei.

It follows that we obtain an estimate of the transmission rate modifier $\pi(t)$ as follows: $\widehat{\pi}(t) = \widehat{\beta}_t/\beta_0$. The estimated $\widehat{\pi}(t)$ and its extrapolation are shown in the left panel of Figure 2. we ran the `eSIR` model (Wang et al., 2020) and obtained the projected infection dynamics in the right panel of Figure 2. Based on results from the `eSIR` model, we obtained the estimated parameters and their 95% confidence intervals, respectively: the transmission rate $\widehat{\beta}_0 = 0.123$ (CI: [0.0422, 0.256]), the removal rate $\widehat{\gamma} = 0.0257$ (CI: [0.0144, 0.0389]), and the basic reproduction number $\mathcal{R}_0 = 4.71$ (CI: [2.20, 8.60]) adjusting the time-varying transmission modifier due to various preventive measures implemented in Hubei Province.

In summary, the authors developed a new framework from which one can estimate a time-varying transmission rate and consequently a time-varying effective reproduction number. Their method can provide a way to derive a transmission rate modifier useful in our `eSIR` model (Wang et al., 2020).

References

Wang L, Zhou Y, He J, Zhu B, Wang F, Tang L, et al. (2020). An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China (with discussion). *Journal of Data Science*, 18(3): 409–454.