Journal of Data Science 18 (3), 480–482
July 2020

DOI: 10.6339/JDS.202007_18(3).0014

Rejoinder: "Tracking Reproductivity of COVID-19 Epidemic 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[*5,6]

[1]*Center for Data Science, Peking University, Beijing, China*
[2]*Department of Statistics, Iowa State University, Ames, Iowa, USA*
[3]*School of Mathematical Sciences, Sichuan University, Chengdu, Sichuan, China*
[4]*Yuanpei College, Peking University, Beijing, China*
[5]*Center for Statistical Science, Peking University, Beijing, China*
[6]*Guanghua School of Management, Peking University, Beijing, China*

We thanks the valuable comments from the research team (Wang et al.) from University of Michigan led by Professor Peter Song, the team (Jiang et al.) from Sun Yat-Sen University and University of Science and Technology of China led by Professor Xueqin Wang, and Professor Lu Tang from University of Pittsburgh. These comments broaden the scope of our work.

Contact Rate Estimation We are very pleased to see the extension by Wang et al. for estimating the contact rate function $\pi(t)$ from the estimated time-varying infection rate function $\hat{\beta}(t)$. Indeed, the contact rate function is directly reflective to the effects of the early COVID-19 control measures in China which were largely designed to reduce human contacts.

Euler and Higher Order Approximations to ODEs Professor Tang's comments on more accurate discretization of the ODEs are valuable. While the estimator for $\beta(t)$ is directly motivated by the ODEs of the vSIR model, it can be also formulated under the discrete Poisson-vSIR model under a fixed Δt (which is daily in our study) as indicated before (1) of the paper. Hence, from the prospect of the statistical Poisson-vSIR model, there is no need to conduct the high order correction to the ODEs, although for practical performance such implementation can reduce the bias of the estimation.

Parametric Form of Infection Rate The comment by Professor Tang regarding the estimation of the infection rate $\beta(t)$ as parametric was perhaps based on the reciprocal model (5) proposed for forecasting purposes (to project the future course of the infection rates). Our in-sample estimator of $\beta(t)$ is the nonparametric kernel estimator.

Gradual Effects of Abrupt Lockdowns Regarding Professor Tang's comments on the smoothed $\beta(t)$ estimates despite many provinces took sudden preventive actions, the epidemic system from contacts to infections then to diagnosis is a convoluted delayed process. The sudden application of the lock-downs is smoothed by the delayed effects of infections, and the time gap from the onset of the disease to diagnosis, which would make the observed $\beta(t)$ smooth without sudden drops.

Moving Average Filters and the Big Revision of Statistics The moving average filter is applied to remove measurement errors. However, the moving average filter is incapable to deal

*The corresponding authors are Song Xi Chen (csx@gsm.pku.edu.cn) and Yumou Qiu (yumouqiu@iastate.edu).

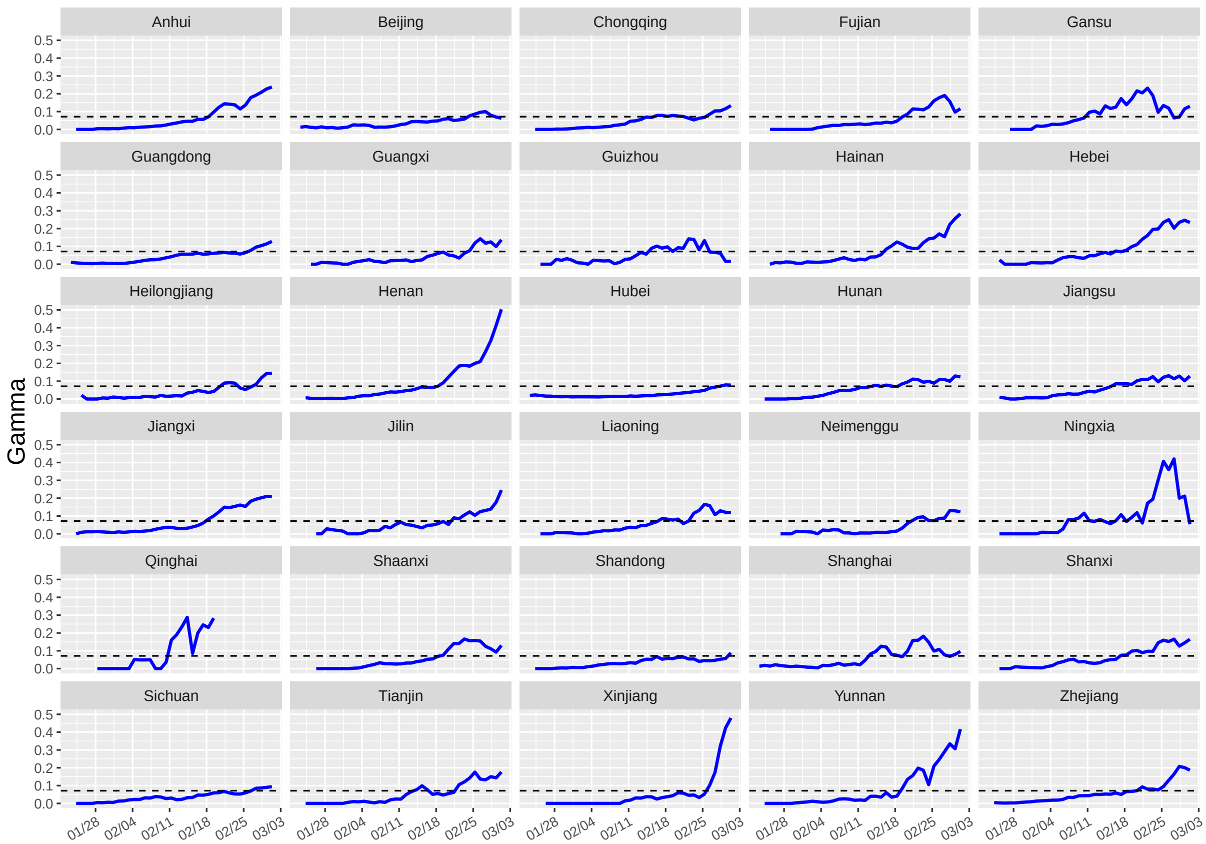

Figure 1: The estimated $\hat{\gamma}(t)$ from the varying coefficient SIR model (solid) and 1/14 (dashed) for the data to Mar 2nd 2020 for 30 provinces.

with the big revision of statistics on February 12th, 2020 in Wuhan. As shown in Section 5, we had applied a one-side linear filter that re-distributes the spikes in the Hubei cities to the previous 7 days with decreasing weights ranging from 7/28 to 1/28.

Roles of Infectious Duration D on R_t Professor Tang's question on D's role on R_t and its variation in D is timely and important. A major challenge in the estimation of R_t is the volatility of $\gamma(t)$ in the early stage of the epidemic. We adopt the formulation of $R_t^D = \beta(t)D$. Rather than estimating D, we assign three sets of D values in our analysis, $D = 7, 10.5$ and 14. The narrow confidence intervals for R_t observed by Wang et al. are likely due to the use of the fixed D. A wider confidence intervals would appear by adopting the $D_t = \beta(t)/\gamma(t)$ version via the parametric bootstrap. We have updated Figure 3 (See Figure 1) on the removal rates $\gamma(t)$ to March 2nd 2020, which shows a general increasing trend in the infection rates.

Regarding the possible over-estimation of R_t^{14} in Hubei cities other than Wuhan raised by Jiang et al., we would say that the idea of using three D values to provide a range of R_t^D values that R_t^{14} serves as the worse case scenario and hence may over-estimate. In this case, one may use $R_t^{10.5}$. Another matter that is in play in the early stage of epidemic is the high volatility in the infection rate estimation despite of adopting a fixed D version, which may explain the high volatility and the diverse ranges of R_t estimates in the literature.

Infection in the Incubation Period and Population Mobility. As rightly pointed out by Jiang et al., the vSIR framework, which consists of only three compartments, does not accommodate infection in the incubation period, neither the four compartments SEIR model. In a following up work (Gu et al., 2020) we have proposed an extended SEIR model, called the vSEIdR model, which allows infections in both the exposed and infection states. Population mobility or migrations are also not considered, which can be accommodated by adding an extra term to the susceptible population counts $S(t)$ to reflect immigration or emigration respectively, and a certain percentage of the immigrants (emigrants) are infected patients (imported or exported cases). Doing so would require population mobility data. We hope to conduct such research in a future project.

References

Gu J, Yan H, Huang Y, Zhu Y, Sun H, Zhang X, et al. (2020). Better strategies for containing COVID-19 epidemics: A study of 25 countries via an extended varying coefficient SEIR model. MedRxiv preprint: https://doi.org/10.1101/2020.04.27.20081232.