Discussion of "Evaluate the Risk of Resumption of Business for the States of New York, New Jersey and Connecticut via a Pre-Symptomatic and Asymptomatic Transmission Model of COVID-19"

Yishu Xue[1], Hou-Cheng Yang[2], Yuqing Pan[3], and Guanyu Hu[4,*]

[1]*Department of Statistics, University of Connecticut, CT, USA*
[2]*Department of Statistics, Florida State University, FL, USA*
[3]*Microsoft, WA, USA*
[4]*Department of Statistics, University of Missouri – Columbia, MO, USA*

Tian et al. (2021) proposed the Susceptible-Unidentified infectious-Self-healing without being confirmed-Confirmed cases (SIHC) model that divides the population into four compartments as opposed to three, which is assumed by the popular Susceptible-Infectious-Recovered model (SIR; Kermack and McKendrick, 1927). Specifically, the authors divided the infectious compartment into those who exhibit symptoms, and asymptomatic carriers. Instead of using a recovered/removed compartment, the authors assumed that individuals in the infectious compartment eventually end up confirmed and hospitalized or quarantined, or self-healed without being confirmed. This novel segregation is of practical value as it matches the current practices in fighting COVID-19. In the rest of this discussion, we comment on the approach that the authors have proposed, and suggest some possible extension of the work for future research.

1 The Proposed SIHC Model

There is a notable amount of work done by statisticians and biostatisticians since the breakout of COVID-19. One of the most frequently used model, the SIR model, segregate the population into three compartments, and used three differential equations to depict the evolution of each. Many variants of the SIR model or its extensions have been used to model the development of COVID-19 from different aspects (see, Wang et al., 2020; Yang et al., 2020; Hu and Geng, 2020). The proposed model takes into consideration hospitalization/quarantine (compartment C) and self-healing (compartment H), and similar to the original SIR model, the evolution of the four compartments can also be described using differential equations:

$$
\begin{aligned}
\frac{dS(t)}{dt} &= -\rho\theta(t)S(t)\frac{I(t)}{N(t)}, \\
\frac{dI(t)}{dt} &= \rho\theta(t)S(t)\frac{I(t)}{N(t)} - \frac{I(t)}{D_C} - \frac{I(t)}{D_H}, \\
\frac{dH(t)}{dt} &= \frac{I(t)}{D_H}, \\
\frac{dC(t)}{dt} &= \frac{I(t)}{D_C},
\end{aligned}
\tag{1}
$$

with ρ being transmissibility, and $\theta(t)$ denoting the time-varying average per-person contact number, D_H being the average duration from infection to self-healing, D_C being the time taken

---

*Corresponding author. Email: guanyu.hu@missouri.edu.

from infection to test confirmation, and N denoting the population, which is assumed to be constant over the duration of the pandemic.

The function $\alpha(t) = \rho\theta(t)$ is designed to reflect policy intervention, which, for states of New York (NY), New Jersey (NJ), Connecticut (CT), and California (CA), means stay at home orders, wear fast mask orders, etc. Among the compartments, $C_{1:T}$ is observable, while $I_{1:T}$ and $J_{2:T}$ are latent, assuming $H_1 = 0$. With Poisson assumption for $I_{t+1} \mid \mathcal{I}_{t+1}(Z_t, \Theta)$ and $C_{t+1} \mid \mathcal{C}_{t+1}(Z_t, \Theta)$, and $H_{t+1} = \mathcal{H}_{t+1}(Z_t, \Theta)$, the posterior distribution of the parameters can be specified, which can be approximated using MCMC.

2 Real World Implications

The parameters that are used to specify the curve $\alpha(t)$ are estimated for the four states separately, and the estimated I_t per 100,000 for the four states are visualized together with their respective confidence bands in Figure 6. It is clearly seen that after the stay at home orders were enforced, I_t for NY, NJ and CT began to decrease, while CA showed an increasing trend, which was rather concerning. This concern turned out to be rather serious - after the early reopen and a summer surge, CA has become one of the hotspot states, together with Arizona, Florida, and Texas. The five different scenarios corresponding to different re-open dates in Figure 7 are also interesting.

Using these four states as benchmarks, a few interesting observations can be made. As we know, NY, NJ and CT have already experienced their peaks before their resumption of business. Some other states, such as CA, have not reached the peak yet still resumed their business. One problem of interest is how parameters are adjusted with the data of a new state coming in. If the state or region of interest is not quite similar to any of NY, NJ, CT and CA, such as states in the midwest area or the Great Lakes region, it becomes challenging as to how to select the benchmark. Second, the authors indicated that the initial work completed in May 2020 before widespread protests of social injustice occurred. The cases in NY and NJ, however, still showed decreasing trend in June despite there have been protests in these two states. It would be more interesting if further analysis could extend the study period to around the school re-opening time. Finally, several recent works (Thomas et al., 2020; Hu and Geng, 2020; Pan et al., 2020) studied spatial patterns in COVID-19 data from different perspectives. Spatial heterogeneity leads to dramatic difference in social exposure to COVID-19, and stresses local healthcare systems differently in terms of timing and severity. Population migration between nearby states could be caused by the re-opening of schools. Detection of heterogeneity patterns among spatial regions could reveal interesting insights for the study of COVID-19.

3 Conluding Remarks

Tian et al. (2021) proposed a modified version of the SIR model which has finer-partitioned compartments, which provides more accurate depiction of the role of quarantine policies in the fight against the COVID-19 pandemic. Four U.S. states are used as examples. The potential impacts of choosing to re-open businesses on different dates are assessed. As the U.S. is currently seeing another wave of COVID-19, the proposed approach can be further calibrated, and used to assess future risks for more states.

References

Hu G, Geng J (2020). Heterogeneity learning for SIRS model: an application to the COVID-19. Statistics and Its Interface. Forthcoming.

Kermack WO, McKendrick AG (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, 115(772): 700–721.

Pan T, Shen W, Hu G (2020). Spatial homogeneity learning for spatially correlated functional data with application to COVID-19 growth rate curves. ArXiv preprint: https://arxiv.org/abs/2008.09227.

Thomas LJ, Huang P, Yin F, Luo XI, Almquist ZW, Hipp JR, et al. (2020). Spatial heterogeneity can lead to substantial local variations in COVID-19 timing and severity. *Proceedings of the National Academy of Sciences*, 117(39): 24180–24187.

Tian T, Tan J, Jiang Y, Wang X, Zhang H (2021). Evaluate the timing of resumption of business for the states of New York, New Jersey, and California via a pre-symptomatic and asymptomatic transmission model of COVID-19. *Journal of Data Science*, 19. In this issue.

Wang L, Zhou Y, He J, Zhu B, Wang F, Tang L, et al. (2020). An epidemiological forecast model and software assessing interventions on the COVID-19 epidemic in China (with discussion). *Journal of Data Science*, 18(3): 409–432.

Yang HC, Xue Y, Pan Y, Liu Q, Hu G (2020). Time fused coefficient SIR model with application to COVID-19 epidemic in the United States. ArXiv preprint: https://arxiv.org/abs/2008.04284.