TIME-VARYING REPRODUCTION NUMBER FOR COVID-19 PANDEMIC ON A COLLEGE CAMPUS

Elizabeth Amona*, Yuan Wang, Elissa J. Schwartz

*Virginia Commonwealth University & Washington State University

Introduction

The rapidly spreading coronavirus pandemic has now become a threat to the world at large. The susceptible-infectious-removed (SIR) model and related models have been used to model the pandemic. However, models that assume time-invariant parameters may not capture the dynamics of transmission well or account for inaccuracies in the reported cases. We used a Poisson model with time-varying transmission and removal rates (Hong & Li (2020)) and estimated the time-varying reproduction number of the coronavirus pandemic in the university town of Pullman in Washington state. This method accounts for randomness in disease transmission and possible misreporting in data. We used a simulation method to generate missing recovery counts, and we applied the B-spline approximation method to get a smooth estimates of the reproduction number. Unlike the deterministic ODE based SIR models, this method doesn't require Infectious I(t) & removed R(t) to be known, but estimate them using the data. Our estimates of the time-varying reproduction number help us understand the spread of coronavirus in small high-density rural towns. It also provides a forecast for the spread of COVID in the future.

Model

The following ordinary differential equations (ODE) describe the change of s(t), i(t), and r(t):

$$\frac{ds(t)}{dt} = -\beta(t)s(t)i(t), \tag{1}$$
$$\frac{di(t)}{dt} = \beta(t)s(t)i(t) - \gamma(t)i(t), \tag{2}$$
$$\frac{dr(t)}{dt} = \gamma(t)i(t), \tag{3}$$

with initial conditions: $i(0) = i_0$ and $r(0) = r_0$, where $i_0 > 0$ in order to let the epidemic develop.

Here, $\beta(t) > 0$ is the time-varying transmission rate of an infection at time t, and $\gamma(t) > 0$ is the time-varying removal rate at time t. Diekmann et al.(2000).

Furthermore, the time-dependent reproduction number is

$$\mathcal{R}_0(t) = \frac{\beta(t)}{\gamma(t)} \tag{4}$$

Under the fixed population size assumption, $s(t) + i(t) + r(t) = 1$.

A Poisson Model Based on Time-Varying SIR

To link the reported number (observed daily numbers) of infectious and removed, denoted by $Z_I(t)$ & $Z_R(t)$ and the true numbers of infectious and removed, denoted by I(t) & R(t), a Poisson model is proposed as follows:

$$Z_I(t) \sim Pois\{I(t)\}, \tag{5}$$
$$Z_R(t) \sim Pois\{R(t)\}.$$

It is assumed that given $I(t)$ and $R(t)$, the observed daily number $\{Z_I(t), Z_R(t)\}$ are independent across $t = 1, \cdots, T$.
Thus, given the data $\{Z_I(t), Z_R(t)\},\ t = 1, \cdots, T$, we obtain the estimates of (β, γ), that is $(\hat{\beta}, \hat{\gamma})$, by maximizing the following likelihood,

$$\mathcal{L}(\beta, \gamma) = \prod_{t=1}^{T} \frac{e^{-I(t)} I(t)^{Z_I(t)}}{Z_I(t)!} \times \prod_{t=1}^{T} \frac{e^{-R(t)} R(t)^{Z_R(t)}}{Z_R(t)!},$$

or equivalently, maximizing the log likelihood function

$$\ell(\beta, \gamma) = \mathbf{N} \sum_{\mathbf{t=1}}^{\mathbf{T}} \{-\mathbf{r(t)} + \mathbf{Z_R(t)} \log \mathbf{r(t)} - \mathbf{i(t)} + \mathbf{Z_I(t)} \log \mathbf{i(t)}\} + C. \tag{6}$$

Where C is a constant free of β and γ.

Model Estimation

A B-spline is a linear combination of a set of basis functions that are determined by the number and location of specified knots or cut-points, as well as the (polynomial) degree of curvature. A degree of one implies a set of straight lines, degree of two implies a quadratic curve, three a cubic curve, etc. In this study, the B-spline basis function is over [0,T] associated with the knots $0 = w_0 < w_1 < ... < w_{q-1} = T$. For flexibility, we allow the number of knots (q_1 and q_2) to differ between $\beta(t)$ & $\gamma(t)$, which is defined as

$$\log \beta(t) = \sum_{j=1}^{q_1} \beta_j B_j(t), \tag{7}$$
$$\log \gamma(t) = \sum_{j=1}^{q_2} \gamma_j B_j(t).$$

When $\beta_1 = \beta_{q_1}$ and $\gamma_1 = \gamma_{q_2}$, the model reduces to a constant SIR model, Parast et al. (2019).

Let N be the size of the population of the Pullman community. The date when the first case was reported is set to be the starting date with $t = 1,\ i_0 = Z_I(1)/N$, and $r_0 = Z_R(1)/N$.

Results

The actual data we obtained from WSU had some missing recovery data from Aug 1st-Sept 9th, 2021. So, we simulated recovered data in order to account for the missing actual dataset. We considered three different times to recovery; we assumed ranges between $(5-10$ days), $(5-11$ days), and $(5-12$ days). Afterwards, we run 10 replicates on each of the three simulated recovered data, hence, we had thirty replicates in total. Furthermore, the least square method was implemented to choose the best model (closest to the actual data).

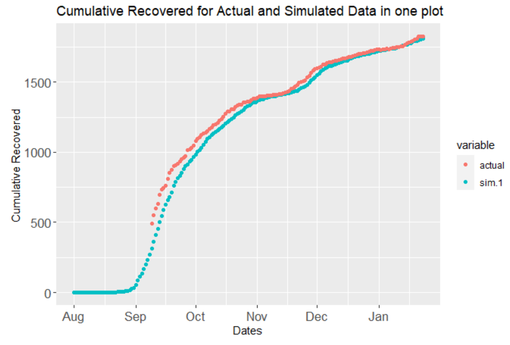

Fig. 1: Cummulative Recovered (Individuals)

Estimating the time-varying reproduction number from the time-varying transmission rates and recovery rates

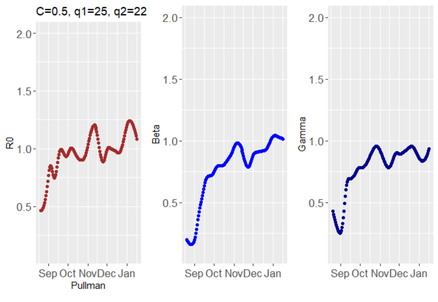

Fig. 2: $\mathbf{c = 0.5,\ q_1 = 25}$ **and** $\mathbf{q_2 = 22}$..

The Effect of Basis Function on the Model

The choice of basis functions is important in the B-spline approximation method. These can either be obtained by using cross-validation or guessing different values within [0, T], where T is the duration for estimation remaining for validation. In this study, we varied the number of knots (q_1 and q_2) within the interval [0, T] and observed the results obtained. We noticed that when q_1 and q_2 are between [15, 50] we had a better result compared to smaller values, as shown in Fig 3 and Fig 4 below. Furthermore, increasing the value of the basis functions beyond the above interval makes no difference in the result, therefore, we did not include those results.

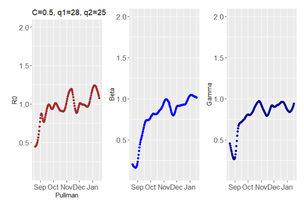

Fig. 3: $\mathbf{c = 0.5,\ q_1 = 28}$ **and** $\mathbf{q_2 = 25}$.

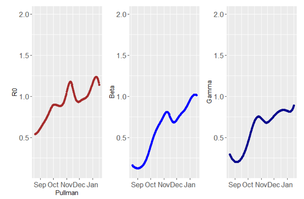

Fig. 4: $\mathbf{c = 0.5,\ q_1 = 10}$ **and** $\mathbf{q_2 = 8}$.

Conclusions, Limitations, and Future Work

1. The reproduction number for the Washington State University COVID-19 epidemic was found to vary over time. This result is consistent with the observed report of daily cases.
2. The major limitation in this study is that the B-spline method does not consider the behavior of the disease, hence the dynamics of R_0 is limited. Thus, we could modify our model by adding a constraint on R_0 structure.
3. Furthermore, we could extend this study by using the SEIR model instead of SIR model.
4. Lastly, we could extend this study by considering multiple communities within the campus.

References

[1] Hong, Hyokyoung G and Li, Yi, *Estimation of time-varying reproduction numbers underlying epidemiological processes: A new statistical tool for the COVID-19 pandemic*. Public Library of Science San Francisco, CA USA,PloS one,15,7,e0236464,2020.

[2] Diekmann, Odo and Heesterbeek, Johan Andre Peter, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. John Wiley & Sons,5,2000.

[3] Parast, Layla and Mathews, Megan and Friedberg, Mark W, *Dynamic risk prediction for diabetes using biomarker change measurements*. BMC medical research methodology, BioMed Central,19,1,1–12,2019.