SEIRD MODEL FOR QATAR: A CASE STUDY.

Aiman Hanif†

†Virginia Commonwealth University, Richmond, Virginia 23284

An Important Problem

Coronavirus Disease 2019 (COVID-19) is a severe pandemic affecting the whole world with a fast spreading regime, requiring to perform strict precautions, such as social distancing and mask-wearing, to keep it under control. The virus has spread from China to 196 other countries across the globe. The State of Qatar was one of the countries that were affected by COVID-19 spreading; the first infected case was reported on February 29, 2020, and it could be considered the 2nd highest in the Arab World with the number of confirmed cases 28,272 as of May 14, 2020. For effectively specifying such security measures, it is essential to have a real-time monitoring system of the infection, recovery and death rates. We develop, implement and deploy a data-driven forecasting model for use by stakeholders in the State of Qatar to deal with the Covid-19 pandemic. The model will focus on infected, deaths and recovered as those are the only data available at this time.

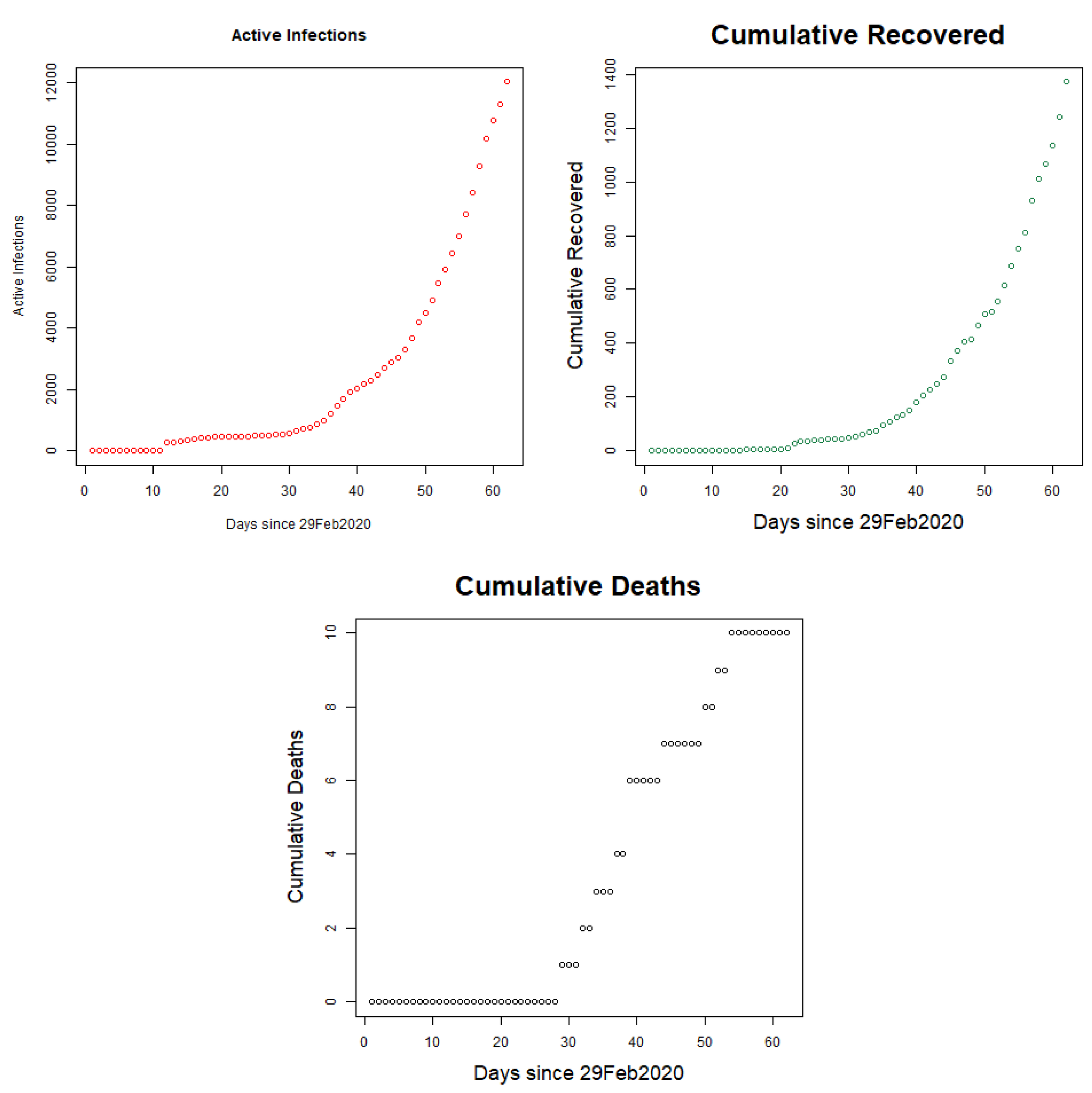

Fig. 1: Covid-19 Infected, Recovered and Deaths for Qatar.

Model

Let $S(t)$ be the number of people Susceptible at time t, $E(t)$ be the number of people Exposed at time t, $I(t)$ be the total number of Infected at time t, $R(t)$ be the cumulative number of recovered at time t and $D(t)$ be the cumulative number of Deaths at time t. Furthermore, α is the transmission rate (per day×individual^2) from Susceptible to Exposed, β is the rate (per day) at which Exposed become Infected, γ is the rate (per day) at which Infected become recovered and η is the mortality rate (per day) for those Infected. This can be modeled with the following system of ordinary differential equations:

$$\begin{aligned}
\frac{d\lambda_S(t)}{dt} &= -W(t)^T\boldsymbol{\alpha}\lambda_S(t)\lambda_E(t)\\
\frac{d\lambda_E(t)}{dt} &= W(t)^T\boldsymbol{\alpha}\lambda_S(t)\lambda_E(t) - \beta\lambda_E(t) - \gamma\lambda_E(t) - \beta_A\lambda_E(t)\delta(t-\tau)\\
\frac{d\lambda_I(t)}{dt} &= \beta\lambda_E(t) + \beta_A\lambda_E(t)\delta(t-\tau) - \gamma\lambda_I(t) - \eta\lambda_I(t)\\
\frac{d\lambda_R(t)}{dt} &= \gamma\lambda_I(t)\\
\frac{d\lambda_D(t)}{dt} &= \eta\lambda_I(t)
\end{aligned} \tag{1}$$

where $\lambda_S(t)$, $\lambda_E(t)$, $\lambda_I(t)$, $\lambda_R(t)$ and $\lambda_D(t)$ are the means of $S(t)$, $E(t)$, $I(t)$, $R(t)$ and $D(t)$, respectively and the parameters have the same definition as provided in the system given in the equation. Since there is no data for $S(t)$ and $E(t)$, these compartments will be latent variables and will not directly factor into the likelihood. The likelihood for $I(t)$, $R(t)$ and $D(t)$ are given by:

$$\begin{aligned}
I(t) &\sim Poisson\,(\lambda_I(t))\\
R(t) &\sim Poisson\,(\lambda_R(t))\\
D(t) &\sim Poisson\,(\lambda_D(t))
\end{aligned} \tag{2}$$

Interventions

The model needs to be able to handle interventions made by the Government of the State of Qatar. The main parameter that policy can influence is α, the rate of transmission from Susceptible to Exposed. One way to implement this is the use of indicator functions $W_k(t)$ defined as:

$$W_k(t) = \begin{cases} 1 & \text{if } t > t_k \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where t_k is the time where the k^{th} intervention is taken and index $k = 1, 2, .., K$. For each intervention, there needs to be a change to the value of α, denoted α_k, that captures the impact of the intervention. Let the vector $W(t) = (1, W_1(t), W_2(t), ..., W_K(t))^T$ be the vector of the values of each $W_k(t)$ at time t. Let $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, ...\alpha_K)^T$. This formulation gives the following transitions rates between $S(t)$ and $E(t)$:

$$\alpha(t) = \begin{cases} \alpha_0 & \text{if } 0 < t < t_1 \\ \alpha_0 + \alpha_1 & \text{if } t_1 < t < t_2 \\ \alpha_0 + \alpha_1 + \alpha_2 & \text{if } t_2 < t < t_3 \\ \vdots & \vdots \\ \alpha_0 + \alpha_1 + ...\alpha_K & \text{if } t_K < t \end{cases} \tag{4}$$

which will require the following constraints due to the fact that $\alpha(t) > 0$ for all t:

$$\begin{aligned}
\alpha_0 &> 0\\
\alpha_0 + \alpha_1 &> 0\\
\alpha_0 + \alpha_1 + \alpha_2 &> 0\\
\vdots &> \vdots\\
\alpha_0 + \alpha_1 + ... + \alpha_K &> 0
\end{aligned} \tag{5}$$

Let $\mathcal{A}$ be the set defined by the constraints above.
In addition to changes in infection rates $\boldsymbol{\alpha}$, impulse functions can be used to model dramatic one time shifts in transitions between states. A Dirac delta function defined by

$$\delta(x) = 0, \text{ if } x \neq 0 \tag{6}$$

which satisfies $\int_{-\infty}^{\infty} \delta(x)dx = 1$. This can be integrated in the model to capture spikes in the number of cases. In our case, the State of Qatar data exhibits this type of behavior at day 12 where one can clearly see a large jump in the number of infections. This is incorporated into the model presented by a Dirac delta function, $\delta(t-\tau)$, in transition rate between Exposed and Infected, which is coupled with a coefficient to β_A to capture the impact of the jump.

Bayesian Computation

To specify the prior distributions for $\boldsymbol{\alpha}$, β_A, β, γ and η one must incorporate the following constraints $\alpha > 0$, $\beta > 0$, $\gamma > 0$ and $\eta > 0$. Hence the following prior distributions are set:

$$\begin{aligned}
\boldsymbol{\alpha} &\sim MVN(a\mathbf{1}, \sigma^2\mathbf{I})C(\mathcal{A})\\
\beta_A &\sim Exp(1)\\
\beta &\sim Exp(1)\\
\gamma &\sim Exp(1)\\
\eta &\sim Exp(1)
\end{aligned} \tag{7}$$

where $C(\mathcal{A})$ is an indicator function takes the value 1 if $\boldsymbol{\alpha} \in \mathcal{A}$. This serves to truncate the normal distribution in order to keep $\boldsymbol{\alpha}$ in the feasible range of values. The likelihood and prior distributions specifications lead to the following posterior distribution when $a = 1$ and $\sigma^2 = 1$:

$$\begin{aligned}
\pi(\boldsymbol{\alpha}, \beta_A, \beta, \gamma, \eta|\mathbf{D}) &\propto \pi(\boldsymbol{\alpha})\pi(\beta)\pi(\gamma)\pi(\eta)L(\mathbf{D}|\boldsymbol{\alpha}, \beta_A, \beta, \gamma, \eta)\\
&= C(\mathcal{A})e^{-\frac{1}{2}(\boldsymbol{\alpha}-\mathbf{1})^T(\boldsymbol{\alpha}-\mathbf{1})-\beta_A-\beta-\gamma-\eta}\\
&\times \prod_{t=1}^{T} \frac{\lambda_I(t)^{I(t)}\lambda_R(t)^{R(t)}\lambda_D(t)^{D(t)}e^{-\lambda_I(t)-\lambda_R(t)-\lambda_D(t)}}{I(t)!R(t)!D(t)!}
\end{aligned} \tag{8}$$

The posterior distribution does not lend to any analytic solution, hence Markov chain Monte Carlo (MCMC) technique Metropolis-Hastings sampling is used to obtain samples from the posterior distribution. The tuned sampler was used to generate 5,000 samples from $\pi(\boldsymbol{\alpha}, \beta_A, \beta, \gamma, \eta|\mathbf{D})$, and trace plots were visually examined for convergence and deemed to be acceptable. All inferences will be made from these 5,000 samples. The model and sampling algorithm is custom programmed in the R statistical programming language version 3.6.3. The computation takes approximately 290 seconds using an AMD A10-9700 3.50GHz processor with 16GB of RAM to obtain 5,000 samples from the posterior distribution.

Results

To apply the model, the following initial conditions are specified: $S(0) = 2,782,000$, $E(0) = 3$, $I(0) = 1$, $R(0) = 0$ and $D(0) = 0$. Furthermore, model interventions were placed at days $t_1 = 12,\ t_2 = 19,\ t_3 = 36,\ t_4 = 40$ and $t_5 = 59$ with an Dirac delta impulse at time $\tau = 12$.
To assess the fit of the model, the posterior predictive distribution was used and is given by:

$$\begin{aligned}
\pi(I_{new}(t), R_{new}(t), D_{new}(t)|\mathbf{D}) &= \int L(I_{new}(t), R_{new}(t), D_{new}(t)|\boldsymbol{\alpha}, \beta_A, \beta, \gamma, \eta)\\
&\times \pi(\boldsymbol{\alpha}, \beta_A, \beta, \gamma, \eta|\mathbf{D})\, d\boldsymbol{\alpha} d\beta_A d\beta d\gamma d\eta.
\end{aligned} \tag{9}$$

Using the 5,000 samples from the posterior distribution, 5,000 samples were generated from the posterior predictive distribution. At each time t, the median 0.025 and 0.975 quantiles were obtained to form a posterior predictive interval.

Figure below shows the model fits for Active Infections, Recovered and Deaths with posterior predictive bands as well as a hold out sample of 10 days. Notice that the model does quite well at fitting the dynamics of the Active Infections including the jump at day 12 and captures the plateau and the exponential growth after the plateau as well. The Recovered model fits well as does the Deaths data. To assess the explained variance, a pseudo-R^2 was formed using the median from the posterior predictive distribution at each time as the point estimates. This resulted in a pseudo-R^2 of 0.998 which indicates the fitted model explains approximately 99.8% of the variance in the data. Based on this, the model is deemed to fit well.

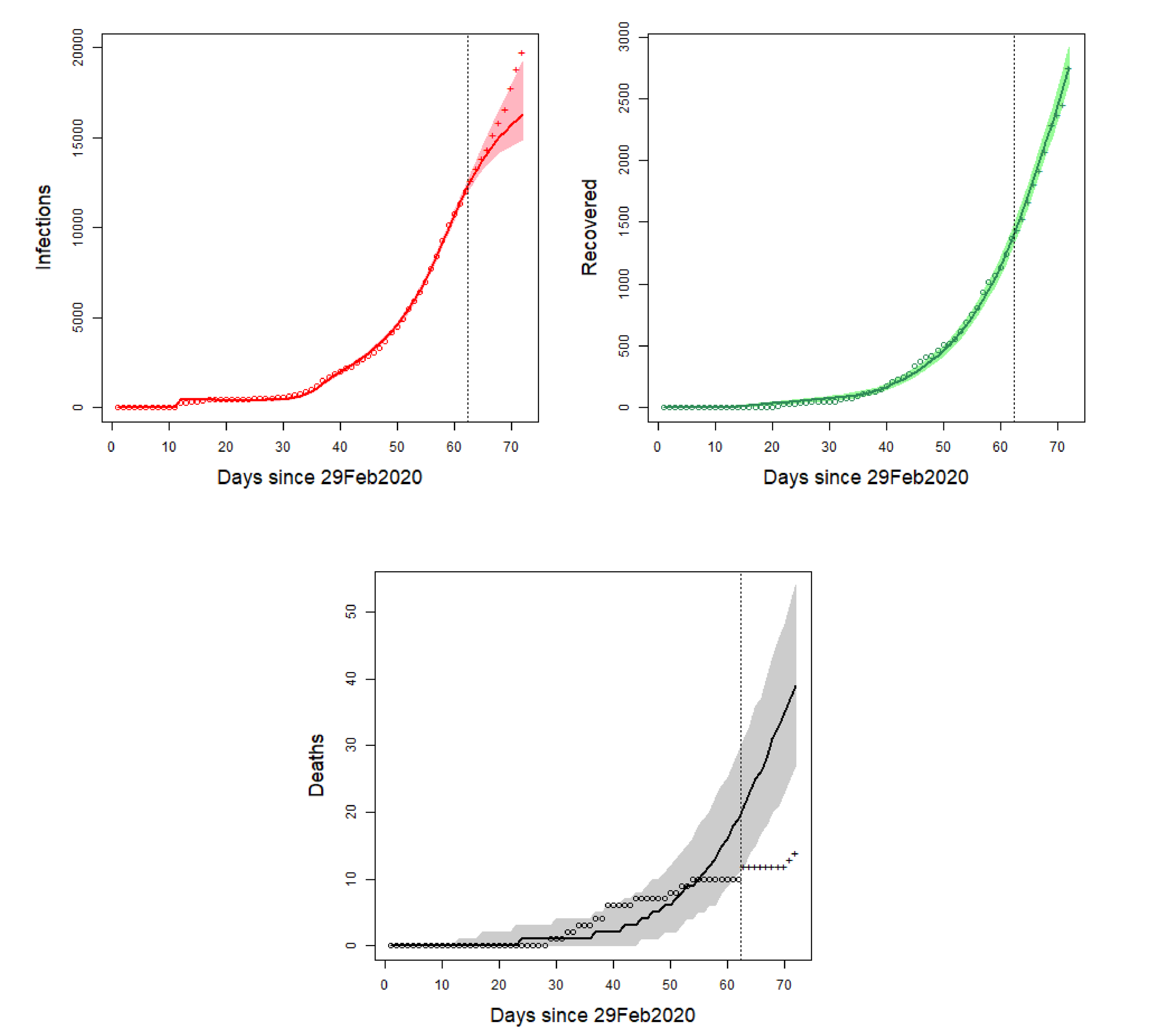

Fig. 2: Covid-19 Infected, Recovered and Deaths for Qatar.

Discussion

This work has demonstrated how to build a SEIRD model for the Covid-19 outbreak in the State of Qatar, include interventions, estimate model parameters and generate posterior predictive intervals using a Bayesian framework. The modeling paradigm is quite flexible for modeling the Covid-19 data as it easily incorporates interventions into the system and can quantify the impact of the intervention. Future work could be to add an overdispersion parameter into the model to allow for the more accurate capture of uncertainty. Furthermore, one can perform simulation studies to better understand how the model may perform under various scenarios. Feature selection methods could be employed to select where the interventions should be placed as well as other forms of interventions could be included in the model. To address any deviations from the standard model, a semi-parametric technique could be studied as well.