PIDE
COVID-19 BLOG
No. 27

Forecasting the Incidence of COVID-19 in Pakistan

The outbreak of global pandemic COVID-19 has also hit Pakistan and critically exposed vulnerable segments of the population to significant economic and health shocks. Objective of this note is to apply basic Susceptible, Infected and Recovered (SIR) epidemiological model and forecast the spread of COVID-19 in Pakistan[1]. Based on the observed patterns of the pandemic spread and the policy response in Pakistan, we estimate that if interacted with the infected persons, the probability of the susceptible getting infected from that interaction is around 60%. Reproductive rate, i.e. the expected number of new infections from a single infection is around 4. Lastly, for the period of 180 days starting from February 2020 when the first case of corona positive was reported in Pakistan, we predict that the incidence of COVID-19 will gain its peak around fourth week of May, 2020. Thereafter the incidence of the pandemic starts declining. Our findings warrant further extension in social distancing and lockdown policy of the government in order to flatten the curve in the predicted time period and restrict the incidence of COVID-19 pandemic in Pakistan.

Developing SIR Model for Pakistan

Making use of the seminal work of the Kermack and McKendrick (1927) where they developed a set of general differential equations which can be invoked for forecasting the spread of an epidemic, Breda et al. (2012) applied these equations to analyze the dynamics of an infectious disease spreading through a susceptible population. Following a general mathematical depiction, they suggested a set of practical assumptions which lead to a simple SIR model of ordinary differential equations. The model divides total population N into Susceptible (S), Infected (I) and Recovered (R) as follows:

$$\frac{dS}{dt} = -\beta\left(\frac{S}{N}\right)I \qquad (1)$$

$$\frac{dI}{dt} = \beta\left(\frac{S}{N}\right)I - \gamma I \qquad (2)$$

$$\frac{dR}{dt} = \gamma I \qquad (3)$$

Considering 2 cases were reported for the first time on 26th February 2020 in Pakistan, our initial conditions will be $S(0) = N - 2,\ I(0) = 2, R(0) = 0;$ at $t = 0$ and N is the total population of Pakistan. Given that $\frac{dN}{dt} = 0$ as N = S+I+R (Constant) where S is the number of Susceptible population, I is the number of Infected, R is the Recovered population, and N is the sum of these three.

[1] Recently, Stock (2020) available at https://www.nber.org/papers/w26902 and Li et al. (2020) available at https://science.sciencemag.org/content/early/2020/03/24/science.abb3221 have made use of the SIR type models and estimated the probable spread of the COVID-19 in USA and China respectively.

β is the contact rate times the probability of transmission given a contact between a susceptible and an infectious individual. γ is the rate of recovery. We can also determine the reproductive number $\left(R_0 = \frac{\beta}{\gamma}\right)$, where R_0 represents the expected number of new infections from a single infection.

The Eq.1 above represents rate of interaction between susceptible and infected people. The more people interact with each other, the higher is the chance of getting infected from a set of susceptible people causing a reduction in the susceptible and increase in the infected.

The Eq.2 is of particular importance. It gives the rate at which people are infected. A stationary point is that where the term $\beta\left(\frac{S}{N}\right)I - \gamma I$ becomes zero and the epidemic gains its peak. Before this point the graph of estimated Infected (via SIR) should be increasing and after that it has a decreasing trend.

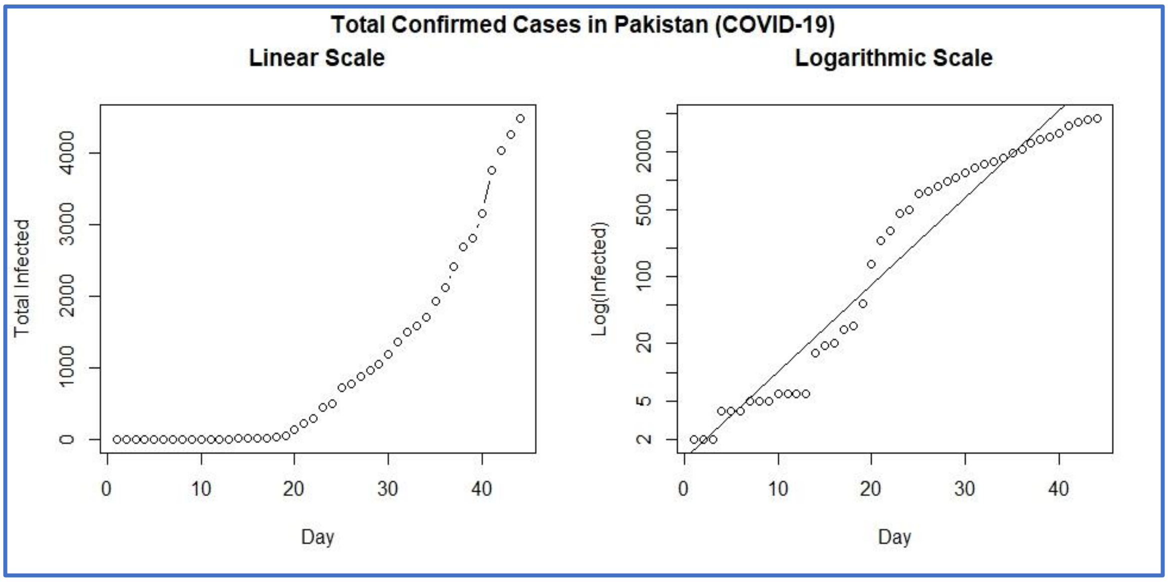
The Eq.3 represents the rate at which infected people are getting recovered. These differential equations govern the rate of change between the different compartments, and can be used to predict future spread of epidemic. Before we can predict the future of the epidemic, we first need to find the best fit of the parameters of the model (β and γ) on basis of the number of confirmed cases so far observed in Pakistan.

Key assumptions of the model are:

i. The infection circulates in a population of size N, with a per capita “background” death rate which is in this case assumes to be zero as there is no death co firmed initially.
ii. The model assumes that recovered individuals are immune from reinfection for life.
iii. Infected individuals move directly into the infectious class and remains there for an average infectious period of 1/γ.
iv. Transmission of infection from infectious to susceptible individuals is controlled by a bilinear contact term $\beta\left(\frac{S}{N}\right)I$. This stems from the assumption that the I infectious individuals are independently and randomly mixing with all other individuals, so the fraction S/N of the encounters is with susceptible individuals.

Findings of the Model

Looking at the observed trend in the spread of COVID-19 pandemic in Pakistan first, one can see an exponential increase starting from the third week of March 2020. The following graphs explain the spread of COVID-19 in Pakistan from 26th February to 9th April:



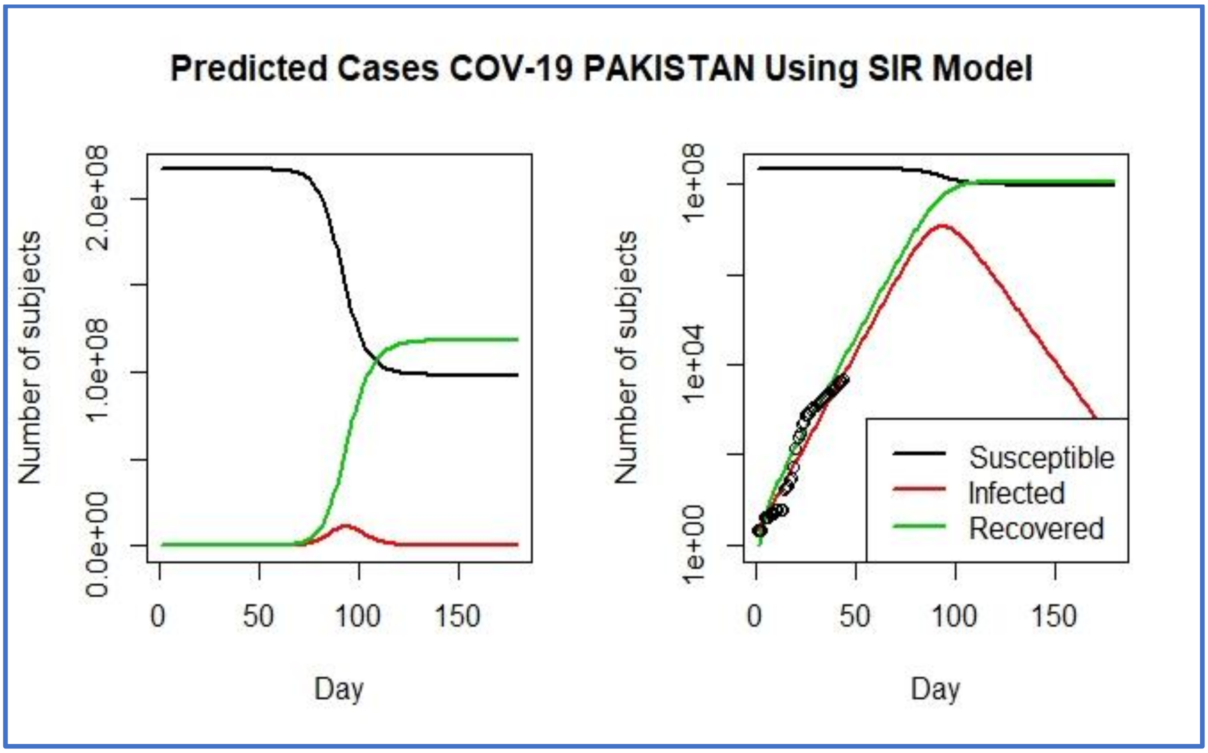
At time of writing this note, the total no. of confirmed cases in Pakistan stands at 4601 as on 9th April, 2020[2]. To fit a model on the set of values we need two things to perform; firstly, a solution for above written differential equations and secondly, an optimization procedure to optimize our solution. For this the *Limited Memory Broyden-Fletcher-Goldfarb-Shanno with boundaries* (L-BFHS-B) method is used. (Liu et al. 1989) provided a detailed review on mathematical programming of the said optimization. Concretely, we will minimize the sum of the squared differences between the number of infected *I* at time *t* and the corresponding number of predicted cases by our model $\hat{I}(t)$:

$$RSS(\beta,\gamma) = min \sum_{t} \left(I(t) - \hat{I}(t)\right)^2$$

Our model estimates the parameter of SIR model β and γ under the optimum boundary constrains to be found as follows:

SIR Model Estimates		
Parameter	β	γ
Estimate	0.6008	0.4155

The following graph on the right plots predicted spread of the pandemic on log(infected) for 180 days. On the left we show the actual incidence of COVID-19 pandemic as well the forecasted behavior of susceptible and recovered population for 180 days[3]. One important number is the so-called basic reproduction number (also basic reproduction ratio) denoted by R_0 which basically shows how many healthy people get infected by a sick person on average and can be estimated using ration $\frac{\beta}{\gamma}$:



[2] http://covid.gov.pk/

In case of Pakistan the value of R_0 is 1.4462 which is less than 2, highlighting the fact that the epidemic is not as much acute as in case of other hardly hit countries such as Italy, Spain and USA. Another important aspect of the model is that the peak of the epidemic COVID-19 would be about 90 days ahead from the first case reported in Pakistan and possibly it would be around 25th May and after that there is a decline in rate of infected cases. The slope $\frac{dI}{dt} < 0$ after this point predicts the same result.

Our findings suggest a further extension in social distancing and lockdown policy of the government in order to flatten the curve in the predicted time period and restrict the incidence of COVID-19 in Pakistan.

References:

Breda, D., Diekmann, O., De Graaf, W., Pugliese, A., & Vermiglio, R. (2012). On the formulation of epidemic models (an appraisal of Kermack and Mckendrick). Journal of Biological Dynamics, 6(Suppl. 2), 103–117.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 115(772), 700–721.

Li, R. et. al. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science, published online March 16, 2020.

Available at https://science.sciencemag.org/content/early/2020/03/13/science.abb3221

Liu, D. C., & Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization methods. Mathematical Programming (45), 503-528.

Stock, J. H. (2020). Data Gaps and the policy response to the Novel Corona Virus. NBER, Working Paper 26902.

Available at http://www.nber.org/papers/w26902

By: Muhammad Jawad, PIDE, and Muhammad Saleh, SDGs Unit, Ministry of Planning Development and Special Initiatives.

Pakistan Institute of Development Economics

Web: www.pide.org.pk, Twitter: @PIDEpk, Facebook: PIDEIslamabad