BTP: Intermediate Progress Report

-Ojaswi Jain

1. Goals

- Calculating precise betas, gammas for each day: Possibly formulating a model to take into account the underreporting/missed reporting on a few days to fill in the holes
- Design a robust prediction mechanism/model for the prediction of long-term trends of betas.

2. Basic SIR Model

- Three differential equations:

  1. The rate of change of susceptible individuals:
     $dS/dt = -\beta * S * I / N$

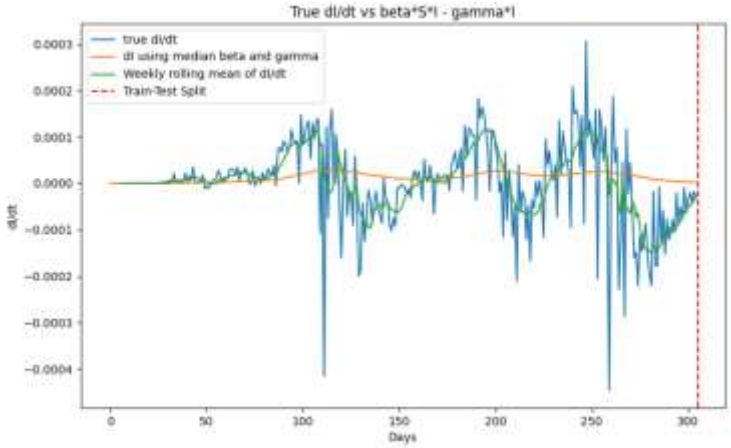
  2. The rate of change of infected individuals:
     $dI/dt = \beta * S * I / N - \gamma * I$

  3. The rate of change of recovered individuals:
     $dR/dt = \gamma * I$

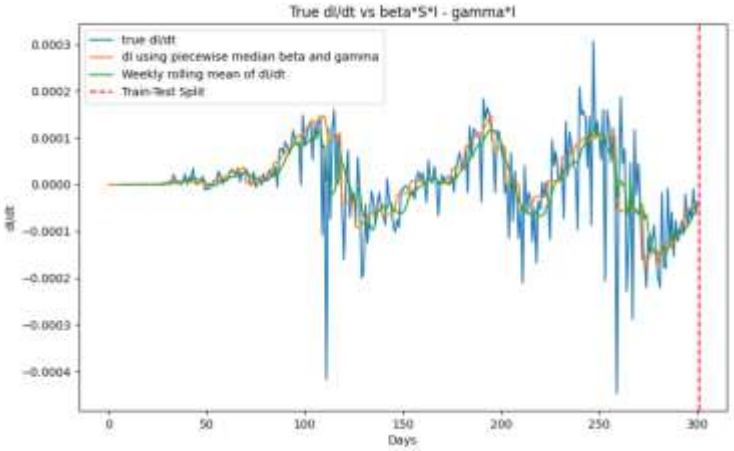
- Assumptions:
  1. The Population remains constant
  2. People don't move from R to I compartment
  3. In cases when the recovery data is not available, assume the recovery period to be a constant (now refuted)

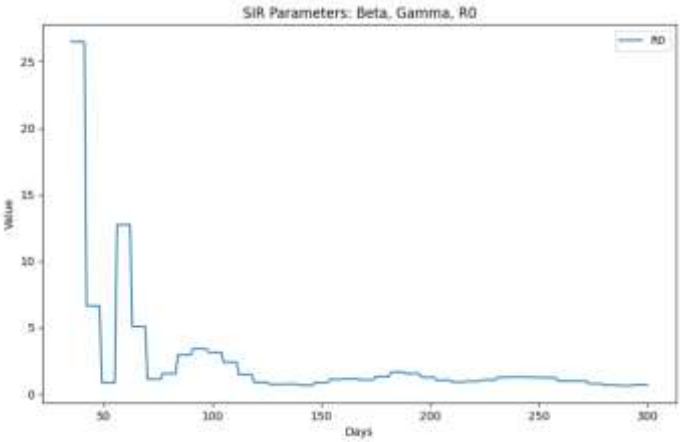
3. Preliminary Implementations

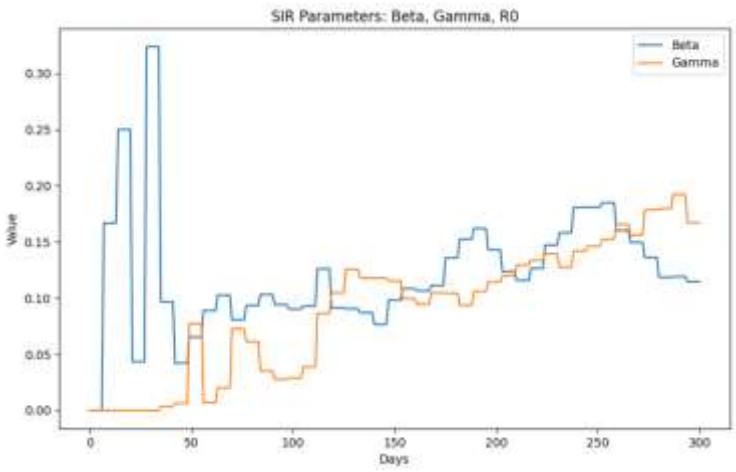
- Keeping both beta and gamma constant, using their overall medians



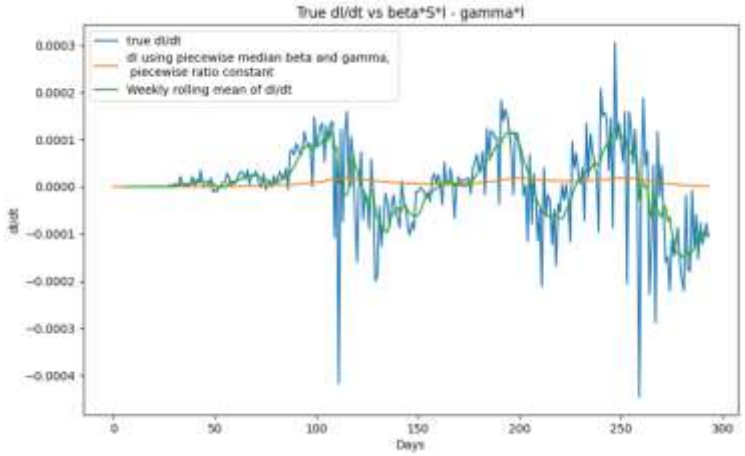
- Moving on to week-wise median betas, which performed better

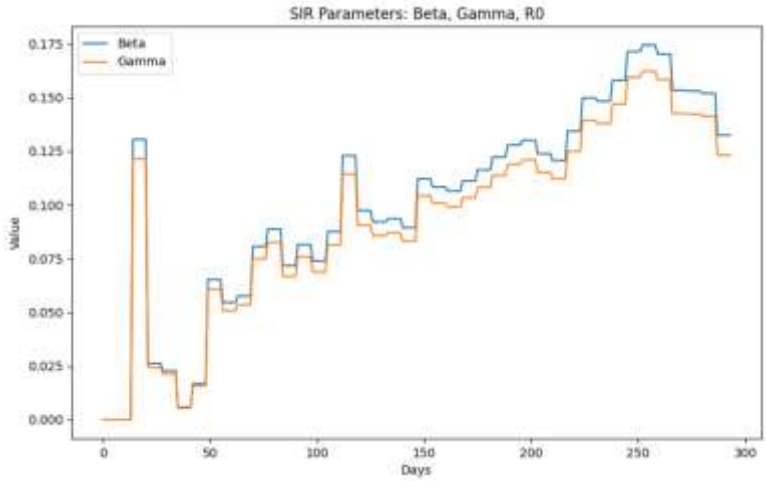






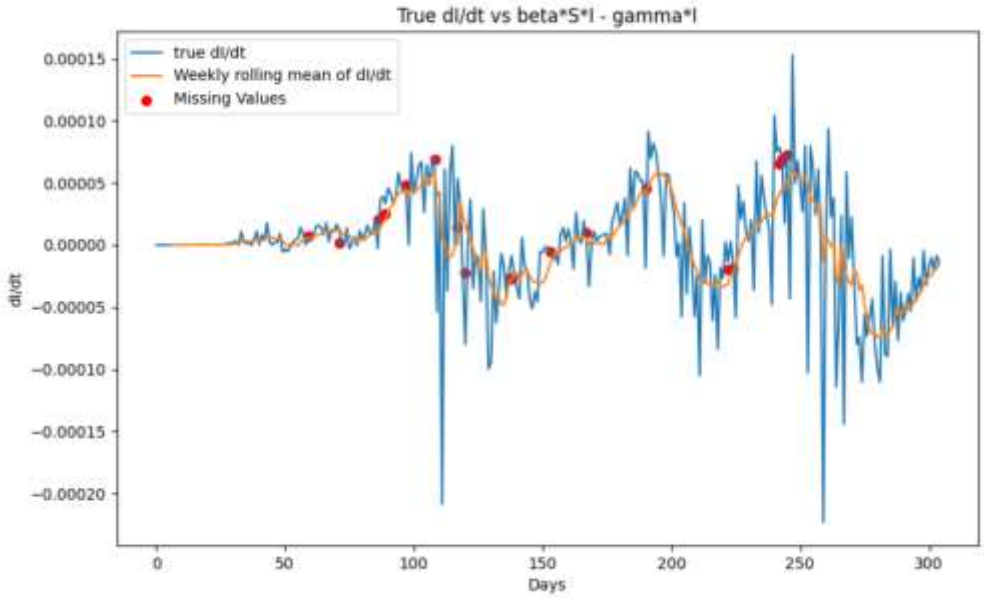
- Convex-optimized betas for each week (weighted median). Identical results
- Additional constraints: Keeping the ratio constant (overall, week-wise). keeps the parameters tied together (de-noising), while ensuring the fit doesn't worsen.





4. Key experiments performed

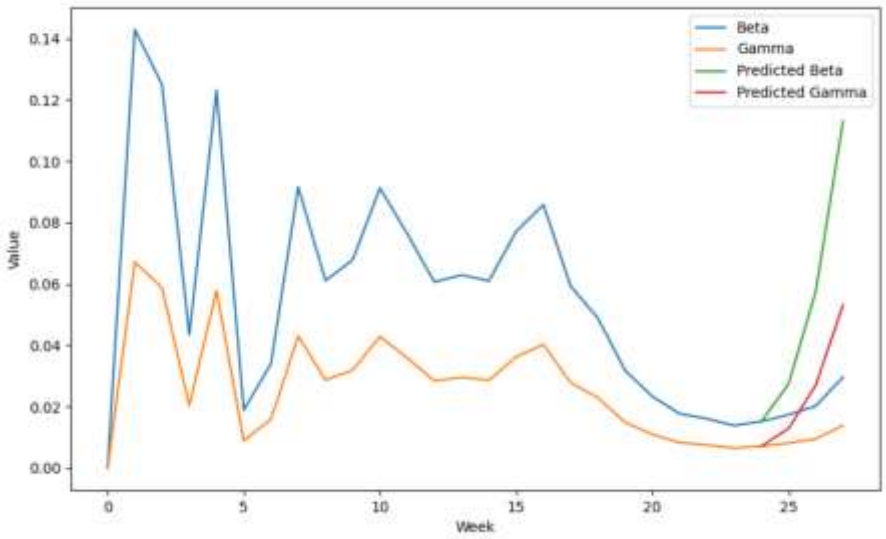
- Filling in the holes (Interpolation of S, I, R)




```
Error from mean [0.04866340296518897, 0.04866340296518897, 0.08523433584521443, 0.08523433584521443, 0.0858871038
0755044, 0.09077276828762115, 0.12120691939686565, 0.17828892695381202, 0.19013788513086252, 0.19240196078385924,
 0.3700363440556748, 0.3984558961742575, 0.5167053536523936, 0.6043747977237248, 0.644302306765254, 0.71372428878
7325, 0.8470710966168619, 1.1873521377866967, 2.274330928342603, 12.877785148743497]
50th percentile error from the mean:  0.281219152419767
```

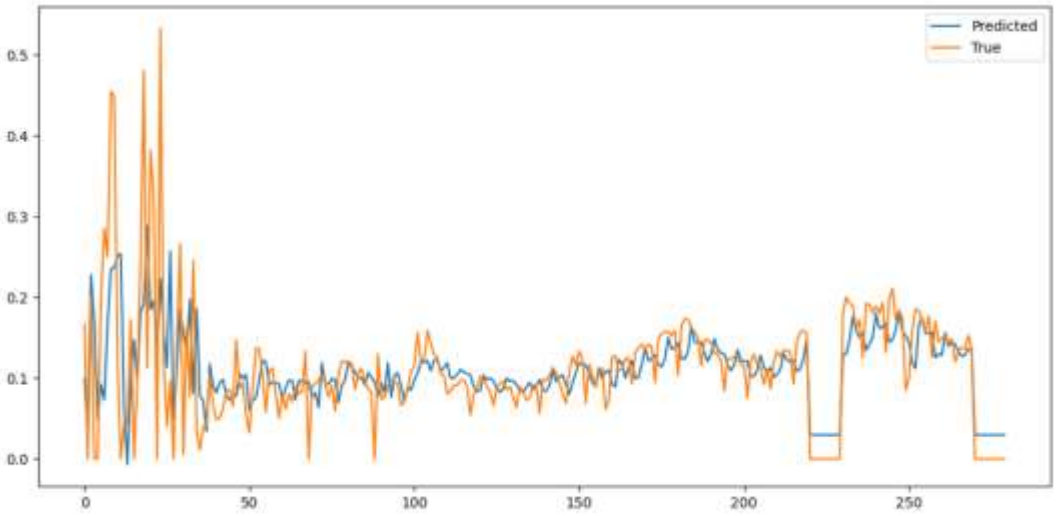
While the day wise values of dI are volatile, being around the mean would be a fair indicator of accuracy.
The (fractional) mean absolute error is shown above. The median of these is 0.28, which seems like a decent number considering the volatility of the values.

- Extrapolation of beta and gamma using splines (failed)



- Linear Regression over betas and gammas
    - Performed linear regression over daily beta, using the last 10 days' beta values to predict the next day's beta values
    - Obtained an R2 score of 0.77 on the test set, following are the weights designated from d-10 to d-1
    - Performance is sub-optimal.

```
[ 0.05519877  0.19720729  0.00961157 -0.07841104  0.01801923  0.10247489
 -0.23156516  0.19574378  0.19144277  0.26682347]
```



- Reproduction of results on other datasets: Looking at the data from other states, our assumptions remain consistent in their performances.

- Examining correlation of betas and gammas: To be compiled

5. Ideas obtained from other papers

- Separate the dead and the recovered, calculate their parameters separately (hence 4 differential equations)
- Use the SEIR model. Challenge: E is an abstract compartment: - Defining it is hard.
- Separate undetectable (asymptomatic) patients from the truly infected. Challenge: Finding datasets categorizing them as such.
- Modelling I and dI/dt as Poisson distributions linked with a common parameter.
- State space model ensuring the flow of people from the R compartment back into the I compartment.

Linear regression for betas and gammas

Some other model

Try on different datasets