Forecasting the COVID-19 Pandemic in the 5th District Appendix: Model and Data Details

Paul Ho Thomas A. Lubik Christian Matthes

April 22, 2020

1 Model

We specify a model in which the growth rate of the cumulative number of cases depends on the current cumulative number of cases. In particular, denoting the cumulative number of cases in period t by C_t, we consider:

$$\Delta \log C_t = \frac{\log(1+\gamma)}{\phi(10^{-5};\beta_0,\beta_1,\zeta)}\phi(C_{t-1};\beta_0,\beta_1,\zeta)\exp(\varepsilon_t), \tag{1}$$

$$\phi(C_{t-1};\beta_0,\beta_1,\zeta) \equiv \exp[-C_{t-1}^{-\beta_0} - (\zeta - C_{t-1})^{-\beta_1}], \tag{2}$$

where $\varepsilon_t \sim \mathcal{N}(0,\sigma^2)$.

The model is set up to flexibly match the trajectory of cases. If there are no existing cases, the growth rate is zero. When a fraction 10^{-5} of the population has been infected, the growth rate in the absence of shocks is γ. The parameter β_0 determines how the growth rate of C_t increases or decreases in the early stages of the pandemic, capturing the appearance of large clusters or the effects of social distancing measures. The parameters β_1 and ζ determine the long-run number of cumulative cases and the speed at which society converges to that number, which could depend on factors such as demographics or policies. Finally, ε_t is a shock that allows for deviations from the model predictions, arising due to randomness in how the virus spreads.

We estimate the model independently for each locality using Bayesian methods, obtaining 2.5 million Monte Carlo draws from locality-specific posterior distributions of each of the parameters $\{\gamma,\beta_0,\beta_1,\zeta,\sigma\}$. For our forecasts, we simulate paths for $\{\varepsilon_t\}$ for a sub-sample of the Monte Carlo draws.

2 Data

Data for the United States is taken from the European Centre for Disease Prevention and Control[1], which has data for the daily number of cases in different countries since December 31, 2019. The state level and Fifth District data are taken from The Atlantic's COVID Tracking Project[2] and Google's COVID Map[3], which consolidate data from state governments.

We use data up until and including April 20, 2020. Starting dates for the sample are chosen such that there are a sufficient number of cases for the model to be a plausible representation of the data. For the United States, we start the sample when the cumulative number of cases exceeds three hundred. For the Fifth District and individual states, we start the sample when the cumulative number of cases in the given locality exceeds fifteen.

[1] https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases
[2] https://covidtracking.com/
[3] https://news.google.com/covid19/map