

Restarting after COVID-19: A Data-driven Evaluation of Opening Scenarios

Ashwin Aravindakshan, Jörn Boehnke, Ehsan Gholami, Ashutosh Nayak

University of California, Davis

Supplementary Material

Model Configuration and Initialization

Following (1), the SEIR transmission model integrates the effect of human mobility on disease spread through a set of evolutionary equations. The model is given below:

$$\frac{dS_i}{dt} = -\frac{(1-\gamma^*sd_i)\beta S_i I_i^d}{N_i} - \frac{(1-\gamma^*sd_i)\mu\beta S_i I_i^u}{N_i} + \sum_v \left(\frac{\theta_v \sum_j M_{ij}^v S_j}{N_j - I_j^d} - \frac{\theta_v \sum_j M_{ji}^v S_i}{N_i - I_i^d} \right) \quad (1)$$

$$\frac{dE_i}{dt} = \frac{(1-\gamma^*sd_i)\beta S_i I_i^d}{N_i} + \frac{(1-\gamma^*sd_i)\mu\beta S_i I_i^u}{N_i} - \frac{E_i}{Z} + \sum_v \left(\frac{\theta_v \sum_j M_{ij}^v E_j}{N_j - I_j^d} - \frac{\theta_v \sum_j M_{ji}^v E_i}{N_i - I_i^d} \right) \quad (2)$$

$$\frac{dI_i^d}{dt} = \frac{\alpha_i E_i}{Z} - \frac{I_i^d}{D} \quad (3)$$

$$\frac{dI_i^u}{dt} = \frac{(1-\alpha_i)E_i}{Z} - \frac{I_i^u}{D} + \sum_v \left(\frac{\theta_v \sum_j M_{ij}^v I_j^u}{N_j - I_j^d} - \frac{\theta_v \sum_j M_{ji}^v I_i^u}{N_i - I_i^d} \right) \quad (4)$$

$$N_i = N_i + \sum_v (\theta_v \sum_j M_{ij}^v - \theta_v \sum_j M_{ji}^v) \quad (5)$$

where S_i , E_i , and N_i represent the susceptible, exposed, and the total population of state i , respectively. I_i^d represents the documented infected individuals which is the subset of the infected population that have symptoms severe enough to be diagnosed with the illness. I_i^u is the rest of the infected population known as the undocumented infected individuals. We consider values for SEIR metapopulation state variables on day t . For the rest of this paper, we omit the time index t that represents the time dependency of variables. β is the transmission rate of the disease from a

documented infected individual to a susceptible individual under normal population mobility. The transmission rate due to undocumented infected is captured by $\mu\beta$ with μ being the reduction coefficient ($\mu < 1$). Additionally, the variable sd_i ($sd_i \in [-1, 1]$) defines the daily change in social mobility (or degree of social distancing) in the state i , and the coefficient $(1 - \gamma * sd_i)$ is the decrease/increase in transmission rate due to changes in socializing and population mobility ($\gamma \leq 1$). The ratio of documented to total infected individuals in the state i is α_i , which varies among states with dissimilar population demographics based on age and gender (2). Z is the average incubation time, and D is the infection period. More precisely, D captures the effective period in which the infected individual moves out of the chain of disease transmission by perishing, recovering, or entering quarantine. The number of interstate travelers from state j to state i via transportation network v , is M_{ij}^v on a given day, with θ_v to fix the underreporting of transportation ($\theta_v \geq 1$). Two transportation networks represented by G and A account for ground and air mobility. The ground transportation $M_{ij,G}$ includes movement of individuals by four different sub-networks: $M_{ij}^G = \sum_g M_{ij}^g$. These sub-networks are: cars, trains, trucks and buses. Similarly, M_{ij}^A is the number of people traveling via flights. In this model, we assume documented infected patients do not travel between states, while the asymptomatic undocumented infected individuals have the ability to move from one state to another.

We estimate the parameters of this model using the procedure described in (1). The model parameters of the SEIR model are inferred via iterative filtering of stochastic ensemble Kalman filter (EnKF-IF). The EnKF is a Monte Carlo (MC) approximation of the Kalman filter. We specifically used the Ensemble Adjustment Kalman filter (EAKF) which is suitable for models with a high number of parameters. In this technique, in each iteration, a presumably Gaussian distributed ensemble of state vectors are adjusted to posterior distribution via Bayes rule. The state vector includes model parameters and metapopulation values. We use the maximum likelihood approach to determine the final values of the state vector in the algorithm. The daily documented cases act as observations in the model. We find that a few hundreds of ensembles are sufficient to accurately infer the model parameters.

We introduce a randomly drawn number for detecting delay of each documented case. This additional delay captures the latency period from the onset of symptoms to diagnosis and the time it takes an individual to become contagious from an initial exposure (2). The model randomly adds delay with Gamma-distributed values of shape $a = 1.78$ and average of $T = 6$ days. This distribution fits with the information from confirmed cases in China and South Korea (1,5,6). We examined Gamma distributions with various averages and a constant shape to capture the most accurate average delay time for estimation purposes.

Furthermore, we introduce another delay for the effect of social distancing on daily confirmed cases. There is an average of 6 days delay between the change in human mobility (sd_i) and the corresponding change in the number of daily documented infected cases across all states. This constant shift accounts for the time that the change in mobility starts showing an effect on disease transmission rate.

The configuration of model and initial ranges for parameters and values are represented in Table

1. The EAKF algorithm is not limited by parameter prior range and can move outside of prior range to find the optimum solution. We choose a suitable prior range for state vector initial values to facilitate the convergence while covering most of the possible values for parameters.

The prior range of μ covers a wide range of possible values [0, 1]- The prior range for α is set to include most of its possible values (0,1] with lower bound set to 0.5 to account for the high volume of Covid-19 testing in Germany. The prior range for β is set to cover a wide range of values for R_0 , i.e. [0, 12]. Prior range of Z and D are chosen to cover the known average incubation and infection period for Covid-19 (3, 5). Prior range of θ is set to capture most of the possible range of underreporting of transportation. The range limits the number of each state's travelers to its population (Figure 5,6,7 in the manuscript). The prior range for γ covers most of the possible range for the effect of human mobility and social distancing on transmission rate, i.e. (0, 1]. The lower bound allows for at least a 50% drop in transmission rate if social mobility is dropped 100%.

We use Feb 18, 2020 – Apr 20, 2020 as the time period for model inference. Feb 18, 2020 is the early stages of *COVID-19* epidemic in Germany with only two states with reported cases. The initial susceptible population S_i of each state is set to its initial total population. The initial documented cases, I_i^d , is set to reported cases on Feb 18, 2020. The exposed and undocumented infected initial values are set in the following way. Based on reported daily cases, there are three states that are early hubs of Covid-19 in Germany. Nordrhein-Westfalen (NW) is the pioneer state linked to large carnival events, followed by states Baden-Wurttemberg (BW) and Bayern (BY) via outbreak in Italy (2). The first large cluster of reported cases in Nordrhein-Westfalen was reported on Feb 28, 2020 with 25 cases. Accounting for the average doubling time of 6.4 days [95% CI: 5.8-7.1 days] for cases (4) and an average of 86% [95%CI: 83%, 90%] undocumented cases in early stages (1), we estimate to have [95% CI: 390-825] cases on the initial day of the model in Nordrhein-Westfalen. We set the initial value range for E_{NW} , I_{NW}^u to be [0-400] and uniformly select random values from this range. Similarly, we set the initial range of E_{BW} , I_{BW}^u , and E_{BY} , I_{BY}^u for Baden-Wurttemberg and Bayern to [0, 200]. Excluding the aforementioned states, the rest of the German states' initial values for E and I^u were drawn uniformly from [0, C_{max}] with C_{max} being the aggregated number of undocumented infected travelers from the three hub states Nordrhein-Westfalen, Baden-Wurttemberg, and Bayern on the day of Feb 18:

$$C_{i_{max}}^{t=0} = \sum_v \sum_{i \in \{BW, BY, NW\}} M_{ji}^v E_i / N_i \quad (6)$$

All ensembles' initial state vector values are drawn randomly via Latin hypercube sampling with uniform distribution from the initial ranges.

The histograms of inferred model parameters are shown in Figure S1. The model parameter values selected from the set of inferred parameters are depicted by red lines in Figure S1. Please note that parameters sets are complex and while one parameter might have a high inferred frequency in a set, another parameter in that set might not have the highest inferred count.

Observations of confirmed COVID-19 Cases

The daily number of confirmed COVID-19 cases in Germany have been collated from data released by the Robert Koch Institute (RKI) in Berlin, Germany. Figure S2 shows how COVID-19 spread across Germany over time. When individuals contract COVID-19, they may remain asymptomatic. During this period, they are active carriers of the virus. Symptomatic and asymptomatic individuals are identified through clinical testing. In Germany, a primary care physician's referral is required for COVID-19. More than 100 laboratories were contracted to conduct testing for the COVID-19 virus. The daily cases in RKI includes the number of confirmed positive cases.

The RKI published the daily case reports at 10:00 prior to March 1, 2020. Between March 1, 2020 and March 9, 2020, the RKI reported cases at 10:00 and at 15:00. Starting March 10, 2020, due to continuously rising case numbers, RKI switched to adopting the numbers that were electronically transmitted from testing centers across Germany. The new case numbers were published once at 15:00. Starting March 17, 2020, the daily cases are published at midnight for the previous day. To account for the discrepancy in reporting time, we used a constrained cubic spline interpolation method to obtain the number of cases prior to March 17, adjusted for midnight reporting. Daily cases using cubic spline interpolation is shown in Figure S3.

Community Mobility Data Trends

Google aggregates data from users' anonymized location history (for users who switched on the location history settings in their android mobile phones) to estimate foot traffic across six different location categories -- retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential (Figure S4). With the outbreak of COVID-19, Google released data for changes in foot traffic (in percent points) for the six location categories from February 15, 2020. These percentage changes in foot traffic are reported as community mobility trends in the reports. We use trends in retail and recreation to measure social distancing.

Mobility Data

COVID-19 started in Wuhan, China, and spread to Germany and other parts of the world through cross-border human movement. We use air and different types of ground transportation to accurately collect the movement data across different states in Germany. For international travel into Germany from other countries we consider traffic from 142 countries (including nine countries that share international borders with Germany) through ground and air transportation. Next, we discuss the movement data for different models.

Car Mobility Data and Truck Mobility Data

We collected detailed five-year highway traffic data from Jan 1, 2013 to Dec 31, 2018, provided by the German Bundesanstalt für Straßenwesen (Federal Institute for Roadways). The dataset contains the hourly count of vehicles passing through one of about 2,800 automatic counting stations along highways and state streets. The dataset contains geographical coordinates of the locations for these checkpoints. Sensors were used to identify the vehicles as cars, buses and trucks. To extrapolate the data to 2020, we construct a linear regression model to predict hourly traffic for Jan 1, 2020 to May 7, 2020. The model uses year, public holidays, day of the week, and state population as control variables. The linear regression model is shown in Equation 7, where $T_{ii'}(y)$ is the number of cars moving from state i to i' during year $y \in \{2013, 2014, \dots, 2020\}$. $h_i = 1$ if it was a state holiday in state i , and $w = 1$ if a given day was a weekend. pop_{iy} is the population of state i in year y and $y_0 = 2013$ is the origin year of model. We build a similar model for trucks to predict the number of trucks moving from Jan 1, 2020 to May 7, 2020.

$$T_{ii'}(y) = \beta^v + \beta^{vh}h_i + \beta^{vw}w + \beta^{vp}pop_{iy} + \beta^{vy}(y - y_0) \quad (7)$$

To adjust for changes in car movement due to COVID-19, we use the daily google mobility trends for workplaces. We multiply the projected car movement for 2020 with google mobility to obtain adjusted car movement for 2020. Projected car movement and truck movement for 16 states are shown in Figure 5(a) and Figure 5(b). As part of the essential services to keep the supply chain from breaking, there were no restrictions on the truck movement. Therefore, we do not adjust for changes in truck movement for the period of our study.

Train mobility data

We use Deutsche Bahn's public timetable to determine all major train routes in Germany. We first identify the 110 biggest cities in Germany and their respective states. We use these major cities to identify movement across states and neighboring countries. Each train station has two timetables (See Figure 6(a) and (b)) for that station – one table shows the arrival time of all the trains to that station (including the departure time from its previous stations) and the other table shows the departure time of all the trains from that station (including the arrival time of its previous stations). For each train, using both the schedules, we find a complete route. The number of passengers boarding a train for the next station is kept proportional to the sequence of the cities in the train route. We identified 14,712 trains. There are 33 types of trains (based on speed, distance travel and capacity). We assume the total number of passengers in a train to be 400. For ICE, THA, and TGV trains, we assume the number of passengers in a train to be 600.

Due to COVID-19 and state policies, several trains were canceled, and train movement declined. To account for these changes, we adjust the number of passengers using Google mobility data. Similar to the adjustments made to the car movement data, we adjust the number of passengers traveling by train after the outbreak of COVID-19 using Google mobility trends at transit stations.

Bus mobility data

We use travel search history provided by a large third-party European bus and train price comparison and booking company to estimate the number of passengers moving across cities (states) in Germany and passengers traveling to Germany from neighboring countries. The bus data contains the number of searches for a route (departure city to arrival city) aggregated by the day. For example, 33 people searched for buses from Frankfurt to Heidelberg on January 25, 2020. The data does not show the actual number of travelers in the bus, but we use this data as an indicator for bus movement across Germany and its neighboring countries. We only include connections that *arrive* at or *depart from* Germany. We define a bus route as the tuple [departure city, arrival city, date]. The dataset contains the history for 857,159 unique connections (aggregated by day) from December 1, 2019 to May 7, 2020. Of these, 191,356 routes had either their origin or destination in Germany. A subset of 116,706 routes originated and ended in Germany. We assume a capacity of 20 passengers in each bus to estimate the number of travelers. On March 16, 2020, all bus trips were halted in Germany.

Flight mobility data

We use flight transportation information from <https://opensky-network.org>. It is an open source platform containing historical information of all airborne flights. The database uses Automatic Dependent Surveillance – Broadcast (ADS-B) trajectories and maps it with airport International Civil Aviation Organization (ICAO) codes to identify the departure and arrival airport of a flight. The database also maintains the UNIX timestamp for each contact signal (trajectory recorded). We use the *last* UNIX timestamp of a flight to identify its date of arrival and departure. We only consider flights with arrival or departure airports in Germany. We use ICAO airport codes to identify the state for each airport. We ignore flights for which neither the departure nor the arrival airport can be established. The dataset has 187526 flights (with arrival or departure in Germany) for the period of December 1, 2019 to May 7, 2020. The variation in the number of flights over time is shown in Figure S5. We assume a capacity of 200 passengers for a domestic flight and 500 passengers for an international flight to determine the number of individuals on a flight.

Effect of Non-Pharmaceutical Interventions (NPI) on Social Distancing

To understand the effect of policy interventions on social distancing, we build a penalized linear regression model (Lasso regression). Regression coefficients for different policies in Lasso regression model are shown in Table S2. The penalty term makes the coefficients of insignificant (or non-contributing) variables zero. We adjust the parameter λ^p such that we could keep reducing the number of insignificant variables until model performance drops significantly. We use $\lambda^p = 0.25$ for our Lasso regression model. We observe that the coefficients for Border Closure is zero. Henceforth in our analysis of the SEIR model, we do not consider Border Closures.

We use different controls to isolate the effects of these NPIs. We model Google community mobility for retail and recreation to create a measure for social distancing. We also use Google

Trends, weather data and dissatisfaction as additional controls in the model to account for latent awareness levels in the population, the proclivity to leave one's shelter due to higher temperatures and Spring in Germany, and the distress felt due to confining oneself at home. We consider Google Trends data for the search term “*COVID-19 in Deutschland*”. Google Trends is an indicator of search interest of a topic over time. It calculates the proportion of all other searches at the same time and normalizes it to the range of 0 - 100. High Google Trend numbers indicate high interest for the topic during that time. As more cases were observed around the world and Germany, public interest in COVID-19 increased. Increased search is also an indicator for public awareness towards increasing social distancing. Google Trends data for all the sixteen states is shown in Figure S6. We use maximum temperature recorded in a day to account for increased public interest in going out as summer approaches. Finally, we model PTV (Propensity To Violate) on a day t using Equation 8 as an arctan function applied to the number of days since the Feb 18 (first day of our study period, represented by t_o). We scale the PTV such that we get the model with best r-square value. Numerator inside arctan function is used to shift the curve while the denominator is used to change the slope of the curve.

Lasso regression places a penalty on the sum of absolute value of coefficients as shown in Equation 9. We estimate daily community mobility C_i . We use a binary variable $x_{ip} = 1$ if policy p is active in state i . l_i is the Google Trends number for search term in state i , PTV is the dissatisfaction metric and $t_{max,i}$ is the daily maximum temperature in degree Celsius.

$$PTV = 30 \arctan((t - t_o - 50)/8) \quad (8)$$

$$C_i = \sum_{o=1}^7 \beta_o^p x_{io} + \beta^{pl} l_i + \beta^{pt} t_{max,i} + \beta^{pd} diss + \lambda^p \left(\left(\sum_o^7 |\beta_o^p| \right) + |\beta^{pl}| + |\beta^{pt}| + |\beta^{pd}| \right) \quad (9)$$

Predicted and Google Community mobility numbers C_i for different states in Germany is shown in Figure S7.

Generating Different What-If Scenarios

To understand the contribution of different policies in containing the spread of COVID-19, we consider different sets of scenarios for which policies were implemented and relaxed. State governments started introducing different policies around Mid-March. They started relaxing some of these policies on April 20, 2020 (Figure 3). We use data from Feb 18, 2020 to May 7, 2020. We also observe that the coefficients for the Border closure Policy is zero in Lasso regression). Therefore, we do not consider it when creating scenarios. We create test scenarios with— contact restriction order lifted, initial businesses opened, stay at home orders lifted, non-essential services opened, retail outlets opened, and all policies in place.

Figure 3 shows the timeline for policy introduction across different states in Germany. Some states did not introduce all the seven policies and in some cases introduced certain policies on varying dates, allowing for a quasi-experimental set-up to test the effect of policies on social distancing and spread of disease subsequently. Next, we create five counterfactual scenarios (one for each policy) to study what will happen if the states lift the focal policy.

Different states across Germany started relaxing some policies on Apr 20, 2020. To determine the effect of lifting a policy we examine policy relaxation in two scenarios: What would have happened if the policies were relaxed on April 21 or April 28, by easing one policy at a time to estimate the marginal effect of a policy. The week-long delay helps determine the increase in cases by relaxing a policy one week earlier.

The differential effects allow for rank-ordering the policies by order of their impacts on disease spread. If a policy was relaxed on April 21 (without changing other policies), the social distancing would decrease after April 21. Similarly, if a policy was relaxed on April 28 (without changing other policies), social distancing will start decreasing after April 28 (without changing policies from April 21 to April 27 and keeping it as it is). The social distancing under different scenarios for different states are shown in Figure S8 and Figure S9.

Effective Reproduction Number

The effective reproduction number R_e is the average number of new infected cases caused by a single infected case on a given day. The effective reproduction number R_e for the SEIR compartmental model can be calculated with the method introduced in (7) and later expanded in (8). The value of R_e changes by an increase or decrease in social mobility of both the infected and susceptible population. Other factors impacting the value of R_e include environmental conditions and the drop in susceptible population over time.

In the SEIR model, $X = [E, I^d, I^u]$ are considered the infected compartments of the model. R_e can be calculated as the leading Eigen value of the next generation matrix (NGM), $K = FV^{-1}$. $F = \frac{\partial \mathcal{F}_i(x)}{\partial X_j}$ is the Jacobian matrix of the rate of new infections in infected compartments and $V = \frac{\partial \mathcal{V}_i(x)}{\partial X_j}$ is the Jacobian matrix of the rate of transitions between infected compartments. Using equations (1-4), we have:

$$F = \begin{bmatrix} \frac{(1-\gamma*sd)\beta SI^d}{N} + \frac{(1-\gamma*sd)\mu\beta SI^u}{N} & & \\ & 0 & \\ & & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \frac{E}{Z} & & \\ \frac{I^d}{D} - \frac{\alpha E}{Z} & & \\ \frac{I^u}{D} - \frac{(1-\alpha)E}{Z} & & \end{bmatrix} \quad (10)$$

$$F = \begin{bmatrix} 0 & \frac{(1-\gamma*sd)\beta S}{N} & \frac{(1-\gamma*sd)\mu\beta S}{N} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \frac{1}{Z} & 0 & 0 \\ -\frac{\alpha}{Z} & \frac{1}{D} & 0 \\ -\frac{1-\alpha}{Z} & 0 & \frac{1}{D} \end{bmatrix} \quad (11)$$

And calculating R_e :

$$R_e = ((1 - \gamma * sd)\alpha\beta D + (1 - \gamma * sd)(1 - \alpha)\mu\beta D) \frac{S}{N} \quad (12)$$

Figure S10 shows the effective reproduction number over time in different states of Germany using inferred parameters values. These values are in-line with the reported values for Germany 0.79 [95%CI: 0.66 – 0.90] using nowcasting approach as a moving 4-day average by Robert Koch Institute (2). As depicted in this figure, the reproduction number at the beginning of the epidemic in Germany is above 1 in all states. Once the social distancing policies are in effect, this rate drops to below 1 in all states. When $R_e > 1$, the disease starts to spread throughout the population. Equation (9) shows that R_e also decreases with a drop in the susceptible fraction of population. A large proportion of population must transition from susceptible to infected, immune, or dead for this factor to be considerable.

To understand the effects of social distancing on R_e , we simulate the effect of lifting different policies on the value of R_e in Figures S11 and S12. Different social distancing policies are simulated to be removed on April 21, 2020 and April 28, 2020. As illustrated in this figure, removing some policies would increase R_e to above 1 in some states. For example, in large states such as Baden-Wurttemberg (BW) and Bayern (BY), R_e will rise above 1 if the contact restriction order is lifted or initial businesses are opened. Certain policies must be kept in effect until the susceptible proportion of population drops significantly for R_e to remain below 1.

The dependency of R_e to different combinations of parameters α , γ , and sd is depicted in Figure S13. These figures represent the basic reproduction number where all population is assumed to be susceptible. Figure S13 (a) represents the changes in R_e with respect to possible values of parameters α and γ . High α and low γ combinations result in the highest R_e value. Figure S13 (b) shows that R_e monotonically increases with decrease in sd and increase in α . Figure S13 (c) shows that R_e will increase in combinations of high γ and low sd , while its value drops with high γ and high sd combinations. As γ decreases, the dependency of R_e to sd also declines.

Figures and Tables

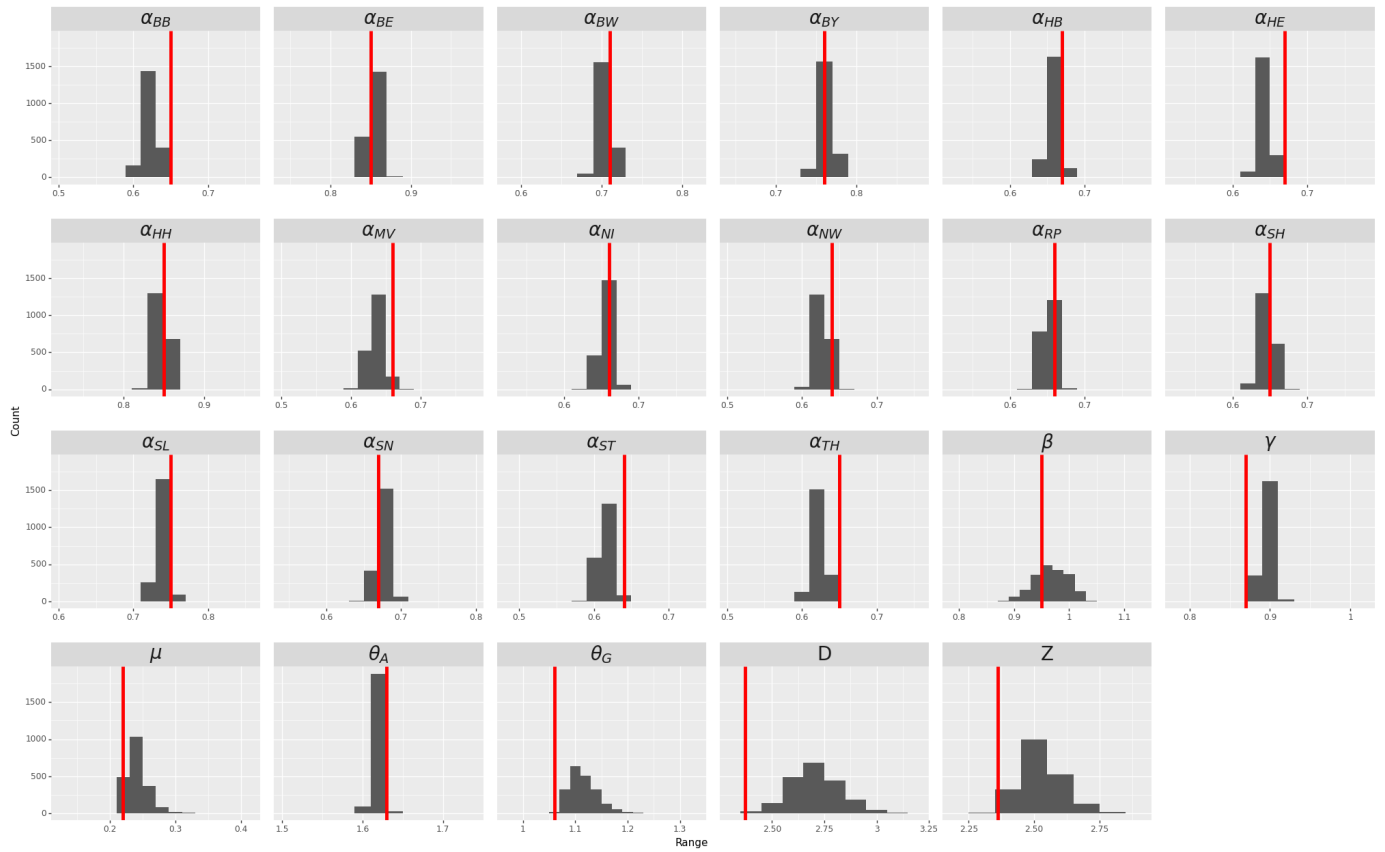


Figure S1. Model parameter inference. Initial day for the model is set to Feb 18, 2020. Histograms show distribution of model parameters inference over 2000 runs, each with 500 ensembles. Each run provides a set of parameters as a group that is optimized together and should be used as one. Red lines show the set of inferred parameters of a randomly selected run used for estimation.

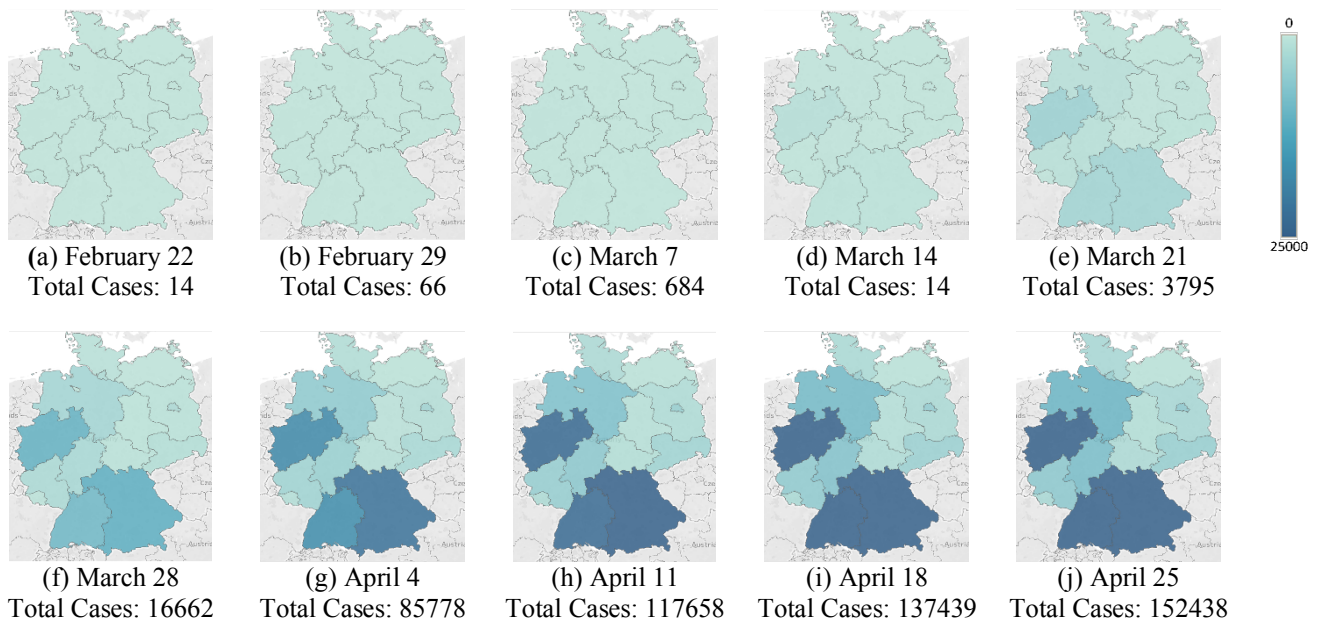


Figure S2. Spread of COVID-19 across Germany over time

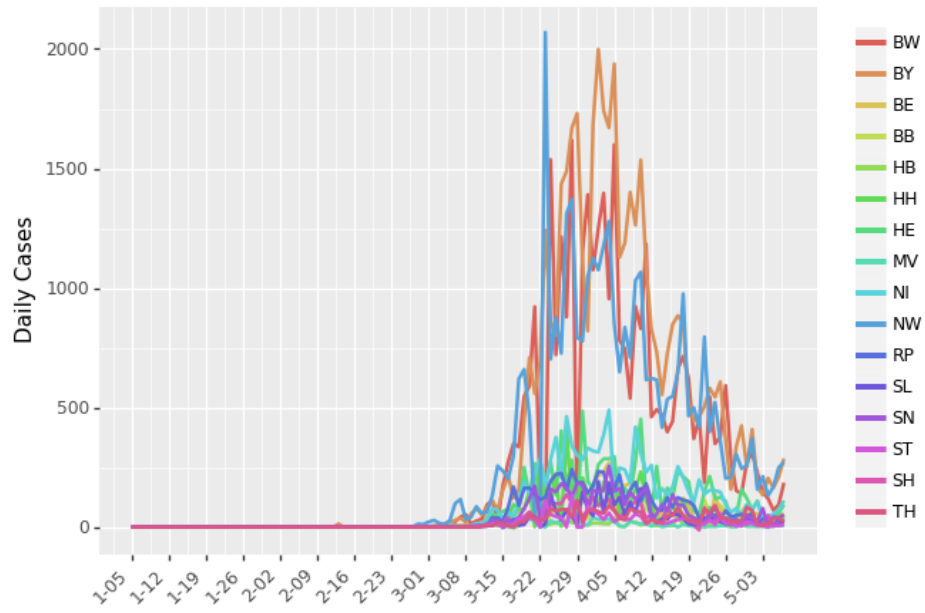


Figure S3. Daily New Cases from RKI (Adjusted for Midnight Reporting)

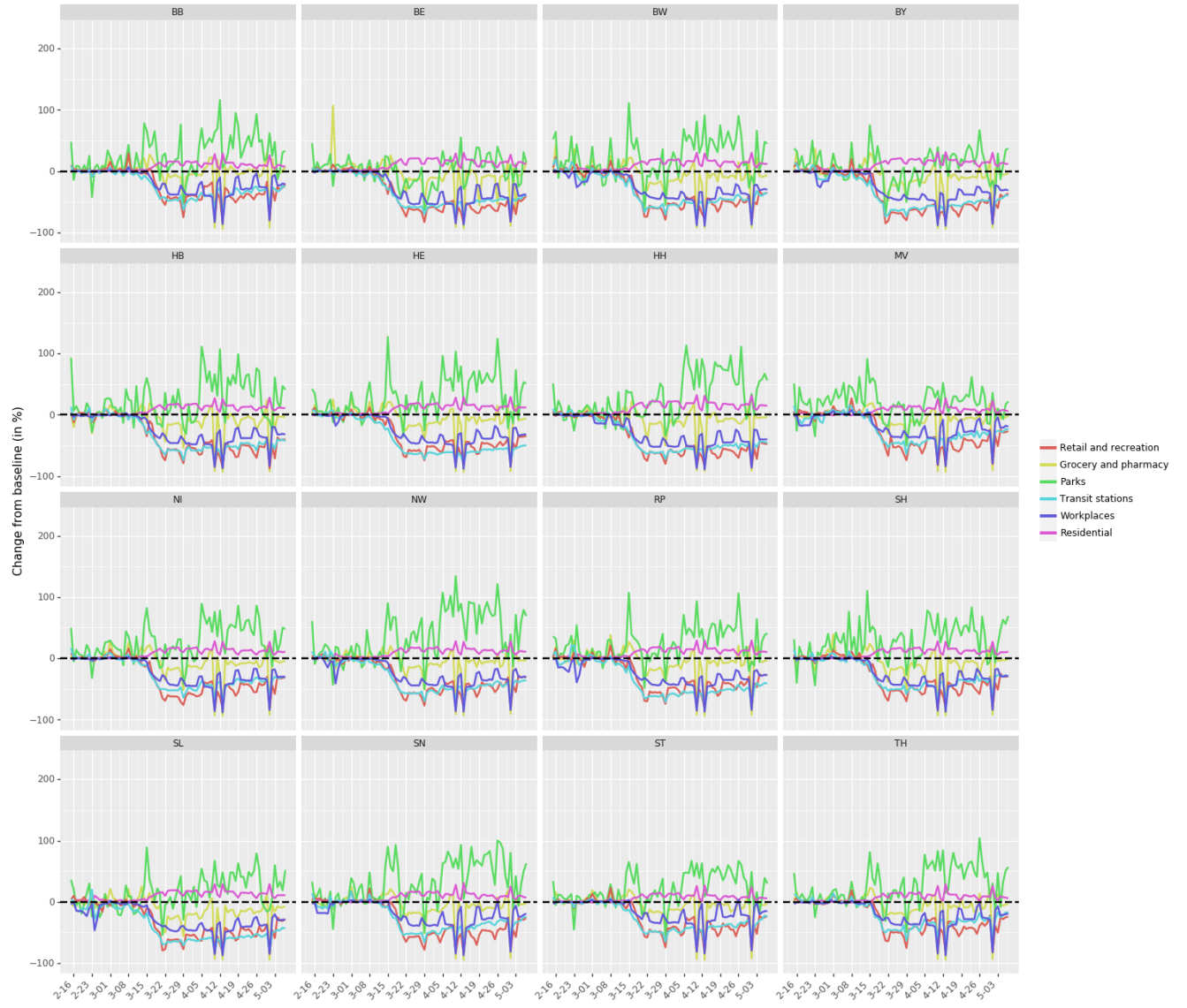


Figure S4. Google Community Mobility

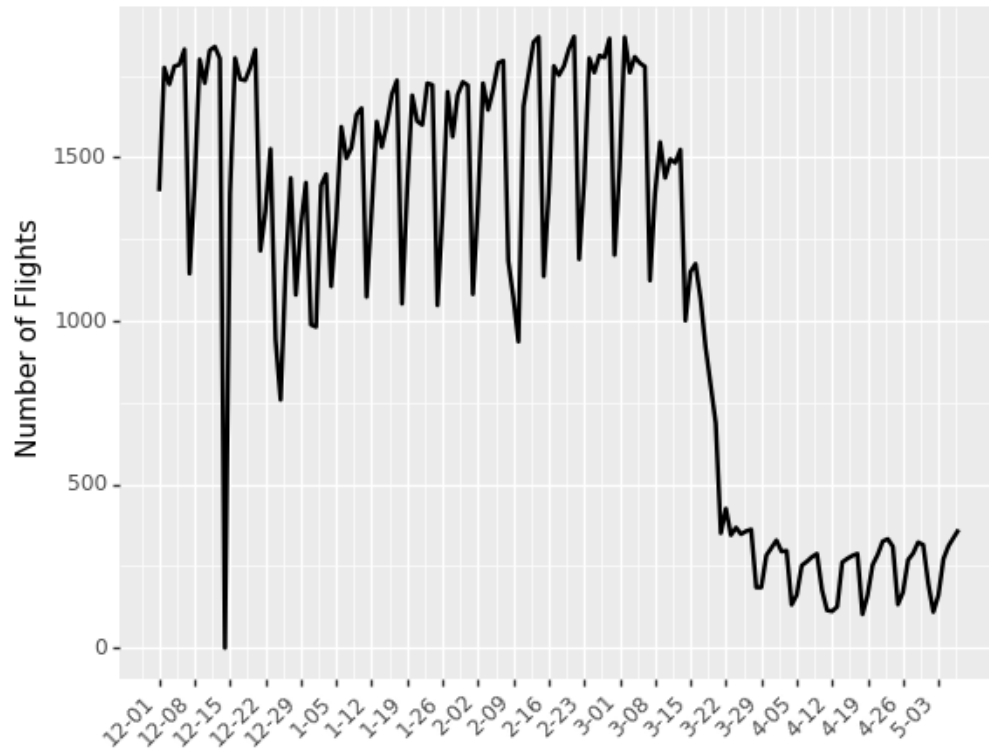


Figure S5. Daily Flights Arriving to Different Airports in Germany

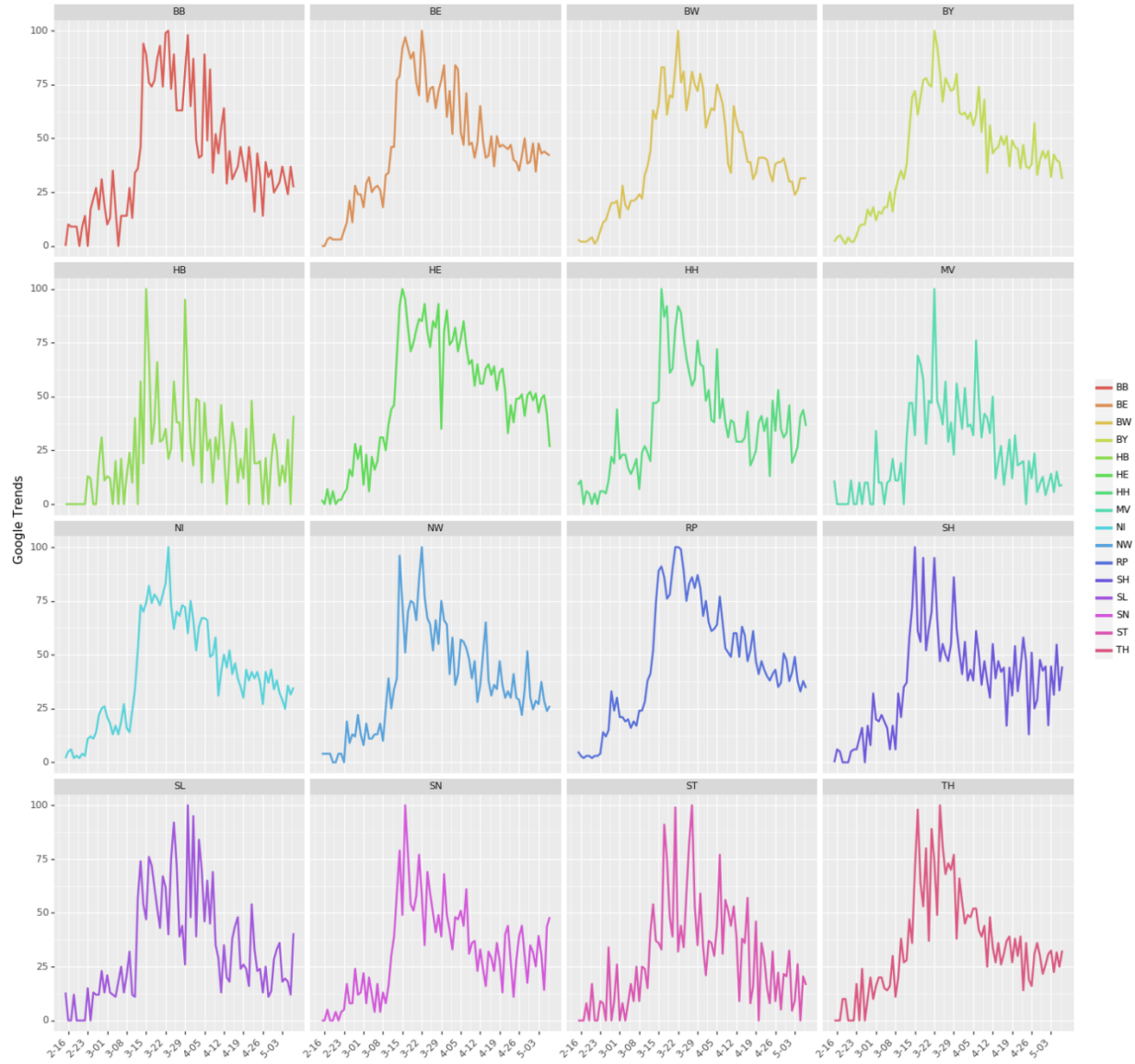


Figure S6. Google Trends Data for Different States in Germany

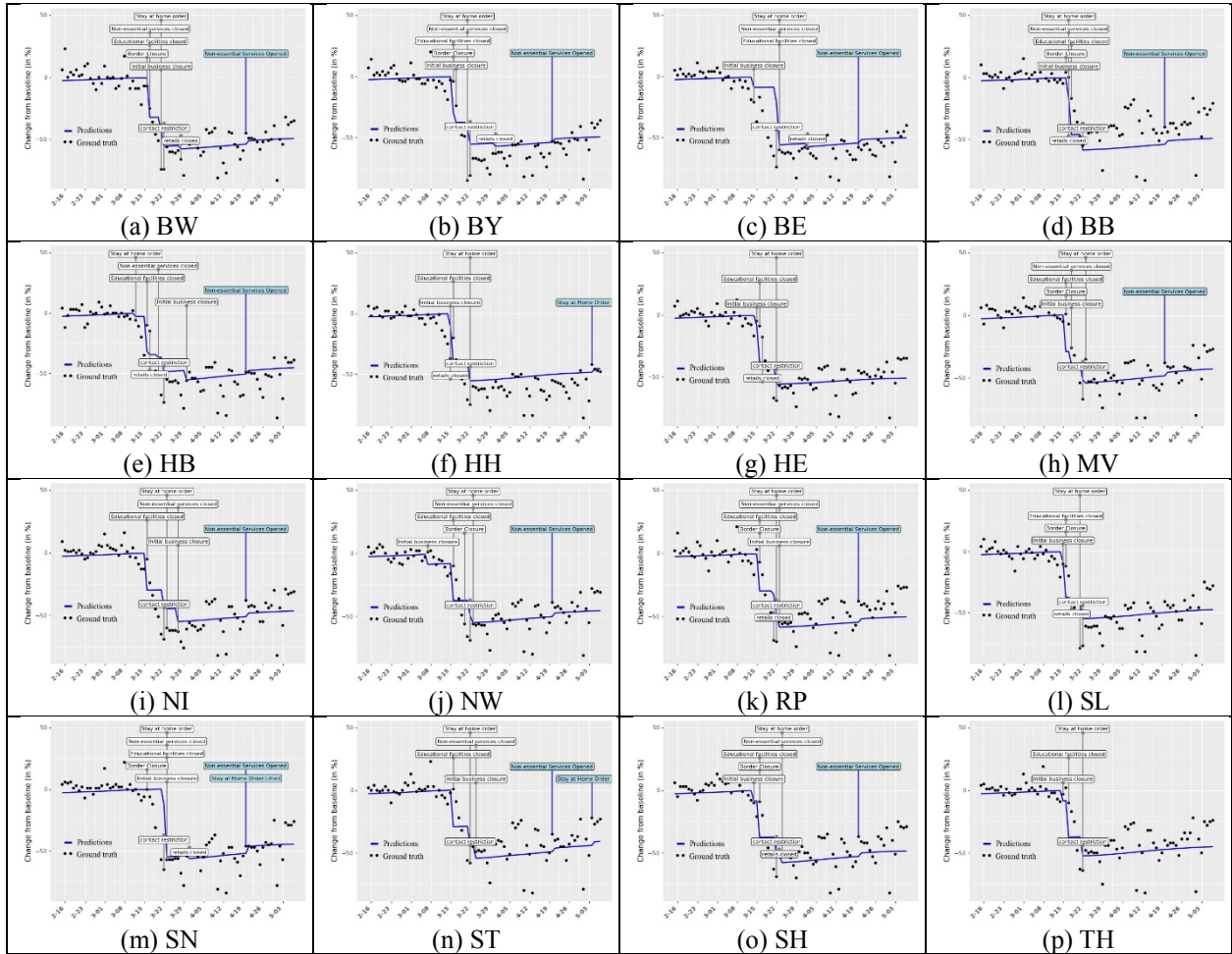


Figure S7. Predicting Social Distancing for Different States in Germany

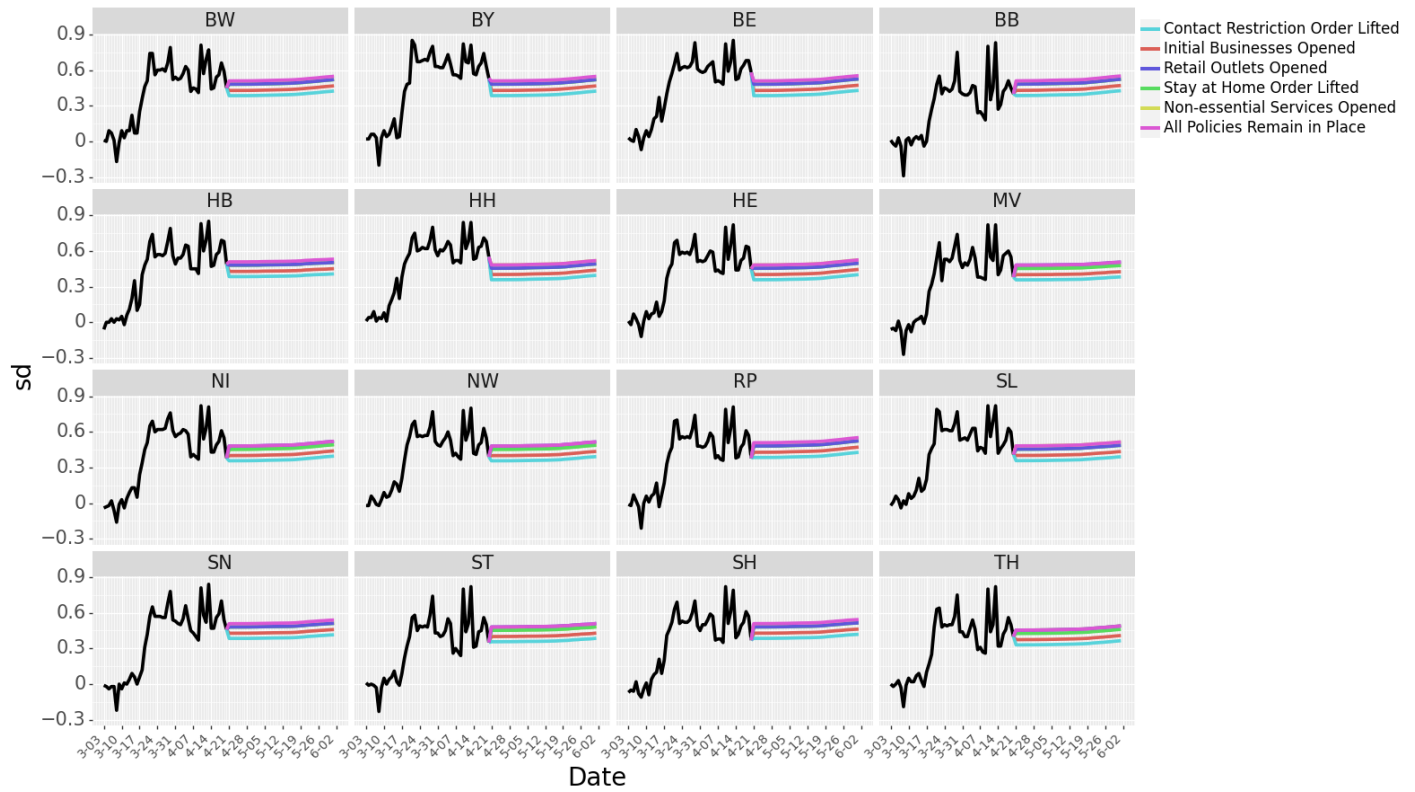


Figure S8. Scenario 2: Social distancing sd_i when policies are relaxed on April 21, 2020

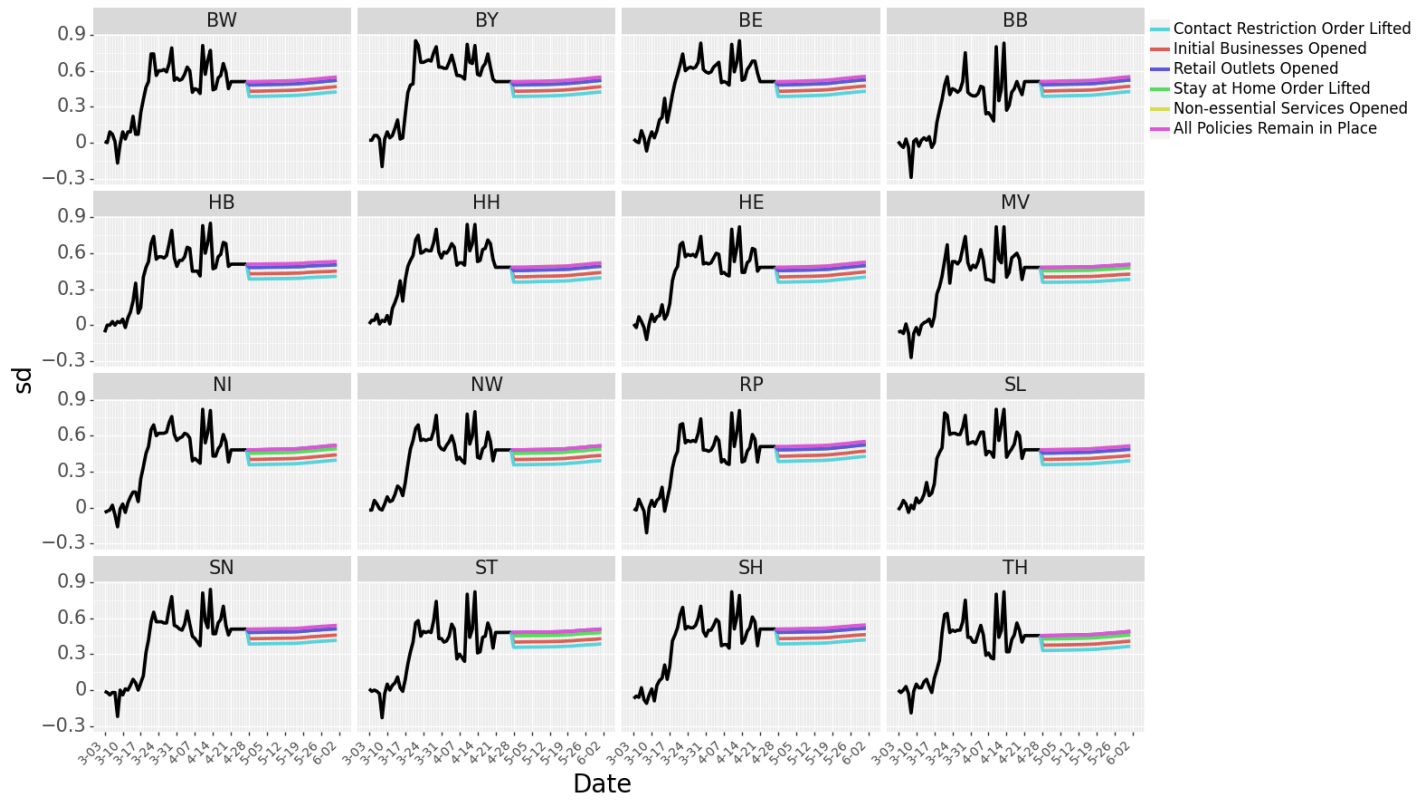


Figure S9. Scenario 2: Social distancing sd_t when policies are relaxed on April 28, 2020

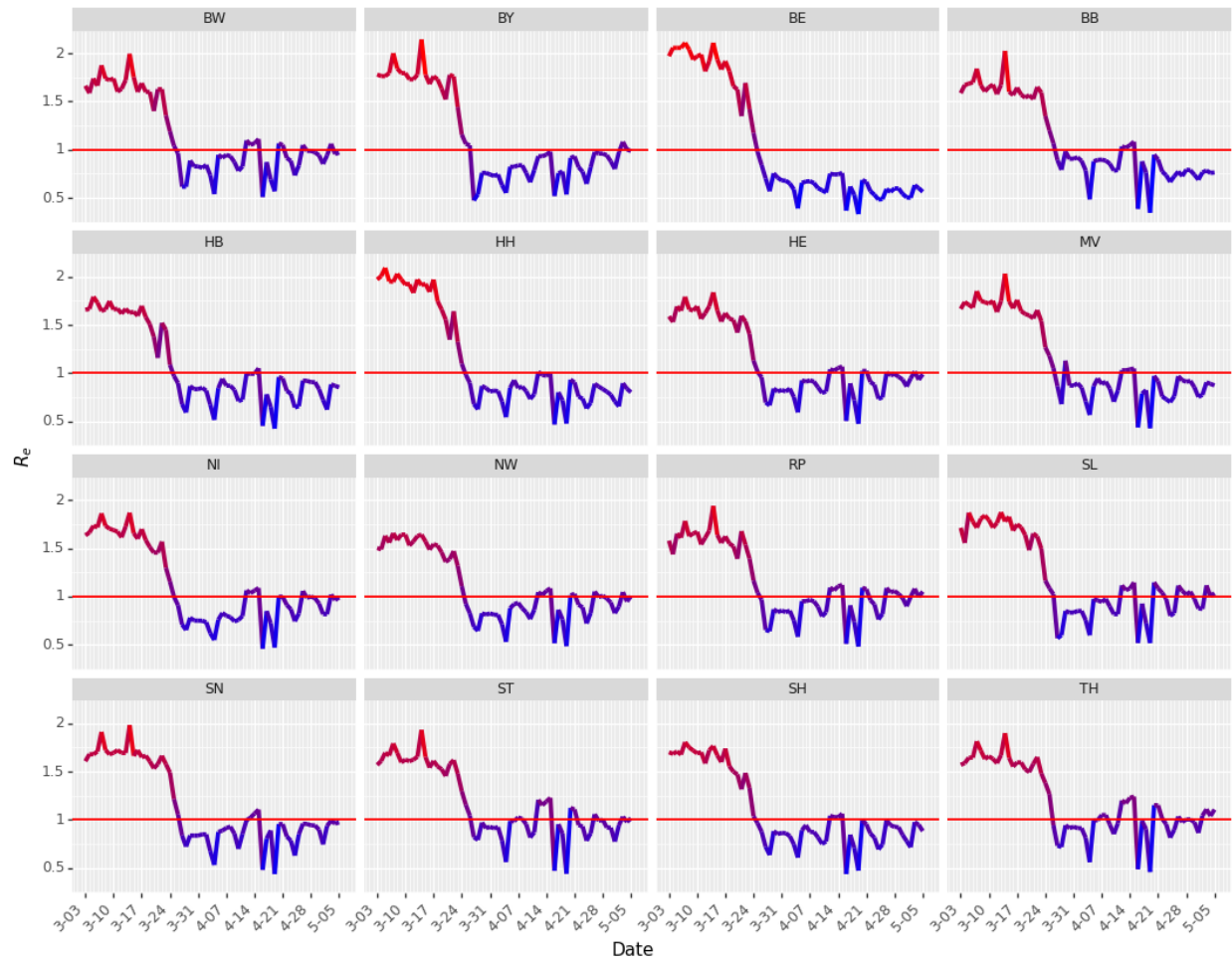


Figure S10. Effective reproduction number. The values of R_e drops from above the red line 1 to below 1 after policies are in effect.

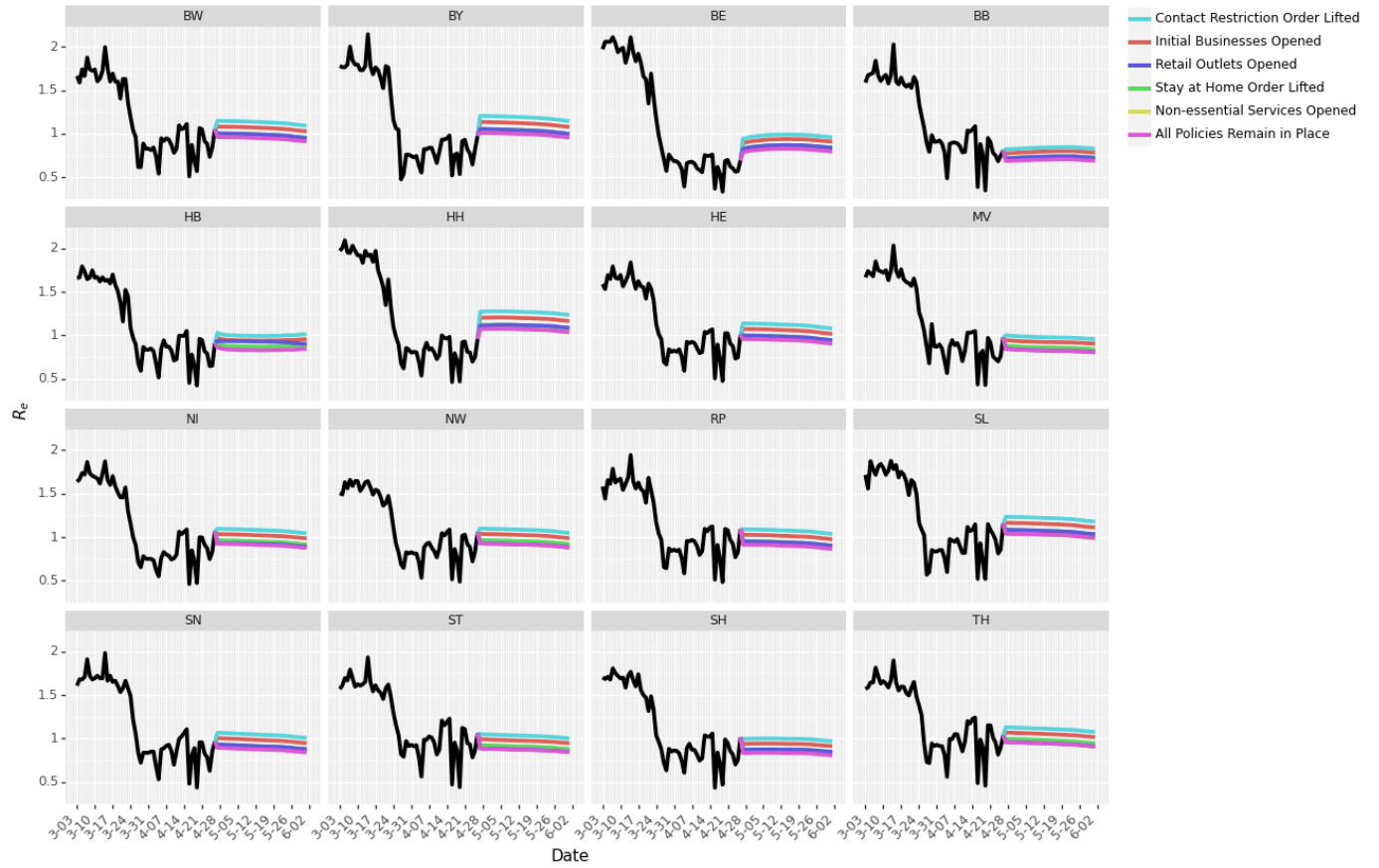


Figure S11. Scenario 1: The effective reproduction number R_e when policies are relaxed on April 21, 2020. Under some policies, the R_e rises to above 1.

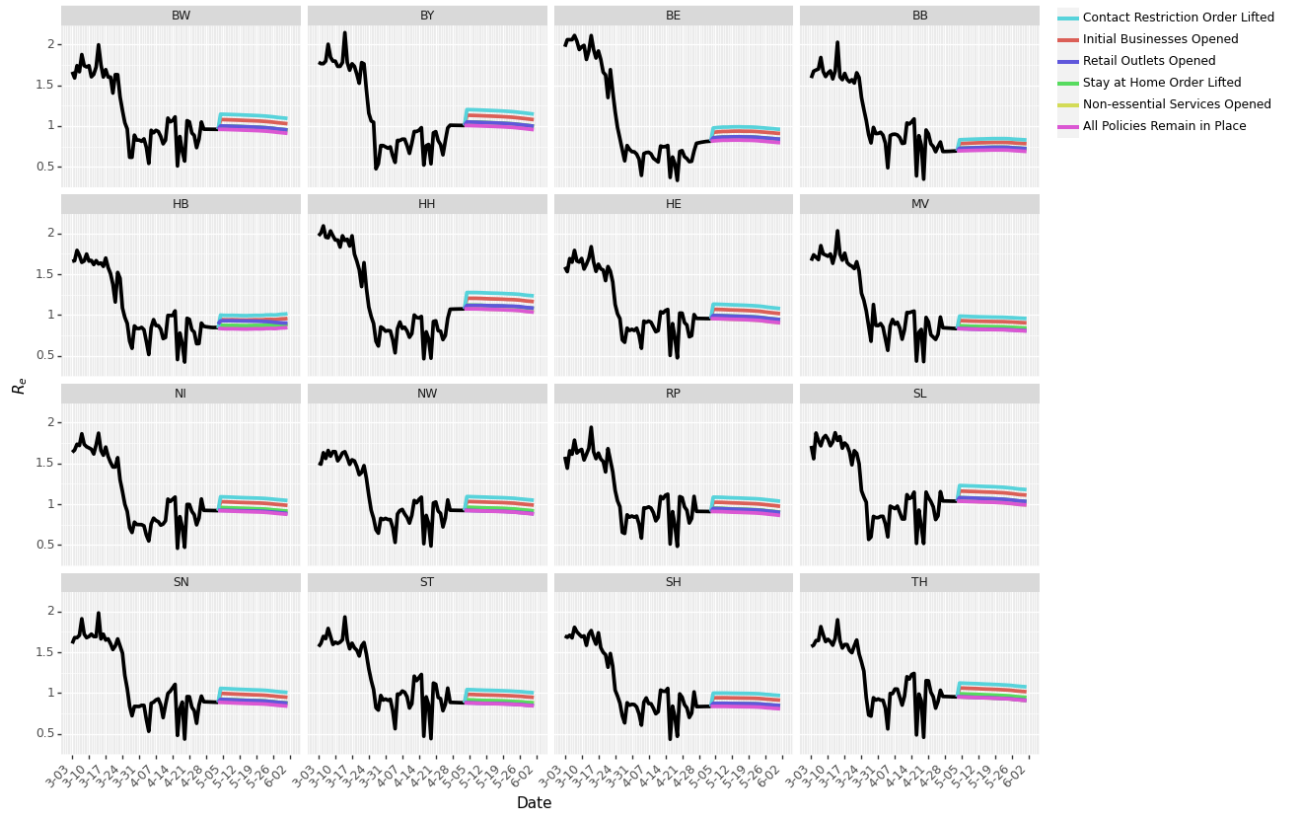


Figure S12. Scenario 2: The effective reproduction number R_e when policies are relaxed on April 28, 2020. Under some policies, the R_e rises to above 1.

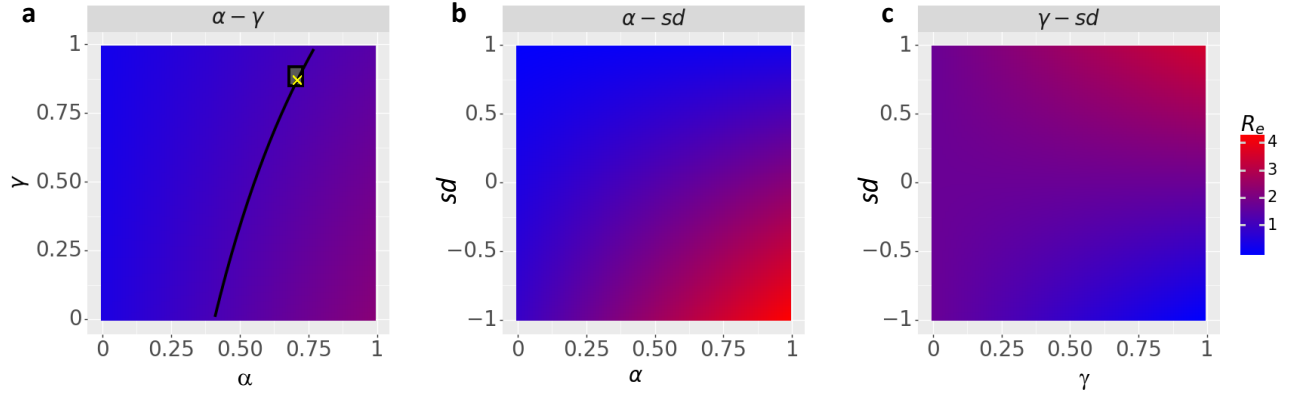


Figure S13. The dependency of R_e on model parameters α and γ and social mobility sd . All other variables are held constant with values represented in Table S1 for state BW. $\alpha - \gamma$: The impact of α and γ on R_e . The value of sd is set to $sd = 0.35$. Black curve represents the constant $R_e = 1.21$. The box represents the interval of inferred parameters that have the maximum likelihood and the yellow 'X' mark shows the parameter combination used for estimation ($\alpha = 0.71$, $\gamma = 0.89$). $\alpha - sd$: The impact of α and social mobility sd on R_e . Note that sd is not a model parameter. $\gamma - sd$: The impact of γ and social mobility sd on R_e .

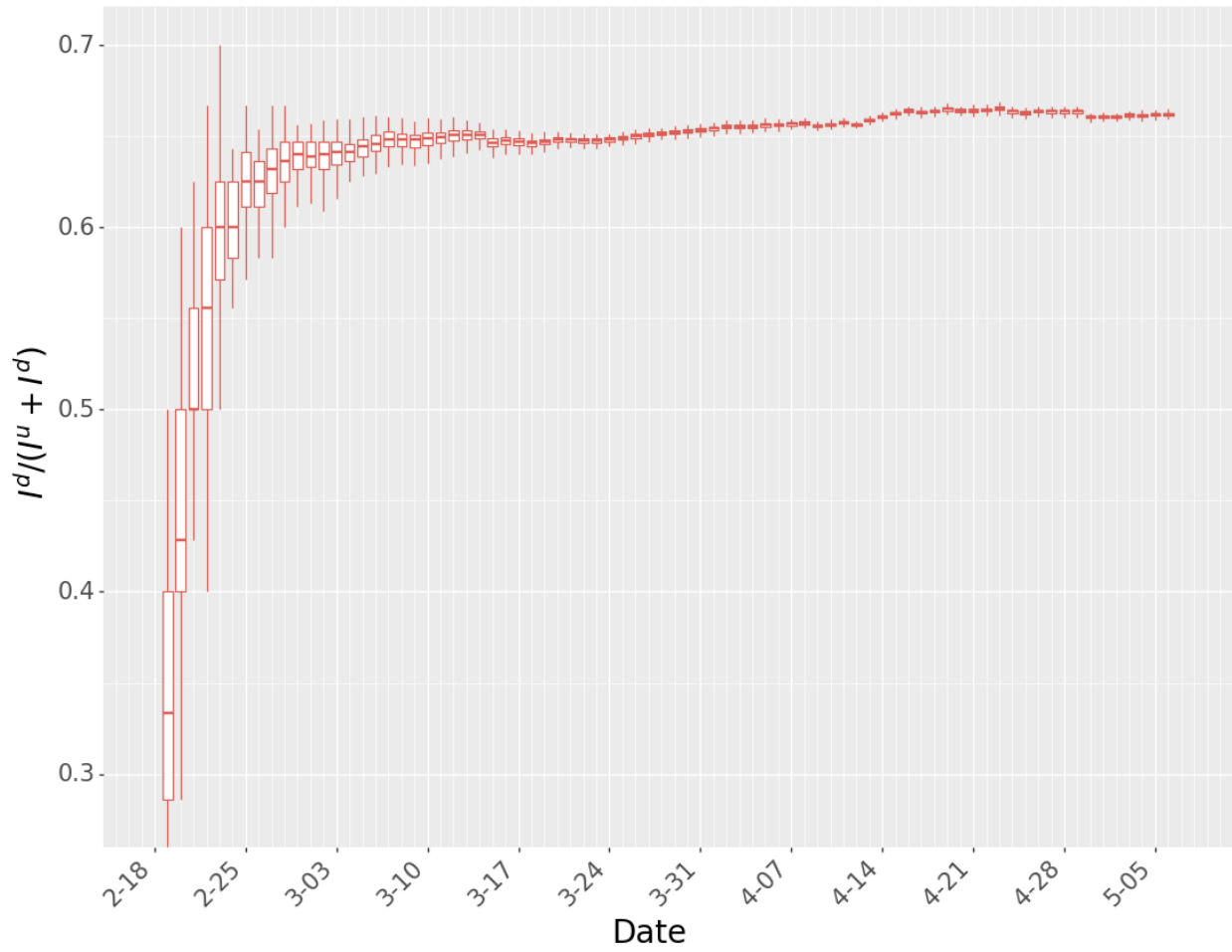


Figure S14. The ratio of documented infected cases to total daily infected cases for Germany.

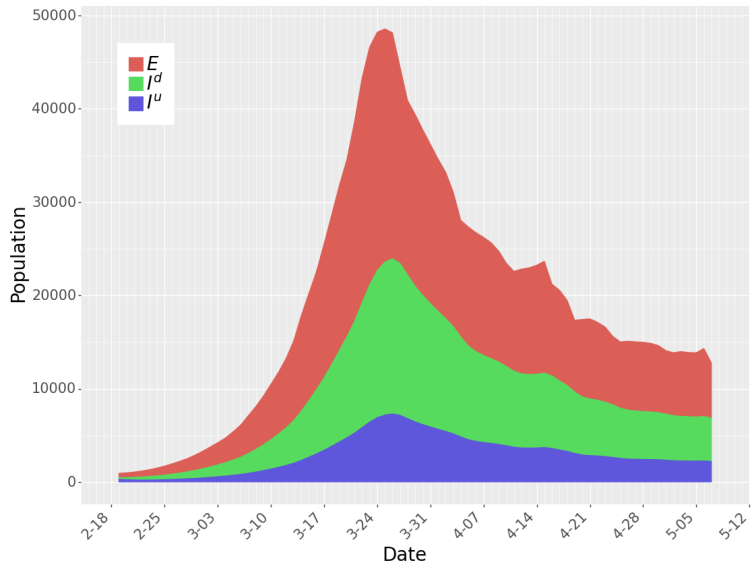


Figure S15. The population count of exposed, documented infected and undocumented infected cases for Germany.

Parameters	Initial Range	Inferred Value [95% Confidence Intervals]
β	0.8-1.2	0.95 [0.92 – 1.01]
μ	0.2-1.0	0.22 [0.22 – 0.27]
θ_G	1.0-1.8	1.08 [1.08 – 1.16]
θ_A	1.0-1.8	1.63 [1.61 – 1.63]
Z	2.0-5.0	2.41 [2.41 – 2.67]
γ	0.6-1.0	0.89 [0.89 – 0.91]
D	2.0-5.0	2.52 [2.52 – 2.91]
α_i	0.5-1.0	(Per State Estimate Follows)
α_{BW}		0.71 [0.69 – 0.72]
α_{BY}		0.76 [0.75 – 0.77]
α_{BE}		0.85 [0.84 – 0.87]
α_{BB}		0.64 [0.61 – 0.64]
α_{HB}		0.67 [0.65 – 0.67]
α_{HH}		0.85 [0.84 – 0.86]
α_{HE}		0.65 [0.63 – 0.65]
α_{MV}		0.65 [0.62 – 0.65]
α_{NI}		0.66 [0.64 – 0.67]
α_{NW}		0.64 [0.61 – 0.64]
α_{RP}		0.66 [0.64 – 0.66]
α_{SL}		0.75 [0.73 – 0.75]
α_{SN}		0.67 [0.66 – 0.69]
α_{ST}		0.63 [0.6 – 0.63]
α_{SH}		0.65 [0.63 – 0.66]
α_{TH}		0.64 [0.61 – 0.64]

Table S1. Model parameters inference and prior range. The initial range is the prior range for parameters in EAKF algorithm. The selected set of inferred values along with 95% confidence interval is reported.

Policy	Description	Coefficients
Border Closure	Closure of International Borders. Border between Germany and Netherlands were never closed.	β_1^p
Initial Business Closure	Closure of dance events, trade fairs, exhibitions, special markets, arcades, casinos, betting shops and similar companies; amusement places; prostitution institutions, concert halls, fairs, leisure and animal parks, providers of leisure activities and similar facilities	β_2^p
Educational Facilities Closed	Closure of All Educational Facilities, K-12 and University	β_3^p
Non-Essential Services Closed	Closure of bars, clubs, cinemas, theatres, museums, florists, garages, fashion stores and churches.	β_4^p
Stay at Home Order	Residents asked to shelter-in-place	β_5^p
Contact Restriction	Gatherings limited to no more than 5 people, unless family	β_6^p
Retail Outlets Closed	Does not apply to retail for food, weekly markets, pick-up and delivery services, beverage markets, pharmacies, medical supply stores, drug stores, petrol stations, banks and savings banks, post offices, hairdressers, dry cleaners, Laundromats, newspaper sales, DIY and garden centers, pet supplies and wholesale, craftsmen and craft like trades.	β_7^p

Table S2. Description of different NPIs

References

- (1) Li R, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*. 2020. DOI:10.1126/science.abb3221
- (2) Robert Koch Institute “Current Situation Report of the RKI to COVID-19”
https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Gesamt.html Accessed: May 9, 2020
- (3) He, X., Lau, E.H.Y., Wu, P. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* **26**, 672–675 (2020). <https://doi.org/10.1038/s41591-020-0869-5>
- (4) J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet* **395**, 689–697 (2020). DOI:10.1016/S0140-6736(20)30260-9 Medline
- (5) Lauer, Stephen A. et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9), 577-582 (2020). doi: 10.7326/M20-0504
- (6) Li, Qun et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*, 382, 1199-1207 (2020) doi: 10.1056/NEJMoa2001316.
- (7) Diekmann, O., Heesterbeek, J.A.P. & Metz, J.A.J. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382 (1990). <https://doi.org/10.1007/BF00178324>
- (8) Van Den Driessche P., Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*. 2002;180:29–48.