# Overview of DELPHI Model V3 - COVIDAnalytics

Michael Lingzhi Li[1], Hamza Tazi Bouardi[1], Omar Skali Lami[1], Nikolaos Trichakis[1], Thomas Trikalinos[2], Mohammad Fazel Zarandi[1] and Dimitris Bertsimas[1]

{mlli, htazi, oskali, nitric, fazel, dbertsim}@mit.edu, {thomas_trikalinos}@brown.edu
[1]Massachusetts Institute of Technology, [2]Brown University

July 2020

## 1 Introduction to the Model

The model underlying our predictions is DELPHI (Differential Equations Leads to Predictions of Hospitalizations and Infections). DELPHI is a compartmental model that is based on the widely successful SEIR model, but with many additions to account for realistic effects. In particular there are two important effects that we consider:

- **Underdetection**: In any pandemic, a lot of cases go undetected due to many factors (failure to record, unable to test, mistaken for other disease, etc). This is an important factor that if not appropriately accounted for, would underestimate the real reach and spread of the epidemic.
- **Governmental Response**: No epidemic exists in a world where it is allowed to spread completely freely. As the epidemic spreads, governments start to respond by enacting policies designed to limit the spread of the virus, and we explicitly design a framework to take such policies into account in the model.

The model separates people into 11 possible states of being in the epidemic:

- **Susceptible (S)**: The general populace who have not been infected.
- **Exposed (E)**: People who are currently infected, but are not contagious and within the incubation period.
- **Infected (I)**: People who are currently infected and contagious.
- **Undetected (AR) & (AD)**: People who are infected, and self-quarantined themselves at home due to the effects of the disease, but was not confirmed due to lack of testing. Here, we model it in a way that some of these people recover (AR) and some of these die (AD).
- **Detected, Hospitalized (DHR) & (DHD)**: People who are infected, confirmed, and hospitalized. Again, we model it in two separate states: some of these people recover (DHR) and some of these die (DHD).
- **Detected, Quarantine (DQR) & (DQD)**: People who are infected, confirmed, and home-quarantined rather than hospitalized. Similar as before, we have two states: (DQR) for those that recover, and (DQD) for those that die.
- **Recovered (R)**: People who have recovered from the disease (and assumed to be immune).
- **Death (D)**: People who have perished from the disease.

The separation of recovery and death states in the detection phase (including AR/AD, DQR/DQD, DHR/DHD) is required so that recovery and deaths can be tuned separately.

The interactions between the different states are summarized in the picture below, where the arrows indicate possible flow between the states:

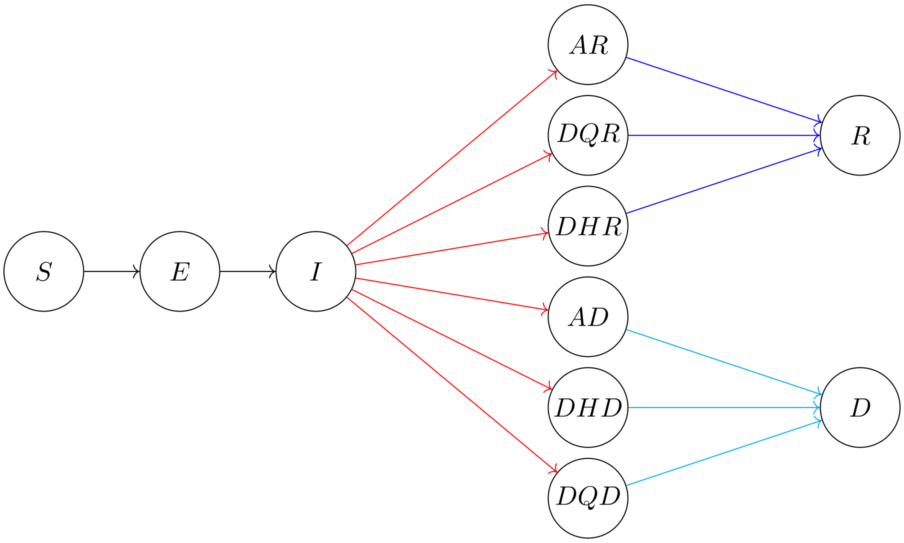

Figure 1: Simplified Flow Diagram of DELPHI (V2)

## 2 Detailed Model Formulation

In addition to main functional states, we also introduce helper states to help us calculate a few useful quantities. This includes Total Hospitalized (TH), Total Detected Deaths (DD) and Total Detected Cases (DT). The mathematical formulation of the model, along with these helper states, is as followed:

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\alpha\gamma(t)S(t)I(t)\\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \alpha\gamma(t)S(t)I(t) - r_iE(t)\\
\frac{\mathrm{d}I}{\mathrm{d}t} &= r_iE(t) - r_dI(t)\\
\frac{\mathrm{d}AR}{\mathrm{d}t} &= r_d(1-p_{dth}(t))(1-p_d)I(t) - r_{ri}AR(t)\\
\frac{\mathrm{d}DHR}{\mathrm{d}t} &= r_d(1-p_{dth}(t))p_dp_hI(t) - r_{rh}DHR(t)\\
\frac{\mathrm{d}DQR}{\mathrm{d}t} &= r_d(1-p_{dth}(t))p_d(1-p_h)I(t) - r_{ri}DQR(t)\\
\frac{\mathrm{d}AD}{\mathrm{d}t} &= r_dp_{dth}(t)(1-p_d)I(t) - r_{dth}AD(t)\\
\frac{\mathrm{d}DHD}{\mathrm{d}t} &= r_dp_{dth}(t)p_dp_hI(t) - r_{dth}DHD(t)\\
\frac{\mathrm{d}DQD}{\mathrm{d}t} &= r_dp_{dth}(t)p_d(1-p_h)I(t) - r_{dth}DQD(t)\\
\frac{\mathrm{d}TH}{\mathrm{d}t} &= r_dp_dp_hI(t)\\
\frac{\mathrm{d}DD}{\mathrm{d}t} &= r_{dth}(DHD(t) + DQD(t))\\
\frac{\mathrm{d}DT}{\mathrm{d}t} &= r_dp_dI(t)\\
\frac{\mathrm{d}R}{\mathrm{d}t} &= r_{ri}(AR(t) + DQR(t)) + r_{rh}DHR(t)\\
\frac{\mathrm{d}D}{\mathrm{d}t} &= r_{dth}(AD(t) + DQD(t) + DHD(t))
\end{aligned}
$$

We define each of the parameters in the previous equations below; in red are the parameters that are being fitted against historical data for each country/province, and in black the parameters considered to be constant for each country:

- $\alpha$ is the infection rate and is considered to be constant across all countries.
- $\gamma(t)$ measures the government response and is defined as:

  $$\gamma(t) = \frac{2}{\pi}\arctan\left(\frac{-b(t-a)}{20}\right) + 1 + j_0 \exp\left(-\frac{(t-t_{jump})^2}{2\sigma^2}\right)$$

  The details for choosing this function is explained in the sub-section below.

- $r_d$ is the rate of detection. This equals to $\frac{\log 2}{T_d}$ where $T_d$ is the median time to detection (assumed to be 2 days).
- $r_i$ is the rate of infection leaving incubation phase. This equals to $\frac{\log 2}{T_i}$ where $T_i$ is the median time to leave incubation (assumed to 5 days).
- $r_{ri}$ is the rate of recovery not under hospitalization. This equals to $\frac{\log 2}{T_{ri}}$ where $T_{ri}$ is the median time to recovery not under hospitalization (assumed to be 10 days).
- $r_{rh}$ is the rate of recovery under hospitalization. This equals to $\frac{\log 2}{T_{rh}}$ where $T_{rh}$ is the median time to recovery under hospitalization (assumed to be 15 days).
- $r_{dth}$ is the rate of death. This equals to $\frac{\log 2}{T_{dth}}$ where $T_{dth}$ is the time till death for dying patients.

- $p_{dth}(t)$ measures the mortality percentage over time. It is set to be a declining function due to the improved ability to detect milder cases and increased standard of care for COVID-19 patients. The functional form is defined as:

  $$p_{dth}(t) = (p_{dth0} - \underline{p_{dth}})\left(\frac{2}{\pi}\arctan\left(-\frac{t}{20}\cdot r_{ddec}\right) + 1\right) + \underline{p_{dth}}$$

  Where $p_{dth0}$ is the initial mortality percentage, $\underline{p_{dth}}$ is the lower bound in mortality percentage assuming perfect detection and perfect treatment, and $r_{ddec}$ is the rate of decay of mortality percentage.

- $p_d$ is the percentage of infection cases detected. This percentage is constant and is set to 0.2.
- $p_h$ is the percentage of detected cases hospitalized. This percentage is also constant.

Therefore, in total, we fit 9 parameters on which the equations depend directly, as well as two other "internal" parameters (called $k_1$ and $k_2$ in our implementation) used for the initial conditions, which brings the total to **6 parameters per province**. Now let us explain how we model the Societal-Governmental response for all countries in the world first, and then more specifically for the United States to take into account the possible policies for re-opening states.

## 3 Modeling Societal-Governmental Response

The rate of infection is never constant in an epidemic. As governments start responding to an epidemic, the rate of infection would start decreasing due to the measures being put in place. We have decided to model the response by multiplying an initial infection rate with an arctan curve (to model the initial three phases of governmental intervention) along with an exponential jump correction to model the resurgence in cases in many places.

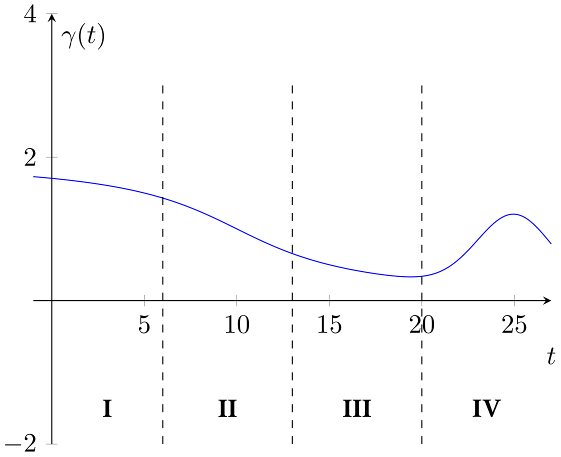

Figure 2: Illustration with $\gamma(t) = \frac{2}{\pi}\arctan(-\frac{4(t-10)}{20}) + 1 + \exp(-\frac{(x-25)^2}{8})$ , so $b = 4$, $a = 10$, $t_{jump} = 25$, $\sigma = 2$, and $j_0 = 1$.

- **Phase I**: This phase models the initial response when the government has just started to consider implementing policies for the epidemic. Some portion of the populace would have already changed their behavior responding to the reports of an epidemic, but a large portion of the population continues to experience life normally.
- **Phase II**: This phase is characterized by the sharp decline in infection rate as policies to control the spread go into full force (e.g. closing down part of the economy), and the society in whole experiences a shock event.
- **Phase III**: This phase models the inevitable flattening out of the response as the measures reach saturation. This is represented by the diminishing marginal returns (i.e. convexity) in the decline of infection rate.
- **Phase IV**: This phase models the resurgence in the cases caused by lifting the social distancing restrictions (prematurely) and people returning to normal behavior. This is represented by the discrete normal-like jump in the infection rate centered around a certain date. The normal-like jump is used as we assume that once the resurgence reaches a certain level, strong restrictions would be reintroduced, and the infection rate would decline again.

Using parameters $a$ and $b$, we are able to control, respectively, when the measures start, and the strength of such measures; we can therefore interpret $a$ as the median day of action, and $b$ as the median rate of action. The parameter $j_0$ represents the magnitude of the jump, $t_{jump}$ the median day when the jump occurred, and $\sigma$ the rate of which the resurgence in the cases occurred. This formulation allows us to model a wide variety of policies that different governments impose, and then later retract under the same framework, including social distancing, stay-at-home policies, quarantines, among many others. This model is currently used for all provinces in the world to model government response.

# 4 Policy Predictions using DELPHI

## 4.1 Future Policy and Re-Opening Modeling

The DELPHI model assumes that restrictions would be reintroduced in the magnitude consistent with how large the resurgence is. However, in reality, we are unsure with what policy is going to be introduced by the various authorities in different areas, and thus it is extremely useful to understand what would be the potential impact of various policies on the future infections.

| *Policy* | $\overline{\gamma(t)}$ |
|---|---|
| No Measure | 1.0 |
| Restrict Mass Gatherings | 1.0 |
| Restrict Others Only | 0.668 |
| Restrict Mass Gatherings and Schools Only | 0.479 |
| Restrict Mass Gatherings and Others Only | 0.794 |
| Restrict Mass Gatherings, Schools, and Others | 0.423 |
| Lockdown | 0.239 |

Table 1: Normalized $\gamma(t)$ values of various policies

To do so, we would need to know the effect on $\gamma(t)$ for each measure that has been implemented, so that when the measure is lifted, we can reverse the effect. The optimal solution would be trying to learn how much each measure impacted $\gamma(t)$ *per state*, but since many states implemented multiple measures at the same time, it is impossible to extract the effect of a single measure from the observed trend. Luckily, however, many states implemented similar policies, so we can instead aim to estimate the *average* effect of each measure as implemented across states. Then we can utilize these predicted values to correct our estimation of $\gamma(t_p)$ in the case of a change of policy. We explain the methodology below.

### 4.1.1 Creation of the Regression Tree

We use policy data collected for each predicted area around the world[1], each of those features being a time dependent binary variable, and collated the policies into a MECE (Mutually Exclusive, Collectively Exhaustive) combination corresponding to the 6 following variables:

- **No Measure**: if the state is not implementing any measure
- **Restrict Mass Gatherings**: if the state restricts mass gatherings only
- **Mass Gatherings Authorized but Others Restricted**: if the state authorizes mass gatherings but is restricting other things (among schools, work, travel and non-essential services)
- **Restrict Mass Gatherings and Schools**: if the state is restricting Mass Gatherings and Schools **only**
- **Authorize Schools but Restrict Mass Gatherings and Others**: if the state is authorizing schools but restricting at least Mass Gatherings (and maybe others among work, travel and non-essential services)
- **Restrict Mass Gatherings and Schools and Others**: if the state is restricting Mass Gatherings and Schools and others among work, travel and non-essential services

The training data consists in those binary variables, and the target corresponds to the “historical” values for $\gamma(t)$ predicted by DELPHI V2.0 (in previous document) since early March. We voluntarily chose not to include any time lagged values of $\gamma(t-k)$ as features because most of the feature importance was taken by these values, instead of the policies themselves, while interpretability here is key (although it managed to achieve much higher out of sample $R^2$). The resulting values of $\gamma(t)$, normalized by the no-measure policy, are shown below:

---

[1] Data gathered by the Institute for Health Metrics and Evaluation, and the Oxford COVID-19 Government Policy Tracker

#### 4.1.2 Correction of Government Response estimation

From the tree in Figure 3, we have to find a way to correct the current predicted value of $\gamma(t)$ once there is a change in policy using the values predicted by the tree. One way to do that was to first compute a **normalized** offset (*w.r.t* the largest value of $\gamma(t)$ predicted by the tree, *i.e.* the No Measure value) between different pairs of policies, i.e. the normalized difference between tree-predicted values of $\gamma(t)$ if we move from one policy to another. A few examples can be found in Table 1; notice that the offset is positive if the new policy is more lenient, and negative if it is more stringent: Having computed this table, now we can use this normalized offset as follows to correct the government response, from time $t_c$ when there is a change of policy[2]:

| *Policy Change* | $k_0'$ |
|---|---|
| Lockdown → No Measure | $1 - 0.239 = +0.761$ |
| Lockdown → Restrict Mass Gatherings & Schools & Others | $0.423 - 0.239 = +0.184$ |
| Restrict Mass Gatherings and Schools Only → Lockdown | $0.239 - 0.479 = -0.240$ |

Table 2: Examples of normalized offset computations to correct estimation of $\gamma(t)\ \forall t \geq t_c$

$$\boxed{\forall t \geq t_c \quad \gamma'(t) = \frac{2}{\pi}\arctan\left(\frac{-b(t-a)}{20}\right) + 1 + j_0 \exp\left(-\frac{(t-t_{jump})^2}{2\sigma^2}\right) + k_0' \cdot \min\left\{\frac{2-\gamma(t_c)}{1-p_0}, \frac{\gamma(t_c)}{p_0}\right\}}$$

Where $p_0$ is the normalized tree-predicted value of the current policy, and $\gamma$ is the government response in the first stage (so without the correction term). For example, if we are currently in Lockdown and are moving to No measure, then we'd have $k_0' = 0.787$, $p_0 = 0.329/1.544 = 0.213$ and $\gamma(t_c) = \frac{2}{\pi}\arctan\left(-\frac{t_c-a}{b}\right) + 1$.

The term $\Delta := k_0' \cdot \min\left\{\frac{2-\gamma(t_c)}{1-p_0}, \frac{\gamma(t_c)}{p_0}\right\}$ represents the combination of three intuitive assumptions:

- **If there is No Measure, $\gamma(t_c) \leq 2$**: Since we are using the original $\arctan$ curve to fit the change ($\Delta$) for $\gamma(t)$, immediately after the change at $t_c$ we should conform to the bounds of the original $\arctan$ curve, which are $[0, 2]$.
- **Under the Lockdown policy, $\gamma(t_c) \geq 0$**: Reasoning same as above.
- **The effect of each change in policy is additive and proportional to the offset $k_0'$ estimated by the tree**: This assumption limits $\Delta = \beta k_0'$ for some linear coefficient $\beta$. We can then see that the maximum $\beta$ we can take while respecting the two assumptions above is $\beta = \min\left\{\frac{2-\gamma(t_c)}{1-p_0}, \frac{\gamma(t_c)}{p_0}\right\}$, as required.

---

[2] In implementation, we take $\max\{\gamma(t), 0\}$ to eliminate the negative cases.