

Appendix 1: Epidemiological sub-model

This appendix presents the epidemiological sub-model in detail. The model is an extension and refinement of one described in Kemp-Benedict (2020). To keep the presentation in this report self contained, some of the text from that working paper is repeated verbatim.

Matrix model

The core model is a matrix-model variant of the susceptible-exposed-infected-recovered (SEIR) model (Grant, 2020). The standard SEIR model operates in continuous time, while the matrix model operates in discrete time with a daily time step. People move from the susceptible to the exposed category at a model-dependent rate that we denote r_e . There is a maximum number of days for incubation, m , and a maximum number of days for remaining infected, n . It is possible for individuals to move out of the exposed category or the infective category earlier than that, as given by fractions of the population transitioning from one category to the next. These fractions are indicated below with the symbols r_{it} and r_{rt} . Infected individuals may also die from the disease, at a daily rate d_i . (The model ignores the possibility for reinfection or loss of immunity). The matrix model is given by the following equation, where a “+1” subscript means a subsequent day, and parenthesized numbers in superscripts are the days since exposure or infection.

$$\begin{pmatrix} N_{s,+1} \\ N_{e,+1}^{(1)} \\ N_{e,+1}^{(2)} \\ \vdots \\ N_{e,+1}^{(m)} \\ N_{i,+1}^{(1)} \\ N_{i,+1}^{(2)} \\ \vdots \\ N_{i,+1}^{(n)} \\ N_{r,+1} \end{pmatrix} = \begin{pmatrix} 1-r_e & & & & & & & & \\ r_e & & & & & & & & \\ & 1 & & & & & & & \\ & & \ddots & & & & & & \\ & & & 1-r_{im} & & & & & \\ & & & r_{im} & 1 & & & & \\ & & & & & 1-r_{r1}-d_1 & & & \\ & & & & & & \ddots & & \\ & & & & & & & 1-r_{rn}-d_n & \\ & & & & & & & r_{rn} & 1 & 1 \end{pmatrix} \begin{pmatrix} N_s \\ N_e^{(1)} \\ N_e^{(2)} \\ \vdots \\ N_e^{(m)} \\ N_i^{(1)} \\ N_i^{(2)} \\ \vdots \\ N_i^{(n)} \\ N_r \end{pmatrix}. \quad (1)$$

In this equation, N_s is the total number of susceptible individuals, while N_r is the total number of recovered individuals. The total exposed, N_e , and infected, N_i , are given by

$$N_e = \sum_{j=1}^m N_e^{(j)}, \quad N_i = \sum_{j=1}^n N_i^{(j)}. \quad (2)$$

The incubation period of COVID-19 has been estimated to have a median of 5.1 days, with 95% of cases exhibiting symptoms between 2.2 and 11.5 days after exposure (Lauer et al., 2020). We approximated the curve in Lauer et al. by setting r_{it} and d_i to different values when passing from day 1 to day 2, day 2 to day 3, and so on. The recovery period (or time to death) covers a wider range. Wang et al. (2020) documented a comparatively small number (17) of fatal cases and found a range of 6 to 41 days from the onset of symptoms to death, with a median of 14 days and a mean of $T_i = 16$ days. Of those, 90% of cases lasted between 7 and 32 days. Lauer et al. do not distinguish between recovery and death; so, to calibrate to their curve, the total of r_{rt} and d_i , denoted by r_{rt}^{tot} , was set to different values to mimic that curve. We assume that patients either recover or die along that curve in the same proportion. Thus, we allocate between recovery and mortality with a fixed ratio d , such that

$$d_t = dr_{rt}^{\text{tot}}, \quad r_{rt} = (1-d)r_{rt}^{\text{tot}}. \quad (3)$$

The key step is the passage from the susceptible to the exposed pool. The rate coefficient for that step is r_e . From equation (1),

$$N_{s,t+1} = N_s - r_e N_s. \quad (4)$$

We introduce a variation on this basic model that allows for an “at-risk” population to be defined with a higher mortality rate. The parameters for the at-risk population are the fraction of the population at risk ϕ_r , the case fatality rate of the population at risk d_r , the recovery rate for the at-risk population as a fraction of the recovery rate for the not-at-risk population ρ_r , and the fraction of observed cases (the “visible” cases discussed later in this appendix) requiring hospitalization. The model calculates total rates for the at-risk and not-at-risk categories as $r_{rt}^{\text{tot}(r)}$ and $r_{rt}^{\text{tot}(nr)}$ with the following assumptions:

First the case fatality rate for the population not at risk, d_{nr} , is calculated from the average value, d , and the value for the population at risk:

$$d = \phi_r d_r + (1 - \phi_r) d_{nr} \Rightarrow d_{nr} = \frac{d - \phi_r d_r}{1 - \phi_r}. \quad (5)$$

Next, the relative recovery rate is assumed to be the same across the infective period. From equation (3), this means that

$$(1 - d_r) r_{rt}^{\text{tot}(r)} = \rho_r (1 - d_{nr}) r_{rt}^{\text{tot}(nr)}. \quad (6)$$

Finally, by definition, the average total rate is

$$r_{rt}^{\text{tot}} = \phi_r r_{rt}^{\text{tot}(r)} + (1 - \phi_r) r_{rt}^{\text{tot}(nr)}. \quad (7)$$

Taken together, equations (5), (6), and (7) are sufficient to calculate the rates $r_{rt}^{\text{tot}(r)}$ and $r_{rt}^{\text{tot}(nr)}$.

The structure of the matrix equation is the same for both the at-risk and not-at-risk populations. For the remainder of this appendix they are not distinguished.

Standard SEIR model

The standard SEIR model assumes that infected individuals are spread evenly across the population and the probability of infection is independent of the number of infective cases. In that case, the probability of being in the infected population is N_i/N , where $N = N_s + N_e + N_i + N_r$ is the total population. Next, the question is how many susceptible people meet an infected person and are actually exposed. That can be written as

$$N_{e,t+1}^{(1)} = P(\text{exposed} | \text{transmitted}) \times P(\text{transmitted} | \text{meet}) \times P(\text{meet} | \text{infected}) \times \frac{N_i}{N} \times N_s. \quad (8)$$

That is, it is the number of susceptible people, multiplied by the probability that someone is infected, multiplied by the probability of meeting someone who is infected, multiplied by the probability that the virus is transmitted given that an individual comes into contact with an infected person, multiplied by the probability that the susceptible person moves into the “exposed” stage of the disease given that the virus is transmitted. The factor $P(\text{exposed} | \text{transmitted})$ is a characteristic of the disease. It is conventionally designated by a factor β_0 , which in turn is equal to the basic reproduction number R_0 , divided by the mean infected time T_i . The other factors can be modified through public health measures:

- $P(\text{meet} \mid \text{infected})$ is affected by
 - Stay-at-home orders;
 - Store closings;
 - Immigration restrictions;
 - Isolating symptomatic individuals;
 - Isolating asymptomatic individuals identified through testing and tracing.
- $P(\text{transmitted} \mid \text{meet})$ is affected by
 - Social distancing;
 - Masks;
 - Ventilation;
 - Duration of exposure.

In the model, the product of those probabilities is denoted by an overall public health factor $\Phi \leq 1$, giving an effective rate of exposure per infected and susceptible people

$$\Phi\beta_0 = \Phi \frac{R_0}{T_i}. \quad (9)$$

For the standard SEIR model, the rate r_e can now be written

$$r_e = \Phi\beta_0 \frac{N_i}{N}. \quad (10)$$

The k factor

For many viruses, and in particular for COVID-19, the rate of exposure depends on the number of infected cases. When the number of infected cases is comparatively large relative to the susceptible population (although usually much smaller in absolute terms), the average rate β_0 is a reasonable approximation. However, when the number of infected cases is small, particularly when the population is first exposed, the probability may be much less. This is conventionally represented by a “k-factor” in a negative binomial model of disease spread (Blumberg et al., 2014). For the purposes of the present model, the relevant parameter is the probability $p_s(n_i)$ that a set of $n_i < N_i$ cases in some local area that is not yet experiencing a spread of the disease results in at least $n_i + 1$ additional cases. While those cases might afterward fail to propagate, this criterion is adopted as the start of spread, thereby moving that local area from the “not spreading” to the “spreading” category. This probability is one minus the probability of up to n_i cases, a probability that depends on the effective reproduction factor $R_{\text{eff}} = \Phi R_0$ and the measure of over-dispersion k . From equation (4) in the Methods Supplement of Blumberg et al. (2014), the probability follows a negative binomial distribution, which has the regularized incomplete Beta function as a cumulative distribution $I(x; \alpha, \beta)$, giving

$$p_s(n_i) = 1 - I(x; kn_i, n_i + 1), \quad x = \frac{k}{\Phi R_0 + k}. \quad (11)$$

In the standard SEIR model, k is effectively infinite, so the probability is exactly the same for any value of n_i . Because k is rather small for COVID-19, perhaps around 0.1 (Endo et al., 2020), $p_s(n_i)$ depends significantly on n_i up to a few tens of infected people. As shown in Figure 1, the asymptotic value depends significantly on Φ , and low values of n_i strongly dampen the probability of spread. If public health measures are sufficient to bring $R_{\text{eff}} = \Phi R_0$ below one, then the probability of spread is strongly damped for any number of infective individuals.

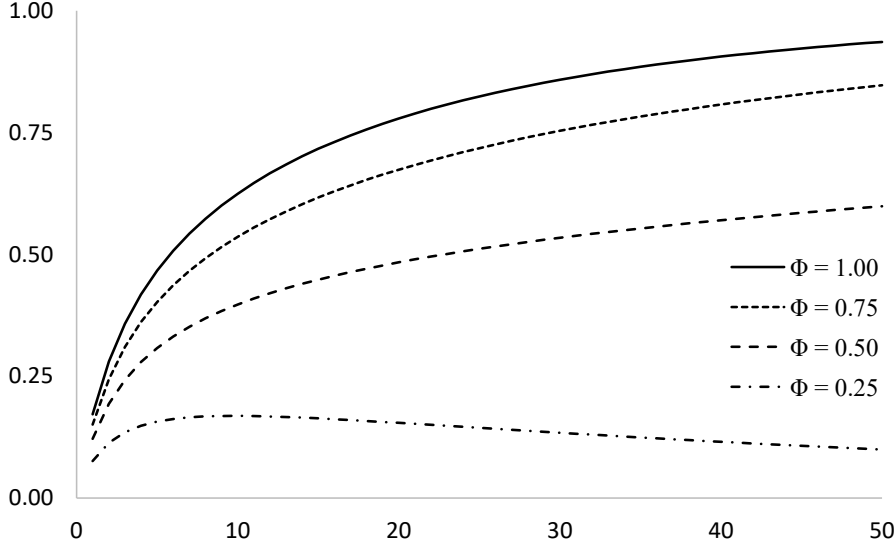


Figure 1: $p_s(n_i)$ for $k = 0.1$, $R_0 = 2.5$, and different values of Φ

Capturing inhomogeneity

COVID-19 is spreading very unevenly around the world, and it also spreads unevenly within countries. This uneven spread can affect the progress of the disease. The goal in this note is to construct an aggregate model that reflects the unevenness at a local level. The model is built by starting from a disaggregated picture and then making simplifying assumptions where necessary to construct an aggregate model with a small number of parameters.

To represent uneven distribution of the disease, we treat the total population as separated into a large number of localities. Furthermore, each locality is labeled by a category (called a “region” in the model configuration file), such as rural or urban. For the j th locality, the total population is n_j , the susceptible population by n_{js} , and so on. The set of localities in category c is denoted C . Following conventional notation, the number of localities within a particular category (the cardinality of the set) is denoted by vertical bars: $|C|$. In practice, these might be counts of cities or urban agglomerations for urban areas, and counts of rural districts for rural areas. Within a locality, the rate of exposure per member of the susceptible population, r_{ej} , is given by the product of the probability of encountering an infective individual per encounter – assumed as in the standard SEIR model to be the ratio of the infected to the total population – and the probability of becoming exposed given that encounter. That probability is given by $\Phi\beta_0$. In the model, common values are assumed for localities within a category, which is indicated here by the notation $c(j)$ for the category to which locality j belongs,

$$r_{ej} = \beta_0 \Phi_{c(j)} \frac{n_{ji}}{n_j}. \quad (12)$$

Applying the matrix equation (1), the susceptible population in the subsequent day is

$$n_{js,+1} = (1 - r_{ej}) n_{js} = n_{js} - \beta_0 \Phi_{c(j)} \frac{n_{js} n_{ji}}{n_j} + \Delta n_{js}, \quad (13)$$

where Δn_{js} captures changes due to causes other than the progress of the disease (such as migration or background mortality).

1 An aggregate model

2 For the second, nonlinear term in equation (13), note that the sum of susceptible, exposed, infected, and
3 recovered must equal the total population, so n_{js} can be replaced with

$$n_{js} = n_j - n_{je} - n_{jr} - n_{ji}. \quad (14)$$

5 Substituting into equation (12) and summing over all localities gives

$$N_{s,+1} = \sum_j n_{js,+1} = N_s - \beta_0 \sum_j \Phi_{c(j)} n_{ji} + \beta_0 \sum_j \Phi_{c(j)} \frac{(n_{je} + n_{jr}) n_{ji}}{n_j} + \beta_0 \sum_j \Phi_{c(j)} \frac{n_{ji}^2}{n_j} + \Delta N_s. \quad (15)$$

7 The sum can be expressed in terms of population categories,

$$N_{s,+1} = N_s - \sum_c \left(\beta_0 \Phi_c \sum_{j \in C} n_{ji} + \beta_0 \Phi_c \sum_{j \in C} \frac{(n_{je} + n_{jr}) n_{ji}}{n_j} + \beta_0 \Phi_c \sum_{j \in C} \frac{n_{ji}^2}{n_j} \right) + \Delta N_s. \quad (16)$$

9 For any given category c , we can write

$$N_{c,s,+1} = N_{c,s} - \beta_0 \Phi_c \sum_{j \in C} n_{ji} + \beta_0 \Phi_c \sum_{j \in C} \frac{(n_{je} + n_{jr}) n_{ji}}{n_j} + \beta_0 \Phi_c \sum_{j \in C} \frac{n_{ji}^2}{n_j} + \Delta N_{c,s}, \quad (17)$$

11 Where $N_{c,s}$ is the susceptible population in category c , and $\Delta N_{c,s}$ is the change in the susceptible
12 population due to causes other than the disease.

13 At this point it is helpful to distinguish between those localities where there is community spread of
14 COVID-19, and those where there is no spread. Where it is not spreading, any infected individuals will
15 have arrived from elsewhere and will be a negligible proportion of the population. In those localities, $n_{ji} \approx$
16 0. Areas with and without community spread are distinguished by a variable s_j , which is equal either to
17 zero (no spread) or one (spread). With this definition, $s_j n_{ji} \approx n_{ji}$, because $s_j = 0$ whenever n_{ji} is negligible and
18 $s_j = 1$ otherwise. The same is true for the exposed and recovered populations n_{je} and n_{jr} . Moreover, $s_j^2 = s_j$.
19 In terms of this variable, equation (17) can be written

$$N_{c,s,+1} = N_{c,s} - \beta_0 \Phi_c \sum_{j \in C} s_j n_{ji} + \beta_0 \Phi_c \sum_{j \in C} s_j \frac{(n_{je} + n_{jr}) n_{ji}}{n_j} + \beta_0 \Phi_c \sum_{j \in C} s_j \frac{n_{ji}^2}{n_j} + \Delta N_{c,s}. \quad (18)$$

21 The second term involves the total infected population in the category

$$N_{c,i} = \sum_{j \in C} s_j n_{ji}. \quad (19)$$

23 The share of the population in areas where there is community spread within category c , σ_c , is given by

$$\sigma_c = \frac{\sum_{j \in C} s_j n_j}{\sum_{j \in C} n_j} = \frac{\sum_{j \in C} s_j n_j}{N_c}. \quad (20)$$

25 In the final expression, N_c is the total population in category c . Next, consider the sum in the third term on
26 the right:

$$\sum_{j \in C} s_j \frac{(n_{je} + n_{jr})n_{ji}}{n_j} = \sum_{j \in C} s_j n_j \frac{n_{je} + n_{jr}}{n_j} \frac{n_{ji}}{n_j}. \quad (21)$$

To simplify the following expressions, define $x_j \equiv n_{je} + n_{jr}$ and $y_j \equiv n_{ji}$. Then the sum becomes

$$\sum_{j \in C} s_j \frac{(n_{je} + n_{jr})n_{ji}}{n_j} = \sum_{j \in C} s_j n_j \frac{x_j}{n_j} \frac{y_j}{n_j}. \quad (22)$$

Next, define X_c and Y_c as the sums of x_j and y_j across the category and construct the following sum:

$$\begin{aligned} \sum_{j \in C} s_j n_j \left(\frac{x_j}{n_j} - \frac{X_c}{\sigma_c N_c} \right) \left(\frac{y_j}{n_j} - \frac{Y_c}{\sigma_c N_c} \right) &= \sum_{j \in C} s_j n_j \frac{x_j}{n_j} \frac{y_j}{n_j} - \frac{Y_c}{\sigma_c N_c} \sum_{j \in C} s_j n_j \frac{x_j}{n_j} - \frac{X_c}{\sigma_c N_c} \sum_{j \in C} s_j n_j \frac{y_j}{n_j} \\ &\quad + \frac{Y_c}{\sigma_c N_c} \frac{X_c}{\sigma_c N_c} \sum_{j \in C} s_j n_j. \end{aligned} \quad (23)$$

The terms on the right can be simplified using this definition, to give

$$\sum_{j \in C} s_j n_j \left(\frac{x_j}{n_j} - \frac{X_c}{\sigma_c N_c} \right) \left(\frac{y_j}{n_j} - \frac{Y_c}{\sigma_c N_c} \right) = \sum_{j \in C} s_j n_j \frac{x_j}{n_j} \frac{y_j}{n_j} - \frac{X_c Y_c}{\sigma_c N_c}. \quad (24)$$

The expression on the left is the correlation between the x_j and y_j across the localities where there is community spread – that is, it is the correlation between the number of infected individuals and the sum of exposed and recovered individuals. Because COVID-19 emerges after a delay, there are lags between infection, exposure, and recovery, so that in any given location these variables might be positively or negatively correlated at different points in time. The same holds true across localities, at least once there has been sufficient spread.

For simplicity, the model assumes that the correlation is sufficiently close to zero so that the left-hand side of equation (24) can be assumed to be zero. Setting the right-hand side of the equation to zero, substituting from equation (22), and replacing X_c and Y_c with their definitions in terms of population sizes, result in

$$\sum_{j \in C} s_j \frac{(n_{je} + n_{jr})n_{ji}}{n_j} \approx \frac{(N_{c,e} + N_{c,r})N_{c,i}}{\sigma_c N_c}. \quad (25)$$

Combining this equation with equation (19), the expression for the change in the susceptible population within the category, equation (18), can now be written

$$N_{c,s,t+1} \approx N_{c,s} - \beta_0 \Phi_c N_{c,i} + \beta_0 \Phi_c \frac{(N_{c,e} + N_{c,r})N_{c,i}}{\sigma_c N_c} + \beta_0 \Phi_c \sum_{j \in C} s_j \frac{n_{ji}^2}{n_j} + \Delta N_{c,s}. \quad (26)$$

The second and third terms can be simplified further. Aside from the factor $\beta_0 \Phi_c$, they are

$$-N_{c,i} + \frac{(N_{c,e} + N_{c,r})N_{c,i}}{\theta_c N_c} = \frac{(N_{c,e} + N_{c,r} - \theta_c N_c)N_{c,i}}{\theta_c N_c}, \quad (27)$$

Because the entire population is susceptible in locales without community spread, the total susceptible population is given by

$$N_{c,s} = (1 - \theta_c) N_c + \theta_c N_c - N_{c,e} - N_{c,r} - N_{c,i}. \quad (28)$$

Substituting into equation (26) gives

$$N_{c,s,+1} \approx N_{c,s} - \frac{\beta_0 \Phi_c}{\sigma_c} \frac{N_{c,s} N_{c,i}}{N_c} - \frac{\beta_0 \Phi_c}{\sigma_c} \frac{N_{c,i}^2}{N_c} + \frac{\beta_0 \Phi_c}{\sigma_c} (1 - \sigma_c) N_{c,i} + \beta_0 \Phi_c \sum_{j \in C} s_j \frac{n_{ji}^2}{n_j} + \Delta N_{c,s}. \quad (29)$$

The last term in this equation can be put into a more recognizable form by constructing the following sum:

$$\begin{aligned} \sum_{j \in C} s_j n_j \left(\frac{n_{ji}}{n_j} - \frac{N_{c,i}}{\sigma_c N_c} \right)^2 &= \sum_{j \in C} s_j \frac{n_{ji}^2}{n_j} - 2 \frac{N_{c,i}}{\sigma_c N_c} \sum_{j \in C} s_j n_{ji} + \left(\frac{N_{c,i}}{\sigma_c N_c} \right)^2 \sum_{j \in C} s_j n_j \\ &= \sum_{j \in C} s_j \frac{n_{ji}^2}{n_j} - \frac{N_{c,i}^2}{\sigma_c N_c}. \end{aligned} \quad (30)$$

The expression on the left is related to the coefficient of variation $c_{v,c}$ of the distribution of counts of infected individuals across localities within population category c :

$$\sum_{j \in C} s_j n_j \left(\frac{n_{ji}}{n_j} - \frac{N_{c,i}}{\sigma_c N_c} \right)^2 = \sigma_c N_c \left(\frac{N_{c,i}}{\sigma_c N_c} \right)^2 c_{v,c}^2 = \frac{N_{c,i}^2}{\sigma_c N_c} c_{v,c}^2. \quad (31)$$

Substituting from equations (30) and (31) into equation (29) then gives

$$N_{c,s,+1} \approx N_{c,s} - \frac{\beta_0 \Phi_c}{\sigma_c} \frac{N_{c,s} N_{c,i}}{N_c} - \frac{\beta_0 \Phi_c}{\sigma_c} \frac{N_{c,i}^2}{N_c} + \frac{\beta_0 \Phi_c}{\sigma_c} (1 - \sigma_c) N_{c,i} + \frac{\beta_0 \Phi_c}{\sigma_c} \frac{N_{c,i}^2}{N_c} (1 + c_{v,c}^2) + \Delta N_{c,s}. \quad (32)$$

The third term cancels with one of the terms in the final expression, giving

$$N_{c,s,+1} \approx N_{c,s} - \frac{\beta_0 \Phi_c}{\sigma_c} \frac{N_{c,s} N_{c,i}}{N_c} + \frac{\beta_0 \Phi_c}{\sigma_c} (1 - \sigma_c) N_{c,i} + \frac{\beta_0 \Phi_c}{\sigma_c} \frac{N_{c,i}^2}{N_c} c_{v,c}^2 + \Delta N_{c,s}. \quad (33)$$

For ease of comparison with the standard SEIR expression, this can be written

$$N_{c,s,+1} \approx N_{c,s} - \frac{\beta_0 \Phi_c}{\sigma_c} \frac{N_{c,s} N_{c,i}}{N_c} \left[1 - \frac{N_{c,i}}{N_{c,s}} c_{v,c}^2 - (1 - \sigma_c) \frac{N_c}{N_{c,s}} \right] + \Delta N_{c,s}. \quad (34)$$

From this we conclude that the rate of exposure per susceptible person in population category c is

$$r_{e,c} = \frac{\beta_0 \Phi_c}{\sigma_c} \frac{N_{c,i}}{N_c} \left[1 - \frac{N_{c,i}}{N_{c,s}} c_{v,c}^2 - (1 - \sigma_c) \frac{N_c}{N_{c,s}} \right]. \quad (35)$$

To make sense of the magnitudes, note that $(1 - \sigma_c) N_c \leq N_{c,s}$, because the total susceptible population is at least as large as the population in areas where there is not yet community spread. Furthermore, the

number of infected individuals at any time will generally be much smaller than the total population in areas where the disease is spreading, so $N_{c,i} \ll \sigma_c N_c$.

When there is very little community spread and $\sigma_c \approx 0$, the susceptible population is almost equal to the total population, and

$$r_{e,c} = \beta_0 \Phi_c \frac{N_{c,i}}{N_c} \left(1 - \frac{N_{c,i}}{\sigma_c N_c} c_{v,c}^2 \right), \quad \text{when } \sigma_c \approx 0. \quad (36)$$

This is smaller than the standard SEIR result because of inhomogeneity as captured by the coefficient of variation. The square of the coefficient of variation is multiplied by the average infection rate in communities where infection is spreading.

Hospital crowding and excess mortality

Assuming that only “visible” cases with evident symptoms can end in death, the mortality rate d in equation (3) can be expressed as a product of three factors: the fraction of visible cases v , the hospitalization rate h per visible case, and the mortality rate per hospitalized patient m . For the hospitalization rate per visible case, one less than the demographically weighted average rate of mildly symptomatic cases using age-specific rates from (Moghadas et al., 2020 Table A5) comes to $h = 38\%$.

The mortality rate has been observed to rise when hospitals overflow, and beds become unavailable (Moghadas et al., 2020). When beds are available, hospitalized patients experience a baseline mortality rate m_0 . However, for those who cannot be provided a bed due to overflow, the mortality rate rises to a higher level. This appears to have been the case in Hubei Province in China (Ji et al., 2020), which experienced an average mortality rate of 2.9%, strikingly higher than the 0.7% rate reported in other provinces. Baud et al. (2020) estimated confirmed case fatality rates within and outside China. They found that case fatality rates have been exceedingly high, around 15%, outside China, but that the rates were converging over time to a value close to 5.7% -- a finding that again suggests higher mortality when health care systems are under stress. The baseline mortality rate per hospitalized patient can be estimated as

$$m_0 = \frac{5.7\%}{h} = 0.15. \quad (37)$$

The bed-overflow mortality rate is expressed in the model as a multiple θ of the baseline. The findings reported by Ji et al. (2020) and Baud et al. (2020) suggest a factor of three, but practices have improved since the first cases, so it is likely lower. The multiplier is a model parameter that can be set by the user. Finally, calibration suggests a value for the fraction of cases that are displaying symptoms, v , of 13%. Multiplying that value by h and m_0 gives a baseline mortality rate of 0.74% per infected individual.

In any particular locality, there will be a normal level of hospital bed occupancy from causes other than the pandemic. As some of those cases are postponed – for example, for elective surgery – the stock of available beds will expand. The model assumes the number of hospital beds per person to be the same value, b_c , in each locality within category c . Of those, there is normally an excess fraction ε_c , which rises during the pandemic (e.g., by postponing elective surgery) to a higher level ε'_c . The number of available beds per person is then equal to $\varepsilon'_c b_c$. There is overflow if the number of hospitalized infected patients exceeds that value. For locality j , that happens when

$$\eta \frac{n_{ji}}{n_j} > \varepsilon'_{c(j)} b_{c(j)}. \quad (38)$$

This gives a category-specific threshold value for the infected fraction $x_{ji} = n_{ji}/n_j$ of

$$x_c^{\text{thresh}} = \frac{\varepsilon'_c b_c}{\eta}. \quad (39)$$

Overflow occurs when this threshold is exceeded.

Above, the coefficient of variation of the infected fraction across localities within category c was denoted $c_{v,c}$. To estimate bed overflow it is necessary to specify the probability distribution. The variable of interest – the infected fraction – must lie between zero and one, so we assume a Beta distribution $B(x; \alpha, \beta)$ with mean μ_c . For a Beta distribution with mean μ and coefficient of variation c_v , it is possible to show that

$$\alpha = a\mu, \quad \beta = a(1 - \mu), \quad \text{where } a = \frac{1 - \mu}{\mu c_v^2} - 1. \quad (40)$$

These parameters depend on the number of infected individuals through the mean,

$$\mu = \frac{N_{c,i}}{N_c}. \quad (41)$$

The mean exceedance of the threshold per infected person, E_i , is given by

$$E_c = \frac{1}{\mu_c} E(x - x_c^{\text{thresh}} | x > x_c^{\text{thresh}}) = \int_{x_c^{\text{thresh}}}^1 dx \left(\frac{x}{\mu_c} - \frac{x_c^{\text{thresh}}}{\mu_c} \right) B(x; \alpha_c, \beta_c). \quad (42)$$

The integral can be expressed in terms of the cumulative Beta distribution, which is the regularized incomplete Beta function,

$$E_c = 1 - I(x_c^{\text{thresh}}; \alpha_c + 1, \beta_c) - \frac{x_c^{\text{thresh}}}{\mu} (1 - I(x_c^{\text{thresh}}; \alpha_c, \beta_c)). \quad (43)$$

The average mortality rate, which now depends on the category, can be calculated as

$$d_c = (1 - E_c + \theta E_c) v h m_0. \quad (44)$$

Movement of the infected population

The development in the previous section gives an independent evolution of COVID-19 within each population category,

$$N_{c,s,t+1} = (1 - r_{e,c}) N_{c,s,t}, \quad (45)$$

where $r_{e,c}$ is given by equation (35). That value depends on the number of infected people in localities within the population category (for example, in urban areas, ports of entry, or rural areas). In an area with no infections, the disease can be introduced by infected visitors. The model assumes that these visitors are not treated for their symptoms, and that the effect on the mortality rate in equation (44) is negligible. However, the rate of exposure is affected.

On the assumption that the number of infected visitors is small compared to the population, the number of infected visitors is added to the infected population already present, the expression for $r_{e,c}$ is changed to

$$r_{e,c} \rightarrow \frac{\beta_0 \Phi_c}{\sigma_c} \frac{N_{c,i} + N_i^{\text{visit}}}{N_c} \left[1 - \frac{N_{c,i} + N_i^{\text{visit}}}{N_{c,s}} c_{v,c}^2 - (1 - \sigma_c) \frac{N_c}{N_{c,s}} \right]. \quad (46)$$

This is unproblematic in a population category where the disease has already spread sufficiently. However, where $\sigma_c = 0$, this equation will be inconsistent. To accommodate this potential problem, σ_c is replaced by an adjusted value σ_c^{adj} ,

$$r_{e,c} \rightarrow \frac{\beta_0 \Phi_c}{\sigma_c^{\text{adj}}} \frac{N_{c,i} + N_i^{\text{visit}}}{N_c} \left[1 - \frac{N_{c,i} + N_i^{\text{visit}}}{N_{c,s}} c_{v,c}^2 - (1 - \sigma_c^{\text{adj}}) \frac{N_c}{N_{c,s}} \right]. \quad (47)$$

Defining the effective value presents a challenge. While the number of visitors may be small compared to the total population, it may be large compared to the average population across localities. For example, some districts or counties may contain major cities, while others contain dispersed rural populations. The model makes two assumptions for a “virgin” population with no or little spread: 1) that all visitors arrive in the largest locality; 2) that locality size follows a conventional rank-size rule.

Rank-size distribution of localities

The rank-size rule, which is often but not uniformly observed (Parr, 1976), says that the size of cities decreases in a steady progression from the first rank, to the second rank, to the third, and so on. More concretely, the model assumes what Carroll (1982) calls the “restrictive” rank-size rule in which

$$n_{j(c)} = \frac{n_c^{\text{largest}}}{R_j}. \quad (48)$$

The rank R_j is an integer that is equal to 1 for the largest locality, 2 for the second-largest, and so on, up to the total number of localities $|C|$ (that is, the cardinality of the set of localities in category c). Summing across all localities in the category gives an estimate for the number of people in the largest locality as

$$n_c^{\text{largest}} = \frac{N_c}{\sum_{r=1}^{|C|} r^{-1}}. \quad (49)$$

The adjusted value σ_c^{adj} is then defined to be

$$\sigma_c^{\text{adj}} = \begin{cases} \sigma_c, & \sigma_c > 0, \\ \frac{N_i^{\text{visit}}}{n_c^{\text{largest}}}, & \sigma_c = 0. \end{cases} \quad (50)$$

With the expression, the model is fully determined.

Appendix 2: Macroeconomic sub-model

This appendix presents the macroeconomic sub-model in detail. It contains an input-output (I-O) matrix at its core. The model is made dynamic by specifying demand for goods and services, and then having the investment decisions depend on deviation from a target growth rate. All sectors have the same target, so the model exhibits “balanced growth” in the absence of demand shocks. In the presence of demand shocks, economic growth may fall below the target level, leading to a slower pace of investment. When demand exceeds the capacity to supply, a “forced saving” mechanisms comes into play, driving household consumption below a desired level.

Derivation of core equation

The presentation of the model begins with a basic balance. Using \mathbf{Y} to represent a vector of output, \mathbf{A} the symmetric I-O matrix, \mathbf{X} a vector of exports, \mathbf{M} of imports, \mathbf{H} of household demand, \mathbf{G} government demand, and \mathbf{J} demand for investment goods, the balance is

$$\mathbf{Y} = \mathbf{A} \cdot \mathbf{Y} + \mathbf{H} + \mathbf{G} + \mathbf{J} + \mathbf{X} - \mathbf{M}. \quad (51)$$

Investment goods are supplied in part from domestic sources and in part from imports. In developing countries, in particular, most machinery is imported, but construction is locally sourced. Given total investment demand I , demand for investment goods is given by vectors $\boldsymbol{\theta}_{\text{dom}}$ and $\boldsymbol{\theta}_{\text{imp}}$,

$$\mathbf{J} = (\boldsymbol{\theta}_{\text{dom}} + \boldsymbol{\theta}_{\text{imp}}) I. \quad (52)$$

Desired imports of household goods are given by a sector-specific fraction \mathbf{m}_H of demand in excess of a level \mathbf{H}_0 that is always supplied domestically. That quantity represents all household expenditures on non-tradeables (most services) and a certain minimum quantity of tradeables (e.g., domestically sourced food). A similar coefficient \mathbf{m}_A applies to intermediate goods (with no domestic minimum). Allowing for residual, undesired imports $\Delta\mathbf{M}$, or a reduction in domestic household expenditure $\Delta\mathbf{H}_{\text{dom}}$ when domestic production is insufficient, equation (51) becomes

$$\mathbf{Y} = (\mathbf{I} - \tilde{\mathbf{m}}_A) \cdot \mathbf{A} \cdot \mathbf{Y} + (\mathbf{I} - \tilde{\mathbf{m}}_H) \cdot (\mathbf{H} - \mathbf{H}_0) + \mathbf{H}_0 - \Delta\mathbf{H}_{\text{dom}} + \mathbf{G} + \boldsymbol{\theta}_{\text{dom}} I + \mathbf{X} - \Delta\mathbf{M}. \quad (53)$$

In this expression, a tilde over a vector indicates a diagonal matrix with the vector values along the diagonal, while \mathbf{I} is the identity matrix.

Much of the model works with an aggregate quantity, desired final demand (excluding investment goods), denoted \mathbf{F} . It is equal to

$$\mathbf{F} \equiv (\mathbf{I} - \tilde{\mathbf{m}}_H) \cdot (\mathbf{H} - \mathbf{H}_0) + \mathbf{H}_0 + \mathbf{G} + \mathbf{X}. \quad (54)$$

The model assumes that export demand, purchases of investment goods, and government consumption expenditures are all satisfied. However, the supply of household demand may deviate from the desired level, a change which may be accommodated either through lower domestic demand or imports,

$$\Delta\mathbf{F} \equiv \Delta\mathbf{H}_{\text{dom}} + \Delta\mathbf{M}. \quad (55)$$

Equation (53) can then be written

$$\mathbf{Y} = (\mathbf{I} - \tilde{\mathbf{m}}_A) \cdot \mathbf{A} \cdot \mathbf{Y} + \mathbf{F} - \Delta\mathbf{F} + \boldsymbol{\theta}_{\text{dom}} I. \quad (56)$$

Solving for \mathbf{Y} gives

$$\mathbf{Y} = [\mathbf{I} - (\mathbf{I} - \tilde{\mathbf{m}}_A) \cdot \mathbf{A}]^{-1} \cdot (\mathbf{F} - \Delta\mathbf{F} + \boldsymbol{\theta}_{\text{dom}} I) \equiv \mathbf{\Lambda} \cdot (\mathbf{F} - \Delta\mathbf{F} + \boldsymbol{\theta}_{\text{dom}} I), \quad (57)$$

where $\mathbf{\Lambda}$ is a generalized Leontief inverse that takes import propensities into account. This is the core equation, in that it propagates variations in demand throughout the economy.

Working with the core equation

It is easiest from this point onward to write the core equation in terms of indices, with explicit summations. Equation (57) becomes

$$Y_i = \sum_j \Lambda_{ij} (F_j - \Delta F_j + \theta_{\text{dom},j} I). \quad (58)$$

Next, we write output Y_i as potential output Y_i^* multiplied by utilization u_i , where the potential is scaled such that normal utilization is at $u_i = 1$,

$$u_i Y_i^* = \sum_j \Lambda_{ij} (F_j - \Delta F_j + \theta_{\text{dom},j} I). \quad (59)$$

Potential output is increased through investment with a marginal output coefficient v_i , and depreciates at a rate δ_i , so that (in discrete time),

$$Y_i^* = Y_{i,-1}^* + v_i I_i - \delta_i Y_{i,-1}^*, \quad (60)$$

where a “-1” subscript indicates a value from a previous time step. It is convenient to express investment in terms of a gross investment rate g_i ,

$$g_i \equiv \frac{v_i I_i}{Y_{i,-1}^*}. \quad (61)$$

Then equation (59) can be written

$$u_i (1 + g_i - \delta_i) Y_{i,-1}^* = \sum_j \Lambda_{ij} (F_j - \Delta F_j + \theta_{\text{dom},j} I), \quad \text{where } I = \sum_k \frac{g_k Y_{k,-1}^*}{v_k}. \quad (62)$$

Below, we specify desired household expenditure, export demand, and government expenditure. Once they are calculated, \mathbf{F} is fully determined. Depreciation rates and the previous-period potential output are also known. The coefficients entering the Leontief matrix $\mathbf{\Lambda}$ are similarly presumed not to change. That leaves as unknowns to be determined the values of u_i , g_i , and ΔF_j .

To determine these values within the model, it is necessary to define some behavioral assumptions. The model assumes the following approach:

Step 1: Investment expenditure

First, investment expenditure is determined by firms who anticipate: 1) that utilization will tend towards its normal level at $u_i = 1$; and 2) that economic growth will tend towards an exogenously specified normal rate γ . On a balanced-growth path, all sectors will grow at the common rate γ . Specifically, expected utilization is

$$u_i^e = 1 - \varphi(1 - u_{i,-1}), \quad (63)$$

while investment expenditure is set such that

$$\frac{u_i^e}{u_{i,-1}}(1 + g_i - \delta_i) = 1 + \gamma. \quad (64)$$

This gives a value for g_i of

$$g_i = \frac{u_{i,-1}(1 + \gamma)}{1 - \varphi(1 - u_{i,-1})} - 1 + \delta_i. \quad (65)$$

If the ratio falls below one, then investment will fall below the replacement rate. The model assumes that this does not happen and sets

$$g_i = \max \left[0, \frac{u_{i,-1}(1 + \gamma)}{1 - \varphi(1 - u_{i,-1})} - 1 \right] + \delta_i. \quad (66)$$

The value for utilization at which net investment becomes zero, termed the “threshold” utilization u_{thresh} , is a user-defined parameter that determines the value of φ . It satisfies

$$\frac{u_{\text{thresh}}(1 + \gamma)}{1 - \varphi(1 - u_{\text{thresh}})} = 1 \Rightarrow \varphi = 1 - \frac{u_{\text{thresh}}}{1 - u_{\text{thresh}}} \gamma. \quad (67)$$

However, φ cannot fall below zero, corresponding to a maximum allowable value for u_{thresh} of $1/(1 + \gamma)$. If the threshold value exceeds this level, the model will silently set u_{thresh} to the maximum so that φ does not become negative.

Step 2: Find utilization

Next, unconstrained utilization u_i^* is calculated, whereby “unconstrained” is the utilization required to meet *desired* levels of household expenditure, committed levels of investment and government expenditure, and expressed demand for exports. That is, it solves

$$u_i^*(1 + g_i - \delta_i)Y_{i,-1}^* = \sum_j \Lambda_{ij}(F_j + \theta_{\text{dom},j}I). \quad (68)$$

Utilization of $u_i = 1$ is defined as full utilization at normal capacity, with existing shift structures and normal maintenance downtime. Buildings and equipment can be operated at higher-than-normal capacity, so the model allows values for u_i up to a user-specified maximum utilization level for each sector $u_i^{\text{max}} \geq 1$. Values for unconstrained utilization may exceed these levels, so actual utilization is given by

$$u_i = \min(u_i^{\text{max}}, u_i^*). \quad (69)$$

Comparing to equation (62), it is possible to see that

$$(u_i^* - u_i)(1 + g_i - \delta_i)Y_{i,-1}^* = -\sum_j \Lambda_{ij}\Delta F_j. \quad (70)$$

This is a forced-saving mechanism in which production constraints lead to a fall in household demand (and therefore undesired saving).

Step 3: Finding the shift in final demand

Solving equation (70) is a linear algebra problem that is best expressed in matrix form:

$$(\tilde{\mathbf{u}}^* - \tilde{\mathbf{u}}) \cdot (\mathbf{I} + \tilde{\mathbf{g}} - \tilde{\mathbf{\delta}}) \cdot \mathbf{Y}_{-1}^* = \mathbf{\Lambda} \cdot \Delta \mathbf{F}. \quad (71)$$

This is solved by

$$\Delta \mathbf{F} = \mathbf{\Lambda}^{-1} \cdot (\tilde{\mathbf{u}}^* - \tilde{\mathbf{u}}) \cdot (\mathbf{I} + \tilde{\mathbf{g}} - \tilde{\mathbf{\delta}}) \cdot \mathbf{Y}_{-1}^*. \quad (72)$$

Because the generalized Leontief inverse is already an inverse, this can also be written

$$\Delta \mathbf{F} = [\mathbf{I} - (\mathbf{I} - \tilde{\mathbf{m}}_A) \cdot \mathbf{A}] \cdot (\tilde{\mathbf{u}}^* - \tilde{\mathbf{u}}) \cdot (\mathbf{I} + \tilde{\mathbf{g}} - \tilde{\mathbf{\delta}}) \cdot \mathbf{Y}_{-1}^*. \quad (73)$$

That gives the shortfall arising from forced saving. In principle, it is made up by a reduction in household demand (for non-tradeables) or a rise in imports (for tradeables), or a mix of the two, but the distinction is not drawn in the model.

Updating the components of final demand

At each time step, the components of final demand are computed in the following way: First, baseline household demand \mathbf{H}_0 and a portion of government expenditure are assumed to grow at the target growth rate γ . Thus, they represent sources of autonomous demand. The rest of government expenditure grows with a smoothed GDP growth rate, with a one-year smoothing time.

The wage bill W_k within each sector is assumed to expand in proportion to output,

$$W_k = \frac{Y_k}{Y_{k,-1}} W_{k,-1}. \quad (74)$$

The total wage bill is the sum across all sectors. The excess of household demand over the baseline level, $\mathbf{H} - \mathbf{H}_0$, is assumed to grow (or shrink) in line with the wage.

Export demand grows according to the specified global GDP growth trajectory and export elasticities ε with respect to global GDP growth. Denoting the global GDP growth rate by γ_{glob} , with a reference value γ_{glob}^* , export demand is calculated as

$$X_k = [1 + \gamma + \varepsilon_k (\gamma_{\text{glob}} - \gamma_{\text{glob}}^*)] X_{k,-1}. \quad (75)$$

With this specification, if global GDP growth never changes, then exports grow at the balanced-growth rate γ . If they decline, then they do so in proportion to the user-specified elasticities.

Finally, desired final demand is adjusted for social distancing, given user-specified social distancing sensitivities, and the effect of travel bans. It is also adjusted for low or high demand for hospital services, due to a combination of avoided elective care and surges in the number of COVID-19 patients.

