Lecture 9: Full Calibration of SIR Models

Radu Balan

Department of Mathematics, NWC
University of Maryland, College Park, MD

Version: February 27, 2024

SIR Model Calibration

Recall the model:

$$\left\{ \begin{array}{lcl} \frac{dS}{dt} & = & -\beta S \frac{I}{N} \quad , \quad S(0) \\ \frac{dI}{dt} & = & \beta S \frac{I}{N} - \alpha I \quad , \quad I(0) \\ \frac{dR}{dt} & = & \alpha I \quad , \quad R(0) \end{array} \right.$$

with the sub-compartments $X(t) = (1 - \gamma)R(t)$ for "recovered" and $Y(t) = \gamma R(t)$ for deaths.
Before making useful predictions (testing), the model has to be calibrated. Last time we analyzed estimators for γ from the time series of *cumulative detected infections*, $\{V(0), \cdots, V(T_{max})\}$, and the time series of *cumulative deaths*, $\{Y(0), \cdots, Y(T_{max})\}$. The idea was to minimize over $\tau \geq 0, \gamma \in [0, 1]$ the norm $\|Y(\cdot + \tau) - \gamma V\|_p$, scaled by the number of terms in each sum.
Now we integrate these estimators into a scheme that performs full calibration of the SIR model.

Optimization based on Forward Model Simulation

Pre-processing steps

Step 1: t_0. First detect the onset of infections, and reset the time origin to match this starting time t_0.
Step 2: $I(t)$. The time series of cumulative detected infections $\{V(0), \cdots, V(T_{max})\}$ should be turned into a rate of infections. The daily rate may not be relevant to infection transmissions. Instead use a time window to convert the cumulative count into a rate:

$$I(t) = V(t + t_0 + \tau_0) - V(t + t_0 - \tau_0) \quad , \quad t = 0, 1, 2, \cdots, T_{max}$$

where $\tau_0 > 0$ is chosen so that τ_0 accounts for average infection period. For Covid-19 it is somewhere between 5 days and 10 days. A possible value is $\tau_0 = 7$. Thus $I(t)$ measures the number of invections in a 2-week period centered around t.
Step 3: $Y_{measured}$. Align the time series of cumulative death with the onset time t_0:

$$Y_{measured}(t) = Y(t + t_0) \quad , \quad t = 0, 1, 2, \cdots, T_{max}.$$

Optimization based on Forward Model Simulation

The Meta Loop

A natural choice for initial conditions is given by: $S(0) = N$ and $R(0) = 0$, where we assumed $I(0) \ll N$:

$$\begin{cases} \frac{dS_{sim}}{dt} & = & -\beta S_{sim} \frac{I_{sim}}{N} \quad , \quad S_{sim}(0) = N \\ \frac{dI_{sim}}{dt} & = & \beta S_{sim} \frac{I_{sim}}{N} - \alpha I \quad , \quad I_{sim}(0) = I(0) > 0 \\ \frac{dR_{sim}}{dt} & = & \alpha I_{sim} \quad , \quad R_{sim}(0) = 0 \\ Y_{sim} & = & \gamma R_{sim} \end{cases}$$

At this point, the parameters that need to be estimated are: $\{\alpha, \beta, \gamma\}$.
The Forward Model based calibration works like this:

1. Construct a search set Ω of the "free" parameters (α, β)
2. For each pair $(\alpha, \beta) \in \Omega$: (i) run the forward SIR model using an Euler scheme that produces $(S_{sim}, I_{sim}, R_{sim})$. (ii) Fit $\hat{\gamma}$ to match the observed time series $Y_{measured}$. (iii) Fit $\hat{\rho}$ that measures the undercounting factor in detected infections I. (iv) Compute the value of the objective function $J_p(\alpha, \beta, \hat{\gamma}, \hat{\rho})$.
3. Select $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\rho})$ that minimize J_p.

Optimization based on Forward Model Simulation

Details: The objective function

The objective function J_p is chosen to measure residuals $I - \rho I_{sim}$ and $Y_{measured} - Y_{sim}$. Fix cost coefficients c_I and c_Y. For $1 \leq p < \infty$ define

$$J_p(\alpha,\beta,\gamma,\rho) = c_I \sum_{t=0}^{T_{max}} |I(t) - \rho I_{sim}(t)|^p + c_Y \sum_{t=0}^{T_{max}} |Y_{measured}(t) - \gamma R_{sim}(t)|^p$$

For $p = \infty$ define

$$J_\infty(\alpha,\beta,\gamma,\rho) = c_I \max_{0\leq t\leq T_{max}} |I(t) - \rho I_{sim}(t)| + c_Y \max_{0\leq t\leq T_{max}} |Y_{measured}(t) - \gamma R_{sim}($$

Typical choices: $p = 1, 2, \infty$ and $(c_I, c_Y) = (0,1)$ (if the cumulative detected infections are unreliable) or $(c_I, c_Y) = (1,1)$ (if detected infections is a reliable measure).

The Calibration Algorithm for SIR Models

Algorithm (Meta-Algorithm for SIR Calibration)

Inputs: *Time series* $\{V(0), \cdots, V(T)\}$, $\{Y(0), \cdots, Y(T)\}$. *Parameters:* V_{min} *(default,* $V_{min} = 5$*),* τ_0 *(default,* $\tau_0 = 7$*),* $N =$ *Population,* $p \in [1, \infty]$ *(default,* $p = 2$*),* $c_I, c_Y > 0$. *Search set* Ω.

1. *Detect the onset of the infection* t_0 *as the first time so that* $V(t_0) \geq V_{min}$*, reset the time origin, and create the time series of infection rates* $I(t) = V(t + t_0 + \tau_0) - V(t + t_0 - \tau_0)$*, and aligned cumulative death* $Y_{measured}(t) = Y(t + t_0)$, $0 \leq t \leq T_{max}$.
2. *For each* $(\alpha, \beta) \in \Omega$ *repeat:*
   1. *Simulate a SIR model with parameters* (α, β) *and initial condition* $S_{sim}(0) = N$, $I_{sim}(0) = I(0)$, $R_{sim}(0) = 0$*, and obtain* *daily* *time series* $(S_{sim}, I_{sim}, R_{sim})$.
   2. *Solve* $\hat{\gamma} = argmin_{\gamma} \|Y_{measured} - \gamma R_{sim}\|_p$.
   3. *Solve* $\hat{\rho} = argmin_{\rho} \|I - \rho I_{sim}\|_p$.
   4. *Compute the objective function* $J = J_p(\alpha, \beta, \hat{\gamma}, \hat{\rho})$.
3. *Determine the minimum and the minimizer of* J.

Outputs: *Estimated* $\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\rho}$ *and minimum value* J_{min}.

Numerical results (1)

Analysis of DC data for 2020

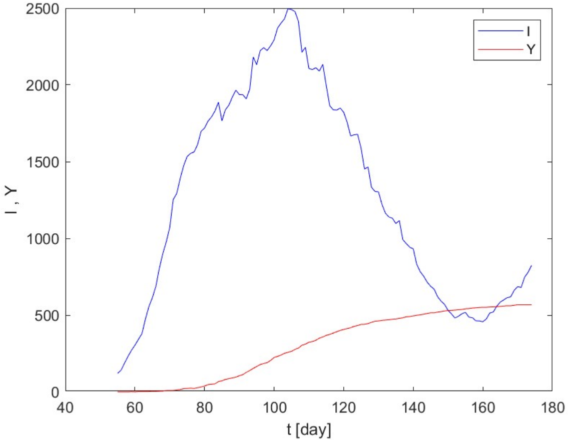
$V_{min} = 5$, $\tau_0 = 7$ [days]. Onset $t_0 = 55$ (Monday: March 16, 2020), $T_{max} = 119$, $N = 705749$.

Notable dates:

Local dips: day=152 (Sunday: 6/21/2020), day=159 (Sunday: 6/28/2020).

Memorial weekend: Sunday May 24, 2020: day = 124 .

4th of July: Saturday July 4, 2020: day = 165.



Numerical results (2)

Analysis of DC data for 2020. $p = 1$, $c_I = 0$, $c_Y = 1$

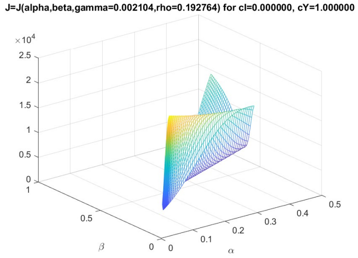
Results:
$\hat{\alpha} = 0.5$, $\hat{\beta} = 0.62$,

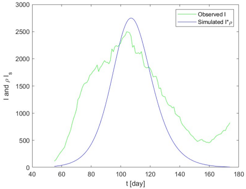
$\hat{R}_0 = 1.24$,

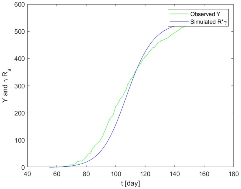
$\hat{\gamma} = 0.21\%$.

$\hat{\rho} = 19\%$

$J_{optim} = 2690$.







Numerical results (3)

Analysis of DC data for 2020. $p = 2$, $c_I = 0$, $c_Y = 1$

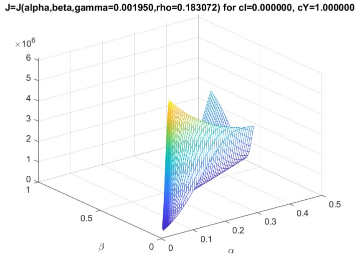
Results:
$\hat{\alpha} = 0.49$, $\hat{\beta} = 0.617$

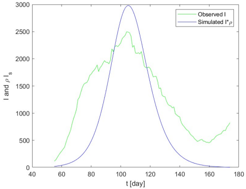
$\hat{R}_0 = 1.26$,

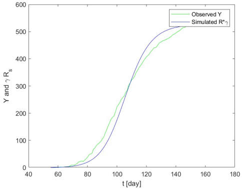
$\hat{\gamma} = 0.195\%$.

$\hat{\rho} = 18\%$

$J_{optim} = 89365$.







Numerical results (4)

Analysis of DC data for 2020. $p = \infty$, $c_I = 0$, $c_Y = 1$

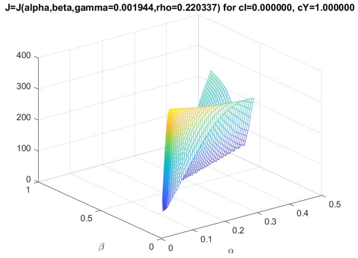
Results:
$\hat{\alpha} = 0.5$, $\hat{\beta} = 0.63$,

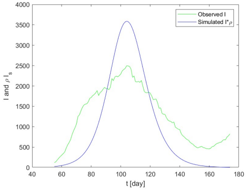
$\hat{R}_0 = 1.26$,

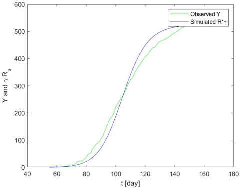
$\hat{\gamma} = 0.19\%$.

$\hat{\rho} = 22\%$

$J_{optim} = 44.98$.







Numerical results (5)

Analysis of DC data for 2020. $p = 1$, $c_I = 1$, $c_Y = 1$

Results:

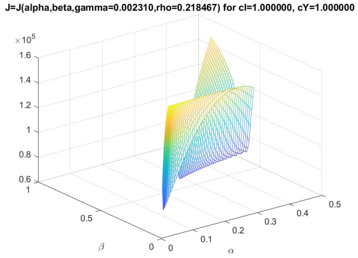
$\hat{\alpha} = 0.5$, $\hat{\beta} = 0.61$,

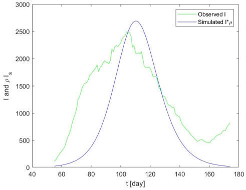
$\hat{R}_0 = 1.22$,

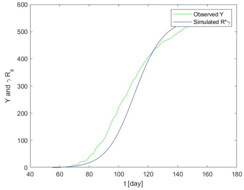
$\hat{\gamma} = 0.23\%$.

$\hat{\rho} = 22\%$

$J_{optim} = 67250$.

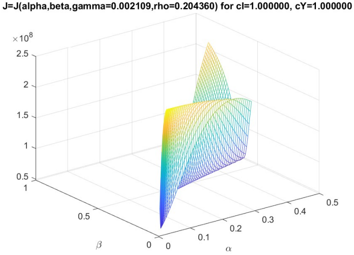






Numerical results (6)

Analysis of DC data for 2020. $p = 2$, $c_I = 1$, $c_Y = 1$



Results:

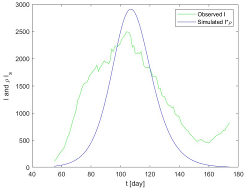
$\hat{\alpha} = 0.5$, $\hat{\beta} = 0.62$

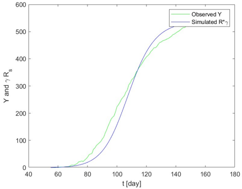
$\hat{R}_0 = 1.24$,

$\hat{\gamma} = 0.21\%$.

$\hat{\rho} = 20\%$

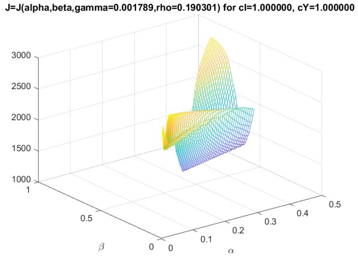
$J_{optim} = 48340173$.





Numerical results (7)

Analysis of DC data for 2020. $p = \infty$, $c_I = 1$, $c_Y = 1$



Results:

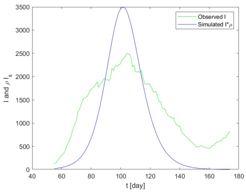
$\hat{\alpha} = 0.5$, $\hat{\beta} = 0.64$,

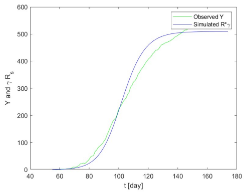
$\hat{R}_0 = 1.28$,

$\hat{\gamma} = 0.18\%$.

$\hat{\rho} = 19\%$

$J_{optim} = 1232$.





Next Model: SEIR

The Susceptible-Exposed-Infected-Removed (SEIR) model is obtained from SIR by introducing a compartment between Susceptible and Infected of population that has been exposed to the virus but are not yet contagious:

$$\left\{ \begin{array}{rcl} \frac{dS}{dt} & = & -\beta \frac{SI}{N} \quad , \quad S(0) \\ \frac{dE}{dt} & = & \beta \frac{SI}{N} - \delta E \quad , \quad E(0) \\ \frac{dI}{dt} & = & \delta E - \alpha I \quad , \quad I(0) \\ \frac{dR}{dt} & = & \alpha I \quad , \quad R(0) \end{array} \right. \qquad \text{(SEIR Model)}$$

where $\delta \geq 0$ is the rate of transition from exposed to infected. Its reciprocal $1/\delta$ represents the *average incubation period*.
If data is selected from the onset of infections, a natural initial condition is: $R(0) = N \gg I(0)$, $R(0) = 0$. The initial exposed population $E(0)$ may be set to $I(0)$, or can be fine tuned to fit the data. Assumin $E(0) = I(0)$ is known, the parameters that need to be calibrated are: $\alpha, \beta, \gamma, \delta, \rho$.