New Approach to the SIR Inversion Problem:
From the 1905-1906 Plague Outbreak in the Isle of Bombay
To the 2021-2022 Omicron Surges in New York City and Los Angeles County

Jeffrey E. Harris MD PhD*†§ March 13, 2023

Abstract. We describe a novel approach to recovering the underlying parameters of the SIR dynamic epidemic model from observed data on case incidence or deaths. We formulate a discrete-time approximation to the original continuous-time model and search for the parameter vector that minimizes the standard least squares criterion function. We show that the gradient vector and matrix of second-order derivatives of the criterion function with respect to the parameters adhere to their own systems of difference equations and thus can be exactly calculated iteratively. Applying our new approach, we estimate four-parameter SIR models on two types of datasets: (1) daily reported cases of COVID-19 during the SARS-CoV-2 Omicron/BA.1 surges of December 2021 - March 2022 in New York City and Los Angeles County; and (2) weekly deaths from a plague outbreak on the Isle of Bombay during December 1905 - July 1906, originally studied by Kermack and McKendrick in their now-classic 1927 paper. The estimated parameters from the COVID-19 data suggest a duration of persistent infectivity beyond that reported in small-scale clinical studies of mostly symptomatic subjects. The estimated parameters from the plague data suggest that the Bombay outbreak was in fact driven by pneumonic rather than bubonic plague. (198 words)

Key words. SIR model; inversion problem; Newton-Raphson algorithm; nonlinear least squares; COVID-19; SARS-CoV-2; Omicron variant; bubonic plague; pneumonic plague; structural models; reduced form models; parameter identification; underreporting; heterogeneous mixing

* Professor of Economics Emeritus, Massachusetts Institute of Technology, Cambridge MA USA 02139; and Physician, Eisner Health, Los Angeles CA USA 90015.
† Email: jeffrey@mit.edu
§ See Author Declarations at the end of the manuscript for additional details.

Introduction

Nowadays our sophisticated graphic software can draw attractive plots showing how many people have fallen victim to a highly contagious disease over the course of days, weeks or months. But our graphs alone don't teach us how to reliably determine the underlying risk of transmission from an infected to a susceptible person, or the amount of time that an infected individual remains contagious to others, or what proportion of the population was already infected at the critical point in time when the epidemic wave took off, or how many people remain at risk of infection.

We've just described what mathematicians call the inversion problem [1-4]: how to work backwards from limited data on incident cases or deaths to recover the key parameters underlying our dynamic epidemic models. The problem was born nearly a century ago when Kermack and McKendrick fit a curve derived from their now-classic model to datapoints of weekly deaths from a plague outbreak on the Isle of Bombay [5]. Since then, scores of investigators have searched for a robust, workable method of estimating the parameters of what has famously come to be known as the SIR (Susceptible-Infected-Removed) model, and the race to find a solution has accelerated with the arrival of the COVID-19 epidemic.

What has made the inversion problem so difficult is that, with some possible exceptions [6-10], the SIR model of coupled differential equations does not admit a closed-form mathematical solution that can be readily used to test the model's predictions against the observed data. That major stumbling block has left us with a motley collection of second-best alternatives.

One idea has been to back out the parameters from the certain salient characteristics of the observed epidemic curve, such as the initial exponential rate of increase of cases [11], the time to reach the peak incidence [12], the rate of decline after the peak [13], and the proportion of the population that is ultimately infected [14-16]. This approach may give us point estimates of the key parameters, but it does not provide any uncertainty ranges around the estimates.

Another idea is to pare down the set of parameters to be identified by making judicious use of prior information on some parameters [17-20]. Perhaps the most traveled road to a solution has been the use of various parameter search algorithms [17, 21-27] which, when it comes down to it, offer only a marginal improvement over brute force search [28]. Bayesian estimation may be better able to integrate prior information into our search procedures [29-34],

but its computational burden is usually even greater. Last but not least, we can resort to trial and error combined with visual inspection [35].

The present study, we suggest, offers an easily workable solution to the inversion problem. Rather than seeking a closed-form, analytical solution to Kermack and McKendrick's system of differential equations, we pursue an alternate strategy. First, following the lead of other investigators [33, 36-38], we develop a discrete-time version of their classic, continuous-time SIR model. This step allows us to write their dynamic system in terms of difference equations rather than differential equations. Second, similarly following in others' footsteps [24, 39-42], we define a minimum least squares objective function to test our SIR model's predictions against the observed data.

Third, in what appears to be an innovation, we show that both the gradient vector and Hessian matrix of second-order derivatives of our objective function with respect to the parameters follow their own systems of difference equations. As a result, both the gradient and Hessian can be rapidly and exactly computed by straightforward iteration, an approach that is computationally far superior to numerical approximation [43].

Fourth, once we have calculated the gradient and Hessian, we can use the well-known Newton-Raphson [44] or Gauss-Newton [45] algorithms to find the global optimum. Fifth, relying on the minimum least squares criterion, we can then calculate the variance-covariance matrix of the parameters and thus determine their confidence intervals [46]. Sixth, our approach permits us to readily determine what parameters are in fact identified when we have only time-series data on new cases or deaths.

We apply our strategy to the estimation of a four-parameter SIR model to two types of datasets. First, we study COVID-19 incidence over a 99-day interval from the December 4, 2021, through March 12, 2022, during the Omicron/BA.1 wave in two separate, large metropolitan areas in the United States: New York City and Los Angeles County. Second, we study 31 weeks of data on deaths from the plague from December 17, 1905, through July 15, 1906, during an outbreak in the Isle of Bombay, originally studied by Kermack and McKendrick [5].

Statistical Methods

Discrete-Time SIR Model

Following the lead of other investigators [4, 33, 36-38], we adopt a discrete-time approach. In our SIR epidemic model, the time axis is marked off in equally spaced intervals $t = 0,1,\ldots,T$, where the duration of each interval is sufficiently small as to adequately approximate the classical, continuous-time version [5, 47-50]. At any time t, individuals within this closed population can be in one of three mutually exclusive states: *susceptible* (*S*), *infected* (*I*), or *removed* (*R*). The latter state, which includes both recovered living individuals and decedents, is assumed to be absorbing. To minimize possible complications arising from the non-identifiability of multiple parameters [51], we assume a fixed, demographically closed population of size N.

Let S_t, I_t, and R_t denote the respective numbers of individuals in each of the three states at time t. The dynamic path of the epidemic is governed by the following coupled difference equations:

$$S_t = S_{t-1} - \beta S_{t-1} I_{t-1}/N \tag{1a}$$

$$I_t = I_{t-1} + \beta S_{t-1} I_{t-1}/N - \gamma I_{t-1} \tag{1b}$$

$$R_t = R_{t-1} + \gamma I_{t-1} \tag{1c}$$

$$S_t + I_t + R_t = N \tag{1d}$$

Apart from the population size N, this dynamic system has two parameters: β and γ. In equations (1a) and (1b), $\beta > 0$ is an infection transmission parameter, while in equations (1b) and (1c), the parameter $\gamma > 0$ gauges the proportion of infected individuals who transition to the removed state at each time period. The multiplicative term $\beta S_{t-1} I_{t-1}/N$ in equations (1a) and (1b) reflects the law of mass action [52], whereby susceptible individuals become infected in proportion to their frequency of contact with currently infected individuals. All individuals within the population are assumed to mix homogeneously, with no subgroup of individuals mixing preferentially with any other subgroup.

The final equation (1d) reflects the constant size N of the population and is consistent with equations (1a) through (1c). Strictly speaking, one should adjust the size N of the mixing

population to take account of removals by death. Unless the overall death rate is substantial, this adjustment is usually ignored in model implementations.

To further simplify matters, we assume that initially all individuals are either susceptible or infected. Specifically, at $t = 0$, we have:

$$S_0 = (1 - i_0)N \tag{2a}$$

$$I_0 = i_0 N \tag{2b}$$

$$R_0 = 0 \tag{2c}$$

In equations (2a) and (2b), the additional parameter i_0 represents the proportion of the entire population of size N that is initially infected. We address the more general case where $R_0 > 0$ below.

In the dynamic system (1), the mean duration of infection is time-invariant and equal to $1/\gamma$. Given the initial conditions (2), the system will result in an epidemic wave so long as $(1 - i_0)\,\beta/\gamma > 1$, a well-known result known as the epidemic threshold theorem [22, 49, 53]. Assuming that $1 - i_0 \approx 1$, most authors write this epidemic threshold condition as $\mathcal{R}_0 > 1$, where $\mathcal{R}_0 = \beta/\gamma$ is the basic reproduction number [22, 50, 54, 55].

Parameter Estimation Problem

We do not have direct observations on the underlying state variables S_t, I_t, and R_t. If we had such data, our inversion problem would border on trivial [37, 56, 57]. Instead, we observe only the reported counts y_t of new infections at each time t. These counts represent observations on latent variables X_t, $t = 1, \ldots, T$, which from (1) correspond to:

$$X_t = \beta S_{t-1} I_{t-1}/N \tag{3}$$

Define the parameter $\alpha = 1 - \gamma$ as the proportion of infected individuals who remain infected from one time period to the next, so that the probability an infected individual remains infectious after τ time periods is α^τ. Given the definition of the latent variable X_t in (3), our dynamic system (1) can be redefined as:

$$S_t = S_{t-1} - X_t \tag{4a}$$

$$I_t = X_t + \alpha I_{t-1} \tag{4b}$$

So long as the state variables adhere to the condition that that $S_t + I_t + R_t = N$, an explicit difference equation for R_t is unnecessary.

We can now characterize our parameter estimation problem. Given observations y_t on the latent variables X_t, we want to estimate the unknown parameters β, α, i_0, and N. Our stumbling block is that we cannot express X_t as a closed-form function of these parameters. We know only that the latent variables X_t adhere to the dynamic system defined by (2), (3) and (4), which in turn depends on the parameters β, α, i_0, and N.

Solution Strategy: Nonlinear Least Squares

To proceed, we need some additional notation. Let $\boldsymbol{y} = (y_1, \dots, y_T)'$ and $\boldsymbol{X} = (X_1, \dots, X_T)'$, respectively, denote column vectors of the observed incidence data and the corresponding latent variables at each time t, where we use boldface symbols denote vectors or matrices. Let $\boldsymbol{\theta} = (\beta, \alpha, i_0)'$ denote the column vector of the unknown parameters excluding the population size N. For now, we condition on N, but as we discuss below, this additional parameter can be separately identified once $\boldsymbol{\theta}$ has been estimated. Let $\boldsymbol{X}(\boldsymbol{\theta})$ represent the functional dependence of the latent variables on the parameters, suppressing for now that the fact that $\boldsymbol{X}$ also depends on N. Finally, let the operator $\boldsymbol{D}$ denote the gradient of partial derivatives with respect to the parameters. For example, the column vector $\boldsymbol{DX_t} = \left(\frac{\partial X_t}{\partial \beta}, \frac{\partial X_t}{\partial \alpha}, \frac{\partial X_t}{\partial i_0}\right)'$ represents the gradient of partial derivatives of the latent variable X_t at time t.

We want to optimize some objective function $V\big(\boldsymbol{y}, \boldsymbol{X}(\boldsymbol{\theta})\big)$ with respect to $\boldsymbol{\theta}$. While $\boldsymbol{X}(\boldsymbol{\theta})$ has no closed-form expression, the time-specific gradient vectors $\boldsymbol{DX_t}$ of partial derivatives of X_t with respect to the elements of $\boldsymbol{\theta}$ follow a system of difference equations and can thus be readily calculated iteratively. The same goes for the Hessian matrices $\boldsymbol{D^2X_t}$ of second-order partial derivates with respect to the elements of $\boldsymbol{\theta}$. Once we have computed the gradient vectors $\boldsymbol{DX_t}$ and Hessian matrices $\boldsymbol{D^2X_t}$, we can use the Newton-Raphson [44] or Gauss-Newton [45] algorithms to optimize V. While some authors have gone so far as to recognize the significance of these gradient vectors [39], to date none appear to have recognized that they could be calculated iteratively.

To see how this solution strategy works, we specifically consider the nonlinear least squares optimization criterion $V\big(\boldsymbol{y}, \boldsymbol{X}(\boldsymbol{\theta})\big) = \big(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{\theta})\big)'\big(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{\theta})\big)$, which can be written in summation notation as:

$$V = \sum_{t=1}^{T}\left(y_t - X_t(\boldsymbol{\theta})\right)^2 \tag{5}$$

This criterion has been widely used in attempts to fit the SIR model to incidence data [24, 39-42]. It is well known that minimization of V is equivalent to computing the maximum likelihood estimate of $\boldsymbol{\theta}$ under the assumption that the residual errors $\varepsilon_t = y_t - X_t(\boldsymbol{\theta})$ are independently and identically normally distributed with mean zero. Below, we return to the more general maximum likelihood framework and consider other error distributions. For now, we focus on the minimum least squares criterion V defined in (5).

Calculating the Gradient of the Least Squares Optimization Criterion V

The gradient of the least-squares criterion V in equation (5) with respect to the parameter vector $\boldsymbol{\theta}$ is:

$$\boldsymbol{DV} = -2\sum_{t=1}^{T}(y_t - X_t)\,\boldsymbol{DX_t} \tag{6}$$

From (6), we learn that the gradient $\boldsymbol{DV}$ depends on the gradients $\boldsymbol{DX_t}$. Our next step is to derive specific expressions for the latter gradient terms. Taking the derivative of (3) with respect to $\boldsymbol{\theta}$, we get:

$$\boldsymbol{DX_t} = \left(S_{t-1}\frac{I_{t-1}}{N}\right)\boldsymbol{D\beta} + \left(\frac{\beta}{N}\right)I_{t-1}\boldsymbol{DS_{t-1}} + \left(\frac{\beta}{N}\right)S_{t-1}\boldsymbol{DI_{t-1}} \tag{7},$$

where the notation $\boldsymbol{D\beta}$ represents the unit column vector $\left(\frac{\partial\beta}{\partial\beta}, \frac{\partial\beta}{\partial\alpha}, \frac{\partial\beta}{\partial i_0}\right)' = (1,0,0)'$. Taking the derivative of (1a), we get:

$$\boldsymbol{DS_t} = \boldsymbol{DS_{t-1}}\left(1 - \left(\frac{\beta}{N}\right)I_{t-1}\right) - S_{t-1}\frac{I_{t-1}}{N}\boldsymbol{D\beta} - S_{t-1}\left(\frac{\beta}{N}\right)\boldsymbol{DI_{t-1}} \tag{8}$$

From (1b) and (3), we can write $I_t = \alpha I_{t-1} + X_t = \alpha I_{t-1} + \beta S_{t-1}I_{t-1}/N = \left(\alpha + \frac{\beta}{N}S_{t-1}\right)I_{t-1}$. Taking the derivative of this expression gives:

$$\boldsymbol{DI_t} = \left(\alpha + \frac{\beta}{N}S_{t-1}\right)\boldsymbol{DI_{t-1}} + I_{t-1}\left(\boldsymbol{D\alpha} + \frac{S_{t-1}}{N}\boldsymbol{D\beta}\right) + I_{t-1}\left(\frac{\beta}{N}\right)\boldsymbol{DS_{t-1}} \tag{9},$$

where $\boldsymbol{D\alpha}$ similarly represents the unit vector $(0,1,0)'$.

We next need to compute the initial values of the gradients $\boldsymbol{DS_0}$ and $\boldsymbol{DI_0}$. Since $I_0 = i_0 N$ and $S_0 = (1 - i_0)N$, we can write the gradients $\boldsymbol{DS_0} = -N\boldsymbol{Di_0}$ and $\boldsymbol{DI_0} = N\boldsymbol{Di_0}$, where the gradient $\boldsymbol{Di_0}$ simplifies to $(0,0,1)'$. Given these initial gradient values, we can use (8) and (9) to iteratively compute the vectors $\boldsymbol{DS_t}$ and $\boldsymbol{DI_t}$ for all $t = 1, \ldots, T$. Once these gradients have been

computed, we can then apply (7) to iteratively compute the corresponding gradients $\boldsymbol{DX_t}$. Those quantities in turn yield the gradient $\boldsymbol{DV}$ of our objective function through equation (6).

Calculating the Hessian Matrix

To obtain the Hessian matrix of second-order derivatives of V, we first need some additional notation. In general, for two column vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ of dimension $n \times 1$, we define the outer product $\boldsymbol{A} \cdot \boldsymbol{B}$ as the $n \times n$ matrix $\boldsymbol{AB'}$ formed by the cross-products of the elements of $\boldsymbol{A}$ and $\boldsymbol{B}$. (We note that $\boldsymbol{B} \cdot \boldsymbol{A}$ is the transpose of $\boldsymbol{A} \cdot \boldsymbol{B}$.) For $n = 3$, we have

$$\boldsymbol{A} \cdot \boldsymbol{B} = \boldsymbol{AB'} = \begin{bmatrix} a_1b_1 & a_1b_2 & a_1b_3 \\ a_2b_1 & a_2b_2 & a_2b_3 \\ a_3b_1 & a_3b_2 & a_3b_3 \end{bmatrix}$$

We further denote $\boldsymbol{A} \odot \boldsymbol{B} = (\boldsymbol{A} \cdot \boldsymbol{B}) + (\boldsymbol{B} \cdot \boldsymbol{A})$ as the symmetric matrix formed by the sum of the outer product and its transpose. That is,

$$\boldsymbol{A} \odot \boldsymbol{B} = \begin{bmatrix} a_1b_1 + b_1a_1 & a_1b_2 + b_1a_2 & a_1b_3 + b_1a_3 \\ a_2b_1 + b_2a_1 & a_2b_2 + b_2a_2 & a_2b_3 + b_3a_2 \\ a_3b_1 + b_3a_1 & a_3b_2 + b_3a_2 & a_3b_3 + b_3a_3 \end{bmatrix}$$

We now take the derivative of the gradient in (6) to obtain:

$$\boldsymbol{D^2V} = 2\sum_{t=1}^{T}(\boldsymbol{DX_t}) \cdot (\boldsymbol{DX_t}) - 2\sum_{t=1}^{T}(y_t - X_t)\boldsymbol{D^2X_t} \tag{10}$$

In equation (10), each term $(\boldsymbol{DX_t}) \cdot (\boldsymbol{DX_t})$ in the first summation is a 3×3 square matrix of the form

$$(\boldsymbol{DX_t}) \cdot (\boldsymbol{DX_t}) = \begin{bmatrix} \frac{\partial X_t}{\partial \beta}\frac{\partial X_t}{\partial \beta} & \frac{\partial X_t}{\partial \beta}\frac{\partial X_t}{\partial \alpha} & \frac{\partial X_t}{\partial \beta}\frac{\partial X_t}{\partial i_0} \\ \frac{\partial X_t}{\partial \alpha}\frac{\partial X_t}{\partial \beta} & \frac{\partial X_t}{\partial \alpha}\frac{\partial X_t}{\partial \alpha} & \frac{\partial X_t}{\partial \alpha}\frac{\partial X_t}{\partial i_0} \\ \frac{\partial X_t}{\partial i_0}\frac{\partial X_t}{\partial \beta} & \frac{\partial X_t}{\partial i_0}\frac{\partial X_t}{\partial \alpha} & \frac{\partial X_t}{\partial i_0}\frac{\partial X_t}{\partial i_0} \end{bmatrix} \tag{11}$$

From (10), we similarly learn that the Hessian $\boldsymbol{D^2V}$ depends on the gradients $\boldsymbol{DX_t}$, which we have already derived, and the Hessian matrices $\boldsymbol{D^2X_t}$.

We turn to the corresponding iterative formulas for the Hessian matrices $\boldsymbol{D^2X_t}$. Taking the derivative of $\boldsymbol{DX_t}$ as displayed in (7), and making use of the previous section's results, we get:

$$\boldsymbol{D^2X_t} = \left(\frac{1}{N}\right)(S_{t-1}\boldsymbol{DI_{t-1}} + I_{t-1}\boldsymbol{DS_{t-1}})\odot(\boldsymbol{D\beta}) + \left(\frac{\beta}{N}\right)\left((\boldsymbol{DI_{t-1}})\odot(\boldsymbol{DS_{t-1}}) + S_{t-1}\boldsymbol{D^2I_{t-1}} + I_{t-1}\boldsymbol{D^2S_{t-1}}\right) \tag{12}$$

With the gradient $\boldsymbol{D\beta}$ equal to $(1,0,0)'$, the first term on the right-hand side of (12) becomes the symmetric 3×3 matrix

$$\left(\frac{1}{N}\right)\begin{bmatrix} 2\left(S_{t-1}\frac{\partial I_{t-1}}{\partial \beta} + I_{t-1}\frac{\partial S_{t-1}}{\partial \beta}\right) & \left(S_{t-1}\frac{\partial I_{t-1}}{\partial \alpha} + I_{t-1}\frac{\partial S_{t-1}}{\partial \alpha}\right) & \left(S_{t-1}\frac{\partial I_{t-1}}{\partial i_0} + I_{t-1}\frac{\partial S_{t-1}}{\partial i_0}\right) \\ \left(S_{t-1}\frac{\partial I_{t-1}}{\partial \alpha} + I_{t-1}\frac{\partial S_{t-1}}{\partial \alpha}\right) & 0 & 0 \\ \left(S_{t-1}\frac{\partial I_{t-1}}{\partial i_0} + I_{t-1}\frac{\partial S_{t-1}}{\partial i_0}\right) & 0 & 0 \end{bmatrix} \tag{13}$$

Following the same procedure, we obtain the corresponding expressions for the Hessian matrices $\boldsymbol{D^2S_t}$ and $\boldsymbol{D^2I_t}$:

$$\boldsymbol{D^2S_t} = \left(1 - \left(\frac{\beta}{N}\right)I_{t-1}\right)\boldsymbol{D^2S_{t-1}} - \left(\frac{\beta}{N}\right)(\boldsymbol{DI_{t-1}})\odot(\boldsymbol{DS_{t-1}}) - \left(\left(\frac{\beta}{N}\right)S_{t-1}\right)\boldsymbol{D^2I_{t-1}} - \left(\frac{1}{N}\right)(S_{t-1}\boldsymbol{DI_{t-1}} + I_{t-1}\boldsymbol{DS_{t-1}})\odot(\boldsymbol{D\beta}) \tag{14}$$

$$\boldsymbol{D^2I_t} = \left(\alpha + \frac{\beta}{N}S_{t-1}\right)\boldsymbol{D^2I_{t-1}} + \left(\frac{\beta}{N}\right)(\boldsymbol{DI_{t-1}})\odot(\boldsymbol{DS_{t-1}}) + \left(\left(\frac{\beta}{N}\right)I_{t-1}\right)\boldsymbol{D^2S_{t-1}} + \left(\frac{1}{N}\right)(S_{t-1}\boldsymbol{DI_{t-1}} + I_{t-1}\boldsymbol{DS_{t-1}})\odot(\boldsymbol{D\beta}) + \boldsymbol{DI_{t-1}}\odot(\boldsymbol{D\alpha}) \tag{15}$$

To complete our calculations, we note that the initial values $\boldsymbol{D^2S_0}$ and $\boldsymbol{D^2I_0}$ are simply null matrices.

Taken together, equations (10), (12), (14) and (15), along with the initial null values of $\boldsymbol{D^2S_0}$ and $\boldsymbol{D^2I_0}$, permit us to iteratively compute the Hessian $\boldsymbol{D^2V}$, just as we outlined for the gradient $\boldsymbol{DV}$ above. Equipped with these calculations, we can then estimate the value of the parameter vector $\boldsymbol{\theta}$ that minimizes V through either the Newton-Raphson [44] or Gauss-Newton [45] method. We address the possibility of multiple local optima below.

Confidence Intervals

Our reliance on the nonlinear least squares criterion V in equation (5) allows us to readily compute confidence intervals for the estimated parameters [46]. In particular, the variance-covariance matrix of the parameters $\boldsymbol{\theta} = (\beta, \alpha, i_0)$ is given by $\boldsymbol{C} = s^2(\boldsymbol{DX_t} \cdot \boldsymbol{DX_t})^{-1}$, where $s^2 = V/T$ and the quantities $\boldsymbol{DX_t}$ and V are evaluated at the optimum. The standard errors are

then the square roots of the diagonal elements of $\boldsymbol{C}$. The symmetric 95 percent confidence intervals, based upon the assumption of an asymptotic normal distribution, would then be evaluated as ± 1.96 standard errors about the estimates. We can then use the Delta method [58] to compute the standard errors and confidence intervals around nonlinear functions of the parameters. These include the basic reproduction number $\mathcal{R}_0 = \beta/\gamma = \beta/(1-\alpha)$ and the mean duration of infection $1/\gamma = 1/(1-\alpha)$.

Multiple Local Optima; Highly Correlated Parameter Estimates

Our objective function V may not be globally convex. On the one hand, the classical SIR model predicts a single-peaked wave of incident cases X_t over time, so long as $\mathcal{R}_0 > 1$. On the other hand, the data y_t may have multiple peaks. In that case, an iterative search algorithm to locate the minimum least squares V may end up at a local rather than a global optimum. A good example is the two-peak plot of the 2001 Dengue fever outbreak in Havana [53]. Accordingly, care needs to be taken to run our search algorithms from various starting values for the parameters. To address the possibility of nonconvexities and ensure that we ended up at a global optimum, we plotted the least squares criterion V as well as its gradient $\boldsymbol{DV}$ as functions of the individual parameters.

While our objective function V may indeed have a global optimum, and while the neighborhood surrounding the optimum may indeed be convex, we may still encounter an inverted ridge or ravine where the parameters are highly correlated. To some extent, we can check for such parameter correlation by inspecting the variance-covariance matrix $\boldsymbol{C}$. Here, we go further and construct heat plots of the level sets of V for various pairs of parameters.

Recovering the Population Size Parameter N

While $\boldsymbol{X}(\boldsymbol{\theta})$ has no closed-form expression, it turns out that each its elements X_t is a linear function of the population size parameter N. That is, X_t can be written in the form $\varphi_t(\boldsymbol{\theta})N$, where $\varphi_t(\boldsymbol{\theta})$ is a function of the remaining parameters $\boldsymbol{\theta}$.

To prove this result, we only need to show that the state variables S_t and I_t are themselves proportional to N. Thus, if for all times $t = 1, \ldots, T$, we can show that $S_t = f_t(\boldsymbol{\theta})N$ and $I_t = g_t(\boldsymbol{\theta})N$ for some functions $f_t(\boldsymbol{\theta})$ and $g_t(\boldsymbol{\theta})$, then from equation (3), we would have $X_t = \beta S_{t-1} I_{t-1}/N = \beta f_{t-1}(\boldsymbol{\theta}) g_{t-1}(\boldsymbol{\theta}) N$. Thus, $\varphi_t(\boldsymbol{\theta}) = \beta f_{t-1}(\boldsymbol{\theta}) g_{t-1}(\boldsymbol{\theta})$.

We can prove that S_t and I_t are proportional to N by mathematical induction. From equations (2a) and (2b), respectively, we known that $S_0 = (1 - i_0)N$ and $I_0 = i_0 N$. So, the proposition is true for $t = 0$. Now suppose that $S_t = f_t(\boldsymbol{\theta})N$ and $I_t = g_t(\boldsymbol{\theta})N$, at time t. We claim that $S_{t+1} = f_{t+1}(\boldsymbol{\theta})N$ and $I_{t+1} = g_{t+1}(\boldsymbol{\theta})N$ necessarily hold for some functions $f_{t+1}(\boldsymbol{\theta})$ and $g_{t+1}(\boldsymbol{\theta})$. From (1a), we can write $S_{t+1} = S_t\left(1 - \frac{\beta}{N} I_t\right)$. We have $S_{t+1} = S_t\left(1 - \frac{\beta}{N} I_t\right) = f_t(\theta)N\left(1 - \frac{\beta}{N} g_t(\theta)N\right) = \left(f_t(\theta) - \beta g_t(\theta)\right)N$, and so $f_{t+1}(\theta) = f_t(\theta) - \beta g_t(\theta)$. Similarly, $I_{t+1} = I_t\left(\alpha + \frac{\beta}{N} S_t\right) = g_t(\theta)N\left(\alpha + \frac{\beta}{N} f_t(\theta)N\right) = g_t(\theta)\left(\alpha + \beta f_t(\theta)\right)N$, and so $g_{t+1}(\theta) = g_t(\theta)\left(\alpha + \beta f_t(\theta)\right)$. ■

The fact that each latent variable X_t is proportional to N permits us to employ an iterative procedure based upon the expectation-maximization (or EM) algorithm [59] to recover an estimate of that parameter. Let's say that at iteration n of the algorithm, we already have parameter estimates $\boldsymbol{\theta}^{(n)}$ and $N^{(n)}$, which in turn give us estimates $X_t^{(n)} = \varphi_t\left(\boldsymbol{\theta}^{(n)}\right)N^{(n)}$ for all t. We now compute:

$$\kappa^{(n+1)} = \boldsymbol{y}'\boldsymbol{X}^{(\boldsymbol{n})}/\boldsymbol{X}^{(\boldsymbol{n})\prime}\boldsymbol{X}^{(\boldsymbol{n})} = \sum_{t=1}^{T} y_t X_t^{(n)} / \sum_{t=1}^{T}\left(X_t^{(n)}\right)^2 \tag{16}$$

That is, we compute an adjustment factor $\kappa^{(n+1)}$ by regressing $\boldsymbol{y}$ on $\boldsymbol{X}^{(\boldsymbol{n})}$. For the expectation step, we update the population size as $N^{(n+1)} = \kappa^{(n+1)} N^{(n)}$. Then, for the maximization step, we use our previously described procedure to estimate $\boldsymbol{\theta}^{(n+1)}$ conditional upon the updated population size $N^{(n+1)}$. This iterative procedure will work for any optimization criterion V and not just for the least squares criterion described above.

Non-Identifiability of the Initially Resistant Population

We have thus far assumed in equation (2c) that no one in the population was initially resistant to infection, that is, $R_0 = 0$. Now let's assume more generally that a non-negative fraction $1 > r_0 \geq 0$ of the population is already resistant at $t = 0$. We thus replace the restricted initial conditions (2) with the following more general initial conditions:

$$S_0 = (1 - i_0 - r_0)N \tag{17a}$$

$$I_0 = i_0 N \tag{17b}$$

$$R_0 = r_0 N \tag{17c}$$

So long as the R_0 initially resistant individuals mix homogeneously with those in the susceptible and infected states, equation (3) governing the latent variables X_t as well as the equations of motion (4) for the state variables S_t and I_t remain unchanged.

Apart from our original parameter vector $\boldsymbol{\theta} = (\beta, \alpha, i_0)$ and the population size N, it appears that we now have an additional parameter r_0 to be estimated. It turns out, however, that r_0 cannot be identified in our discrete SIR model from the observed data $\boldsymbol{y}$ alone [51].

To see why, let's suppose that we estimate our more general model conditional upon some fixed value of r_0. We denote the resulting minimum least squares estimates as $\boldsymbol{\theta}^*(r_0)$ and $N^*(r_0)$ to show their dependence on the value of r_0 chosen. We denote the corresponding individual components of $\boldsymbol{\theta}^*(r_0)$ as $\beta^*(r_0)$, $\alpha^*(r_0)$, and $i_0^*(r_0)$. Our original parameter estimates conditional upon $r_0 = 0$ thus correspond to $\beta^*(0)$, $\alpha^*(0)$, $i_0^*(0)$, and $N^*(0)$. Then the following conditions hold for all $1 > r_0 \geq 0$:

$$\beta^*(r_0) = \beta^*(0)/(1 - r_0) \tag{18a}$$

$$\alpha^*(r_0) = \alpha^*(0) \tag{18b}$$

$$i_0^*(r_0) = i_0^*(0)(1 - r_0) \tag{18c}$$

$$N^*(r_0) = N^*(0)/(1 - r_0) \tag{18d}$$

What's more, the paths of the latent variable X_t and the state variables S_t and I_t at the optimum values of the parameters, but not the state variable R_t, will be independent of r_0. Accordingly, the optimum value of V will likewise be independent of r_0.

We again rely upon mathematical induction to prove that the state variables S_t and I_t at the optimum are independent of r_0. First consider $t = 0$. For any value of r_0, we have from (17a) that $S_0 = (1 - i_0^*(r_0) - r_0)N^*(r_0)$. Applying (18c) and (18d), we get $S_0 = \left(1 - i_0^*(0)\right)N^*(0)$, which does not depend on r_0. Similarly for any value of r_0, we have from (17b) that $I_0 = i_0^*(r_0)N^*(r_0)$. Again applying (18c) and (18d), we get $I_0 = i_0^*(0)N^*(0)$, which is likewise independent of r_0. Now assume that S_t and I_t are independent of r_0. We show that S_{t+1} and I_{t+1} must also be independent of r_0. From (1a), we have $S_{t+1} = S_t(1 - \beta^*(r_0)I_t/N^*(r_0))$. Applying (18a) and (18d), this expression resolves to $S_{t+1} = S_t(1 - \beta^*(0)I_t/N^*(0))$. From (1b) and (3), we have $I_{t+1} = I_t(\alpha + \beta^*(r_0)S_t/N^*(r_0))$. Again applying (18a) and (18d), this expression similarly resolves to $I_{t+1} = I_t(\alpha + \beta^*(0)S_t/N^*(0))$. ■

Interpreting the Estimated Population Size N: Underreporting and Incomplete Mixing

While the additional parameter r_0 is not identifiable from the data $\boldsymbol{y}$ alone, we might still be able to identify it from other data. Suppose, for example, that we had sharp prior information on the basic reproduction number $\mathcal{R}_0 = \beta/\gamma = \beta/(1-\alpha)$. Let's denote this prior estimate $\bar{\mathcal{R}}_0$. We first compute the basic reproductive number $\mathcal{R}_0^*(0) = \beta^*(0)/\big(1-\alpha^*(0)\big)$ implied by our parameter estimates conditional upon $r_0 = 0$. So long as $\bar{\mathcal{R}}_0 \geq \mathcal{R}_0^*(0)$, we can then use our prior information to conclude that $r_0 = 1 - \mathcal{R}_0^*(0)/\bar{\mathcal{R}}_0$.

Why couldn't we similarly take advantage of prior information on the population size to ascertain the initial proportion r_0 of resistant individuals? Suppose we knew from census data that the population contained $\bar{N}$ individuals. We would first estimate $N^*(0)$ conditional upon $r_0 = 0$, and then, so long as $\bar{N} \geq N^*(0)$, we would have the estimate $r_0 = 1 - N^*(0)/\bar{N}$. Unfortunately, reliance solely on census data for prior estimates of population size N is complicated by two other phenomena: underreporting and incomplete mixing.

In many contexts, particularly in recent applications of the SIR and related compartmental models to COVID-19 incidence, it has been widely recognized that a significant number of incident cases may have gone unreported [60]. In the absence of reliable information on the temporal pattern of such underreporting, the most parsimonious approach to this phenomenon has been to assume a *case identification ratio* equal to a constant $p < 1$ [61-63]. In that case, our model would need to be modified to accommodate the reality that our reported incidence data y_t are in fact estimates of pX_t rather than X_t. We've already learned, however, that $X_t = \varphi_t(\boldsymbol{\theta})N$, where $\varphi_t(\boldsymbol{\theta})$ is a function of the remaining parameters $\boldsymbol{\theta}$, and this is the case even in our more general model admitting $r_0 > 0$. Accordingly, our reported incidence data y_t are really estimates of $pX_t = \varphi_t(\boldsymbol{\theta})(pN)$, and the estimated population size derived from our model is really an estimate of pN.

We thus have a knotty problem of confounding. We can estimate population size $N^*(0)$ assuming that there are no initially resistant individuals and no underreporting. If we had census data $\bar{N}$, we could account for both phenomena, writing $N^*(0)/\bar{N} = p(1-r_0)$. Without more information, we cannot separately identify p and r_0.

We would ordinarily interpret the parameter N to gauge the size of the population at risk for contagion. This population would consist of all individuals who homogeneously mix with each other in accordance with the law of mass action embodied in equations (1) and (3). Many

investigators, however, have properly recognized that the underlying assumption of homogeneous mixing may not apply to the entire population [18, 64-69].

Let's say we're analyzing an outbreak on a college campus with known student enrollment $\bar{N}$. When we run our model with $r_0 = 0$, we obtain an estimate $N^*(0)$ that is, say, only about 10 percent of $\bar{N}$. This finding does not necessarily imply that 90 percent of the student body was resistant or that only one in ten cases was reported. Instead, it may mean that only small fraction of the student body was directly involved in the mixing process that generated the outbreak. A good example would be the COVID-19 outbreak at the campus of the University of Wisconsin-Madison in September 2020, where total student enrollment was $\bar{N} = 44{,}640$, but where the large fraction of cases was concentrated in two on-campus student residence halls with a combined population of about 3,590 [70].

Any comparison between the estimated population size $N^*(0)$ and the census statistic $\bar{N}$ thus entails three confounded interpretations. Some cases have gone unreported. Some individuals may be initially resistant. And other individuals may be susceptible but remain outside the core population through which the infectious agent has spread.

Other Optimization Criteria

While we relied upon the minimum least squares criterion V, as defined in equation (5), to develop our estimation strategy, we could have employed other optimization criteria. To illustrate, consider the case where the observed observations $\boldsymbol{y}$ consist of count data on the number of reported cases, and assume that each observed data point y_t is independently Poisson distributed [62] with mean X_t. The joint likelihood of combined observations will be:

$$\mathcal{L} = \prod_{t=1}^{T} \frac{1}{y_t!} e^{-X_t} {X_t}^{y_t} \tag{19}$$

As before, we assume that each latent variable X_t is a function of the vector of underlying parameters $\boldsymbol{\theta}$. The log likelihood will then be:

$$\ell(\boldsymbol{\theta}) = -\sum_{t=1}^{T} X_t(\boldsymbol{\theta}) + \sum_{t=1}^{T} y_t \log X_t(\boldsymbol{\theta}) - K \tag{20}$$

In equation (20), the additional term $K = -\log(y_t!)$ does not depend on the latent variable X_t or the underlying parameters $\boldsymbol{\theta}$. The gradient of the log likelihood function will then be:

$$\boldsymbol{D\ell} = -\sum_{t=1}^{T} \boldsymbol{DX_t} + \sum_{t=1}^{T} \frac{y_t}{X_t} \boldsymbol{DX_t} = \sum_{t=1}^{T} \frac{1}{X_t}(y_t - X_t)\boldsymbol{DX_t} \tag{21}$$

The gradient expressed in (21) differs from the gradient in equation (6) only in terms of the weighting factors $\frac{1}{X_t}$ in each summation term. The Hessian matrix of second-order derivatives is:

$$\boldsymbol{D^2\ell} = \sum_{t=1}^{T} \frac{1}{X_t}(y_t - X_t)\boldsymbol{D^2X_t} - \sum_{t=1}^{T}\left(\frac{y_t}{X_t^2}\right)(\boldsymbol{DX_t}) \cdot (\boldsymbol{DX_t}) \tag{22}$$

While the Hessian matrix in equation (10) above was positive definite reflecting an objective function V to be minimized, here the Hessian matrix is negative definite reflecting an objective function ℓ to be maximized. The remaining computations of the gradients $\boldsymbol{DX_t}$ will be computed exactly as described above in the least squares case.

Adapting the Model to Data on Removals

Our analysis has thus far assumed that each data point y_t is an observation on the corresponding latent variable X_t capturing the number of individuals who transitioned from the *susceptible* to the *infected* state at time t. Alternatively, we can construct an analogous model where the data point y_t is instead an observation on a distinct latent variable Z_t capturing the number of individuals who transitioned from the *infected* to the *removed* state at time t. This approach would be important if we had data on the number of deaths during an epidemic of a disease with a near 100 percent fatality rate, as Kermack and McKendrick observed in the case of 1905-1906 plague outbreak on the Isle of Bombay [5].

In this alternative model, the latent variable Z_t is:

$$Z_t = R_t - R_{t-1} = (1-\alpha)I_t \tag{23}$$

Continuing with the least squares framework, the objective function is now:

$$V = \sum_{t=1}^{T}\left(y_t - Z_t(\boldsymbol{\theta})\right)^2 \tag{24}$$

To minimize this objective function employing the same strategy above, we will need to compute the gradient vector $\boldsymbol{DZ_t}$ and the Hessian matrix $\boldsymbol{D^2Z_t}$. From (24), we immediately have:

$$\boldsymbol{DZ_t} = (1-\alpha)\boldsymbol{DI_t} - I_t\boldsymbol{D\alpha} \tag{25}$$

The corresponding Hessian matrix is:

$$\boldsymbol{D^2Z_t} = (1-\alpha)\boldsymbol{D^2I_t} - \boldsymbol{DI_t} \odot \boldsymbol{D\alpha} \tag{26}$$

To compute these two expressions, we will need the gradient $\boldsymbol{DI_t}$ and Hessian $\boldsymbol{D^2I_t}$, but these respective quantities were already computed in equations (9) and (15) above. To complete our estimation procedure, we note that the latent variables are similarly linear functions of the population size parameter N, so that the EM algorithm described above likewise applies.

Working with Aggregate Data

In some applications, we may have only aggregate data reported over a coarse time scale, rather than the finely divided time axis assumed so far. For concreteness, let's assume that the underlying SIR epidemic model is valid when the time axis $t = 0,1, \dots, T$ is marked off in days, but we only have data on the cumulative number of deaths during each 7-day week, indexed by $m = 1, \dots, M$. Days t are mapped into weeks m by the relation $m = h(t) = \left\| \frac{t-1}{7} \right\| + 1$, where the floor operator $\|x\|$ maps into the largest integer j such that $j \leq x$. We further assume that $T = 7M$, so that week M, which corresponds to the last week, ends on day T.

Now define the $M \times T$ aggregation matrix $\boldsymbol{W}$ with element $w_{mt} = 1$ if and only if $m = h(t)$. Otherwise, $w_{mt} = 0$. While our finely disaggregated model generates daily values Z_t of the latent variable from the underlying parameters $\boldsymbol{\theta}$, our data y_m represent observations on the aggregated values $\sum_{t=1}^{T} w_{mt} Z_t$. (From the computational standpoint, we're calculating the integral required to convert a continuous- to a discrete-time SIR model [38].) In vector notation, our least squares criterion would be:

$$V = (\boldsymbol{y} - \boldsymbol{WZ})'(\boldsymbol{y} - \boldsymbol{WZ}) \tag{27}$$

The corresponding gradient is**:**

$$\boldsymbol{DV} = -2\ (\boldsymbol{y} - \boldsymbol{WZ})'\boldsymbol{WDZ} \tag{28}$$

The corresponding Hessian matrix becomes:

$$\boldsymbol{D^2V} = -2\ (\boldsymbol{y} - \boldsymbol{WZ})'\boldsymbol{WD^2Z} + 2(\boldsymbol{WDZ}) \cdot (\boldsymbol{WDZ}) \tag{29}$$

Accordingly, we can run our SIR model to generate the state variables Z_t and then calculate the derivatives $\boldsymbol{DZ_t}$ and $\boldsymbol{D^2Z_t}$ in the disaggregated time scale and then apply the aggregation matrix $\boldsymbol{W}$ to weight the results.

Data

Omicron Wave, December 2021 – March 2022: New York City and Los Angeles County

We studied the reported daily incidence of COVID-19 during the SARS-CoV-2 Omicron/BA.1 wave of December 2021 – March 2022. We analyzed data publicly posted by governmental authorities in two large metropolitan areas in the United States: New York City NY, encompassing Manhattan, Bronx, Brooklyn, Queens, and Staten Island (population 8.49 million) [71]; and Los Angeles County CA, excluding the cities of Pasadena and Long Beach (9.26 million) [72]. In both posted databases, the date of report was intended to be the date when a positive test was performed or when the diagnosis of COVID-19 was otherwise made.

Both databases showed systematic variation in case counts by day of the week, with many fewer cases diagnosed over the weekends. To account for these day-of-the-week-dependent fluctuations, we converted the raw case counts c_t into centered 7-day moving averages, that is, $y_t = \frac{1}{7}\sum_{i=-3}^{+3} c_{t+i}$, where t indexes the date of report. Figure 1A shows the raw counts of reported cases per day c_t (connected gray datapoints) as well as the daily case counts y_t adjusted for the day of the week (red datapoints) for New York City. Figure 1B shows the corresponding raw case counts and day-of-the-week-adjusted case counts for Los Angeles County.

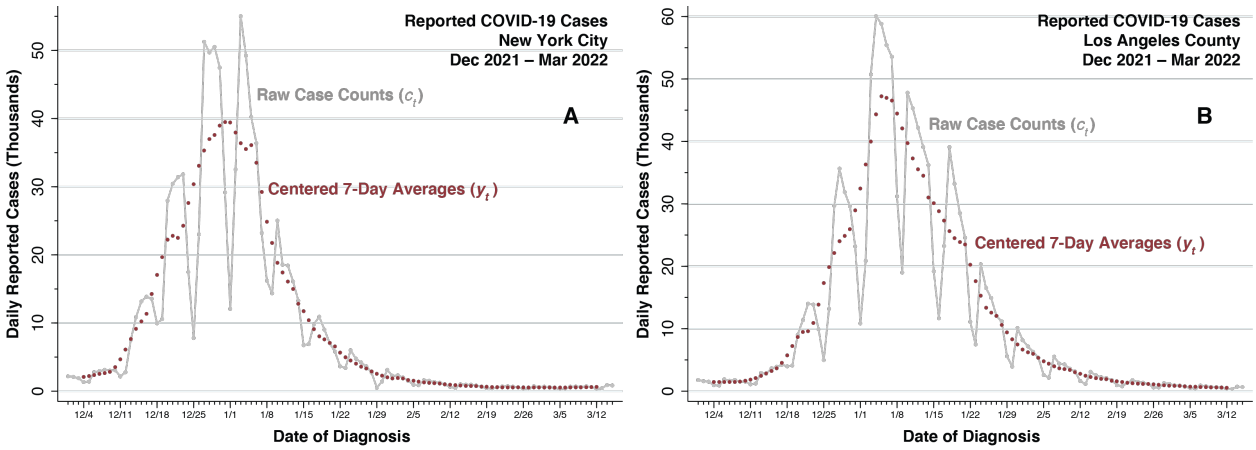

Figure 1. Daily Reported Cases of COVID-19, 12/1/2021 – 3/15/2022, in New York City (Panel A) and Los Angeles County (Panel B). The connected gray datapoints show the raw case counts (c_t). The red datapoints, covering 12/4/2021 – 3/12/2022, show the centered 7-day moving averages (y_t).

While the date of initial upslope in case incidence and the timing of peak incidence varied between the two locations, we employed same 99-day interval from December 4, 2021, through March 12, 2022, to study both datasets. We relied upon the day-of-the-week-adjusted case counts y_t as measures of daily incidence corresponding to the latent variables X_t defined in equation (3) of our SIR model. We estimated the underlying parameters of the model by minimizing the least squares criterion V defined in equation (5) above.

Plague Outbreak, December 1905 – July 1906, Isle of Bombay

Figure 2 displays the average daily number of reported deaths from plague in the Isle of Bombay during each week over a 31-week period, beginning with the week starting Sunday, December 17, 1905, and continuing through the week starting Sunday, July 15, 1906. The average daily deaths were computed from the weekly totals plotted by Kermack and McKendrick in their classic 1927 paper [5]. The height of each bar is the average daily number of deaths during that week. The area of each bar equals the reported number of deaths during that week.

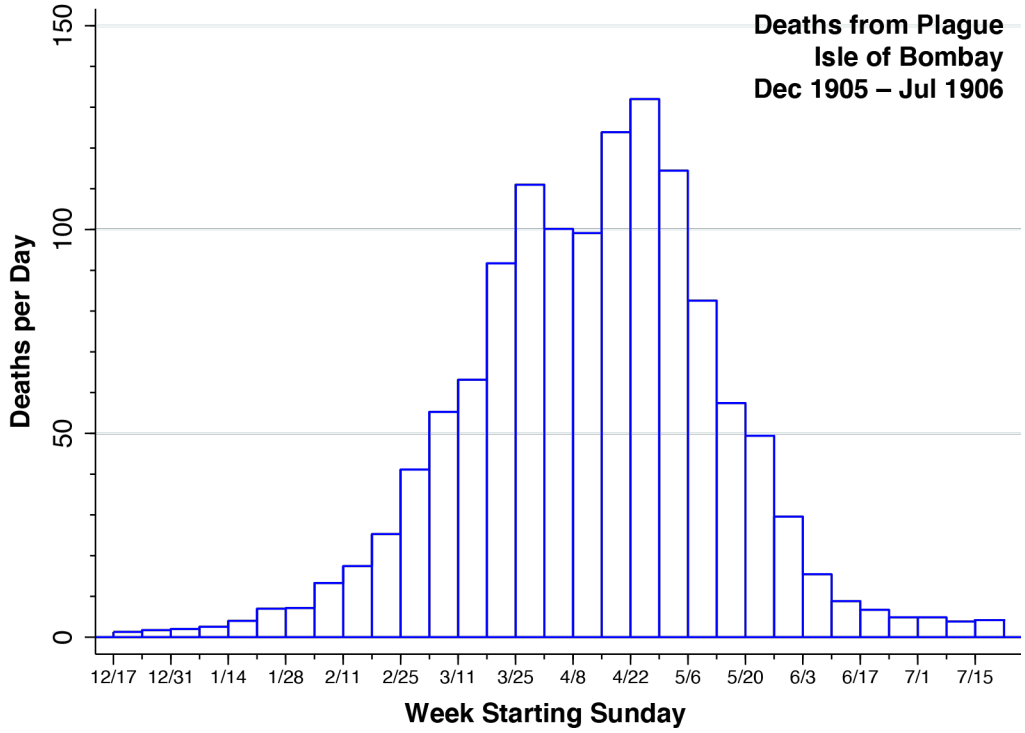

Figure 2. Average Number of Daily Deaths from Plague, Isle of Bombay, Week Starting 12/17/1905 Through the Week Starting 7/15/1906. Each bar corresponds to one week. The height of each bar is the average daily number of deaths during that week. The area of each bar equals the number of deaths during that week. Data adapted from Kermack and McKendrick [5].

In our analysis of the plague death data, we interpreted the weekly death counts y_m, indexed by $m = 1, \ldots, 31$, as observations on the aggregated values $\sum_{t=1}^{T} w_{mt} Z_t$, where Z_t are the underlying latent variables with daily index $t = 1, \ldots, 217$, defined in equation (23), and where w_{mt} are elements of the aggregation matrix $\boldsymbol{W}$ described above. We estimated the underlying parameters of our SIR model by minimizing least squares criterion V defined in equation (27).

Data Processing and Computation

The raw COVID-19 case counts (c_t) were downloaded from the data portals of the four metropolitan areas [71, 72] in comma-separated-value (CSV) format and converted via Stata Statistical Software Release 17 [73] into internal Stata (DTA) format. The 7-day moving averages (y_t), displayed in Figure 1, were computed via Stata programming language [74]. The data on weekly deaths from plague in the Isle of Bombay were taken directly from the graph appearing on page 714 of the original Kermack-McKendrick paper [5]. All code for parameter estimation described in the Statistical Methods section above was written in Mata [75], a matrix programming language embedded within Stata. All calculations were carried out on a MacBook Pro with a 2.3 GHz 8-Core Intel Core i9 processor.

Results

Omicron Wave, December 2021 – March 2022, New York City and Los Angeles County

Figure 3 superimposes the predicted values of the latent variable X_t, drawn as a connected curve, on the observed data y_t, plotted as distinct datapoints, for New York City (Panel A) and Los Angeles County (Panel B). The fit of the data to the four-parameter SIR model is extraordinarily tight for New York City. For Los Angeles County, however, the model undershoots the peak reported incidence in early January 2022 by about 8 percent.

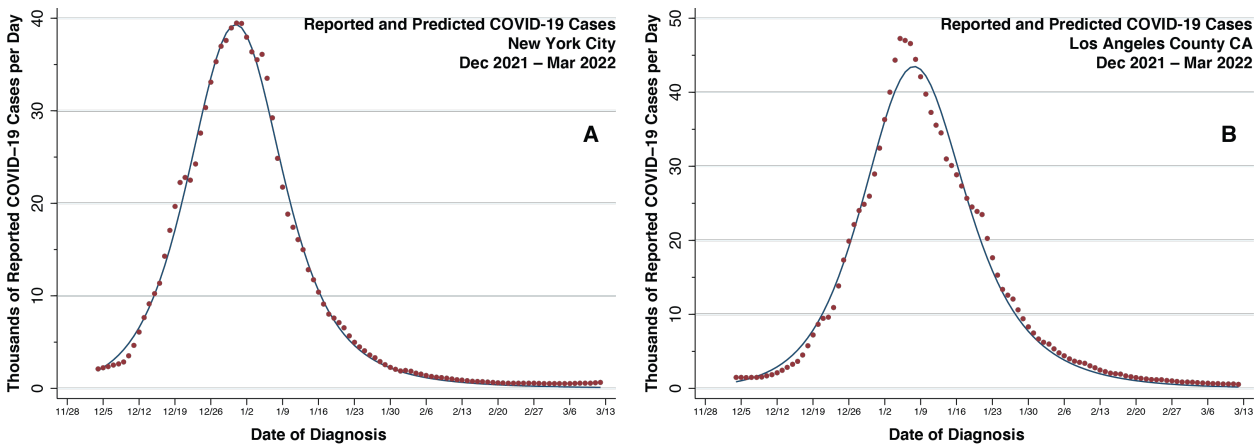

Figure 3. Daily Reported and Predicted Cases of COVID-19, 12/4/2021 – 3/12/2022, for New York City (Panel A) and Los Angeles County (Panel B). The red datapoints show the data y_t reproduced from the corresponding panels in Figure 1, while the curves connect the predicted values of the latent variables X_t.

Figures 4 and 5 focus further on the results for New York City. Figure 4A plots the least squares criterion V (blue curve, left axis) and the first partial derivative $\partial V/\partial\beta$ (red curve, right axis) as functions of the parameter β. The remaining parameters were held constant at their optimum values, to be reported below. The criterion V reaches a minimum at the optimum β = 0.233, at which point $\partial V/\partial\beta$ equals zero. The function V is convex in the interval from β = 0.169, where $\partial V/\partial\beta$ reaches a minimum, to β= 0.352, where $\partial V/\partial\beta$ reaches a maximum.

Figure 4B displays the analogous plot of V and $\partial V/\partial\alpha$ as functions of the parameter α, where the remaining parameters are similarly held constant at their optimum values. The criterion V reaches a minimum at the optimum α = 0.941, at which point $\partial V/\partial\alpha$ equals zero. The function V is convex in the interval from α = 0.868, where $\partial V/\partial\alpha$ reaches a minimum, to α = 1, the boundary of admissible values of α, where $\partial V/\partial\alpha$ remains positive.

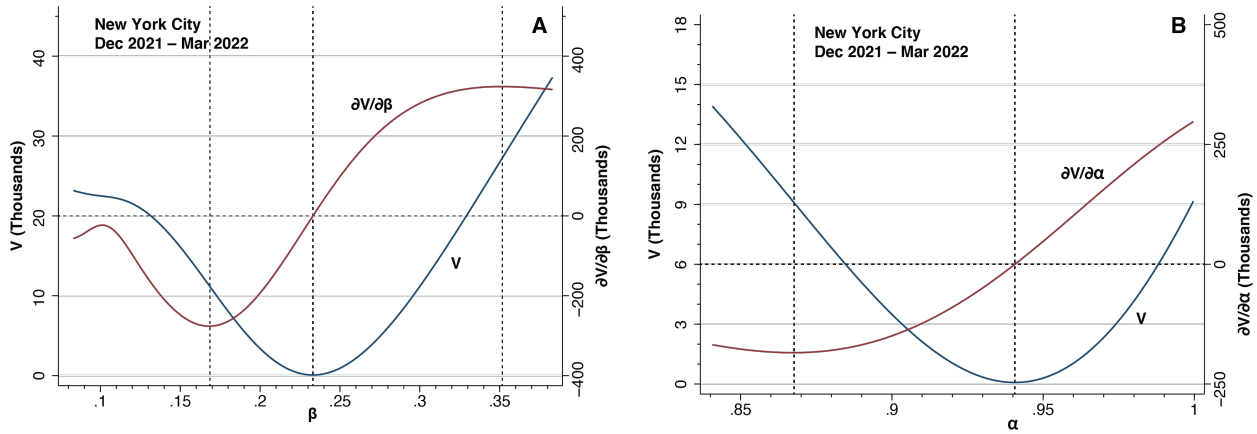

Figure 4. Panel A. Least Squares Criterion V (Left Axis) and First Partial Derivative $\partial V/\partial\beta$ (Right Axis) as Functions of the Parameter β. Panel B. Least Squares Criterion V (Left Axis) and First Partial Derivative $\partial V/\partial\alpha$ (Right Axis) as Functions of the Parameter α.

Figure 5 plots a heat map of the level sets of V in the (α, β) plane. Again, the remaining parameters (i_0, N) were held at their optimum values. The darkest area running diagonally from the upper left to the lower right represents a ravine where V attains its lowest values, while the sky-blue regions at the lower left and upper right corners represent the highest elevations. The yellow point in the center is the optimum $(\alpha, \beta) = (0.941, 0.233)$ where V is minimized.

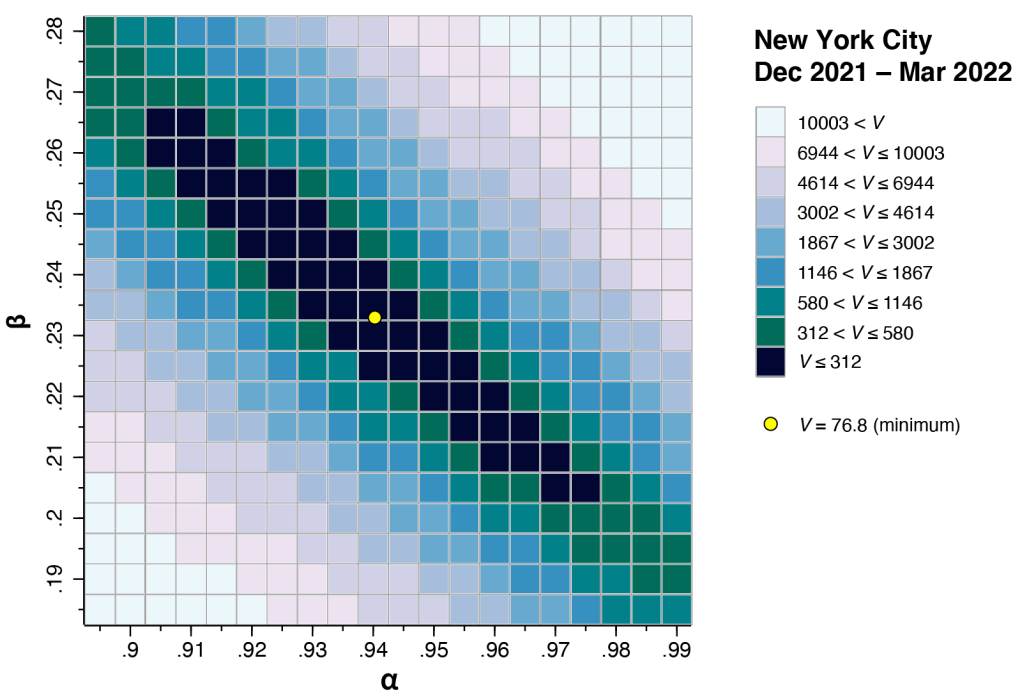

Figure 5. Heat Map of the Level Sets of the Least Squares Criterion V in the (α, β) Plane. The yellow point in the center of the map is the optimum $(\alpha, \beta) = (0.941, 0.233)$ where V is minimized.

Table 1 summarizes the parameter estimation results. By December 3, 2021, which corresponds to date $t = 0$ in both model applications, the estimated infected proportion i_0 was already up to about 0.8 percent in New York City, while in Los Angeles County, the corresponding proportion was still on the order of 0.2 percent. The estimated basic reproduction number $\mathcal{R}_0$ for New York City was significantly larger than the corresponding estimate for Los Angeles County. While the estimated values of the parameter β were comparable across the two sites, the mean duration of infectivity $1/(1-\alpha)$ was significantly higher in New York City. The estimated population-size parameters N were 11.9 percent of the total census population of New York City and 14.5 percent of the total census population of Los Angeles County excluding the cities of Pasadena and Long Beach. On the author's laptop, with a 2.3 GHz 8-Core Intel Core i9 processor, the respective computation times were 0.93 sec for New York City and 1.04 sec for Los Angeles County.

Table 1. Parameter Estimates of an SIR Model of COVID-19 Incidence New York City and Los Angeles, Omicron Wave, December 2021 – March 2022 [a,b]

Parameter	New York City	Los Angeles County
$\boldsymbol{\beta}$	0.233 (0.216, 0.250)	0.272 (0.254, 0.290)
$\boldsymbol{\alpha}$	0.941 (0.928, 0.954)	0.890 (0.876, 0.904)
$\boldsymbol{i_0 \times 10^{-3}}$	8.23 (6.96, 9.51)	2.44 (1.98, 2.90)
$\boldsymbol{N \times 10^6}$	1.013	1.346
$\boldsymbol{\mathcal{R}_0}$	3.93 (3.35, 4.50)	2.49 (2.32, 2.65)
$\boldsymbol{1/(1-\alpha)}$	16.8 (13.2, 20.5)	9.11 (7.93, 10.3)

a. Except for the population size parameter N, 95% confidence intervals are displayed below each estimate.
b. In addition to the four model parameters, the final two rows show the estimated basic reproduction number $\mathcal{R}_0 = \beta/(1-\alpha)$ and the estimated mean duration of infectivity $1/(1-\alpha)$.

Plague Outbreak, December 1905 – July 1906, Isle of Bombay

Figure 6 reports our analysis of data on deaths from plague during December 1905 – July 1906 in the Isle of Bombay, originally reported and studied by Kermack and McKendrick [5].

As in Figure 2 above, we have displayed the raw data on deaths as a bar graph, where each vertical bar covers one week, indexed $m = 1, \ldots, 31$, where the height of each bar is the reported average daily deaths that week, and where the area of each bar is the reported number of weekly deaths (y_m). The superimposed green curve represents the fit to the data, based upon the approximation $\frac{800}{7}\,\mathrm{sech}^2(0.2t - 3.4)$, where sech represents the hyperbolic secant, and where we have divided their published formula by 7 to convert their estimates into daily units. The green curve thus corresponds to the authors' estimates of the underlying latent variable Z_t.

The red curve in Figure 6 connects our predictions of the daily values Z_t for $t = 1, \ldots, 217$. The underlying parameter estimates were: $\beta = 0.610$, with 95% confidence interval (0.537, 0.683); $\alpha = 0.446$ (0.376, 0.515); $i_0 = 2.80 \times 10^{-5}$ ($1.26 \times 10^{-5}, 4.34 \times 10^{-5}$); and $N = 5.103 \times 10^4$. The estimated basic reproduction number was $\mathcal{R}_0 = 1.101$ (1.094, 1.107), while the estimated mean duration in the infected state was $1/(1-\alpha) = 1.804$ (1.579, 2.031).

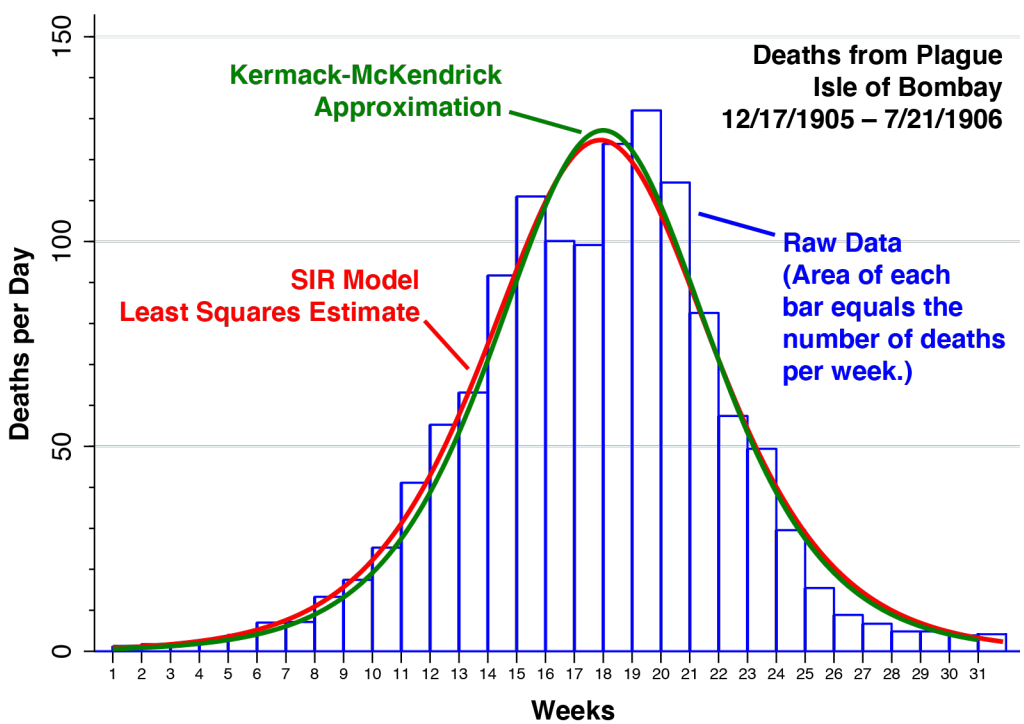

Figure 6. Reported and Predicted Deaths from Plague, Isle of Bombay, 12/17/1905 – 7/21/1906. The time axis is marked off in weeks, with week $m = 1$ starting on 12/17/1905 and week $m = 31$ starting on 7/15/1906. As noted in Figure 2, the height of each bar is the average daily number of deaths reported each week, while the area of each bar equals the number of deaths during that week, adapted from Kermack and McKendrick [5]. The green curve represents the authors' original fit to the data, while the red curve represents our estimates of the latent variable Z_t.

Discussion

Is SIR the Correct Structural Model for Omicron?

The mere fact that we have found a workable solution to the SIR inversion problem does not necessarily mean that the Susceptible-Infected-Removed model is the most appropriate description of the data under study. Our illustrative analysis of the SARS-CoV-2 Omicron/BA.1 wave in New York City and Los Angeles County during December 2021 – March 2022 cogently brings home the point.

Figure 3A shows a tight fit of the SIR model-predicted curve X_t to the case incidence datapoints y_t for New York City. Figures 4A and 4B confirm that the estimated values of the parameters β and α are indeed situated at the minimum of a convex region of our least squares criterion function V. Yet the mean duration of infectivity, computed from our SIR model as $1/(1-\alpha)$, was estimated from the New York City data to be on the order of 17 days, with a 95% confidence interval of two to three weeks. That's way out of line with what is known from direct clinical measurement. In one cohort of 55 symptomatic Omicron-infected patients, only 13.5 percent continued to shed virus ten days after infection [76]. Yet our New York City-based parameter estimate for α would give the proportion remaining infectious after ten days at $\alpha^{10} = 54.4\%$. To be sure, our Los Angeles County data gave a lower mean duration of infectivity on the order of nine days, with a 95% confidence interval of 8 to 10 days. But the estimated proportion remaining infectious after ten days would still be far too high at 31.2%.

Despite its remarkable fit to the data, SIR may thus fail as a *structural* model of the Omicron wave, even if it apparently succeeds as a *reduced form* model [77]. While its parameters β and α can indeed be backed out from the data and the inversion problem solved, these parameters do not necessarily warrant the structural interpretation that we assumed in our exposition of the SIR model in equations (1) through (5) above.

We should not, however, abandon the possibility that SIR may indeed be the correct structural model of the Omicron wave. To the contrary, the parameter estimates may reveal a critically important characteristic of the Omicron variant that clinical studies, which are necessarily biased toward symptomatic patients, have so far missed. In effect, the Omicron wave was so massive not because the variant had such a high per-contact infectivity (through the parameter β), but because a much larger proportion of asymptomatic infected individuals remained persistently infectious (through the parameter α). The divergence in our estimates of α

between New York City and Los Angeles County could mean that the degree of persistent infectivity is not determined solely by biological characteristics of the virus, but also by the variable intensity of interpersonal contact.

It might be argued that there is a third plausible interpretation, namely, that SIR may indeed be the correct structural model, but that the case incidence data analyzed here were inadequate to accurately identify the key parameters β and α. While the dark, diagonal ravine in the heat map of Figure 5 points to a high correlation between the two parameters in the topological neighborhood of the global optimum in the (α, β) plane, the results indicate that they are nonetheless identified from the data. One would have to contend that the estimated confidence intervals surrounding the parameters were derived from an inappropriate assumption of normally distributed errors. If so, a much fatter-tailed error distribution would be required.

What About the Results for the Plague?

Not only do our SIR predictions fit reasonably well to the observed data on Bombay plague deaths in Figure 6, but they virtually coincide with the curve of the approximate closed-form solution drawn by Kermack and McKendrick [5]. While others have fitted SIR-type compartmental models to the 1905–1906 Bombay plague data [78], our study appears to be the first replication of the authors' century-old result.

The plague, caused by the bacterium *Yersinia pestis*, can spread through human populations by various transmission pathways. Bubonic plague, thought to be the cause of a series of Bombay outbreaks that began in 1896 [79, 80], is transmitted to humans through rat fleas [81], but transmission between humans also appears to occur via infected human fleas or body lice [82]. Pneumonic plague, by contrast, is transmitted by direct human-to-human transmission via respiratory droplets and can occur as complication of bubonic plague [83, 84]. Pneumonic plague is more lethal, but requires closer person-to-person contact and is thus thought to have a lower basic reproduction number $\mathcal{R}_0$ [83].

So, does our SIR model similarly succeed as a reduced form model but fail as a structural model of the Bombay plague outbreak? Several structural compartmental models of plague transmission have been tested, typically involving separate states and parameters for humans and vectors [78, 84-86]. In a rat-flea transmission model for bubonic plague, for example, the structural model contained a separate sub-epidemic module for rats with a distinct basic reproduction number for rat-to-rat transmission [86]. A human-flea-human structural model

appeared to have the best fit to nine plague outbreaks during Europe's Second Pandemic from 1348 to 1813 [84]. Our minimalist SIR model has none of these structural features.

The irony, however, is that the parameter estimates derived from our solution to the inversion problem are entirely consistent with pneumonic plague transmission. The estimated basic reproduction number was $\mathcal{R}_0 = 1.1$, squarely in the range for the pneumonic form [84]. The point estimate of the parameter $\alpha = 0.446$ was consistent with high mortality rate of pneumonic plague, while the implied average infectious period of 1.8 days was similarly consistent with pneumonic human-to-human transmission [84]. The plague may indeed have been initially imported into Bombay in its bubonic form through rat-flea-human transmission. Our data, however, support the hypothesis that responsive isolation measures by British authorities [79] may have altered the mode of transmission during the 1905–1906 outbreak from the bubonic rat-flea-human mode to the pneumonic human-human mode, a phenomenon observed in the 17th century when the Derbyshire village of Eyan fell victim to the Black Death [87]. If so, our parsimonious SIR model is in fact the appropriate structural model of the outbreak.

Why are the Estimated Population Sizes N So Small? .

Table 1 reports estimated population size parameters of $N = 1.013$ million for New York City and $N = 1.046$ million for Los Angeles County. As noted, these estimates of N were only 11.9 percent of the total census population of New York City and 14.5 percent of the corresponding census population of Los Angeles County. Where, then, did all the other millions of inhabitants go?

As discussed above, we cannot distinguish between three possible explanations of the shortfall. First, there is concrete evidence of significant underreporting of Omicron infections, due principally to the widespread availability of home rapid antigen testing [60], particularly in New York City [88, 89]. Second, despite the evidence of immune escape on the part of the Omicron variant [90], a substantial fraction of the population may still have retained long-term cellular immunity, particular through the administration of multiple vaccine doses [91]. Third, contrary to the underlying assumption of homogeneous mixing, there is strong evidence that a substantial proportion of the population avoided retail establishments, drinking and eating places, transportation venues, worksites and other high-risk locations [92].

Our analysis of the Bombay outbreak data gave an estimated size parameter $N = 51$ thousand individuals, which comes to about 6.5 percent of estimated Bombay population of 775 thousand in 1900 [93]. The same three factors could well be responsible for the shortfall. In particular, there is significant evidence for the acquisition of at least humoral immunity against *Yersinia pestis* among plague survivors [94].

Limitations and Extensions

Parameter search algorithms such as Newton-Raphson [44] and Gauss-Newton [45] have superior performance when they rely on closed-form estimates of the gradient vector and matrix of second derivates, rather than on numerical approximation [43]. Still, as Figures 4A and 4B teach us, there will nonetheless be a restricted region of the parameter space where the least squares criterion function is convex. Some of these regions may contain local rather than global optima. To avoid such problems of implementation, we plotted the least squares criterion and its first partial derivatives as functions of the parameters. When the search algorithm appeared to escape the allowable parameter space, we reduced the step size. Still, finding the right initial values and keeping the parameter search within bounds remain unavoidable challenges.

As seen in Figures 1A and 1B, striking day-of-the-week effects on COVID-19 testing and case reporting required us to pretreat the raw case counts. Our centered, 7-day moving average appeared to be the most flexible nonparametric approach to this task. Alternative approaches that incorporate parametric models of day-of-the-week effects directly in our model of the latent variable X_t have yet to be tried.

We have focused sharply on the original SIR model, rather than its numerous variations. Still, our basic approach can be extended to these more complex models. Take, for example, the well-studied SIRS model [95], where individuals in the R (recovered) state can transition back to the S (susceptible) state as a result of waning immunity. The SIRS model still has three states, and the definitions of the latent variables X_t and Z_t in equations (3) and (23), respectively, remain unchanged. Now, however, the parameter vector $\boldsymbol{\theta}$ has an additional component governing the rate of transition from R back to S. Since the SIRS model is known to admit oscillations with a stable endemic equilibrium [96], we would ideally run our parameter recovery algorithm on multiple waves of data y_t. That is a task for future research.

Declarations

Author Affiliations and Contact Information

Professor of Economics Emeritus, Massachusetts Institute of Technology, Cambridge MA USA 02139; and Physician, Eisner Health, Los Angeles CA USA 90015. Email: jeffrey@mit.edu.

Sole Authorship

JEH is the sole author of this work. He alone is responsible for the conceptualization of the work, the analysis of the data, the drafting of the manuscript, and the construction of the graphics. He warrants that it is entirely his original work. No requests for permission to use copyrighted material are required.

Acknowledgments

The opinions expressed here are solely those of the author and do not necessarily reflect those of the Massachusetts Institute of Technology, Eisner Health, or any other organization.

Availability of Data and Programs

The data and programs used in this work are publicly available on the Open Science Framework at https://osf.io/3b4hv/.

Human Subjects Declaration

This study relies exclusively on publicly available data that contain no individual identifiers. Links to the data sources are provided in the manuscript references.

Competing Interests Declaration

The author has no competing interests.

Funding Statement

This study did not receive any funding.

References

1. Capistran, M.A., M.A. Moreles, and B. Lara, *Parameter estimation of some epidemic models. The case of recurrent epidemics caused by respiratory syncytial virus.* Bull Math Biol, 2009. **71**(8): p. 1890-901.
2. Pollicott, M., H. Wang, and H.H. Weiss, *Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem.* J Biol Dyn, 2012. **6**: p. 509-23.
3. Marinov, T.T., et al., *Inverse problem for coefficient identification in SIR epidemic models.* Computers and Mathematics with Applications, 2014. **67**(12): p. 2218-2227.
4. Comunian, A., R. Gaburro, and M. Giudici, *Inversion of a SIR-based model: A critical analysis about the application to COVID-19 epidemic.* Physica D, 2020. **413**: p. 132674.
5. Kermack, W.O. and A.G. McKendrick, *A contribution to the mathematical theory of epidemics.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1927. **115**: p. 700-721.
6. Balkew, T.M., *The SIR Model When S(t) is a Multi-Exponential Function*. 2010, https://dc.etsu.edu/etd/1747/: East Tennessee State University, School of Graduate Studies, Electronic Theses and Dissertations, Paper 1747, December.
7. Shabbir, G., H. Khan, and M.A. Sadiq, *A note on Exact solution of SIR and SIS epidemic models*. 2010, https://arxiv.org/abs/1012.5035: ArXiv, Submitted December 22.
8. Harko, T., F.S.N. Lobo, and M.K. Mak, *Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates.* Applied Mathematics and Computation, 2014. **236**(1 Jun 2014): p. 184-194.
9. Kröger, M. and R. Schlickeiser, *Analytical solution of the SIR-model for the temporal evolution of epidemics. Part A: time-independent reproduction factor.* Journal of Physics A: Mathematical and Theoretical, 2020. **53**(505601): p. 1-38.
10. Sadurni, E. and G. Luna-Acosta, *Exactly solvable SIR models, their extensions and their application to sensitive pandemic forecasting.* Nonlinear Dyn, 2021. **103**(3): p. 2955-2971.
11. Harris, J.E., *The Coronavirus Epidemic Curve is Already Flattening in New York City*. 2020, https://www.nber.org/papers/w26917: National Bureau of Economic Research Working Paper No. 26917, Updated April 6, 2020.

12. Magal, P. and G. Webb, *The parameter identification problem for SIR epidemic models: identifying unreported cases.* J Math Biol, 2018. **77**(6-7): p. 1629-1648.
13. Schmitt, F.G., *An algorithm for the direct estimation of the parameters of the SIR epidemic model from the I(t) dynamics.* The European Physical Journal Plus, 2021. **137**(57): p. 1-16.
14. Ma, J. and D.J. Earn, *Generality of the final size formula for an epidemic of a newly invading infectious disease.* Bull Math Biol, 2006. **68**(3): p. 679-702.
15. Arino, J., et al., *A final size relation for epidemic models.* Math Biosci Eng, 2007. **4**(2): p. 159-75.
16. Andreasen, V., *The final size of an epidemic and its relation to the basic reproduction number.* Bull Math Biol, 2011. **73**(10): p. 2305-21.
17. Mkhatshwa, T. and A. Mummert, *Modeling Super-spreading Events for Infectious Diseases: Case Study SARS*. 2010, https://arxiv.org/abs/1007.0908: arXiv, updated October 15.
18. Harris, J.E., *Data from the COVID-19 epidemic in Florida suggest that younger cohorts have been transmitting their infections to less socially mobile older adults.* Rev Econ Househ, 2020. **18**(4): p. 1019-1037.
19. Harris, J.E., *Los Angeles County SARS-CoV-2 Epidemic: Critical Role of Multi-generational Intra-household Transmission.* Journal of Bioeconomics, 2021. **23**: p. 55-83.
20. Rakshit, P., et al., *Modified SIR model for COVID-19 transmission dynamics: Simulation with case study of UK, US and India.* Results Phys, 2022. **40**: p. 105855.
21. Gopal, K., L.S. Lee, and H.-V. Seow, *Parameter Estimation of Compartmental Epidemiological Model Using Harmony Search Algorithm and Its Variants.* Applied Sciences, 2021. **11**(1138): p. 1+25.
22. Nguemdjo, U., et al., *Simulating the progression of the COVID-19 disease in Cameroon using SIR models.* PLoS One, 2020. **15**(8): p. e0237832.
23. Kucharski, A.J., et al., *Early dynamics of transmission and control of COVID-19: a mathematical modelling study.* Lancet Infect Dis, 2020. **20**(5): p. 553-558.
24. Rica, S. and G.A. Ruz, *Estimating SIR model parameters from data using differential evolution: an application with COVID-19 data*. 2020,

https://ieeexplore.ieee.org/document/9277708: IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology (CIBCB)

25. Prodanov, D., *Analytical Parameter Estimation of the SIR Epidemic Model. Applications to the COVID-19 Pandemic.* Entropy, 2021. **23**(59): p. 1-20.
26. Capaldi, A., et al., *Parameter Estimation and Uncertainty Quantification for an Epidemic Model.* Mathematical Biosciences and Engineering, 2012. **9**(3): p. 553-579.
27. Turk, P.J., et al., *Modeling COVID-19 Latent Prevalence to Assess a Public Health Intervention at a State and Regional Scale: Retrospective Cohort Study.* JMIR Public Health Surveill, 2020. **6**(2): p. e19353.
28. Ahmetolan, S., et al., *What Can We Estimate From Fatality and Infectious Case Data Using the Susceptible-Infected-Removed (SIR) Model? A Case Study of Covid-19 Pandemic.* Front Med (Lausanne), 2020. **7**: p. 556366.
29. Li, M., J. Dushoff, and B.M. Bolker, *Fitting mechanistic epidemic models to data: A comparison of simple Markov chain Monte Carlo approaches.* Stat Methods Med Res, 2018. **27**(7): p. 1956-1967.
30. Su, L., et al., *Evaluation of the Secondary Transmission Pattern and Epidemic Prediction of COVID-19 in the Four Metropolitan Areas of China.* Front Med (Lausanne), 2020. **7**: p. 171.
31. Wangping, J., et al., *Extended SIR Prediction of the Epidemics Trend of COVID-19 in Italy and Compared With Hunan, China.* Front Med (Lausanne), 2020. **7**: p. 169.
32. Cauchemez, S. and N.M. Ferguson, *Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London.* J R Soc Interface, 2008. **5**(25): p. 885-97.
33. Gu, J. and G. Yin, *Bayesian SIR model with change points with application to the Omicron wave in Singapore.* Sci Rep, 2022. **12**(1): p. 20864.
34. Koepke, A.A., et al., *Predictive Modeling of Cholera Outbreaks in Bangladesh.* Annals of Applied Statistics, 2016. **10**(2): p. 575-593.
35. Cooper, I., A. Mondal, and C.G. Antonopoulos, *A SIR model assumption for the spread of COVID-19 in different communities.* Chaos Solitons Fractals, 2020. **139**: p. 110057.

36. Finkenstädt, B.F. and B.T. Grenfell, *Time Series Modelling of Childhood Diseases: A Dynamical Systems Approach.* Journal of the Royal Statistical Society. Series C (Applied Statistics), 2000. **49**(2): p. 187-205.
37. Wanduku, D. and C. Rahul, *Complete maximum likelihood estimation for SEIR epidemic models: theoretical development*. 2019, https://arxiv.org/abs/1907.10679: arXiv, Updated July 29.
38. Diekmann, O., et al., *The discrete-time Kermack-McKendrick model: A versatile and computationally attractive framework for modeling epidemics.* Proc Natl Acad Sci U S A, 2021. **118**(39).
39. Cintron-Arias, A., et al., *The estimation of the effective reproductive number from disease outbreak data.* Math Biosci Eng, 2009. **6**(2): p. 261-82.
40. Hooker, G., et al., *Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario.* J R Soc Interface, 2011. **8**(60): p. 961-74.
41. Chowell, G., et al., *Modelling the transmission dynamics of acute haemorrhagic conjunctivitis: application to the 2003 outbreak in Mexico.* Stat Med, 2006. **25**(11): p. 1840-57.
42. Gabor, A. and J.R. Banga, *Robust and efficient parameter estimation in dynamic models of biological systems.* BMC Syst Biol, 2015. **9**: p. 74.
43. Schroeder, C., *Practical course on computing derivatives in code*, in *Proceedings of SIGGRAPH '19 Courses, July 28 - August 01, 2019, Los Angeles, CA, USA*. 2019, ACM: https://dl.acm.org/doi/10.1145/3305366.3328073.
44. Verbeke, J. and R. Cools, *The Newton-Raphson method.* International Journal of Mathematical Education in Science and Technology, 1995. **26**(2): p. 177-193.
45. Hartley, H.O., *The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares.* Technometrics, 1961. **3**(2): p. 269-280.
46. Gallant, A.R., *Nonlinear Regression.* The American Statistician, 1975. **29**(2 (May)): p. 73-81.
47. Kermack, W.O. and A.G. McKendrick, *Contributions to the mathematical theory of epidemics, part II.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1932. **138**(834): p. 55-83.

48. Kermack, W.O. and A.G. McKendrick, *Contributions to the mathematical theory of epidemics, part III.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1933. **141**(843): p. 55-83.
49. Anderson, R.M., *Discussion: the Kermack-McKendrick epidemic threshold theorem.* Bull Math Biol, 1991. **53**(1-2): p. 3-32.
50. Brauer, F., *Mathematical epidemiology: Past, present, and future.* Infect Dis Model, 2017. **2**(2): p. 113-127.
51. Evans, N.D., et al., *The structural identifiability of the susceptible infected recovered model with seasonal forcing.* Math Biosci, 2005. **194**(2): p. 175-97.
52. Guldberg, C.M. and P. Waage, *Studies Concerning Affiinity: Presentation to Academy of Sciences in Christiania (Oslo). Translated by Henry I. Abrash. Originally published by Forhandlinger: Videnskabs-Selskabet i Christiana, 1864.* Journal of Chemical Education, 1986. **63**(12): p. 1044-1047.
53. Weiss, H.H., *The SIR model and the Foundations of Public Health.* Materials Matemàtics, 2013. **3**: p. 1-17.
54. van den Driessche, P. and J. Watmough, *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission.* Math Biosci, 2002. **180**: p. 29-48.
55. Keeling, M.J., et al., *Fitting to the UK COVID-19 outbreak, short-term forecasts and estimating the reproductive number.* Stat Methods Med Res, 2022. **31**(9): p. 1716-1737.
56. Becker, N.G. and T. Britton, *Statistical studies of infectious disease incidence.* J R Statist Soc B, 1999. **61**(2): p. 287-307.
57. Lange, A., *Reconstruction of disease transmission rates: Applications to measles, dengue, and influenza.* J Theor Biol, 2016. **400**: p. 138-53.
58. Oehlert, G.W., *A Note on the Delta Method.* The American Statistician, 1992. **46**(1): p. 27–29.
59. Dempster, A.P., N.M. Laird, and D.B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal Statistical Society, Ser. B, 1977. **39**: p. 1-22.
60. Qasmieh, S.A., et al., *The Importance of Incorporating At-Home Testing Into SARS-CoV-2 Point Prevalence Estimates: Findings From a US National Cohort, February 2022.* JMIR Public Health Surveill, 2022. **8**(12): p. e38196.

61. Abadie, A., et al., *Epidemic Modeling and Estimation*. 2020, https://idss.mit.edu/wp-content/uploads/2020/04/04.28.2020Epidemic_Modeling_A_Memo.pdf: Massachusetts Institute of Technology, Institute for Data Systems and Society (IDSS), April 28.
62. Roda, W.C., et al., *Why is it difficult to accurately predict the COVID-19 epidemic?* Infect Dis Model, 2020. **5**: p. 271-281.
63. Trejo, I. and N.W. Hengartner, *A modified Susceptible-Infected-Recovered model for observed under-reported incidence data.* PLoS One, 2022. **17**(2): p. e0263047.
64. Brauer, F., *Epidemic models with heterogeneous mixing and treatment.* Bull Math Biol, 2008. **70**(7): p. 1869-85.
65. Kong, L., et al., *Modeling Heterogeneity in Direct Infectious Disease Transmission in a Compartmental Model.* Int J Environ Res Public Health, 2016. **13**(3).
66. David, J.F., *Epidemic models with heterogeneous mixing and indirect transmission.* J Biol Dyn, 2018. **12**(1): p. 375-399.
67. Duan, M. and Z. Jin, *The heterogeneous mixing model of COVID-19 with interventions.* J Theor Biol, 2022. **553**: p. 111258.
68. Harris, J.E., *Concentric regulatory zones failed to halt surging COVID-19: Brooklyn 2020.* Front Public Health, 2022. **10**: p. 970363.
69. Hill, A.N., J.W. Glasser, and Z. Feng, *Implications for infectious disease models of heterogeneous mixing on control thresholds.* J Math Biol, 2023. **86**(4): p. 53.
70. Harris, J.E., *Geospatial Analysis of a COVID-19 Outbreak at the University of Wisconsin - Madison: Potential Role of a Cluster of Local Bars.* Epidemiol Infect, 2022: p. 1-31.
71. New York Department of Health and Mental Hygiene, *NYC Coronavirus Disease 2019 (COVID-19) Data*. 2022, https://github.com/nychealth/coronavirus-data: Accessed July 3.
72. Los Angeles County Department of Public Health, *COVID-19 Surveillance Dashboard*. 2022, https://lacdph.shinyapps.io/covid19_surveillance_dashboard/: Accessed July 3.
73. StataCorp, *Stata Statistical Software: Release 17*. 2021, College Station, TX: StataCorp LLC.
74. StataCorp, *Stata Programming Reference Manual Release 17*. 2021, https://www.stata.com/manuals/p.pdf: College Station, Texas: Stata Press.
75. StataCorp, *Mata Reference Manual Release 17*. 2021, https://www.stata.com/manuals/m.pdf: College Station, Texas: Stata Press.

76. Keske, S., et al., *Duration of infectious shedding of SARS-CoV-2 Omicron variant and its relation with symptoms.* Clin Microbiol Infect, 2022.

77. Timmins, C. and W. Schlenker, *Reduced-Form Versus Structural Modeling in Environmental and Resource Economics.* Annual Review of Resource Economics, 2009. **1**: p. 351-380.

78. Monecke, S., H. Monecke, and J. Monecke, *Modelling the black death. A historical case study and implications for the epidemiology of bubonic plague.* Int J Med Microbiol, 2009. **299**(8): p. 582-93.

79. Kidambi, P., *The Making of an Indian Metropolis: Colonial Governance and Public Culture in Bombay, 1890-1920*. 2016, London: Routledge.

80. Frazier, M.W., *Tale of two Mumbai epidemics: Striking parallels between 1896 bubonic plague and 2020 Covid outbreak*. 2020, https://scroll.in/article/976706/tale-of-two-mumbai-epidemics-striking-parallels-between-1896-bubonic-plague-and-2020-covid-outbreak: Scroll.in, November 13.

81. Kugeler, K.J., et al., *Epidemiology of human plague in the United States, 1900-2012.* Emerg Infect Dis, 2015. **21**(1): p. 16-22.

82. Drancourt, M., L. Houhamdi, and D. Raoult, *Yersinia pestis as a telluric, human ectoparasite-borne organism.* Lancet Infect Dis, 2006. **6**(4): p. 234-41.

83. Gani, R. and S. Leach, *Epidemiologic determinants for modeling pneumonic plague outbreaks.* Emerg Infect Dis, 2004. **10**(4): p. 608-14.

84. Dean, K.R., et al., *Human ectoparasites and the spread of plague in Europe during the Second Pandemic.* Proc Natl Acad Sci U S A, 2018. **115**(6): p. 1304-1309.

85. Whittles, L.K. and X. Didelot, *Epidemiological analysis of the Eyam plague outbreak of 1665-1666.* Proc Biol Sci, 2016. **283**(1830).

86. Didelot, X., L.K. Whittles, and I. Hall, *Model-based analysis of an outbreak of bubonic plague in Cairo in 1801.* J R Soc Interface, 2017. **14**(131).

87. Massad, E., et al., *The Eyam plague revisited: did the village isolation change transmission from fleas to pulmonary?* Med Hypotheses, 2004. **63**(5): p. 911-5.

88. Qasmieh, S.A., et al., *The prevalence of SARS-CoV-2 infection and other public health outcomes during the BA.2/BA.2.12.1 surge, New York City, April-May 2022*. 2022, https://www.medrxiv.org/content/10.1101/2022.05.25.22275603v3: medRxiv, July 18.

89. Qasmieh, S.A., et al., *Estimating the period prevalence of SARS-CoV-2 infection during the Omicron (BA.1) surge in New York City (NYC), January 1-March 16, 2022*. 2022, https://www.medrxiv.org/content/10.1101/2022.04.23.22274214v1: medRxiv, April 27.
90. Dejnirattisai, W., et al., *SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses.* Cell, 2022. **185**(3): p. 467-484 e15.
91. Moss, P., *The T cell immune response against SARS-CoV-2.* Nat Immunol, 2022. **23**(2): p. 186-193.
92. Harris, J.E., *Mobility was a significant determinant of reported COVID-19 incidence during the Omicron Surge in the most populous U.S. Counties.* BMC Infect Dis, 2022. **22**(1): p. 691.
93. FIBIwiki, *Bombay (City)*. 2022, https://wiki.fibis.org/w/Bombay_(City): Fibis.org. Last updated November 14, 2022. Accessed February 25, 2023.
94. Andrianaivoarimanana, V., et al., *Short- and long-term humoral immune response against Yersinia pestis in plague patients, Madagascar.* BMC Infect Dis, 2020. **20**(1): p. 822.
95. Munoz-Fernandez, G.A., J.M. Seoane, and J.B. Seoane-Sepulveda, *A SIR-type model describing the successive waves of COVID-19.* Chaos Solitons Fractals, 2021. **144**: p. 110682.
96. Bestehorn, M., et al., *Simple model of epidemic dynamics with memory effects.* Physical Review E, 2022. **105**(024025): p. 024025.