Tight Fit of the SIR Dynamic Epidemic Model to Daily Cases of COVID-19 Reported During the 2021-2022 Omicron Surge in New York City: A Novel Approach

Jeffrey E. Harris MD PhD*† February 26, 2024

Abstract. We describe a novel approach to recovering the underlying parameters of the SIR dynamic epidemic model from observed data on case incidence. 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 estimated a four-parameter SIR model from data on daily reported cases of COVID-19 during the SARS-CoV-2 Omicron/BA.1 surge of December 2021 - March 2022 in New York City. The estimated SIR model showed a tight fit to the observed data, but less so when we excluded residual cases attributable to the Delta variant during the initial upswing of the wave in December. Our analyses of both the real-world COVID-19 data and simulated case incidence data revealed an important problem of weak parameter identification. While our methods permitted separate estimation of the infection transmission parameter and the infection persistence parameter, only a linear combination of these two key parameters could be estimated with precision. The SIR model appears to be an adequate reduced-form description of the Omicron surge, but it is not necessarily the correct structural model. Prior information above and beyond case incidence data may be required to sharply identify the parameters and thus distinguish between alternative epidemic models.

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

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

† Revision of: *New Approach to the SIR Inversion Problem: From the 1905-1906 Plague Outbreak in the Isle of Bombay to the 2021-2022 Omicron Surge in New York City*. Reference [64] below. See Author Declarations at the end of this 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 (KM) 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, 12], the time to reach the peak incidence [13], the rate of decline after the peak [14], and the proportion of the population that is ultimately infected [15-17]. 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 [18-21], in some cases derived from previous waves of a multi-wave epidemic [22]. Perhaps the most traveled road to a solution has been the use of various parameter search algorithms [18, 23-29] which, when it comes down to it, offer only a marginal improvement over brute force search [30]. Bayesian estimation may be better able to

integrate prior information into our search procedures [31-36], but its computational burden is usually even greater. Last but not least, we can resort to trial and error combined with visual inspection [37].

The present study, we suggest, offers an easily workable solution to the inversion problem. Rather than seeking a closed-form, analytical solution to KM's system of differential equations, we pursue an alternate strategy. First, following the lead of other investigators [35, 38-40], 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 [26, 41-44], we define a 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 superior to numerical approximation [45].

Fourth, once we have calculated the gradient and Hessian, we can use the well-known Newton-Raphson algorithm [46] 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 [47]. Sixth, our approach permits us to readily determine what parameters are in fact identified when we have only time-series data on new cases.

We apply our strategy to the estimation of a four-parameter SIR model to 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 New York City.

Statistical Methods

Discrete-Time SIR Model

Following the lead of other investigators [4, 35, 38-40], we adopt a discrete-time approach. We mark off the time axis 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, 48-51]. 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 [52-55], 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 deterministic, coupled difference equations:

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

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

$$R_t = R_{t-1} + (1-\alpha) I_{t-1} \tag{1c}$$

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

Equations (1a) through (1d) represent the well-known forward Euler approximation to the underlying continuous-time SIR model of coupled differential equations [56]. Apart from the population size N, this dynamic system has two parameters: β and α. In equations (1a) and (1b), $\beta > 0$ is an infection transmission parameter. In equations (1b) and (1c), $1 > \alpha > 0$ is an infection persistence parameter. It gauges the proportion of infected individuals at each time t who remain in the infected state. The quantity $(1-\alpha)$ in equation (1c), which in some treatments is represented by the parameter γ, corresponds to the proportion who transition to the removed state. The multiplicative term $\beta S_{t-1} I_{t-1}/N$ in equations (1a) and (1b) reflects the law of mass action [57], 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 in equations (1a) and (1b) to take account of removals by death. Unless the overall death rate is substantial, this adjustment is usually ignored in model implementations.

We complete our discrete-time model with the specification of initial conditions at $t = 0$:

$$S_0 = (1 - i_0 - r_0)N \tag{2a}$$

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

$$R_0 = r_0 N \tag{2c}$$

The additional parameters i_0 and r_0 represent the respective proportions of the entire population of size N that are initially infected and removed.

In the dynamic system (1) and (2), the mean duration of infection is time-invariant and equal to $1/(1-\alpha)$. The system results in an epidemic wave when $(1 - i_0 - r_0)\,\beta/(1-\alpha) > 1$, a well-known result known as the epidemic threshold theorem [24, 50, 58, 59]. Assuming that $1 - i_0 - r_0 \approx 1$, most authors write this epidemic threshold condition as $\mathcal{R}_0 > 1$, where $\mathcal{R}_0 = \beta/(1-\alpha)$ is defined as the basic reproduction number [24, 51, 60, 61].

Parameter Estimation: Least Squares Minimization Criterion

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 [39, 62, 63]. Instead, we observe only the reported counts of new infections at various intervals. In this exposition, we assume that new infection counts y_t are observed at each discrete time t. The more general case where such counts are observed less frequently has been addressed elsewhere [64].. The counts y_t represent observations on the *output variables* X_t, $t = 1, \ldots, T$, which from (1) correspond to:

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

Given the definition of the output 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 output variables X_t, we want to estimate the unknown parameters β, α, i_0, r_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 output variables X_t adhere to the dynamic system defined by (2), (3) and (4), which in turn depends on these unknown parameters.

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 output variables at each time t, where we use boldface symbols denote vectors or matrices. Let $\boldsymbol{\Theta} = (\beta, \alpha, i_0, r_0, N)$ denote the row vector of the unknown parameters. Let $\boldsymbol{X}(\boldsymbol{\Theta})$ represent the functional dependence of the output variables on these parameters. To estimate $\boldsymbol{\Theta}$ from the data $\boldsymbol{y}$, we introduce the least squares minimization criterion $V(\boldsymbol{\Theta}) = \big(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{\Theta})\big)'\big(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{\Theta})\big)$, which can be written in summation notation as:

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

This criterion has been widely used in attempts to fit the SIR model to incidence data [26, 41-44]. It is well known that minimizing V is equivalent to computing the maximum likelihood estimate of $\boldsymbol{\Theta}$ under the assumption that $y_t = X_t(\boldsymbol{\Theta}) + \varepsilon_t$, where the disturbances ε_t are independently normally distributed $N(0, \sigma^2)$ with homoscedastic variance σ^2.

Non-Identifiability of the Five-Parameter Vector $\boldsymbol{\Theta}$

While the model delineated in equations (1) through (4) is standard in the literature [4, 35, 38-40], it turns out that not all five parameters of the vector $\boldsymbol{\Theta} = (\beta, \alpha, i_0, r_0, N)$ can be identified from the data $\boldsymbol{y}$ on case incidence alone. This conclusion is supported by the following proposition, which, along with all other propositions, is proved in Appendix A.

Proposition 1. Let $\Omega \subseteq \mathbb{R}^5$ be the subspace of admissible values of the five-dimensional parameter vector $\boldsymbol{\Theta}$. Let $\boldsymbol{X}(\boldsymbol{\Theta})$ be the resulting $T \times 1$ column vector of output variables defined by the model of equations (1) through (4). Then there exists a mapping $\phi: \Omega \to \mathbb{R}^4$ with the property that $\boldsymbol{X}(\boldsymbol{\Theta}) = \boldsymbol{X}(\boldsymbol{\Theta}')$ for all vectors $\boldsymbol{\Theta}, \boldsymbol{\Theta}' \in \Omega$ satisfying $\phi(\boldsymbol{\Theta}) = \phi(\boldsymbol{\Theta}')$.

Even if we had perfectly accurate data $\boldsymbol{y}$ on case incidence, we could no better than to measure $\boldsymbol{X}$ without error. Proposition 1, however, establishes that $\boldsymbol{X}(\boldsymbol{\Theta})$ is not a one-to-one mapping from parameters $\boldsymbol{\Theta}$ to output variables $\boldsymbol{X}$. Moreover, for any given set of observations $\boldsymbol{y}$ on case incidence, equation (5) teaches us that the dependence of V on $\boldsymbol{\Theta}$ runs solely through

$\boldsymbol{X}(\boldsymbol{\Theta})$. Accordingly, $V(\boldsymbol{\Theta})$ is likewise not a one-to-one mapping from parameters $\boldsymbol{\Theta}$ to our objective function V.

Appended to the proof of Proposition 1 are some corollary results indicating how additional prior information can be used to identify the model parameters. For example, if we had sharp prior information on the initial proportion of recovered individuals r_0, or the initial population size N, or the basic reproduction number $\mathcal{R}_0$, then we could identify all the parameters in $\boldsymbol{\Theta}$. In what follows, we impose the prior constraint that $r_0 = 0$. To simplify our notation, we drop r_0 from the parameter vector and instead write $\boldsymbol{\Theta} = (\beta, \alpha, i_0, N)$ as a four-parameter vector, denoting by $\Omega \subseteq \mathbb{R}^4$ the four-dimensional space of admissible values of $\boldsymbol{\Theta}$.

Gradient and Hessian of the Least Squares Criterion V

Let the operator $\boldsymbol{D}$ denote the 4×1 gradient of partial derivatives with respect to the elements of $\boldsymbol{\Theta}$, and let the operator $\boldsymbol{D^2}$ denote the corresponding 4×4 Hessian matrix of second order partial derivatives. The following two propositions embody the main methodological innovation of this paper.

Proposition 2. The gradient of the least squares criterion V with respect to the parameter vector $\boldsymbol{\Theta} = (\beta, \alpha, i_0, N)$ is:

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

where each column vector $\boldsymbol{DX_t} = \left(\frac{\partial X_t}{\partial \beta}, \frac{\partial X_t}{\partial \alpha}, \frac{\partial X_t}{\partial i_0}, \frac{\partial X_t}{\partial N}\right)'$ represents the corresponding gradient of partial derivatives of the output variable X_t at time t. The basic equations of our dynamic system (1), in combination with the initial conditions (2), can be used to generate complete, computable difference equations for $\boldsymbol{DX_t}$ for all t, and thus for $\boldsymbol{DV}$.

Proposition 3. Let $\boldsymbol{DX}$ denote the $T \times 4$ matrix whose t-th row is the vector $\boldsymbol{DX_t'}$, as defined in Proposition 2. The Hessian matrix of the least squares criterion V with respect to the parameter vector $\boldsymbol{\Theta} = (\beta, \alpha, i_0, N)$ is:

$$\boldsymbol{D^2V} = 2\, \boldsymbol{DX'DX} - 2 \sum_{t=1}^{T} (y_t - X_t) \boldsymbol{D^2X_t} \tag{7}$$

The basic equations of our dynamic system (1), in combination with the initial conditions (2), can similarly be used to generate complete, computable difference equations for $\boldsymbol{D^2X_t}$ for all t, and thus $\boldsymbol{D^2V}$.

As part of the proofs of these two propositions in Appendix A, we show the details of the difference equations involved in the exact computation of $\boldsymbol{DX_t}$ and $\boldsymbol{D^2X_t}$ for all t.

Newton-Raphson Algorithm for Parameter Estimation

While the objective function $V(\boldsymbol{\Theta})$ has no closed-form expression in terms of $\boldsymbol{\Theta}$, we can still rely on our exact computations of the gradient $\boldsymbol{DV}$ and Hessian $\boldsymbol{D^2V}$ to search for a local interior minimum within the four-dimensional subspace Ω of admissible values of $\boldsymbol{\Theta}$ via the Newton-Raphson algorithm [46].

The Newton-Raphson algorithm is iterative. Denote our choice of initial parameter vector as $\boldsymbol{\Theta}^{(0)}$. Let (k) index successive iterations. The current value of the parameter vector $\boldsymbol{\Theta}^{(k)}$ is repeatedly mapped into an updated value $\boldsymbol{\Theta}^{(k+1)}$ according to the well-known rule:

$$\boldsymbol{\Theta}^{(k+1)} = \boldsymbol{\Theta}^{(k)} - q\,(\boldsymbol{D^2V})^{-1}\,\boldsymbol{DV} \tag{8}$$

Here, the gradient $\boldsymbol{DV}$, as defined in (6), and the Hessian $\boldsymbol{D^2V}$, as defined in (7), are both computed at the current value of the parameter vector $\boldsymbol{\Theta}^{(k)}$, while the step size $0 < q \leq 1$ is under control of the programmer.

This iterative approach would work flawlessly if the objective function V were globally convex with an interior minimum. There is good reason, however, to suspect that V may instead have multiple local optima. While the classical SIR model predicts a single-peaked wave of incident cases $\boldsymbol{X}$ so long as $\mathcal{R}_0 > 1$, the observed data $\boldsymbol{y}$ often display multiple peaks over time. (A good example is the two-peak plot of the 2001 Dengue fever outbreak in Havana [59].) In that case, the search algorithm embodied in (8) may easily end up at a local rather than a global minimum of V. To address this possibility, we need to run the Newton-Raphson search routine from various initial parameter vectors $\boldsymbol{\Theta}^{(0)}$.

What's more, there may be regions of the four-parameter space where the Hessian $\boldsymbol{D^2V}$ is not positive definite because the second term in equation (7) (that is, $-2\sum_{t=1}^{T}(y_t - X_t)\boldsymbol{D^2X_t}$) is not necessarily a positive definite matrix. In that case, the parameter updating rule (8) may not result in a decrease in the objective function V as the algorithm veers away from the optimum. To address this possibility, we can back up to $\boldsymbol{\Theta}^{(k)}$ and reduce the step size q.

Alternative EM-Type Procedure for Parameter Estimation

It turns out that there is an alternative iterative procedure for parameter estimation analogous to the so-called EM algorithm [65]. Rather than searching through the four-dimensional space of admissible values of $\boldsymbol{\Theta} = (\beta, \alpha, i_0, N)$, we can separate the estimation of the population-size parameter N and from our search over the remaining three-dimensional space of the remaining identifiable parameters (β, α, i_0). The procedure is motivated by the following two propositions.

Proposition 4. Each element $X_t(\boldsymbol{\Theta})$ of the output vector $\boldsymbol{X}(\boldsymbol{\Theta})$ 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 identifiable parameters $\boldsymbol{\theta} = (\beta, \alpha, i_0)$.

This result does not require that we derive a closed-form expression for the function $\varphi_t(\boldsymbol{\theta})$. The fact that each output variable X_t is proportional to N implies that the optimization criterion V defined in (5) is a quadratic polynomial in N. That conclusion in turn motivates the following additional proposition.

Proposition 5. Let $\boldsymbol{\Theta}^{(n)} = \left(\boldsymbol{\theta}^{(n)}, N^{(n)}\right)$ denote parameter estimates at iteration n of an iterative estimation algorithm. Let $\boldsymbol{X}^{(\boldsymbol{n})} = \boldsymbol{X}\left(\boldsymbol{\Theta}^{(n)}\right)$ denote the corresponding output variable vector derived from the SIR model (2), (3) and (4) based upon these parameter estimates. Define

$$\kappa^{(n+1)} = \boldsymbol{y}'\boldsymbol{X}^{(\boldsymbol{n})} / \boldsymbol{X}^{(\boldsymbol{n})\prime}\boldsymbol{X}^{(\boldsymbol{n})} = \sum_{t=1}^{T} y_t X_t^{(n)} / \sum_{t=1}^{T}\left(X_t^{(n)}\right)^2 \tag{9}$$

as the regression coefficient of $\boldsymbol{y}$ on $\boldsymbol{X}^{(\boldsymbol{n})}$. Then conditional upon $\boldsymbol{\Theta}^{(n)}$, the updated population-size parameter that minimizes the least squares criterion V is $N^{(n+1)} = \kappa^{(n+1)} N^{(n)}$.

Proposition 5 describes the E (or expectation) step for updating the estimate of the population-size parameter $N^{(n+1)}$ in the iterative algorithm. The minimization (or M) step consists of determining the parameter vector $\boldsymbol{\theta}^{(n+1)}$ that minimizes the least squares criterion V conditional upon the updated value of $N^{(n+1)}$. The latter step can be similarly performed via the Newton-Raphson algorithm, as described above, but where the parameter search is now confined to the three-dimensional space of $\boldsymbol{\theta} = (\beta, \alpha, i_0)$.

Confidence Intervals

Proposition 6. Let $\boldsymbol{\Theta}^* = (\beta^*, \alpha^*, i_0^*, N^*)$ and denote the estimated parameter values that minimize the objective function $V(\boldsymbol{\Theta})$ under the identifying restriction $r_0 = 0$, and let $V^* =$

$V(\boldsymbol{\Theta}^*)$ denote the corresponding minimized value of the least squares criterion V. Then the variance-covariance matrix of the estimated parameters $\boldsymbol{\Theta}^*$ is $\boldsymbol{C}^* = (s^*)^2(\boldsymbol{DX}^{*\prime}\boldsymbol{DX}^*)^{-1}$, where $(s^*)^2 = V^*/T$, and where $\boldsymbol{DX}^*$, as defined in Proposition 3, is also evaluated at the optimum $\boldsymbol{\Theta}^*$.

This result allows us to estimate confidence intervals for our parameter estimates $\boldsymbol{\Theta}^* = (\beta^*, \alpha^*, i_0^*, N^*)$ under the additional assumption that their distribution is asymptotically normal. Thus, the standard errors are the square roots of the diagonal elements of $\boldsymbol{C}^*$, while the symmetric 95 percent confidence intervals can be evaluated as ± 1.96 standard errors about the estimates. We can then use the Delta method [66] to compute the corresponding standard errors and confidence intervals around nonlinear functions of the parameters, such as the basic reproduction number $\mathcal{R}_0 = \beta/(1-\alpha)$ and the mean duration of infection $1/(1-\alpha)$.

Only a Linear Combination of the Parameters β and α is Practically Identifiable.

We have imposed the sharp prior restriction that $r_0 = 0$ to identify the remaining four parameters of our SIR model. If there exists a local interior minimum $\boldsymbol{\Theta}^* = (\beta^*, \alpha^*, i_0^*, N^*)$ of our least squares objective function $V(\boldsymbol{\Theta})$, we know that the gradient $\boldsymbol{DV}(\boldsymbol{\Theta}^*)$ is the zero vector and the Hessian $\boldsymbol{D^2V}(\boldsymbol{\Theta}^*)$ is a non-negative definite matrix. When $\boldsymbol{\Theta}^*$ is a strict local minimum, $\boldsymbol{D^2V}(\boldsymbol{\Theta}^*)$ will be strictly positive definite.

Even if our objective function V is indeed *strictly* locally convex at $\boldsymbol{\Theta}^*$ and thus $\boldsymbol{D^2V}(\boldsymbol{\Theta}^*)$ is strictly positive definite, it turns out that the Hessian $\boldsymbol{D^2V}(\boldsymbol{\Theta}^*)$ is *nearly* positive semidefinite. That is, $V(\boldsymbol{\Theta})$ is *nearly* flat along a ray passing through $\boldsymbol{\Theta}^*$ where the parameters β and α are allowed to vary and the remaining parameters i_0^* and N^* remain fixed. In effect, the three-dimensional plot of V projected onto the (β, α) plane contains a ravine where only a linear combination of β and α is *practically identified* [67].

The following proposition, which formalizes the idea, will guide our empirical implementation below.

Proposition 7. Let $\boldsymbol{\Theta}^*$ be a local interior minimum of $V(\boldsymbol{\Theta})$. Define a single-valued mapping from the real line into the four-dimensional subspace Ω of admissible values of $\boldsymbol{\Theta}$ as follows: $\boldsymbol{\lambda}(u) = (\beta^* + u, \alpha^* - wu, i_0^*, N^*)$, where u is its real-valued argument and $w > 0$ is a positive real parameter. As the parameter w varies, this mapping characterizes a family of rays in four-dimensional space passing through the point $\boldsymbol{\Theta}^*$ at $u = 0$. Then there exists a ray, corresponding to a specific value of w, along which the second-order directional derivative

$\left.\frac{d^2V}{du^2}\right|_{u=0}$ is nearly zero, that is, along which $V(\boldsymbol{\Theta})$ is *nearly* equal to its local minimum value $\boldsymbol{V}(\boldsymbol{\Theta}^*)$.

Consistency Checks with Data Simulated from Known Parameters

Apart from our study of COVID-19 case incidence in New York City, to be described below, we applied our estimation procedure to a simulated database generated from known parameters. Specifically, we computed the output variables X_t from the model of equations (2), (3) and (4) over $T = 100$ time periods with assumed known parameters $\boldsymbol{\Theta} = (\beta, \alpha, i_o, N) = (0.6, 0.7, 0.04, 10^5)$. We then repeatedly simulated the observed case incidence as $y_t^{(j)} = X_t^{(j)} + \varepsilon_t^{(j)}$ for each of $t = 1, \ldots, 100$ time periods and $j = 1, \ldots, 100$ epidemics, where the independent errors $\varepsilon_t^{(j)}$ were randomly drawn from a Gaussian $N(0, \sigma^2)$ distribution with zero mean and standard deviation $\sigma = 100$. For each epidemic j, we then employed our 4-dimensional Newton-Raphson algorithm to estimate the parameters $\boldsymbol{\Theta}^{(j)} = \left(\beta^{(j)}, \alpha^{(j)}, i_0^{(j)}, N^{(j)}\right)$. In addition, for each epidemic j, we computed the 3-dimensional estimates of $\boldsymbol{\theta}^{(j)} = \left(\beta^{(j)}, \alpha^{(j)}, i_0^{(j)}\right)$ conditional upon the population size N known to equal its assumed value of 10^5. Finally, for each of these exercises, we compared the dispersion of the resulting estimates $\boldsymbol{\Theta}^{(j)}$ around their assumed values $\boldsymbol{\Theta}$.

Computational Benchmarks

We employed two versions of our procedure to estimate the parameters of our SIR models: the 4-parameter search Newton-Raphson search method embodied in Propositions 2 and 3; and the EM-type procedure embodied in Propositions 4 and 5. We compared the computational times attained through these two approaches with a brute-force, grid search over the 4-dimensional space of allowable values of $\boldsymbol{\Theta}$. All programs were written in Mata [68], a matrix programming language embedded within the Stata programming language [69]. Calculations were carried out on a MacBook Pro with a 2.3 GHz 8-Core Intel Core i9 processor.

Data

Omicron Wave, December 2021 – March 2022, New York City

We studied the reported daily incidence of COVID-19 during the SARS-CoV-2 Omicron/BA.1 wave of December 2021 – March 2022 in New York City, NY, United States, a city of population 8.5 million. Our data consisted of daily counts of cases reported by the New York City department of health [70], where 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.

Our data showed systematic variation in case counts by day of the week, with many fewer cases diagnosed over the weekends. To account for these fluctuations, and to accommodate delays between symptom onset and testing, 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 1 shows the raw counts of daily reported cases c_t (connected gray datapoints) as well as the daily case counts y_t adjusted for the day of the week (red datapoints). We relied upon these adjusted daily counts y_t to estimate our SIR model parameters $\mathbf{\Theta} = (\beta, \alpha, i_0, N)$.

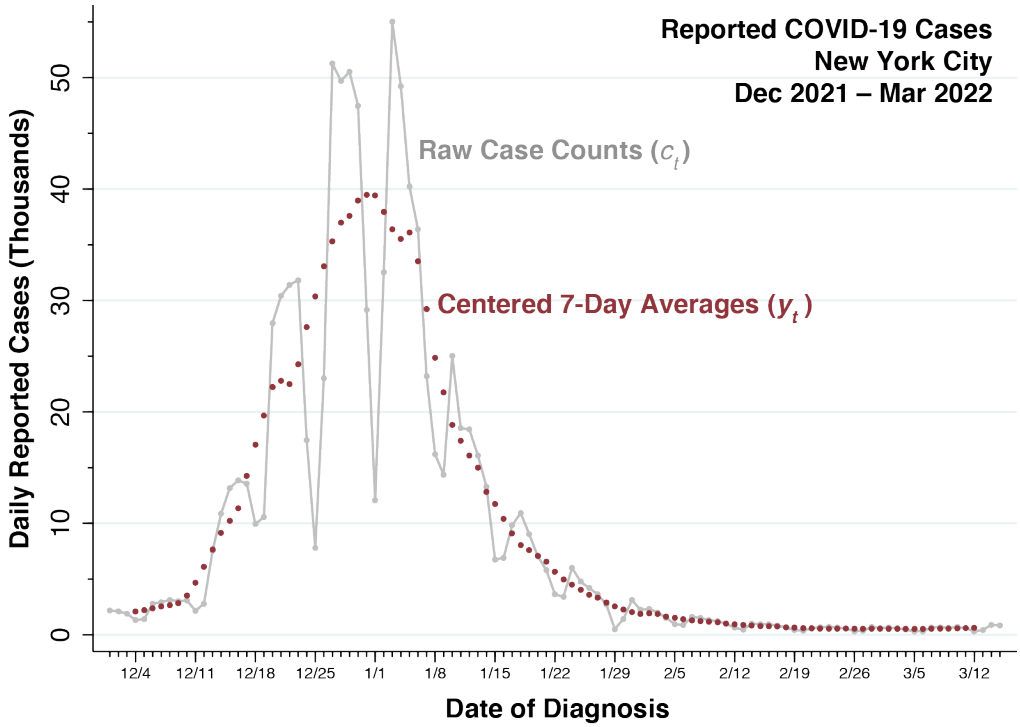

Figure 1. Daily Reported Cases of COVID-19, 12/1/2021 – 3/15/2022, in New York City. The connected gray datapoints show the raw case counts (c_t). The red datapoints, covering 12/4/2021 – 3/12/2022, show the centered 7-day moving averages (y_t).

The Omicron BA.1 variant of the SARS-CoV-2 virus was far-and-away the most important contributor to the massive surge of reported infections observed in Figure 1. Still, the initial phase of the surge overlapped the tail end of the prior Delta wave, while the terminal phase saw the gradual emergence of the Omicron BA.2 variant. Figure 2 plots the estimated daily proportions of the three variants in New York City during our study period.

As an alternative analysis, we re-estimated our SIR model parameters $\mathbf{\Theta} = (\beta, \alpha, i_0, N)$ from the data $\check{y}_t = h_t y_t$, where h_t represents the estimated proportion of BA.1 infections on date t, as derived from the interpolated Omicron BA.1 curve in Figure 2, and where y_t are the moving average counts of all cases displayed in Figure 1.

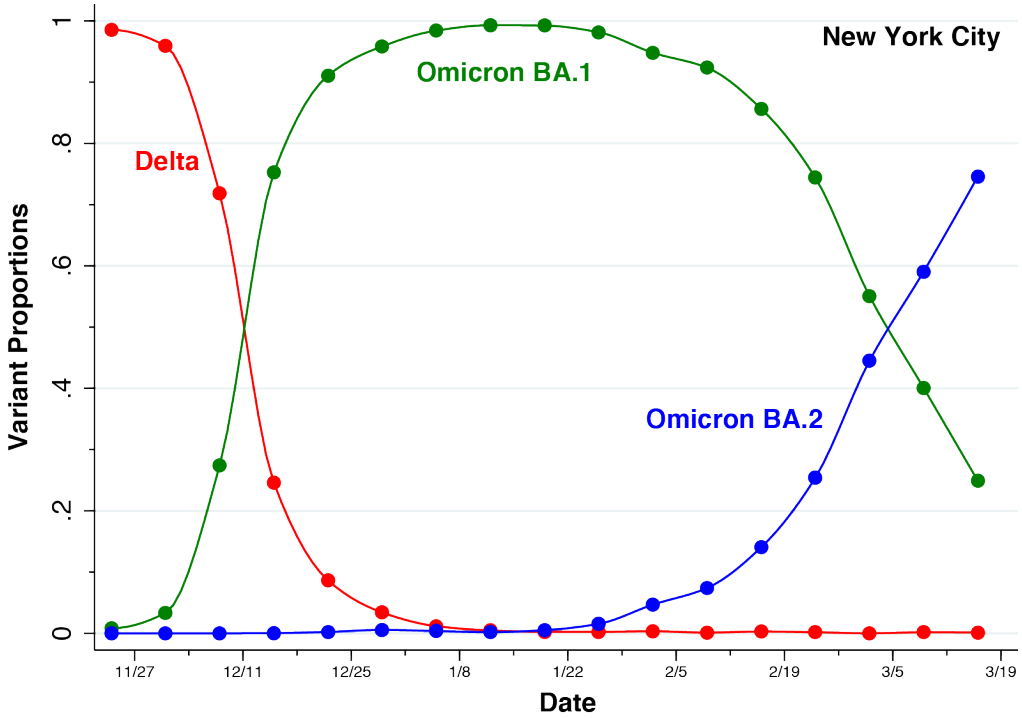

Figure 2. Estimates of the Proportions of the Delta, Omicron BA.1, and Omicron BA.2 Variants in New York City During 11/24/2021 – 3/16/2022. Weekly averages (solid datapoints) were derived from a compilation maintained by the New York City health department [71]. The Stata *pchipolate* interpolation routine [72] was then employed to estimate the intervening days (connecting curves). The Omicron BA.2 category included strains identified as BA.2, BA.2.12.1, and BA.2.75. Not shown are the proportions of all other strains, which represented no more than 0.7 percent of total samples in any one week.

Results

Omicron Wave, December 2021 – March 2022, New York City

Figure 3 shows the predicted values of the output variable X_t as connected curves superimposed on the observed datapoints y_t for all reported cases (shown in blue) and $\check{y}_t = h_t y_t$ for Omicron BA.1 only cases (shown in orange). For all reported cases, the SIR model showed a reasonably tight fit to the data. For Omicron BA.1 only cases, however, the fitted SIR model overpredicted the initial number of cases during the first 10 days of observation (from 12/4 through 12/13/2021). Comparing the two fitted curves, we see that our dropping the Delta variant cases during December 2021 resulted in an attenuation of the initial upswing of epidemic wave. By contrast, elimination of the Omicron BA.2 cases resulted in only a small absolute decrease in the tail of the wave during February and early March of 2022.

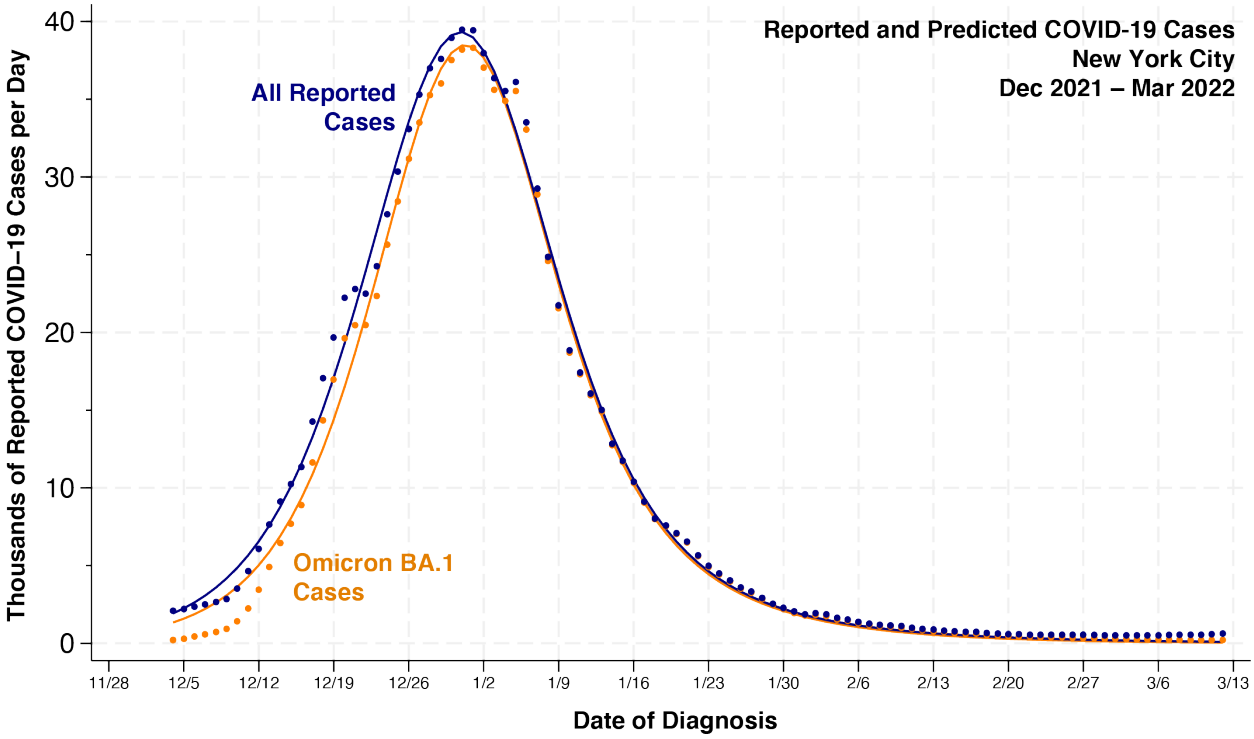

Figure 3. Daily Reported and Predicted Cases of COVID-19, 12/1/2021 – 3/15/2022, in New York City. Estimates Based Upon All Variants (Blue) and Omicron BA.1 Variant Only (Orange). Blue datapoints represent all reported cases y_t, while orange datapoints represent Omicron BA.1 cases $\check{y}_t = h_t y_t$. The respective blue and orange curves connect the predicted values of the output variable X_t. The displayed estimates are based upon the 4-dimensional Newton-Raphson algorithm described in Propositions 2 and 3. The estimates based upon the alternative EM procedure, described in Propositions 4 and 5, were identical.

Table 1 summarizes the resulting parameter estimates. The 95% confidence intervals based upon the 4-dimensional Newton-Raphson approach, shown in parentheses below each point estimate, accounted for the uncertainty in all components of the parameter vector $\mathbf{\Theta} = (\beta, \alpha, i_0, N)$. When we utilized the alternative EM-type algorithm, the final M-step estimated the three parameters (β, α, i_0) *conditional* upon the value of N estimated from the prior E-step. The corresponding conditional confidence intervals, shown in square brackets, were somewhat tighter.

Table 1. Parameter Estimates of an SIR Model of COVID-19 Incidence New York City, December 2021 – March 2022. Data on All Cases Versus Omicron BA.1 Only [a,b,c]

Parameter	All Reported Cases	Omicron BA.1 Only
β	0.233 (0.200, 0.266) [0.216, 0.250]	0.283 (0.189, 0.377) [0.266, 0.299]
α	0.941 (0.912, 0.969) [0.928, 0.953]	0.903 (0.815, 0.991) [0.890, 0.916]
$i_0 \times 10^{-3}$	8.23 (6.17, 10.28) [6.95, 9.50]	4.85 (2.20, 7.50) [4.15, 5.56]
$N \times 10^6$	1.013 (0.980, 1.046) –	0.984 (0.868, 1.100) –
$\mathcal{R}_0$	3.92 (2.58, 5.27) [3.35, 4.50]	2.92 (1.23, 4.61) [2.70, 3.14]
$1/(1-\alpha)$	16.8 (8.7, 25.0) [13.2, 20.5]	10.3 (0.9, 19.7) [8.9, 11.7]
V	76.7686	87.3200
AIC	–15.1780	–2.4284

a. The 4-dimensional Newton-Raphson approach and the alternative EM-algorithm approach gave the same point estimates of the parameters $\mathbf{\Theta} = (\beta, \alpha, i_0, N)$.
b. Symmetric 95% confidence intervals based upon the 4-dimensional Newton-Raphson algorithm are shown in parentheses below each estimate. The corresponding confidence intervals conditional upon the population size parameter N are shown in square brackets. Confidence intervals for the estimated basic reproduction number $\mathcal{R}_0 = \beta/(1-\alpha)$ and the estimated mean duration of infectivity $1/(1-\alpha)$ were based upon the Delta method.
c. The sum-of-squared-residuals V was computed as in equation (5). The Akaike information criterion (*AIC*) was computed as $2k + T\ln(V/T)$, where $k = 5$ parameters (including the parameter s^2) and $T = 99$ observations.

As indicated by the comparative values of the least squares criterion V and the corresponding Akaike information criterion (*AIC*) in Table 1, the model based upon all reported cases had a significantly tighter fit than the model based on cases of Omicron BA.1 alone. Overall, the Omicron BA.1-based model was only $exp((AIC_{All\ Cases} - AIC_{BA.1})/2) = 0.0017$ times as likely as the all cases-based model to minimize the information loss.

Plotting the Criterion V as a Function of the Parameters β and α

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 have been held constant at their estimated values, as shown in Table 1. The criterion V reached a minimum at the optimum β^* = 0.233, where $\partial V/\partial\beta = 0$. The function V was convex in the interval from β = 0.169, where $\partial V/\partial\beta$ reached a minimum, to β= 0.352, where $\partial V/\partial\beta$ reached 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 reached a minimum at the optimum α^* = 0.941, at which point $\partial V/\partial\alpha = 0$. The function V was convex in the interval from α = 0.868, where $\partial V/\partial\alpha$ reached a minimum, to α = 1, the boundary of admissible values of α, where $\partial V/\partial\alpha$ remained positive.

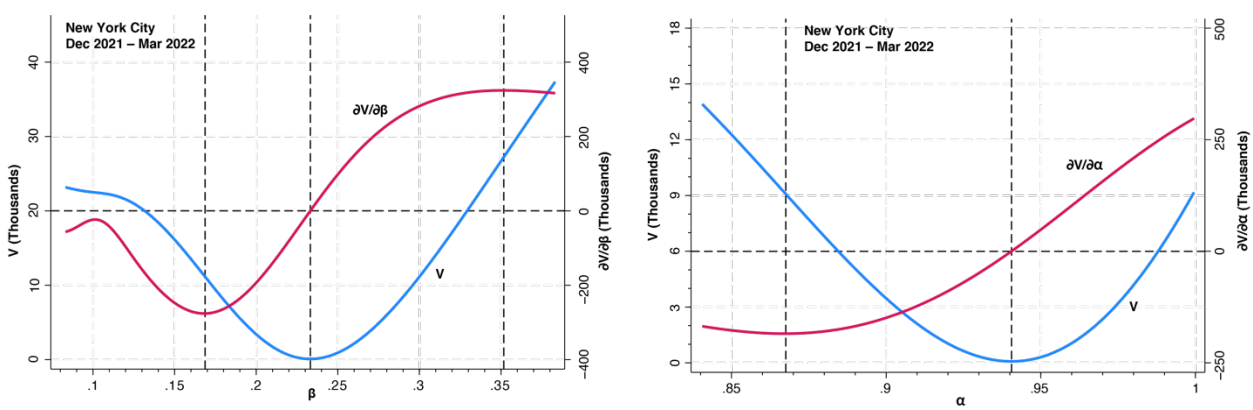

Figure 4. Panel A. Least Squares Criterion V (Left Axis) and First Partial Derivative $\partial V/\partial\beta$ (Right Axis) as Functions of the Parameter β. Panel B. Least Squares Criterion V (Left Axis) and First Partial Derivative $\partial V/\partial\alpha$ (Right Axis) as Functions of the Parameter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

Figure 5 plots the projection of V onto the (β, α) plane, once again based upon the data for all reported cases. As in Figure 4, the remaining parameters (i_0, N) were held at the estimated values given in Table 1 above. The darkest area represents a ravine where V attained its lowest values. The yellow point in the center is the global minimum (β^*, α^*) = (0.233, 0.941).

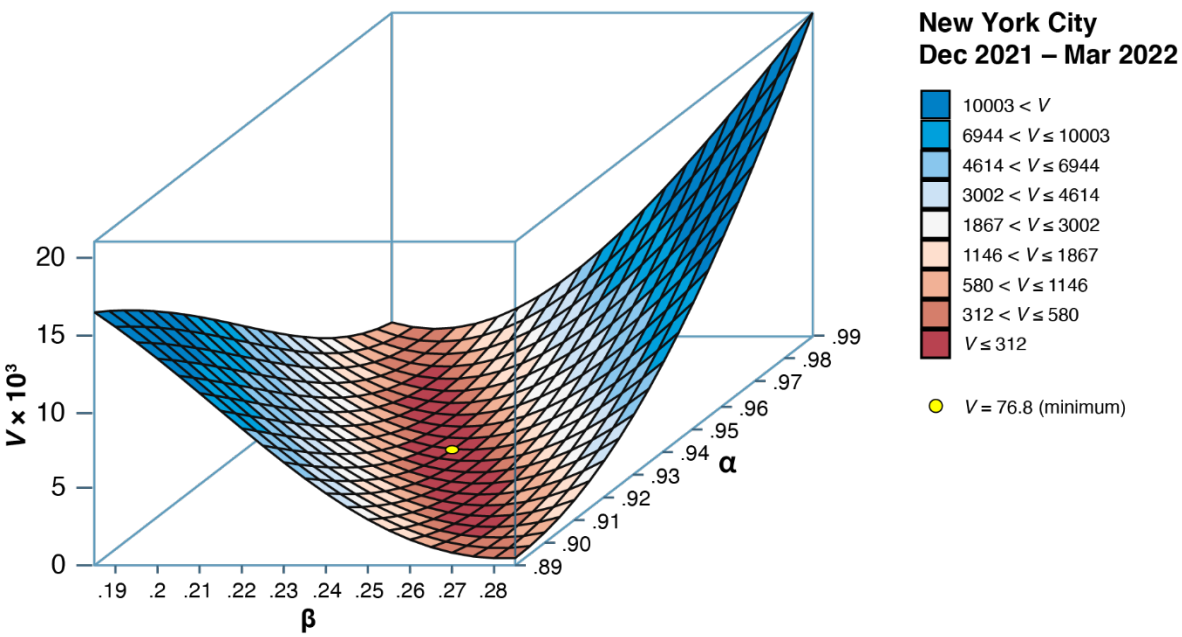

Figure 5. Least Squares Criterion V Projected onto the (β, α) Plane. The yellow point identifies the minimum where (β^*, α^*) = (0.233, 0.941). The parameters i_0 and N were held constant at their optimum values given in Table 1. The plot was produced in part from the Stata program *surface* [73].

Figure 6 further characterizes the ravine identified in Figure 5. To that end, we relied on the results of Proposition 7 to identify the ray through the optimum point (β^*, α^*) along which the second-order directional derivative of V was minimized. Shown in the figure are the contours of V as a function of (β, α) running parallel to each of the two parameter axes, as well as the contour along the ray identified has having the flattest curvature. The ray corresponded to the equation $w^*(\beta - \beta^*) + (\alpha - \alpha^*) = 0$, where w^*= 1.073. The curvature along the ray, as measured by the second-order directional derivative, was less than one-tenth of the corresponding curvatures measured along the contours running parallel to the axes. Conditional on w^*, the 95% confidence interval surrounding the linear combination $w^*\beta + \alpha$ had a very tight range of (1.183, 1.198).

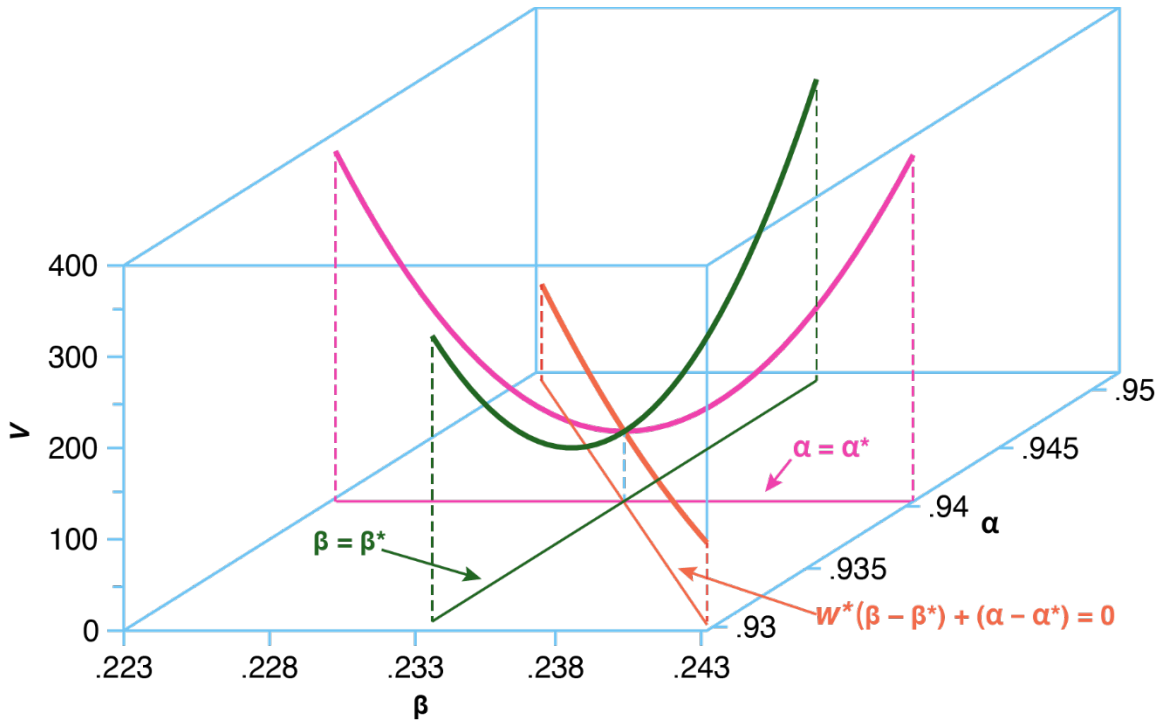

Figure 6. Contours of V Running through the Optimum Point Θ^* Projected onto the (β, α) Plane. All plotted contours run through the optimum $(\beta^*, \alpha^*) = (0.233, 0.941)$, with the parameters i_0 and N held constant at their optimum values. With β also held constant at its optimum value (green contour), the second-order derivative was $\partial^2 V(\boldsymbol{\theta}^*)/\partial\alpha^2 = 4.912\times10^6$. With α also held constant at its optimum value (magenta contour), the second-order derivative was $\partial^2 V(\boldsymbol{\theta}^*)/\partial\beta^2 = 6.075\times10^6$. Along the ray defined by $w^* = 1.073$ (orange contour), the second-order directional derivative was $(d^2V/du^2)|_{u=0} = 0.418\times10^6$.

In Appendix B, we further plot the path of successive iterations of the 4-parameter Newton-Raphson algorithm, showing how the estimate $\boldsymbol{\Theta}^{(k)}$ entered the ravine at iteration $k = 10$ and then stopped at the optimum on iteration $k = 40$.

Consistency Checks with Data Simulated from Known Parameters

In Panel A of Figure 7, we plot the estimates $\beta^{(j)}$ against the estimates $\alpha^{(j)}$ for $j = 1, \ldots 100$ simulated epidemics with known parameters $(\beta, \alpha, i_o, N) = (0.6, 0.7, 0.04, 10^5)$ and thus with known basic reproduction number $\mathcal{R}_0 = \beta/(1-\alpha) = 2$. The gold solid points represent the estimates based upon the 4-parameter Newton-Raphson algorithm, while the superimposed dark blue circles represent the estimates based on a 3-parameter Newton-Raphson algorithm conditional upon the population size N equal to its assumed value of 10^5. In Panel B, we show the corresponding box-and-whisker plots of the estimated basic reproduction numbers $\mathcal{R}_0^{(j)}$ under the same two estimation conditions.

The plots in Figure 7 show that under both conditions, the estimates $\left(\beta^{(j)}, \alpha^{(j)}\right)$ were centered around their assumed values. Their respective sample means were (0.604, 0.696) with all four parameters unconstrained and (0.601, 0.699) conditional upon the known value of the population parameter N. The corresponding sample medians of $\left(\beta^{(j)}, \alpha^{(j)}\right)$ were (0.601, 0.699) and (0.600, 0.700) under the two respective estimation conditions. In addition, the sample means of the basic reproduction numbers $\mathcal{R}_0^{(j)}$ were 2.029 and 1.998 under the two respective estimation conditions, while their respective sample medians were 2.001 and 1.999.

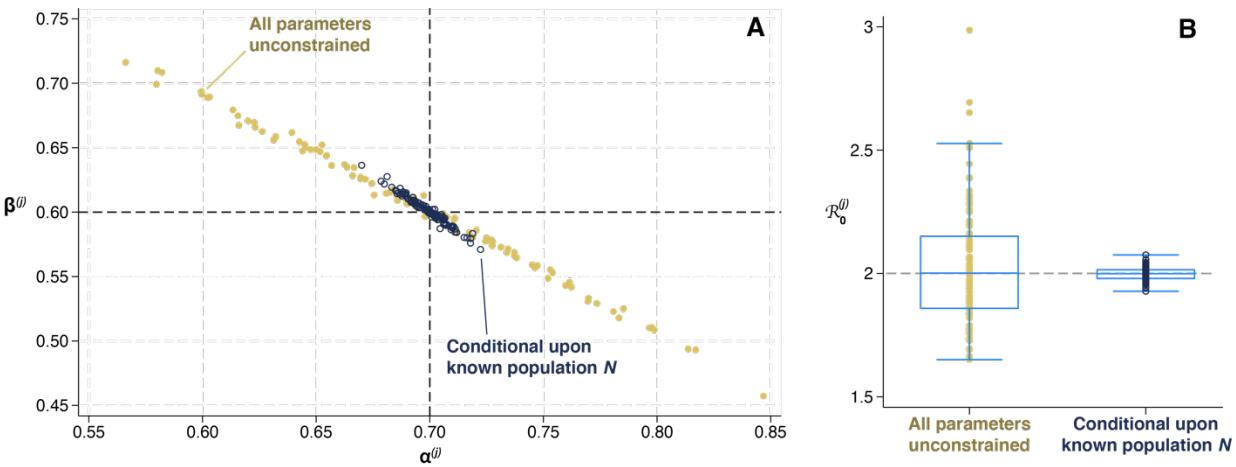

Figure 7. Estimated Parameters Based Upon a Simulated SIR Model with Gaussian Error. The SIR model of (2), (3) and (4) was simulated with assumed known parameters $\mathbf{\Theta} = (\beta, \alpha, i_o, N) = (0.6, 0.7, 0.04, 10^5)$ over $T = 100$ time periods. The observed case incidence was then computed as $y_t = X_t + \varepsilon_t$, where the independent Gaussian errors ε_t were randomly drawn with zero mean and standard deviation $\sigma = 100$. Panel A. Plots of parameter estimates $\beta^{(j)}$ versus $\alpha^{(j)}$ with all parameters unconstrained (gold solid points) and conditional upon a known population N (dark blue circles). Panel B. Box-and-whisker plots of the corresponding estimates $\mathcal{R}_0^{(j)}$ of the basic reproduction number. The boxes show the 25th, 50th, and 75th percentiles while the capped lines show the bounds of interquartile range.

Figure 7 confirms the strong negative correlation of the two estimated parameters, which were tightly situated along a near-linear paths. When we employed the 4-dimensional Newton-Raphson algorithm, the estimates were widely dispersed. Our conditioning upon the known value of the population parameter N substantially narrowed the dispersion, but it did not affect the observed strong negative correlation between the estimates of β and α. Thus, for the gold points in Figure 7A, representing the simulation results with unconstrained parameters, the ordinary least squares regression of $\beta^{(j)}$ against $\alpha^{(j)}$ yielded a slope of -0.911 with 95% confidence interval $[-0.924, -0.899]$. For the dark blue points, representing the simulation results

conditional on a known parameter N, the corresponding slope was -1.184 with 95% confidence interval $[-1.225, -1.142]$.

Benchmarking

Estimation of the parameters $\boldsymbol{\Theta}$ from the New York City Omicron data utilizing the 4-parameter Newton-Raphson method of Propositions 2 and 3 required 0.434 seconds of computational time, whereas estimation of $\boldsymbol{\Theta}$ utilizing the EM-type method of Propositions 4 and 5 required 0.925 seconds. By contrast, estimation of the parameters utilizing an exhaustive grid search routine required 463.682 seconds, about three orders of magnitude longer than the 4-parameter Newton-Raphson routine.

Discussion

We have described a new strategy for estimating the underlying parameters of the Susceptible-Infected-Removed (SIR) model from observations on case incidence alone. Our approach finds the parameter vector that minimizes the sum of squared deviations between observed and predicted case incidence data. It does not require that we derive a closed-form mathematical expression for the predicted case incidence as a function of the parameters. Instead, we show how the gradient vector and the Hessian matrix of the least squares objective function can be exactly calculated. These quantities alone are sufficient to implement the Newton-Raphson algorithm to locate the parameter vector that minimizes the objective function.

Not All Five SIR Parameters Can be Identified from Case Incidence Data Alone.

Our results shed light on the identifiability of the SIR model parameters from observations on case incidence alone. Our full SIR model had five parameters: β, the infection transmission parameter; α, the infection persistence parameter, gauging the proportion of infected individuals who remain in the infected state at each discrete time t; i_0, the initial proportion of the population infected; r_0, the initial proportion of the population in the removed state; and N, the total population.

In Proposition 1 of our theoretical analysis, we found that all five parameters were not jointly identifiable from the incidence data alone. (A related result appears to have been derived from a continuous-time SIR model with seasonal forcing [53].) Accordingly, we estimated the first four parameters under the identifying restriction that $r_0 = 0$. In effect, we assumed that because of immune escape [74, 75], no one at the onset of the Omicron wave was already resistant to infection by the newly emergent SARS-CoV-2 variant.

Only a Linear Combination of β and α is Practically Identifiable.

Our findings based upon the New York City Omicron data confirmed that the remaining four parameters $\boldsymbol{\Theta} = (\beta, \alpha, i_0, N)$ could be indeed identified at a global interior minimum of the least squares objective function V. But we also uncovered a deeper problem of parameter identification. While the bivariate plots of Figure 4 demonstrated that the objective function was locally strictly convex at the optimum point, the three-dimensional plot of Figure 5 revealed that this point was situated along the floor of a ravine where the parameters β and α were highly correlated. In Figure 6, we were able to identify the floor of the ravine as the line in the (β, α)

plane where the second-order derivative of V was minimized. While the individual parameters β and α had wide estimated confidence intervals, the linear combination of β and α along the ravine could be tightly estimated. Appendix Figure B1 further showed how successive estimates $\boldsymbol{\Theta}^{(k)}$ of the four-dimensional Newton-Raphson algorithm followed along the ravine floor until the point of convergence.

What's more, our consistency checks on a simulated database confirmed that the extremely high correlation between the parameters β and α was not an idiosyncrasy of the New York City Omicron data. The results shown in Figure 7, based on simulated data, demonstrated that while the dispersion of the estimates $\left(\beta^{(j)}, \alpha^{(j)}\right)$ were centered around their true values, they similarly tended to demonstrate a near-linear relation between the two parameters. Conditioning on the population size parameter N reduced the dispersion of the estimates in the (β, α) plane as well as the estimated basic reproduction number $\mathcal{R}_0$ – a finding consistent with the narrower confidence intervals shown in Table 1 for the New York City Omicron data – but it did not attenuate the strong negative correlation between the parameters.

Even with the identifying restriction that the initial proportion of recovered individuals r_0 is zero, the remaining four-parameter version of SIR is what some physicists have described as a "sloppy" model [52, 54]. While β and α are, strictly speaking, separately identifiable from case incidence data, we conclude that only a linear combination of the two parameters is *practically* identifiable [67]. (A related result appears to have been derived in [76].)

We Cannot Rely on Census Data Alone to Identify the Parameter N.

Imposing the strong identifying assumption that the initial proportion r_0 of recovered individuals was zero, we estimated the population size parameter N to be 1.013 million, which amounted to only 11.9 percent of New York City's total census population of 8.5 million. This discordant result suggests that it might have been more appropriate to set $N = 8.5$ million as the identifying assumption. In that case, however, relying upon Proposition 1, we would then end up with a rather large estimate of $r_0 = 0.881$, which would in turn imply $1 - r_0 = 0.119$ and thus $N(1 - r_0) = 1.013$ million. Despite the evidence of immune escape on the part of the Omicron variant [75], we would conclude that a very substantial fraction of the population still retained long-term cellular immunity, possibly through the administration of multiple vaccine doses [77].

Unfortunately, such an alternative strategy of relying on census-based population data to identify the parameter N runs into two important complications: 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 [78]. There is concrete evidence of significant underreporting of Omicron infections, due principally to the widespread availability of home rapid antigen testing [78], particularly in New York City [79, 80]. 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 $p < 1$ [76, 81, 82]. 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 learned from Proposition 4, however, that each output variable X_t can be written in the form $\varphi_t(\boldsymbol{\theta})N$, where $\varphi_t(\boldsymbol{\theta})$ is a function of the remaining identifiable parameters $\boldsymbol{\theta} = (\beta, \alpha, i_0)$. Accordingly, our reported incidence data y_t are really estimates of $pX_t = \varphi_t(\boldsymbol{\theta})(pN)$. To identify our model, we thus need a prior estimate of pN. Unless we're prepared to impose additional restrictions on p, census data on N alone will not be enough.

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 [19, 83-88]. This caution applies just as well to the Omicron wave, when a substantial proportion of the population avoided retail establishments, drinking and eating places, transportation venues, worksites and other high-risk locations [89].

Take the case of the COVID-19 outbreak at the campus of the University of Wisconsin-Madison in September 2020, where total student enrollment was 44,640, but where the large fraction of cases was concentrated in two on-campus student residence halls with a combined population of about 2,932 [90]. If we relied on the campus-wide census to identify the parameter N on the campus-wide population, we would have ended up concluding erroneously that about 94 percent of the student population was already immune to SARS-CoV-2.

Should We Exclude the Non-Omicron Case Counts from our Database?

In Figure 3, we saw that exclusion of non-Omicron-variant cases from our database resulted in a fitted model that overshot the initial upswing of cases during December 2021. That is, the predicted output variable X_t exceeded the observed counts $\check{y}_t$ during $t = 1, \ldots, 10$. Our AIC computations in Table 1 confirmed that dropping the residual Delta-variant infections from the previous wave significantly reduced the overall goodness of fit of our SIR model to the new Omicron wave data.

Despite the evidence favoring immune escape by Omicron [75], these residual Delta-variant infections may have conferred partial cross-immunity against the newly emerging Omicron variant and thus reduced the initial proportion of individuals susceptible to Omicron BA.1 [91]. If so, excluding these Delta cases would upwardly bias the model estimates of the numbers of susceptible individuals S_t and, consequently, the predicted output variable X_t.

Still, the apparent preference for inclusion of the non-Omicron cases is hardly clear cut. The unconditional confidence intervals are simply too wide to distinguish sharply between the corresponding parameter estimates in the two columns of Table 1.

Our Estimates of the Parameter α Appear to be Inconsistent with Clinical Observation.

Our estimates of the parameter α appear to be out of line with what is known from direct clinical measurement about the duration of infectivity from the Omicron variant. In one cohort of 55 symptomatic Omicron-infected patients, only 13.5 percent continued to shed virus ten days after infection [92]. Yet our point estimates of the proportion remaining infectious after ten days (computed as α^{10}) would be 54.4 percent from the combined-variant data and 36.0 percent from the Omicron-only data. A meta-analysis of 29 clinical studies suggested that while the mean duration of polymerase chain reaction (PCR) positivity was 10.82 days, the mean duration of viral shedding was only 5.16 days [93]. Yet our point estimates of the mean duration of infectivity (computed as $1/(1-\alpha)$) were 16.8 days from the combined-variant data and 10.3 days from the Omicron-only data.

We need to interpret this apparent discrepancy with caution. The unconditional confidence intervals around our estimates of α were quite wide. A reported short mean duration of infectivity does not rule out a long-tailed distribution. In one study, 25 percent of patients were still shedding virus after 8 days [94]. What's more, an individual's infectivity is not an all-or-nothing characteristic. It is likely to depend on the extent to which an infected individual

isolates himself from others. That behavioral component of the parameter α may have changed during the Omicron wave.

Is SIR the Correct Structural Model for Omicron?

Despite its close fit to the data in Figure 3, the Susceptible-Infected-Recovered framework may thus fail as a *structural* model of the Omicron wave, even if it apparently succeeds as a *reduced form* model [95]. When we imposed an identifying restriction on the parameter r_0 representing the initial proportion of recovered individuals, we could indeed back out the remaining parameters (β, α, i_0, N) from the observed case incidence data and thus solve the inversion problem. But the resulting parameter estimates do not necessarily warrant the *structural* interpretation that we assumed in our exposition of the SIR model in equations (1) through (5) above.

That is not to say, however, that an adequate reduced form model is incapable of making accurate projections [96]. But when it comes to sharply identifying individual parameters and thus distinguishing between alternative epidemic models, some prior information above and beyond case incidence data may be required.

Other Limitations and Extensions

Parameter search algorithms such as Newton-Raphson [46] have superior computational performance when they rely on exact expressions for the gradient vector and Hessian matrix of second derivatives, rather than on numerical approximation or brute-force search [45]. Still, as Figure 4 shows, there may be regions of the parameter space where the least squares criterion function is non-convex and the algorithm does not converge. While Appendix B displays convergence from a suitably chosen starting point, the tasks of finding the right initial values and keeping the parameter search within bounds remain unavoidable challenges.

To approximate a continuous-time dynamic model as closely as possible, we relied on daily counts of reported COVID-19 cases. As Figure 1 shows, however, such a fine level of detail resulted in striking day-of-the-week effects, which 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 output variable X_t have yet to be tried.

Applying our approach to the New York City Omicron wave, we have estimated SIR model parameters from counts of reported incident cases. Our approach can be generalized to estimate SIR model parameters from counts of removed cases, either through recovery or death [64]. Our approach can be further generalized to cases where we have only partial rather than complete temporal observations on the output variable [64].

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. Our findings offer a caution, however, that such models as SEIR, which require an additional parameter governing the transition from an intermediate *exposed* state to the *infected* state, may very well turn out to be sloppy [76].

One exemption may be the well-studied SIRS model [58, 97], where individuals in the R (recovered) state can transition back to the S (susceptible) state due to waning immunity. The SIRS model still has three states, but the parameter vector 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 [58, 98], we would be in a position to run our parameter recovery algorithm on multiple waves of data y_t. That is a task for future research.

Appendix A. Proofs of Propositions

Proposition 1. Let $\Omega \subseteq \mathbb{R}^5$ be the subspace of admissible values of the five-dimensional parameter vector $\boldsymbol{\Theta}$. Let $\boldsymbol{X}(\boldsymbol{\Theta})$ be the resulting $T \times 1$ column vector of output variables defined by the model of equations (1) through (4). Then there exists a mapping $\phi: \Omega \to \mathbb{R}^4$ with the property that $\boldsymbol{X}(\boldsymbol{\Theta}) = \boldsymbol{X}(\boldsymbol{\Theta}')$ for all vectors $\boldsymbol{\Theta}, \boldsymbol{\Theta}' \in \Omega$ satisfying $\phi(\boldsymbol{\Theta}) = \phi(\boldsymbol{\Theta}')$.

Proof. We characterize the mapping ϕ from the five-parameter vector $\boldsymbol{\Theta}$ into a four-parameter vector $\phi(\boldsymbol{\Theta}) = (\kappa, \lambda, \mu, \nu)$ as follows:

$$\kappa = \beta(1 - r_0) \tag{A1a}$$

$$\lambda = \alpha \tag{A1b}$$

$$\mu = i_0/(1 - r_0) \tag{A1c}$$

$$\nu = N(1 - r_0) \tag{A1d}$$

Now consider two distinct parameter vectors $\boldsymbol{\Theta} = (\beta, \alpha, i_0, r_0, N)$ and $\boldsymbol{\Theta}' = (\beta', \alpha', i_0', r_0', N')$, where $\boldsymbol{\Theta} \neq \boldsymbol{\Theta}'$. The condition $\phi(\boldsymbol{\Theta}) = \phi(\boldsymbol{\Theta}')$ requires that:

$$\beta'(1 - r_0') = \beta(1 - r_0) \tag{A2a}$$

$$\alpha' = \alpha \tag{A2b}$$

$$i_0'/(1 - r_0') = i_0/(1 - r_0) \tag{A2c}$$

$$N'(1 - r_0') = N(1 - r_0) \tag{A2d}$$

We need to show that $\boldsymbol{X}(\boldsymbol{\Theta}) = \boldsymbol{X}(\boldsymbol{\Theta}')$. From (3) in the main text, each component X_t of the vector $\boldsymbol{X}(\boldsymbol{\Theta})$ in turn depends on the ratio (β/N) as well as the corresponding time-specific state variables S_t and I_t. Dividing (A2a) by (A2d), we see that $(\beta/N) = (\beta'/N')$. Hence, it remains only to show that $S_t(\boldsymbol{\Theta}) = S_t(\boldsymbol{\Theta}')$ and $I_t(\boldsymbol{\Theta}) = I_t(\boldsymbol{\Theta}')$ for all times $t = 0, \dots, T$.

We proceed by mathematical induction. First consider time $t = 0$. For the respective parameter vectors $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}'$, we have $S_0(\boldsymbol{\Theta}) = (1 - i_0 - r_0)N$ and $S_0(\boldsymbol{\Theta}') = (1 - i_0' - r_0')N'$. From (A2c), we can write $i_0' = i_0 \frac{(1-r_0')}{(1-r_0)}$, and from (A2d), we can write $N' = N \frac{(1-r_0)}{(1-r_0')}$. So, $S_0(\boldsymbol{\Theta}') = \left(1 - i_0 \frac{(1-r_0')}{(1-r_0)} - r_0'\right) N \frac{(1-r_0)}{(1-r_0')} = (1 - i_0 - r_0)N = S_0(\boldsymbol{\Theta})$. Similarly, we have $I_0(\boldsymbol{\Theta}) = i_0 N$ and $I_0(\boldsymbol{\Theta}') = i_0' N'$. Again, it follows from (A2c) and (A2d) that $i_0 N = i_0' N'$ and thus $I_0(\boldsymbol{\Theta}) = I_0(\boldsymbol{\Theta}')$. So, the proposition holds for $t = 0$.

Now assume that $S_t(\boldsymbol{\Theta}) = S_t(\boldsymbol{\Theta}')$ and $S_t(\boldsymbol{\Theta}) = I_t(\boldsymbol{\Theta}')$ for any $t \geq 0$. We show that $S_{t+1}(\boldsymbol{\Theta}) = S_{t+1}(\boldsymbol{\Theta}')$ and $I_{t+1}(\boldsymbol{\Theta}) = I_{t+1}(\boldsymbol{\Theta}')$. From (1a) in the main text, we have $S_{t+1}(\boldsymbol{\Theta}) = S_t(\boldsymbol{\Theta}) - \beta S_t(\boldsymbol{\Theta}) I_t(\boldsymbol{\Theta})/N$. We've already learned from (A2a) and (A2d) that $(\beta/N) = (\beta'/N')$, so we can instead write $S_{t+1}(\boldsymbol{\Theta}) = S_t(\boldsymbol{\Theta}) - \beta' S_t(\boldsymbol{\Theta}) I_t(\boldsymbol{\Theta})/N'$. Since we've assumed that $S_t(\boldsymbol{\Theta}) = S_t(\boldsymbol{\Theta}')$ and $S_t(\boldsymbol{\Theta}) = I_t(\boldsymbol{\Theta}')$, we can therefore write $S_{t+1}(\boldsymbol{\Theta}) = S_t(\boldsymbol{\Theta}') - \beta' S_t(\boldsymbol{\Theta}') I_t(\boldsymbol{\Theta}')/N' = S_t(\boldsymbol{\Theta}')$. Similarly, from (1b), we have $I_{t+1}(\boldsymbol{\Theta}) = \beta S_t(\boldsymbol{\Theta}) I_t(\boldsymbol{\Theta})/N + \alpha I_t(\boldsymbol{\Theta})$. Substituting $(\beta/N) = (\beta'/N')$ and $\alpha' = \alpha$ from equation (A2b) gives $I_{t+1}(\boldsymbol{\Theta}) = I_{t+1}(\boldsymbol{\Theta}')$. ■

Corollary 1. The four components of the parameter vector $\boldsymbol{\Phi} = (\kappa, \lambda, \mu, \nu)$ in the following SIR model can be identified from the data $\boldsymbol{y}$ on case incidence:

$$S_0 = (1 - \mu)\nu \tag{A3a}$$

$$I_0 = \mu\nu \tag{A3b}$$

$$X_t = \kappa S_{t-1} I_{t-1}/\nu \tag{A3c}$$

$$S_t = S_{t-1} - X_t \tag{A3d}$$

$$I_t = X_t + \lambda I_{t-1} \tag{A3e}$$

Proof. We repeat the basic equations of our dynamic model, as articulated in equations (1), (2), and (3) of the main text:

$$S_0 = (1 - i_0 - r_0)N \tag{A4a}$$

$$I_0 = i_0 N \tag{A4b}$$

$$X_t = \beta S_{t-1} I_{t-1}/N \tag{A4c}$$

$$S_t = S_{t-1} - X_t \tag{A4d}$$

$$I_t = X_t + \alpha I_{t-1} \tag{A4e}$$

Now substitute definitions of $(\kappa, \lambda, \mu, \nu)$ given in equations (A1) in the proof of Proposition 1.

To define the new four-dimensional parameter vector $\boldsymbol{\Phi}$, we have essentially dropped the parameter r_0 from the original five-dimensional vector $\boldsymbol{\Theta}$ and then normalized the remaining four

parameters (β, α, i_0, N) by the factor $(1 - r_0)$. The resulting mapping $\boldsymbol{X}(\boldsymbol{\Phi})$, as defined in (A4) is one-to-one. ■

Corollary 2. The sharp prior restriction that $r_0 = 0$ is sufficient to identify the remaining four parameters (β, α, i_0, N) of the five-parameter vector $\boldsymbol{\Theta}$.

Proof. Consider a specific parameter vector $\boldsymbol{\Theta} = (\beta, \alpha, i_0, 0, N)$ with $r_0 = 0$. From equations (A2) in Proposition 1, any other parameter vector $\boldsymbol{\Theta}' = (\beta', \alpha', i_0', r_0', N')$ satisfying $\phi(\boldsymbol{\Theta}') = \phi(\boldsymbol{\Theta})$ and therefore $\boldsymbol{X}(\boldsymbol{\Theta}') = \boldsymbol{X}(\boldsymbol{\Theta})$ must satisfy these conditions:

$$\beta = \beta'(1 - r_0') \tag{A5a}$$

$$\alpha = \alpha' \tag{A5b}$$

$$i_0 = i_0'/(1 - r_0') \tag{A5c}$$

$$N = N'(1 - r_0') \tag{A5d}$$

However, the only vector $\boldsymbol{\Theta}' = (\beta', \alpha', i_0', r_0', N')$ satisfying these conditions *and* adhering to the restriction that $r_0' = 0$ is the vector $\boldsymbol{\Theta}$ itself. Hence, $\boldsymbol{\Theta}$ is uniquely identified. We further note that under the restriction that $r_0 = 0$, the vector of the remaining four parameters (β, α, i_0, N) is exactly equal to the four-parameter vector $\boldsymbol{\Phi} = (\kappa, \lambda, \mu, \nu)$ defined in Corollary 1. ■

Corollary 3. A sharp prior restriction on any one of the four parameters (β, i_0, r_0, N) is sufficient to identify the full five-parameter vector $\boldsymbol{\Theta} = (\beta, \alpha, i_0, r_0, N)$. A sharp prior restriction on the parameter α alone, however, will not identify the remaining parameters.

Proof. We're essentially implementing the proof strategy for Corollary 2. For example, let's impose the sharp prior restriction that the population size N is known and equal to $\widehat{N}$, and consider a specific parameter vector $\boldsymbol{\Theta} = \left(\beta, \alpha, i_0, r_0, \widehat{N}\right)$ with $N = \widehat{N}$. Any other parameter vector $\boldsymbol{\Theta}' = (\beta', \alpha', i_0', r_0', N')$ satisfying $\phi(\boldsymbol{\Theta}') = \phi(\boldsymbol{\Theta})$ and therefore $\boldsymbol{X}(\boldsymbol{\Theta}') = \boldsymbol{X}(\boldsymbol{\Theta})$ must satisfy the conditions in equations (A2) in Proposition 1 as well as the additional restriction that $N' = \widehat{N}$. Since equation (A2d) alone requires that $N'(1 - r_0') = \widehat{N}(1 - r_0)$, the additional restriction would imply that $r_0' = r_0$. From the remaining equations in (A2), we therefore conclude that $\beta' = \beta$, $\alpha' = \alpha$, and $i_0' = i_0$ as well. Hence, $\boldsymbol{\Theta}$ is uniquely identified. We note that imposing a prior restriction that the parameter α alone does no more than repeat the requirement

of equation (A2b) and hence does not identify the remaining parameters. However, a sharp prior restriction on the basic reproductive number $\mathcal{R}_0$ would also be sufficient to identify all five parameters of the vector $\boldsymbol{\Theta}$. ■

Proposition 2. The gradient of the least squares criterion V with respect to the parameter vector $\boldsymbol{\Theta} = (\beta, \alpha, i_0, N)$ is:

$$\boldsymbol{DV} = -2\sum_{t=1}^{T}(y_t - X_t)\,\boldsymbol{DX_t} \tag{A2}$$

where each column vector $\boldsymbol{DX_t} = \left(\frac{\partial X_t}{\partial \beta}, \frac{\partial X_t}{\partial \alpha}, \frac{\partial X_t}{\partial i_0}, \frac{\partial X_t}{\partial N}\right)'$ represents the corresponding gradient of partial derivatives of the output variable X_t at time t. The basic equations of our dynamic system (1), in combination with the initial conditions (2), can be used to generate complete, computable difference equations for $\boldsymbol{DX_t}$ for all t, and thus for $\boldsymbol{DV}$.

Proof. Equation (A2), which repeats equation (6), is derived by taking the derivative of V as defined in equation (5). Taking the derivative of $\boldsymbol{DX_t}$, we get:

$$\boldsymbol{DX_t} = \left(S_{t-1}\frac{I_{t-1}}{N}\right)\boldsymbol{D\beta} + \left(\frac{\beta}{N}\right)I_{t-1}\boldsymbol{DS_{t-1}} + \left(\frac{\beta}{N}\right)S_{t-1}\boldsymbol{DI_{t-1}} - \left(\frac{\beta}{N^2}S_{t-1}I_{t-1}\right)\boldsymbol{DN} \tag{A3}.$$

where $\boldsymbol{D\beta} = (1,0,0,0)'$ and $\boldsymbol{DN} = (0,0,0,1)'$. Taking the derivative of S_t as defined in equation (1a), we get:

$$\boldsymbol{DS_t} = \boldsymbol{DS_{t-1}}\left(1 - \left(\frac{\beta}{N}\right)I_{t-1}\right) - \left(S_{t-1}\frac{I_{t-1}}{N}\right)\boldsymbol{D\beta} - \left(\frac{\beta}{N}\right)S_{t-1}\boldsymbol{DI_{t-1}} + \left(\frac{\beta}{N^2}S_{t-1}I_{t-1}\right)\boldsymbol{DN} \tag{A4}$$

From equations (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:

$$\boldsymbol{DI_t} = \left(\alpha + \frac{\beta}{N}S_{t-1}\right)\boldsymbol{DI_{t-1}} + I_{t-1}\left(\boldsymbol{D\alpha} + \frac{S_{t-1}}{N}\boldsymbol{D\beta}\right) + \left(\frac{\beta}{N}\right)I_{t-1}\boldsymbol{DS_{t-1}} - \left(\frac{\beta}{N^2}S_{t-1}I_{t-1}\right)\boldsymbol{DN} \tag{A5},$$

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

To complete the calculation, we 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 $\boldsymbol{DI_0} = N\boldsymbol{Di_0} + i_0\boldsymbol{DN}$ and $\boldsymbol{DS_0} = -N\boldsymbol{Di_0} + (1 - i_0)\boldsymbol{DN}$, where $\boldsymbol{Di_0} = (0,0,1,0)'$. Given these initial gradient values, we can use (A4) and (A5) to iteratively compute the vectors $\boldsymbol{DS_t}$ and $\boldsymbol{DI_t}$ for all $t = 1, \ldots, T$.

Once vectors $\boldsymbol{DS_t}$ and $\boldsymbol{DI_t}$ have been computed, we can then apply (A3) to iteratively compute the corresponding gradients $\boldsymbol{DX_t}$. Those quantities in turn yield the gradient $\boldsymbol{DV}$ of our objective function through equation (A2). ■

Proposition 3. Let $\boldsymbol{DX}$ denote the $T \times 4$ matrix whose t-th row is the vector $\boldsymbol{DX_t'}$, as defined in Proposition 2. The Hessian matrix of the least squares criterion V with respect to the parameter vector $\boldsymbol{\Theta} = (\beta, \alpha, i_0, N)$ is:

$$\boldsymbol{D^2V} = 2\,\boldsymbol{DX'DX} - 2\sum_{t=1}^{T}(y_t - X_t)\boldsymbol{D^2X_t} \tag{A6}$$

The basic equations of our dynamic system (1), in combination with the initial conditions (2), can similarly be used to generate complete, computable difference equations for $\boldsymbol{D^2X_t}$ for all t, and thus $\boldsymbol{D^2V}$.

Proof. To compute the Hessian matrix $\boldsymbol{D^2V}$, we need to determine the corresponding Hessian matrices $\boldsymbol{D^2X_t}$ for all t. To that end, we adopt the following notation. For two column vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ of dimension $L \times 1$, we define the $L \times L$ symmetric matrix $\boldsymbol{A} \odot \boldsymbol{B} = \boldsymbol{AB'} + \boldsymbol{BA'}$. Now taking the derivative of $\boldsymbol{DX_t}$ in (A3), we get:

$$\begin{aligned}\boldsymbol{D^2X_t} = {} & \left(\frac{1}{N}\right)(S_{t-1}\boldsymbol{DI_{t-1}} + I_{t-1}\boldsymbol{DS_{t-1}}) \odot (\boldsymbol{D\beta}) + \\ & \left(\frac{\beta}{N}\right)\left((\boldsymbol{DI_{t-1}}) \odot (\boldsymbol{DS_{t-1}}) + S_{t-1}\boldsymbol{D^2I_{t-1}} + I_{t-1}\boldsymbol{D^2S_{t-1}}\right) - \\ & \left(\frac{\beta}{N^2}\right)(S_{t-1}\boldsymbol{DI_{t-1}} + I_{t-1}\boldsymbol{DS_{t-1}}) \odot (\boldsymbol{DN})\end{aligned} \tag{A7}$$

Thus, the Hessian matrices $\boldsymbol{D^2X_t}$ depend in turn on the lagged gradient terms $\boldsymbol{DS_{t-1}}$ and $\boldsymbol{DI_{t-1}}$ as well as the lagged Hessian matrices $\boldsymbol{D^2S_{t-1}}$ and $\boldsymbol{D^2I_{t-1}}$. The corresponding expressions for the latter Hessian matrices are:

$$\boldsymbol{D^2S_t} = \boldsymbol{D^2S_{t-1}} - \boldsymbol{D^2X_t} \tag{A8}$$

$$\boldsymbol{D^2I_t} = \alpha\boldsymbol{D^2I_{t-1}} + \boldsymbol{DI_{t-1}} \odot \boldsymbol{D\alpha} + \boldsymbol{D^2X_t} \tag{A9}$$

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 (A6), (A7), (A8) and (A9), 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 $\boldsymbol{DV}$ in the proof of Proposition 2 above. ■

Proposition 4. Each element $X_t(\boldsymbol{\Theta})$ of the output vector $\boldsymbol{X}(\boldsymbol{\Theta})$ 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 identifiable parameters $\boldsymbol{\theta} = (\beta, \alpha, i_0)$.

Proof. 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 $t = 1, \dots, T$, 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 $I_0 = i_0N$. 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 $S_{t+1} = S_t\left(1 - \frac{\beta}{N}I_t\right) = f_t(\theta)N\left(1 - \frac{\beta}{N}g_t(\theta)N\right) = \big(f_t(\theta) - \beta g_t(\theta)\big)N$, and so $f_{t+1}(\theta) = f_t(\theta) - \beta g_t(\theta)$. Similarly, $I_{t+1} = I_t\left(\alpha + \frac{\beta}{N}S_t\right) = g_t(\theta)N\left(\alpha + \frac{\beta}{N}f_t(\theta)N\right) = g_t(\theta)\big(\alpha + \beta f_t(\theta)\big)N$, and so $g_{t+1}(\theta) = g_t(\theta)\big(\alpha + \beta f_t(\theta)\big)$. ■

Proposition 5. Let $\boldsymbol{\Theta}^{(n)} = \left(\boldsymbol{\theta}^{(n)}, N^{(n)}\right)$ denote parameter estimates at iteration n of an iterative estimation algorithm. Let $\boldsymbol{X^{(n)}}$ denote the corresponding output variable vector derived from the SIR model (2)-(4) based upon these parameter estimates. Define

$$\kappa^{(n+1)} = \boldsymbol{y}'\boldsymbol{X^{(n)}}/\boldsymbol{X^{(n)}}'\boldsymbol{X^{(n)}} = \sum_{t=1}^{T} y_t X_t^{(n)} / \sum_{t=1}^{T}\left(X_t^{(n)}\right)^2 \tag{A10}$$

as the regression coefficient of $\boldsymbol{y}$ on $\boldsymbol{X^{(n)}}$. Then conditional upon $\boldsymbol{\Theta}^{(n)}$, the updated population-size parameter that minimizes the least squares criterion V is $N^{(n+1)} = \kappa^{(n+1)}N^{(n)}$.

Proof. From Proposition 4, we know that each output variable 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 $\boldsymbol{\theta} = (\beta, \alpha, i_0)$. Substituting into equation (5) in the main text gives:

$$V(\boldsymbol{\Theta}) = V(\boldsymbol{\theta}, N) = \sum_{t=1}^{T}(y_t - \varphi_t(\boldsymbol{\theta})N)^2 = \sum_{t=1}^{T}(y_t^2 - 2y_t\varphi_t(\boldsymbol{\theta})N - \varphi_t^2(\boldsymbol{\theta})N^2) \tag{A11}$$

Equation (A11), which holds for all $\boldsymbol{\Theta} = (\boldsymbol{\theta}, N)$, teaches us that the least squares criterion V is a quadratic function of the population-size parameter N.

Adhering to the assumptions of the proposition, let's assume that we have parameter estimates $\boldsymbol{\Theta}^{(n)} = \left(\boldsymbol{\theta}^{(n)}, N^{(n)}\right)$ at iteration n of an iterative procedure. Conditional on $\boldsymbol{\theta}^{(n)}$, the first-order condition for minimizing $V\left(\boldsymbol{\theta}^{(n)}, N\right)$ as a function of N is:

$$\frac{\partial}{\partial N}V\left(\boldsymbol{\theta}^{(n)}, N\right) = \sum_{t=1}^{T}\left(-2y_t\varphi_t\left(\boldsymbol{\theta}^{(n)}\right) + 2\varphi_t^2\left(\boldsymbol{\theta}^{(n)}\right)N\right) = 0 \tag{A12}$$

The unique root is:

$$N^{(n+1)} = \sum_{t=1}^{T} y_t\varphi_t\left(\boldsymbol{\theta}^{(n)}\right) / \sum_{t=1}^{T}\left(\varphi_t\left(\boldsymbol{\theta}^{(n)}\right)\right)^2 \tag{A13}$$

From Proposition 4, we have $X_t^{(n)} = \varphi_t\left(\boldsymbol{\theta}^{(n)}\right)N^{(n)}$. Substituting into (A13) and rearranging gives:

$$N^{(n+1)} = N^{(n)}\sum_{t=1}^{T} y_t X_t^{(n)}(\boldsymbol{\theta}) / \left(X_t^{(n)}(\boldsymbol{\theta})\right)^2 \tag{A14}$$

We have $N^{(n+1)} = \kappa^{(n+1)}N^{(n)}$, where $\kappa^{(n+1)}$ is defined in (A10). ■

Proposition 6. Let $\boldsymbol{\Theta}^* = (\beta^*, \alpha^*, i_0^*, N^*)$ and denote the estimated parameter values that minimize the objective function $V(\boldsymbol{\Theta})$ under the identifying restriction $r_0 = 0$, and let $V^* = V(\boldsymbol{\Theta}^*)$ denote the corresponding minimized value of the least squares criterion V. Then the variance-covariance matrix of the estimated parameters $\boldsymbol{\Theta}^*$ is $\boldsymbol{C}^* = (s^*)^2(\boldsymbol{DX}^{*\prime}\boldsymbol{DX}^*)^{-1}$, where $(s^*)^2 = V^*/T$, and where $\boldsymbol{DX}^*$, as defined in Proposition 3, is also evaluated at the optimum $\boldsymbol{\Theta}^*$.

Proof. This result, which follows directly from the general statistical treatment of nonlinear least squares in [47], gives us the *unconditional* 4×4 variance-covariance matrix of the parameter vector $\boldsymbol{\Theta}^* = (\beta^*, \alpha^*, i_0^*, N^*)$ estimated from our 4-dimensional Newton-Raphson algorithm described in Propositions 2 and 3. If we instead employ the EM-type algorithm characterized in Propositions 4 and 5, then at the M-step of the final stage, we can obtain the 3×3 variance-covariance matrix of the parameter vector $\widehat{\boldsymbol{\theta}} = \left(\hat{\beta}, \hat{\alpha}, \hat{\imath}_0\right)$ *conditional* upon $\widehat{N}$, which has already been estimated at the E-step. In that case, we obtain $\widehat{\boldsymbol{C}} = \hat{s}^2\left(\boldsymbol{D}\widehat{\boldsymbol{X}}'\boldsymbol{D}\widehat{\boldsymbol{X}}\right)^{-1}$,

where $\hat{s}^2 = \hat{V}/T$, where $\hat{V} = V(\widehat{\boldsymbol{\theta}}, \widehat{N})$, where and $\boldsymbol{D}\widehat{\boldsymbol{X}}$ is the $T \times 3$ matrix whose t-th row is the gradient of X_t with respect to $\boldsymbol{\theta}$ evaluated at the optimum $\widehat{\boldsymbol{\Theta}} = (\widehat{\boldsymbol{\theta}}, \widehat{N})$. Since both the 4-dimensional Newton-Raphson algorithm and EM-type algorithm result in parameter estimates that minimize the objective function $V(\boldsymbol{\Theta})$, that is, since $\widehat{\boldsymbol{\Theta}} = \boldsymbol{\Theta}^*$, we can readily compute $\boldsymbol{D}\widehat{\boldsymbol{X}}'\boldsymbol{D}\widehat{\boldsymbol{X}}$ as the 3×3 submatrix of $\boldsymbol{D}\boldsymbol{X}^{*\prime}\boldsymbol{D}\boldsymbol{X}^*$. ■

Proposition 7. Let $\boldsymbol{\Theta}^*$ be a local interior minimum of $V(\boldsymbol{\Theta})$. Define a single-valued mapping from the real line into the four-dimensional subspace Ω of admissible values of $\boldsymbol{\Theta}$ as follows: $\boldsymbol{\lambda}(u) = (\beta^* + u, \alpha^* - wu, i_0^*, N^*)$, where u is its real-valued argument and $w > 0$ is a positive real parameter. As the parameter w varies, this mapping characterizes a family of rays in four-dimensional space passing through the point $\boldsymbol{\Theta}^*$ at $u = 0$. Then there exists a ray, corresponding to a specific value of w, along which the second-order directional derivative $\left.\frac{dV^2}{du^2}\right|_{u=0}$ is nearly zero, that is, along which $V(\boldsymbol{\Theta})$ is *nearly* equal to its local minimum value $\boldsymbol{V}(\boldsymbol{\Theta}^*)$.

Proof. Along any such ray, defined by a specific value of the parameter w, the linear combination $w\beta + \alpha$ is constant and equal to $w\beta^* + \alpha^*$, independent of the value of u. Moreover, for all w, we have $\boldsymbol{\lambda}(0) = \boldsymbol{\Theta}^*$ and the composition $V \circ \boldsymbol{\lambda}(0) = V(\boldsymbol{\lambda}(0)) = \boldsymbol{\Theta}^*$. For any specific value of w, the derivative of $\boldsymbol{\lambda}(u)$ with respect to u is $\boldsymbol{D}\boldsymbol{\lambda} = (1, -w, 0, 0)$, which is the directional vector of the ray. Accordingly, the directional derivative of V along the ray $\boldsymbol{\lambda}(u)$ at the minimum $\boldsymbol{\Theta}^*$ is the inner product $\boldsymbol{D}\boldsymbol{\lambda}\, \boldsymbol{D}\boldsymbol{V}(\boldsymbol{\Theta}^*)$, which simplifies to the expression $\left.\frac{dV}{du}\right|_{u=0} = \frac{\partial V(\boldsymbol{\theta}^*)}{\partial \beta} - w\frac{\partial V(\boldsymbol{\theta}^*)}{\partial \alpha}$. Moreover, the second-order directional derivative is the quadratic form $\boldsymbol{D}\boldsymbol{\lambda}\, \boldsymbol{D}^2\boldsymbol{V}(\boldsymbol{\Theta}^*)\, \boldsymbol{D}\boldsymbol{\lambda}'$, which similarly simplifies to:

$$\left.\frac{dV^2}{du^2}\right|_{u=0} = \frac{\partial^2 V(\boldsymbol{\theta}^*)}{\partial \beta^2} - 2w\frac{\partial^2 V(\boldsymbol{\theta}^*)}{\partial \beta \partial \alpha} + w^2\frac{\partial^2 V(\boldsymbol{\theta}^*)}{\partial \alpha^2} \tag{A15}$$

Our objective function V is the flattest along the ray for which this second-order directional derivative is minimized. From (A15), we see that the second-order directional derivative is itself a quadratic function of the parameter w. This quadratic function in turn reaches a minimum at $w^* = \frac{\partial^2 V(\boldsymbol{\theta}^*)}{\partial \beta \partial \alpha} \Big/ \frac{\partial^2 V(\boldsymbol{\theta}^*)}{\partial \alpha^2}$. Plugging this value into (A15) gives the second-

order directional derivative $\left.\frac{d^2V}{du^2}\right|_{u=0} = \left(\frac{\partial^2 V(\boldsymbol{\theta}^*)}{\partial \alpha^2}\right)^{-1} \nabla$, where ∇ is the determinant of the 2×2 submatrix of $\boldsymbol{D^2V}(\boldsymbol{\Theta}^*)$ corresponding to the parameters (β, α). As we have seen in the empirical implementation, this determinant is positive but quite small. ■

Appendix B. Path of the 4-Parameter Newton-Raphson Algorithm: New York City Data

Figure B1 displays the path of each successive iteration $\boldsymbol{\theta}^{(k)}$ generated by the 4-dimensional Newton-Raphson algorithm, along with the corresponding values $\boldsymbol{V}^{(k)}$ of the criterion function. As in Figures 5 and 6, we plot only the two parameter coordinates $(\beta^{(k)}, \alpha^{(k)})$, but we reverse the axes to improve the perspective. The algorithm began at point *A*, where the $(\beta^{(0)}, \alpha^{(0)}, V^{(0)}) = (0.4, 0.8, 3907)$. On iteration $k = 10$ at point *B*, where $(\beta^{(10)}, \alpha^{(10)}, V^{(10)}) = (0.446, 0.730, 101)$, the path of the algorithm had entered the ravine described in Figure 5 above. The path continued along the ravine until the stopping point at $(\beta^{(40)}, \alpha^{(40)}, V^{(40)}) = (0.233, 0.941, 76.7)$.

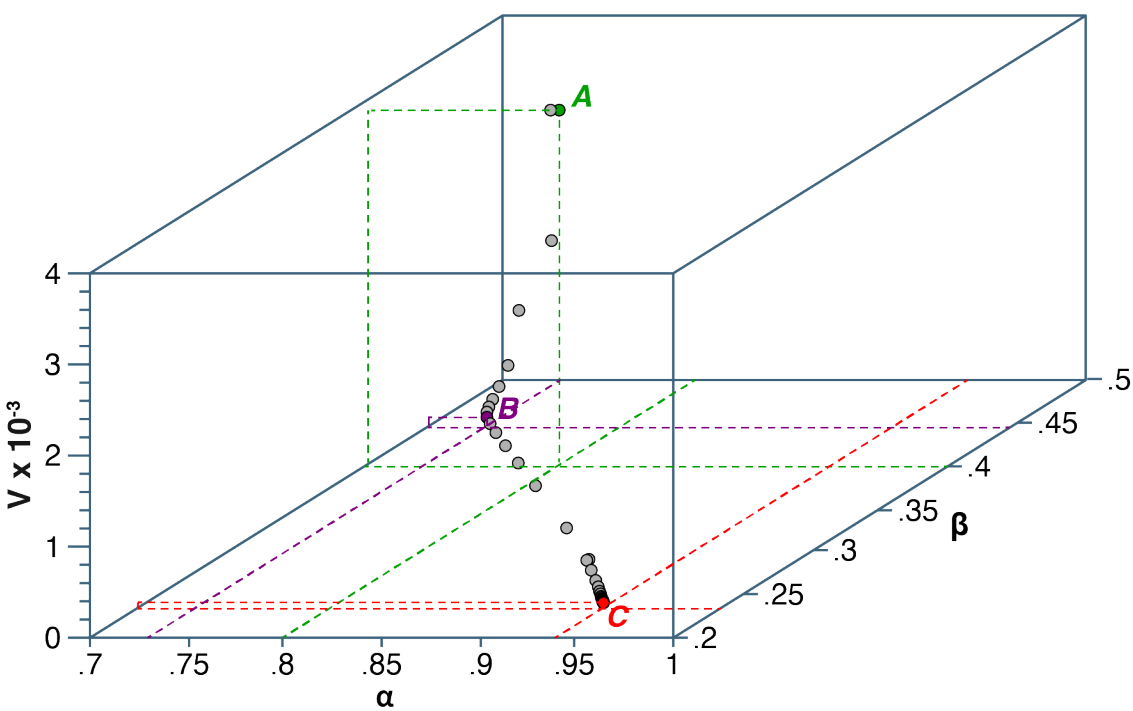

Figure B1. Path of $(\boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{V}^{(k)})$ Through Successive Iterations of the 4-Dimensional Newton-Raphson Algorithm: New York City Omicron Wave. The algorithm began at point *A*, where the $(\beta^{(0)}, \alpha^{(0)}, V^{(0)}) = (0.4, 0.8, 3907)$. The step size for successive iterations, defined in equation (16), was $q = 0.5$. At point *B*, where $(\beta^{(10)}, \alpha^{(10)}, V^{(10)}) = (0.446, 0.730, 101)$, the path of the algorithm had entered the ravine described in Figure 6. The path continued along the ravine until the stopping point at $(\beta^{(40)}, \alpha^{(40)}, V^{(40)}) = (0.233, 0.941, 76.7)$. The stopping criterion was $|V^{(k)} - V^{(k-1)}| < 10^{-4}$. The execution time was 0.436 seconds.

Declarations

Sole Authorship

JEH is the sole author of this original 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. 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 or individual. I gratefully acknowledge the helpful comments of the following individuals on an earlier version of this manuscript: Julio R. Banga (Consejo Superior de Investigaciones Científicas, Spain), Keith Godfrey (University of Warwick, UK), Weiguo Han (University Corporation for Atmospheric Research, USA), Şiran Keske (Koç University School of Medicine, Turkey), Eduardo Massad (University of Sao Paolo, Brazil), Thomas Michelitsch (Sorbonne Université, France), Dimiter Prodanov (IMEC, Belgium), Emerson Sadurni (Benemérita Universidad Autónoma de Puebla, Mexico), Reinhard Schlickeiser (Ruhr-Universität Bochum, Germany), Boris Schmid (University of Oslo, Norway), and James A. Yorke (University of Maryland, USA). The comments of three anonymous referees are likewise appreciated.

Earlier Versions

Earlier versions of this manuscript have been publicly posted on a preprint server [64].

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. Postnikov, E.B., *Estimation of COVID-19 dynamics "on a back-of-envelope": Does the simplest SIR model provide quantitative parameters and predictions?* Chaos, Solitons & Fractals, 2020. **135**(109841).
13. 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.
14. 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.
15. 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.
16. Arino, J., et al., *A final size relation for epidemic models.* Math Biosci Eng, 2007. **4**(2): p. 159-75.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. Schlickeiser, R. and M. Krüger, *Key Epidemic Parameters of the SIRV-Model Determined from Past COVID-19 Mutant Waves*. 2023, https://www.preprints.org/manuscript/202303.0515/v1: Preprints.org, March 30.
23. 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.

24. Nguemdjo, U., et al., *Simulating the progression of the COVID-19 disease in Cameroon using SIR models.* PLoS One, 2020. **15**(8): p. e0237832.
25. 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.
26. 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)
27. Prodanov, D., *Analytical Parameter Estimation of the SIR Epidemic Model. Applications to the COVID-19 Pandemic.* Entropy, 2021. **23**(59): p. 1-20.
28. Capaldi, A., et al., *Parameter Estimation and Uncertainty Quantification for an Epidemic Model.* Mathematical Biosciences and Engineering, 2012. **9**(3): p. 553-579.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.

36. Koepke, A.A., et al., *Predictive Modeling of Cholera Outbreaks in Bangladesh.* Annals of Applied Statistics, 2016. **10**(2): p. 575-593.
37. 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.
38. 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.
39. 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.
40. 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).
41. 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.
42. 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.
43. 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.
44. Gabor, A. and J.R. Banga, *Robust and efficient parameter estimation in dynamic models of biological systems.* BMC Syst Biol, 2015. **9**: p. 74.
45. 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.
46. Verbeke, J. and R. Cools, *The Newton-Raphson method.* International Journal of Mathematical Education in Science and Technology, 1995. **26**(2): p. 177-193.
47. Gallant, A.R., *Nonlinear Regression.* The American Statistician, 1975. **29**(2 (May)): p. 73-81.

48. 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.
49. 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.
50. Anderson, R.M., *Discussion: the Kermack-McKendrick epidemic threshold theorem.* Bull Math Biol, 1991. **53**(1-2): p. 3-32.
51. Brauer, F., *Mathematical epidemiology: Past, present, and future.* Infect Dis Model, 2017. **2**(2): p. 113-127.
52. Brown, K.S. and J.P. Sethna, *Statistical mechanical approaches to models with many poorly known parameters.* Physical Review E, 2003. **68**(021904): p. 1-9.
53. 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.
54. Gutenkunst, R.N., et al., *Universally Sloppy Parameter Sensitivities in Systems Biology Models.* PLOS Computational Biology, 2007. **3**(10 (e189)): p. 1871-1878.
55. Massonis, G., J.R. Banga, and A.F. Villaverde, *Structural identifiability and observability of compartmental models of the COVID-19 pandemic.* Annu Rev Control, 2021. **51**: p. 441-459.
56. Smith, D.A. and L.C. Moore, *Project 2: The SIR Model of the Spread of Disease - Euler's Method for Systems*, in *Calculus: Modeling and Application, 2nd Edition (Interactive Textbook)*. 2010, Mathematical Association of America: http://calculuscourse.maa.org/sample/Chapter5/Projects/SIR%20Model/SIR4.html.
57. 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.
58. Hethcote, H.W., *Qualitative analyses of communicable disease models.* Mathematical Biosciences, 1976. **28**(3-4): p. 335-356.
59. Weiss, H.H., *The SIR model and the Foundations of Public Health.* Materials Matemàtics, 2013. **3**: p. 1-17.

60. 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.
61. 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.
62. Becker, N.G. and T. Britton, *Statistical studies of infectious disease incidence.* J R Statist Soc B, 1999. **61**(2): p. 287-307.
63. Lange, A., *Reconstruction of disease transmission rates: Applications to measles, dengue, and influenza.* J Theor Biol, 2016. **400**: p. 138-53.
64. Harris, J.E., *New Approach to the SIR Inversion Problem: From the 1905-1906 Plague Outbreak in the Isle of Bombay To the 2021-2022 Omicron Surge in New York City*. 2023, https://www.medrxiv.org/content/10.1101/2023.03.13.23287177v2: MedRxiv, version 2 posted May 1.
65. 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.
66. Oehlert, G.W., *A Note on the Delta Method.* The American Statistician, 1992. **46**(1): p. 27–29.
67. Gabor, A., A.F. Villaverde, and J.R. Banga, *Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems.* BMC Syst Biol, 2017. **11**(1): p. 54.
68. StataCorp, *Mata Reference Manual Release 17*. 2021, https://www.stata.com/manuals/m.pdf: College Station, Texas: Stata Press.
69. StataCorp, *Stata Programming Reference Manual Release 17*. 2021, https://www.stata.com/manuals/p.pdf: College Station, Texas: Stata Press.
70. 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.
71. New York City Department of Health and Mental Hygiene, *variant-epi-data*. 2023, https://github.com/nychealth/coronavirus-data/tree/master/variants: Accessed April 21.
72. Cox, N.J., *PCHIPOLATE: Stata module for piecewise cubic Hermite interpolation*. 2012, https://ideas.repec.org/c/boc/bocode/s457561.html: Statistical Software Components S457561, Boston College Department of Economics. Accessed April 21.

73. Mander, A., *SURFACE: Stata module to draw a 3D wireform surface plot*. 2019, https://ideas.repec.org/c/boc/bocode/s448501.html: Boston College Department of Economics, Statistical Software Components S448501, initial version 2005, revised 24 Jun 2019.
74. Harvey, W.T., et al., *SARS-CoV-2 variants, spike mutations and immune escape.* Nat Rev Microbiol, 2021. **19**(7): p. 409-424.
75. 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.
76. Roda, W.C., et al., *Why is it difficult to accurately predict the COVID-19 epidemic?* Infect Dis Model, 2020. **5**: p. 271-281.
77. Moss, P., *The T cell immune response against SARS-CoV-2.* Nat Immunol, 2022. **23**(2): p. 186-193.
78. 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.
79. 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.
80. 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.
81. 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.
82. 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.
83. Brauer, F., *Epidemic models with heterogeneous mixing and treatment.* Bull Math Biol, 2008. **70**(7): p. 1869-85.
84. Kong, L., et al., *Modeling Heterogeneity in Direct Infectious Disease Transmission in a Compartmental Model.* Int J Environ Res Public Health, 2016. **13**(3).

85. David, J.F., *Epidemic models with heterogeneous mixing and indirect transmission.* J Biol Dyn, 2018. **12**(1): p. 375-399.
86. Duan, M. and Z. Jin, *The heterogeneous mixing model of COVID-19 with interventions.* J Theor Biol, 2022. **553**: p. 111258.
87. Harris, J.E., *Concentric regulatory zones failed to halt surging COVID-19: Brooklyn 2020.* Front Public Health, 2022. **10**: p. 970363.
88. 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.
89. 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.
90. 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.
91. Altarawneh, H.N., et al., *Effects of Previous Infection and Vaccination on Symptomatic Omicron Infections.* N Engl J Med, 2022. **387**(1): p. 21-34.
92. Keske, S., et al., *Duration of infectious shedding of SARS-CoV-2 Omicron variant and its relation with symptoms.* Clin Microbiol Infect, 2022.
93. Wu, Y., et al., *Duration of viable virus shedding and polymerase chain reaction positivity of the SARS-CoV-2 Omicron variant in the upper respiratory tract: a systematic review and meta-analysis.* Int J Infect Dis, 2023. **129**: p. 228-235.
94. Boucau, J., et al., *Duration of Shedding of Culturable Virus in SARS-CoV-2 Omicron (BA.1) Infection.* N Engl J Med, 2022. **387**(3): p. 275-277.
95. 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.
96. Hespanha, J.P., et al., *Forecasting COVID-19 cases based on a parameter-varying stochastic SIR model.* Annual Reviews in Control, 2021. **51**: p. 460-476.
97. 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.

98. Bestehorn, M., et al., *Simple model of epidemic dynamics with memory effects.* Physical Review E, 2022. **105**(024025): p. 024025.