Outbreak-size distributions under fluctuating rates

Jason Hindes,[1] Luis Mier-y-Teran-Romero,[2] Ira B. Schwartz,[1] and Michael Assaf[3]
[1]*U.S. Naval Research Laboratory, Washington, DC 20375, USA*
[2]*Leidos, Reston, Virginia 20190, USA*
[3]*Racah Institute of Physics, Hebrew University of Jerusalem, Jerusalem 91904, Israel*

(Received 24 August 2023; accepted 4 November 2023; published 19 December 2023)

We study the effect of noisy infection (contact) and recovery rates on the distribution of outbreak sizes in the stochastic susceptible-infected-recovered model. The rates are modeled as Ornstein-Uhlenbeck processes with finite correlation time and variance, which we illustrate using outbreak data from the RSV 2019-2020 season in the U.S. In the limit of large populations, we find analytical solutions for the outbreak-size distribution in the long-correlated (adiabatic) and short-correlated (white) noise regimes, and demonstrate that the distribution can be highly skewed with significant probabilities for large fluctuations away from mean-field theory. Furthermore, we assess the relative contributions of demographic and reaction-rate noise on the outbreak variance and show that demographic noise becomes irrelevant in the presence of slowly varying reaction-rate noise but persists for large system sizes if the noise is fast. Finally, we show that the crossover to the white-noise regime typically occurs for correlation times that are on the same order as the characteristic recovery time in the model.

DOI: 10.1103/PhysRevResearch.5.043264

I. INTRODUCTION

Epidemic models are useful for understanding the spreading dynamics of general contagious processes and effectively describe a wide variety of phenomena from spreading diseases to rumors, fads, panics, computer viruses, laser systems, and even election dynamics [1–10]. In addition, epidemic models are practically useful, since epidemiologists rely on models to quantify risks of local epidemic outbreaks of various sizes and formulate optimal control strategies for many diseases, including Sars-Cov-2, Ebola, and Dengue [11–16]. Within a given population, outbreak dynamics are typically described in terms of compartmental models [1,4,17]. For example, starting from some small initial infection, over time, individuals in a population make stochastic transitions between some number of discrete disease states (susceptible, exposed, infectious, etc.) based on prescribed probabilities for a particular population and disease [11,14,18–22]. In the limit of large populations and nonfluctuating parameters, the stochastic dynamics are aptly described by deterministic (mean-field) differential equations for the expected fraction of a population in each state [1,4,17,23].

Yet, for real finite populations with time-fluctuating parameters, outbreak dynamics have a wide range of outcomes for each initial condition, which are not predicted by mean-field models [1,19,21,23–27]. For instance, recently we developed a theoretical approach that allows for calculating the distribution of outbreak sizes in well-mixed populations under demographic noise. This approach provided a closed-form expression for the asymptotic outbreak distribution in the susceptible-infected-recovered (SIR) model and more general SIR model extensions with fixed population sizes (N) and static infection and recovery rates [15]. However, many data analyses have shown that, for a multitude of diseases, best-fit epidemic model parameters can fluctuate significantly in time [28–35]. For instance, by measuring the relative changes in reported disease incidence and hospitalization, one can compute an effective infectious contact rate between individuals in a population over time. Doing so, one often finds fluctuating and/or periodic rates, in general [1,33,34,36–40], which in the case of human epidemics correlate with more general social contact rates [41]. For instance, techniques for extracting time-dependent parameters have been applied to the recent COVID-19 pandemic as well [33,35,42,43] to account for fluctuations in contact rates, rendering the usual SIR class of forecasting models time dependent. In addition, here we give an another example based on 2019-2020 hospitalization data of the respiratory syncytial virus (RSV) season in the U.S. [44], and find the data effectively parameterized in terms of two general metrics for quantifying temporal variations about a mean: the infection rate's standard deviation and correlation time.

Despite the theoretical importance of understanding noise effects in canonical nonequilibrium epidemic models, as well as the practical importance for quantifying uncertainty in real epidemics, a general analytical approach for describing small and large fluctuations in outbreak dynamics due to parameter fluctuations is still lacking. Here we develop such an approach within the context of the SIR model with noisy reaction rates with finite variances and correlation times. We motivate our use of these standard noise characteristics by extracting infectious contact rate fluctuations in RSV outbreak data from the

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

U.S. in 2019-2020 using a Bayesian model inference. In terms of general model analysis, we focus on the outbreak-size distribution and quantify the probabilities for outbreaks that differ from mean-field predictions. In particular, we calculate the distribution in the limit of adiabatic and white noises and demonstrate several important properties including the skewness of the outbreak distribution toward unusually small outbreaks and the existence of optimal values of the basic reproductive number that maximize the outbreak variance. We also study the crossover of the outbreak distribution with finite population size and noise-correlation time and analyze when the limiting theories of demographic, adiabatic, and white reaction-rate noise apply.

II. SIR MODEL WITH REACTION-RATE NOISE

We are interested in outbreak dynamics in which the epidemic wave is fast compared to both demographic and reinfection timescales; the latter denotes the possibility for individuals to be infected multiple times [45]. The canonical epidemic model for this regime is the SIR model [1,2], in which individuals are either susceptible (capable of getting infected), infected, or recovered [or removed (deceased)], and can make transitions between these states through two basic processes: infection and recovery. Denoting the total number of susceptibles S, infected I, and recovered R in a population of fixed size N, the probability per unit time that the number of susceptibles decreases by one and the number of infected increases by one is $\beta SI/N$ (for a well-mixed population), where β is the infectious contact rate [1,2,4]. Similarly, the probability per unit time that the number of infected decreases by one is γI, where γ is the recovery rate [1,2,4]. As a result, the deterministic rate equations in the limit of large N describing the mean *densities* of susceptibles, infected, and recovered, $x_s = S/N$, $x_i = I/N$, and $x_r = R/N$, respectively, read

$$\dot{x}_s = -\beta x_s x_i, \quad \dot{x}_i = \beta x_s x_i - \gamma x_i, \quad \dot{x}_r = \gamma x_i, \tag{1}$$

where $x_s + x_i + x_r = 1$. Starting from a small initial infection density, $x_i(t=0) \ll 1$, the final fraction of susceptibles in Eqs. (1) $\overline{x_s^*} \equiv x_0$ satisfies $x_0 = e^{-R_0(1-x_0)}$, where $R_0 = \beta/\gamma$ is the basic reproduction number [1,2]. Hence, in the mean-field theory, the total fraction of the population infected over the whole epidemic wave is $\overline{x_r^*} = 1 - x_0$:

$$\overline{x_r^*} = 1 - e^{-R_0 \overline{x_r^*}}. \tag{2}$$

Note that if $R_0 \leqslant 1$ in Eq. (2) then $\overline{x_r^*} = 0$, giving us the usual condition $R_0 = 1$ as the epidemic threshold.

As noted in Sec. I, in many cases the parameters for the SIR model are time fluctuating. As a simple model, we allow the infection and recovery rates to be generated by independent Ornstein-Uhlenbeck (OU) processes with some correlation times and variances. For simplicity, we assume the correlation times are identical for both rates and equal τ, while the noise variances are σ_β^2 and σ_γ^2 for the infection and recovery rates, respectively. Thus, we write $\beta(t) = \beta_0(1 + \xi_\beta(t))$ and $\gamma(t) = \gamma_0(1 + \xi_\gamma(t))$, and augment Eqs. (1) into the stochastic system:

$$\begin{aligned}
\dot{x}_s &= -\beta_0(1+\xi_\beta)x_s x_i, \\
\dot{x}_i &= \beta_0(1+\xi_\beta)x_s x_i - \gamma_0(1+\xi_\gamma)x_i, \\
\dot{\xi}_\beta &= -\frac{\xi_\beta}{\tau} + \sqrt{\frac{2}{\tau}}\left(\frac{\sigma_\beta}{\beta_0}\right)\eta_\beta(t), \\
\dot{\xi}_\gamma &= -\frac{\xi_\gamma}{\tau} + \sqrt{\frac{2}{\tau}}\left(\frac{\sigma_\gamma}{\gamma_0}\right)\eta_\gamma(t).
\end{aligned} \tag{3}$$

Here, η_β and η_γ are Gaussian white noises, while ξ_β and ξ_γ are OU processes. Note that by construction, $\beta(t)$ and $\gamma(t)$ are assumed to be wide-sense stationary Gaussian processes with $\langle\beta\rangle = \beta_0$, $\langle\gamma\rangle = \gamma_0$, $\langle(\beta(t) - \beta_0)(\beta(t+\Delta) - \beta_0)\rangle = \sigma_\beta^2 e^{-\Delta/\tau}$, and $\langle(\gamma(t) - \gamma_0)(\gamma(t+\Delta) - \gamma_0)\rangle = \sigma_\gamma^2 e^{-\Delta/\tau}$, where $\langle\cdot\rangle$ denotes the expectation operator. In general, one can simulate the system of Eqs. (3) and find the final outbreak-size distribution for fluctuating rates with any magnitude and correlation time.

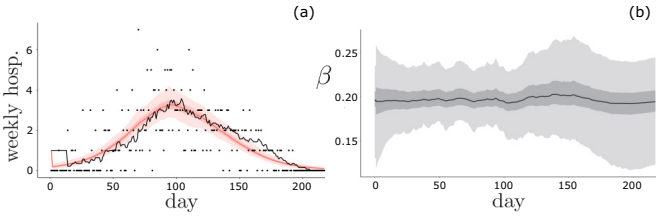

FIG. 1. RSV model inference. (a) Weekly RSV hospitalizations (black dots) and two-week rolling average (black line) from the 2019-2020 season in the U.S. [44]. Results from the Bayesian inference model are overlaid with the data (median: red line, shaded bands represent the interquartile range and the 95% credible intervals). (b) Inferred infectious contact rate obeying a time-discretized version of the OU process (median: black line, shaded bands represent the interquartile range and the 95% credible intervals).

RSV model fit

Finite correlation time and variance are general physical metrics that quantify temporal fluctuations around a mean—the sort of temporal variations observed in many epidemic data analyses [[1,28–40,42,43]]. We can further motivate our study of the SIR model with temporally fluctuating reaction rates by extracting such noise characteristics from data on the 2019-2020 RSV season in the U.S.

We perform a parameter inference from RSV hospitalization data assuming a discretized version of Eqs. (3) with daily time steps and a fixed recovery rate $\gamma = \gamma_0 = 1/7$ days^{-1}. We use the well-known platform `Stan` via the `R` package `rstan` [46,47] to do the numerical Bayesian inference by tying the dynamical model to the number of recorded daily hospitalizations, as obtained from the CDC [44,48]. The parameters for the inference are β_0, the inverse correlation time α, σ_β, the hospitalization rate, and initial conditions for the SIR [49]. Further details are given in Appendix A [50]. Output examples are shown in Fig. 1. In Fig. 1(a), we plot the daily hospitalization numbers and compare to the median prediction of the model (in red) along with its credible intervals.

A similar plot is shown in Fig. 1(b) for the daily infectious contact rate, which drives the predictions for Fig. 1(a).

Our inference uncovers significant temporal fluctuations in the most-likely RSV infectious contact rate. A summary of the output that is relevant for our analysis includes $\hat{R}_0 = 1.37$ in [1.32,1.44], $\hat{\alpha} = 0.11$ days^{-1} in [0.045, 0.20] days^{-1}, and $\hat{\sigma}_\beta = 0.026$ days^{-1} in [0.014, 0.040] days^{-1}, where $\hat{}$ denotes the median within a quartile range specified by the square brackets. From these, we observe a fairly tight value of the inferred time-averaged $R_0 = \beta_0/\gamma_0$, but with substantial temporal fluctuations between $10-20\%$. On the other hand, the correlation time estimate α^{-1} is quite broad, ranging from $5-20$ days. Note, however, that these noise correlation time estimates are significantly smaller than the expected timescale of periodic (seasonal) terms in $\beta(t)$, which are necessary for multiyear predictions [51]. The natural timescales for the latter are on the order of a year. Hence, the noise inference quantifies temporal fluctuations distinct from seasonality.

We can situate the inferred noise characteristics of the RSV season within the results of our analytical theory; see Sec. V. First, we begin by analyzing outbreak statistics driven by the fluctuations in Eqs. (3).

III. LIMIT OF ADIABATIC NOISE

To gain insight into the outbreak distributions generated from the general Eqs. (3) and temporal fluctuations of the sort we inferred from RSV data, we first consider limiting regimes. We start with the limit of adiabatic noise, $\tau \gg 1$. Here, the underlying assumption is that, during the epidemic wave, the rates do not change appreciably. In terms of dynamics, the susceptible, infected, and recovered populations evolve in time according to the mean-field system, Eqs. (1). However, the parameters β and γ are chosen randomly according to some quenched distributions, i.e., Gaussians in our chosen model, Eqs. (3). For simplicity and illustration of the adiabatic limit, here we deal with the case where only β varies and γ is constant, such that $\sigma_\gamma = 0$. To simplify the equations, we take $\gamma_0 = 1$, which merely specifies the time units and results in $R_0 = \beta$.

To find the distribution of the final outbreak size $P(x_r^*)$, we have to compute the following integral [52,53]:

$$P(x_r^*) = \int_{-\infty}^{\infty} P(x_r^*|\beta) P(\beta) d\beta. \tag{4}$$

The conditional probability $P(x_r^*|\beta)$ is a Dirac delta function around the mean-field value of the outbreak at β, namely, $P(x_r^*|\beta) = \delta(x_r^* - \overline{x_r^*})$, where the mean-field final outbreak fraction $\overline{x_r^*}$ satisfies Eq. (2). Taking a Gaussian distribution for $P(\beta)$ with mean β_0 and standard deviation σ_β, Eq. (4) becomes

$$P(x_r^*) = \frac{1}{\sqrt{2\pi\sigma_\beta^2}} \int_1^{\infty} \delta(x_r^* - \overline{x_r^*}(\beta)) e^{-\frac{(\beta-\beta_0)^2}{2\sigma_\beta^2}} d\beta. \tag{5}$$

We point out that for the SIR model to make physical sense, $\beta \geqslant 0$. Therefore, when plugging in an unrestricted Gaussian in Eq. (5), σ_β cannot be too large [54]. Otherwise, other distributions, e.g., that vanish at $\beta = 0$ can be used instead; yet this does not change the results qualitatively. We also note

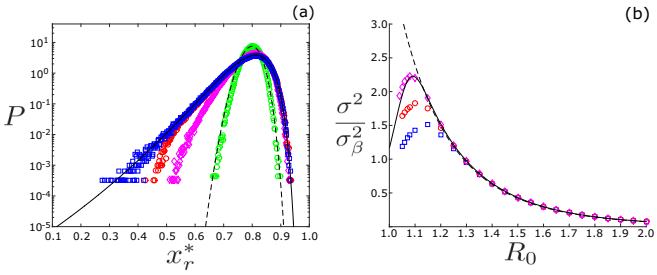

FIG. 2. Outbreak statistics for adiabatic noise. (a) Simulated PDFs of the final outbreak fraction x_r^* from Eqs. (3) in the case of time-correlated reaction-rate noise for $\tau = 10^0$–10^3 (from narrowest to widest) with $\sigma_\beta = 0.1\beta_0$ and $\beta_0 = 2$. The solid and dashed lines denote the adiabatic and white-noise predictions, respectively. (b) Variance of the outbreak PDF (normalized by infection noise) versus $R_0 = \beta_0/\gamma_0$ for $\tau = 10^3$-10^5 (bottom to top) and $\sigma_\beta = 0.04\beta_0$. The predicted variance from the adiabatic distribution, Eq. (6), is shown in solid black, while the small-noise limit, Eq. (7), is shown with a dashed line. For both panels, $\gamma_0 = 1$.

that since $\overline{x_r^*}(\beta)$ vanishes for $\beta \leqslant 1$, the lower boundary in the integral in Eq. (5) can be taken to be 1 without affecting the distribution.

Changing variables from β to x_r^*, and using the fact that $d\beta/dx_r^* = [x_r^* + (1-x_r^*)\ln(1-x_r^*)]/[(1-x_r^*)(x_r^*)^2]$, we can explicitly perform the integration by plugging instead of β, $-\ln(1-x_r^*)/x_r^*$, which is the solution of $x_r^* = \overline{x_r^*}(\beta)$. As a result, Eq. (5) reduces to

$$P(x_r^*) = \frac{d\beta}{dx_r^*} \exp\left[-\frac{\left((1/x_r^*)\ln(1-x_r^*) + \beta_0\right)^2}{2\sigma_\beta^2} \right]. \tag{6}$$

From Eq. (6), we can derive, e.g., the typical fluctuations around the mean-field given by σ_a: the standard deviation associated with the adiabatic outbreak distribution. In particular, in the limit of small σ_β, σ_a becomes

$$\sigma_a = \sigma_\beta \left. \frac{dx_r^*(\beta)}{d\beta} \right|_{\beta=\beta_0} = \frac{\sigma_\beta x_0(1-x_0)}{1-R_0 x_0}. \tag{7}$$

Note that for the adiabatic noise limit, we can repeat our calculation of the outbreak-size distribution and variance for any distribution of β (or γ), and not just the Gaussians implied by Eqs. (3). For instance, if $\beta(t)$ varies slowly compared to the epidemic dynamics, an empirical time series of $\beta(t)$ or $R_0(t)$ can be used to build a histogram, which is substituted into Eq. (4).

Next, we can plot the probability distribution function (PDF) for adiabatic infection-rate noise and explore its qualitative features. An example prediction is shown in Fig. 2(a), with a solid line for fixed values of $\beta_0 = 2$ and $\sigma_\beta = 0.1\beta_0$. The solution from Eq. (6) can be compared to stochastic simulations of Eqs. (3) for large τ. Note that the agreement with simulations is quite good. Qualitatively, one of the most important features that we observe in the PDFs is the high degree of skewness toward small outbreaks. We can get a quantitative measure of this skewness by examining the exponent of Eq. (6), called the *action* (for reasons explained in Sec. IV), for two limiting values of the outbreak fraction: $x_r^* = 0$ and $x_r^* = 1$, i.e., small and large outbreaks. Indeed, the

PDF [Eq. (6)] can be described effectively as $P \sim \exp[-S/\sigma_\beta^2]$, where $S = (\ln(1-x_r^*)/x_r^* + R_0)^2/2$. Importantly, when $x_r^* \to 0$ the action remains finite, i.e., $S \to (R_0-1)^2/2$. On the other hand, when $x_r^* \to 1$, $S \to \infty$. Hence, minimally small outbreaks occur with finite probability for finite R_0, while maximally large outbreaks can never occur when the reaction-rate noise is finite, which is why the outbreak PDF's tails are skewed toward small outbreaks.

In addition to the PDFs, we can examine the variance of the outbreak PDF for adiabatic noise as a function of R_0. Examples are shown in Fig. 2(b), where we plot simulated outbreak variances for three large values of τ with $\gamma = 1$. Here another interesting qualitative feature emerges: the existence of a maximum in the outbreak variance for some value of R_0. On the one hand, as $\sigma_\beta \to 0$, the maximum approaches $R_0 = 1$. On the other hand, as σ_β increases, the maximum variance occurs for an R_0 that is an increasing function of σ_β. For example, in Fig. 2(b) we observe a maximum near $R_0 = 1.1$. Notably, the saddle-point equation for the maximum variance in the adiabatic limit cannot be solved analytically.

In general, we observe good agreement with the predicted variance of Eq. (6) (solid line) and the small-noise limit Eq. (7) (dashed line), including the existence of a maximum, which the former captures. Yet, as $R_0 \to 1$, eventually all the simulation results have discrepancy with both adiabatic predictions. The reason is, as we approach the epidemic threshold, the SIR dynamics slow down, meaning that even a large τ may not be "slow" with respect to the underlying process.

IV. LIMIT OF WHITE NOISE

So far, we have assumed that the dynamics of the noise is slow compared to the epidemic dynamics of Eqs. (1), but what happens if it is fast? In this latter limit, $\tau \to 0$, instead of Eqs. (3) we can write

$$\begin{aligned} \dot{x}_s &= -\beta_0(1+\sigma_1\zeta_\beta(t))x_s x_i, \\ \dot{x}_i &= \beta_0(1+\sigma_1\zeta_\beta(t))x_s x_i - \gamma_0(1+\sigma_2\zeta_\gamma(t))x_i, \end{aligned} \tag{8}$$

which we denote as the white noise reaction-rate limit. Here, ζ_β and ζ_γ are white Gaussian noises. To coincide with Eqs. (3) as $\tau \to 0$, one must demand that

$$\sigma_1 = \sqrt{2\tau}\sigma_\beta/\beta_0, \quad \sigma_2 = \sqrt{2\tau}\sigma_\gamma/\gamma_0. \tag{9}$$

To study the outbreak-size PDF given Eqs. (8), we follow the approach detailed in Ref. [15], and construct the equivalent Fokker-Planck equation for the probability to observe densities x_s and x_i at time t (under Itô calculus):

$$\begin{aligned} \frac{\partial P}{\partial t} = &-\frac{\partial}{\partial x_s}\Big[-\beta_0 x_s x_i P\Big] - \frac{\partial}{\partial x_i}\Big[(\beta_0 x_s x_i - \gamma_0 x_i)P\Big] \\ &+\Big(\frac{\partial^2}{\partial x_s^2} + \frac{\partial^2}{\partial x_i^2} - 2\frac{\partial}{\partial x_s}\frac{\partial}{\partial x_i}\Big)\Big[\frac{1}{2}\beta_0^2 x_s^2 x_i^2 \sigma_1^2 P\Big] \\ &+\frac{\partial^2}{\partial x_i^2}\Big[\frac{1}{2}\gamma_0 x_i^2 \sigma_2^2 P\Big]. \end{aligned} \tag{10}$$

To simplify notation, henceforth, we will assume that $\sigma_2^2 = f\sigma_1^2$, with $f > 0$, and again rescale time $t \to \gamma_0 t$, so β_0 is replaced by the basic reproduction number, $R_0 = \beta_0/\gamma_0$. Next, we employ the WKB approximation $P(x_s, x_i) \sim \exp[-S(x_s, x_i)/\sigma_1^2]$, which is the expected scaling-form for solutions to Eq. (10) in the limit of small noise and large deviations [55–59], and which we observe in simulations of Eqs. (3). Figure 3(a) shows several examples of the expected scaling with noise variance for different values of the final outbreak size. Indeed, the logarithm of the probability tends to straight lines as $1/\sigma_1^2$ is varied, with slopes that change with the outbreak size. Using this insight, we substitute the exponential ansatz into Eq. (10) and arrive, in the leading order in $\sigma_1 \ll 1$, at a Hamilton-Jacobi equation, $\partial S/\partial t + H = 0$, with

$$H = x_i\big[p_i(R_0 x_s - 1) - R_0 x_s p_s\big] + \tfrac{1}{2}x_i^2\big[R_0^2 x_s^2 (p_s - p_i)^2 + f p_i^2\big]. \tag{11}$$

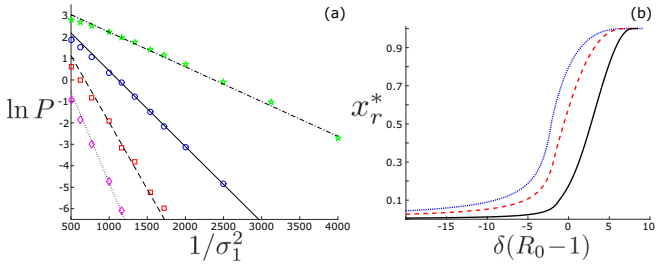

FIG. 3. Scaling of the outbreak-size distribution. (a) The natural log of simulated PDF values for $x_r^* = 0.785$ (green stars), 0.780 (blue circles), 0.775 (red squares), and 0.770 (magenta diamonds) from Eqs. (3) versus the reaction-rate noise variance. The slopes are predictions from the white-noise theory. Other model parameters are $\beta_0 = 2$, $\gamma_0 = 1$, $\tau = 0.1$, and $f = 0$. (b) The final outbreak fraction plotted as a function of the infected momentum initial condition multiplied by $R_0 - 1$, for $R_0 = 1.1$ (solid black), $R_0 = 1.5$ (dashed red), and $R_0 = 2$ (dotted blue). Other model parameters are $\gamma_0 = 1$ and $f = 0.65$.

In this formalism, H is called the Hamiltonian, S is the action, while $p_s = \partial_{x_s} S$ and $p_i = \partial_{x_i} S$ are the conjugate *momenta*, just as in analytical mechanics [55,57,58].

To compute probabilities for different outbreak sizes, we need to find the action S, given by $S = \int p_s dx_s + \int p_i dx_i - \int H dt$, which can be calculated given the Hamilton's equations: $\dot{x}_s = \partial_{p_s} H$, $\dot{x}_i = \partial_{p_i} H$, $\dot{p}_s = -\partial_{x_s} H$, and $\dot{p}_i = -\partial_{x_i} H$. We can simplify the action computation by noting that, first, since we are interested in outbreaks that emerge from initially small levels of infection $x_i(t=0) \simeq 0$, the "energy" of outbreaks is zero, $H(x_i = 0) = 0$. As Hamiltonian (11) has no explicit time dependence, it is a constant of motion, namely, $H(t) = 0$. Second, we can rewrite the Hamiltonian as $H = p_s\dot{x}_s + p_i\dot{x}_i - (1/2)R_0^2 x_s^2 x_i^2 [p_s - p_i]^2 - (1/2) f x_i^2 p_i^2$, using $\dot{x}_s$ and $\dot{x}_i$. Third, by substituting the zero-energy condition into $\dot{p}_i$, we find that $\dot{p}_i = -(1/2)R_0^2 x_s^2 x_i [p_s - p_i]^2 - (1/2) f x_i p_i^2$. Thus, Eq. (11) simplifies to $H(t) = p_s\dot{x}_s + (d/dt)(x_i p_i)$. Integrating both sides of this equation with respect to time over the full course of an outbreak yields $0 = \int p_s dx_s + x_i(t \to \infty)p_i(t \to \infty) - x_i(t=0)p_i(t=0)$. As the fraction of the population infected goes to zero both at short and long times (assuming no reinfection), we derive the useful fact that $\int p_s dx_s = 0$. As a consequence, the action associated with an outbreak in the white-noise limit is simply an integral over the

infected momentum:

$$S = \int p_i dx_i. \tag{12}$$

A. Phase-space trajectories for outbreaks

To compute the outbreak-size distribution, we need to solve Hamilton's equations and substitute the resulting trajectories into Eq. (12). To do so, we must understand the phase-space structure of outbreak paths. First, we recall that in the mean-field system, Eqs. (1), the outbreak dynamics follow a heteroclinic trajectory, which starts at $t=0$ at a fixed point $(x_s=1, x_i=0)$ and ends at the final state $(x_s=x_0, x_i=0)$ as $t\to\infty$. In our Hamiltonian system, this corresponds to a special trajectory with $p_s=p_i=0$, with initial condition $(x_s, x_i, p_s, p_i)=(1, 0, 0, 0)$ for $t=0$. However, in general, there are an *infinite* number of related $x_i=0$ initial conditions with nonzero momenta, which one can find by solving $\dot{x}_s=0$, $\dot{x}_i=0$, $\dot{p}_s=0$, and $\dot{p}_i=0$, given $x_s=1$ and $x_i=0$. It is straightforward to show that the general fixed-point initial conditions are

$$(x_s, x_i, p_s, p_i)_{t=0} \;=\; (1, 0, \delta(R_0-1)/R_0, \delta), \tag{13}$$

where $\delta \equiv p_i(t=0)$ is a free parameter.

As pointed out in Ref. [15] for the case of demographic noise, if we propagate each of the possible initial conditions forward in time, they tend to unique final outbreak values; namely, one x_s^* (x_r^*) for each δ. Examples are shown in Fig. 3(b), where we plot the outbreak fraction as a function of δ for three different values of R_0. A simple algorithm for generating the outbreak distribution numerically for a fixed value of R_0 is to (1) pick a δ, (2) propagate forward with Hamilton's equation given Eq. (11) (assuming some small perturbation from the chosen fixed point), (3) compute the integral in Eq. (12) from the resulting trajectory, and (4) repeat for another value of δ. Each δ results in a unique x_s^* and $S(x_s^*)$.

The slopes of the lines in Fig. 3(a), were computed in just this way, and correspond to numerical solutions for the outbreak paths and associated $S(x_s^*)$ for the chosen values of $x_r^*=1-x_s^*$. Similarly, by sweeping over values of δ we can compute the full white-noise PDF for any x_r^*. Examples are plotted in Fig. 2(a) (dashed line) for infection-rate fluctuations ($f=0$) and in Fig. 4(a) for different combinations of infection and recovery noise ($f\neq 0$). For all Figs. 2(a), 3, and 4(a), the white-noise WKB theory and simulations agree well, which demonstrates the accuracy of our general approach. In fact, by combining the method presented with the results of Ref. [15], we have a complete algorithmic solution for generating the outbreak PDF of the SIR model with general and multicomponent white noise, which we return to in Sec. V.

B. Outbreak variance

In the general case of noise in both β and γ, it seems that Hamilton's equations cannot be solved analytically in a simple manner—apart from constructing a power-series expansion in the momentum initial-condition parameter δ. The primary reason for this, in contrast to Ref. [15], is that for the reaction-rate noise discussed in this paper, there is *no conservation of momentum*. Therefore, we proceed to first calculate the

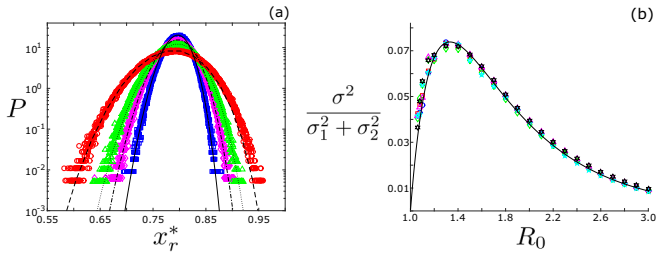

FIG. 4. Outbreak statistics for white noise. (a) The simulated white-noise PDFs from Eqs. (8) for $\sigma_1=0.1$ and $f=0.1$, 1, 2, and 5 (from narrowest to widest) with $\beta_0=2$. White-noise predictions for each combination are shown with curves overlaying the simulation results. (b) Variance of the outbreak PDF (normalized by noise variance) for white reaction-rate noise versus $R_0=\beta_0/\gamma_0$. The noise combinations are blue circles ($\sigma_1=0.1$, $f=0.1$), red squares ($\sigma_1=0.1$, $f=1$), green diamonds ($\sigma_1=0.1$, $f=5$), magenta triangles ($\sigma_1=0.2$, $f=0.1$), and black hexagrams ($\sigma_1=0.04$, $f=1$). The black curve shows the white-noise predictions. In both panels, $\gamma_0=1$.

variance of the outbreak-size distribution, which is related to the lowest order contribution to Eq. (12) in δ. A complete solution for the case of recovery-only fluctuations (not just lowest order in delta) is given in Appendix B.

For the variance calculation, we attempt to find the action in the vicinity of the mean-field final outbreak fraction $1-x_0$. First, let us assume $p_i \ll 1$ to be verified *a posteriori*. Equating $H=0$ yields $p_s=(1-1/(R_0x_s))p_i$, i.e., $p_s \ll 1$ as well. Second, we show that $p_i(t)$ remains small during the entire epidemic duration as long as the initial momentum δ is small. Writing the Hamilton's equation for $\dot{p}_i=-\partial_{x_i}H$, and using Eq. (11) we have $\dot{p}_i=-(1/2)(1+f)x_i(t)p_i^2$. The solution of this differential equation is

$$p_i(t) \simeq \delta\{1+[(1+f)/(2R_0)]\ln(x_s)\delta\}, \tag{14}$$

where $p_i(0)=\delta$ is the initial condition, and we have used the fact that in the leading order in $\delta \ll 1$, $x_r=\int x_i dt = \ln(x_s)/R_0$. This is legitimate as the action will have a δ^2 dependence (see below) so we can substitute in $O(\delta^2)$ terms their mean-field $O(\delta^0)$ approximation.

To compute δ, we can use Hamilton's equations for $\dot{x}_s$, and $\dot{x}_r=-\dot{x}_s-\dot{x}_i$, and compute $\dot{x}_r/\dot{x}_s=dx_r/dx_s$. This yields a differential equation, which can be solved with initial conditions $x_r(t=0)=0$ and $x_s(t=0)=1$, assuming that during the epidemic duration, $p_i(t)$ is almost constant within $O(\delta)$. Using Eq. (14) and that when the outbreak ends $x_r^*=1-x_s^*$, and assuming $x_s^*-x_0 \sim O(\delta)$ (to be confirmed *a posteriori*), we find

$$\delta \simeq \frac{2(1-R_0x_0)}{(1+f)(1-x_0)x_0(2-R_0(1+x_0))}(x_s^*-x_0). \tag{15}$$

This confirms *a posteriori* that $\delta \ll 1$, under the assumption that the final susceptible fraction is close to its mean-field counterpart, i.e., $x_s^*-x_0 \ll 1$.

Finally, to compute the integral in Eq. (12), it is more convenient to change variables to x_s, see Eq. (14). Thus, we write $\int p_i dx_i = \int p_i(dx_i/dx_s)dx_s$. Here, the Jacobian can be found using the Hamilton's equations: $dx_i/dx_s=1/(R_0x_s)-1-[(1+f)/(R_0x_s)(1-x_s+\ln x_s/R_0)]\delta$, where again we have used mean-field results for the $O(\delta)$ terms, namely,

$x_i = 1 - x_s + \ln(x_s)/R_0$. Putting it all together, and using Eqs. (14) and (15), we can perform the integration in Eq. (12) over x_s from 1 to x_s^*, which yields the action, in the leading order in $x_s^* - x_0 \sim \delta$:

$$S = \frac{(x_s^* - x_0)^2}{2v^2} + O\big((x_s^* - x_0)^3\big),$$
$$v^2 = \frac{(1+f)(x_0-1)R_0x_0^2[2 - R_0(1+x_0)]}{2(1-R_0x_0)^2}. \quad (16)$$

Indeed, having obtained a δ^2 dependence of the action corroborates our assumptions *a posteriori*. Here, $v = v(R_0)$ is the (rescaled) variance of the outbreak-size distribution; namely, remembering that we have sought the outbreak-size PDF as $P(x_s^*) \sim \exp[-S(x_s^*)/\sigma_1^2]$, its variance in the limit of white reaction-rate noise, σ_w^2, is

$$\sigma_w^2 = \sigma_1^2 v^2. \quad (17)$$

We can test our predictions for the outbreak variance in the white-noise regime by performing stochastic simulations of Eqs. (8) with different values of R_0 and different combinations of noise. Results are shown in Fig. 4(b). First, one can see an interesting behavior where the variance receives a maximum at $R_0 \simeq 1.33$, similar to the adiabatic regime shown in Fig. 2(b). Here, however, the maximum variance occurs for an R_0 that is independent of the noise amplitude and noise combination, unlike adiabatic noise. The reason for the maximum appearing around 1.33 is that for this value of R_0 the mean outbreak fraction is approximately obtained at $x_r^* \simeq 0.5$, which maximizes the variance possibility. In addition, we note that the predicted outbreak variance in the white-noise regime only depends on R_0 and the total variance of the reaction-rate noise, $\sigma_1^2 + \sigma_2^2$. For example, in Fig. 4(b) we show the predicted scaling collapse to a single function of R_0 of the simulated outbreak variance resulting from different combinations of noise (i.e., different values of f). In general, we would expect infection-rate and recovery-rate noise to produce additive variance (since the two noise sources are independent), but the fact that their prefactor dependence on R_0 is identical is interesting. On the other hand, one can check that this symmetry between infection and recovery noise disappears for higher-order statistics, e.g., by repeating the above calculation to $O(\delta^3)$ for the third central moment.

C. Dynamics for small fluctuations

In addition to the outbreak statistics, one can evaluate how the populations evolve in time during the course of an epidemic. In the case of adiabatic noise, the populations follow the mean-field dynamics with constant parameters, drawn for each outbreak from a Gaussian (or any other) distribution. In the case of white noise, the picture is more involved. Yet, as in Sec. IV B, one can gain insight by looking at small-fluctuation corrections to the mean field due to noise, which correspond to small δ. At lowest order in δ, we have shown that $p_i \approx \delta$ and $p_s \approx \delta(1 - 1/(R_0x_s))$. Substituting these momenta into the Hamilton's equations for x_s and x_i, we arrive at

$$\dot{x}_s \approx -R_0x_sx_i(1+\delta x_i),$$
$$\dot{x}_i \approx R_0x_sx_i(1+\delta x_i) - x_i(1-\delta f x_i). \quad (18)$$

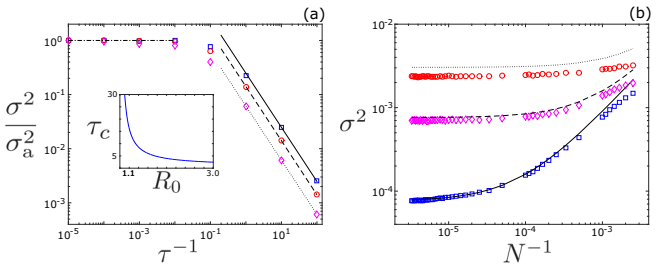

FIG. 5. Variance of the final outbreak size versus the inverse of the correlation time (left) and N (right). (a) The variance (rescaled by the predicted adiabatic-limit) as function of τ^{-1} for $\sigma_\beta = 0.04\beta_0$ and $\beta_0 = 2.0$, 1.5, and 1.2 (top to bottom). The lines are the white-noise predictions. The inset shows the crossover time, Eq. (19) versus R_0. (b) The variance versus N^{-1} for $\tau = 10$, 1 and 0.1 (top to bottom) with $\beta_0 = 2.0$ and $\sigma_\beta = 0.1\beta_0$. The lines are the white-noise predictions, which are the sum of the variances from reaction-rate and demographic noise. For both panels, $\gamma_0 = 1$.

Examining Eqs. (18), we see that for small fluctuations, the noise effectively changes the mean-field model parameters with state-dependent boosts, which are linear in x_i at lowest order in δ; namely, noise produces updated parameters $\tilde{R}_0 = R_0(1+\delta x_i)$ and $\tilde{\gamma} = 1 - \delta f x_i$ where $\tilde{R}_0$ and $\tilde{\gamma}$ are the effective reproduction number and recovery rate that must go into the mean-field system [Eqs. (1)] to account for noise. Recalling that $\delta = 0$ gives the mean-field dynamics, we see that for outbreaks larger than the mean-field ($\delta > 0$), the dynamics behave as if the reproduction number grows with the number of infected and the recovery rate decreases with the number of infected. The opposite is the case for outbreaks less than the mean field ($\delta < 0$).

V. CROSSOVER WITH CORRELATION TIME AND SYSTEM SIZE

Now that we have analyzed the outbreak distribution and dynamics in limiting cases (including Appendix B), we next address when the various limiting regimes apply. In particular, we examine the crossover behavior of the stochastic SIR model as a function of the reaction-rate noise correlation time and population size; the latter has been assumed infinite so far. We use as our metric the variance of the outbreak-size distribution since it is the lowest order statistic not captured by mean-field theory.

First, we remain in the $N \to \infty$ limit and try to understand how small (large) τ has to be to produce effectively white (adiabatic) outbreak statistics. To do so, we plot in Fig. 5(a) the variance of the outbreak-size PDF found from simulating Eqs. (3) versus the (inverse) correlation time τ for three values of R_0 and fixed σ_β. Note that the outbreak variance is normalized by the adiabatic limit, Eq. (7), so each simulation series approaches unity for small τ^{-1}. In addition to the adiabatic limit, for comparison we plot predictions for white noise, Eq. (16), with lines. In the latter case, the τ dependence comes from the definition of the white-noise variance, Eq. (9).

Figure 5(a) has several important features. First, we point out that the outbreak variance has a maximum in the adiabatic limit, meaning that for fixed infection-rate noise variance,

the SIR model dynamics is most sensitive to slow noise. This effect is observed in other population models as well [53,60,61]. For the SIR model, the primary reason is that even relatively small fluctuations in β can bring an epidemic closer to the $R_0=1$ threshold. If the noise is slowly varying, in particular, the effect is felt over the full time-course of the epidemic wave, resulting in a potentially much smaller outbreak than the mean field. As mentioned in Sec. III, this produces highly skewed PDFs with significant probabilities for small outbreaks, and hence large outbreak variance. In contrast, in Fig. 5(a) we can see that the white-noise predictions (the lines) are accurate for quite large values of τ. In fact, for each value of $R_0=2,\ 1.5$, and 1.2 (from top to bottom), we can see that the white noise prediction remains valid for correlation times on the order of the recovery time, $\tau\sim\gamma_0^{-1}=1$.

In general, the crossover point in τ between white and adiabatic-noise regimes has some R_0 dependence: the smaller R_0, the larger τ can be for the white-noise results to be valid, since effectively as the epidemic gets closer to threshold the dynamics slows down, making even slowly-varying noise potentially fast. An estimate for the crossover time, τ_c, can be found by solving $\sigma_a^2=\sigma_w^2$, which yields

$$\tau_c=\frac{R_0^2}{2v^2}\frac{\sigma_a^2}{\sigma_\beta^2}=\frac{R_0(x_0-1)}{(1+f)(2-R_0(1+x_0))}, \tag{19}$$

where we have used Eqs. (7) and (16), valid for small noise. Evidently, τ_c depends only on R_0 and not, e.g., on the noise variance for small noise. The crossover time is plotted in the inset of Fig. 5(a), which for the typical model parameters of $R_0-1\sim O(1)$ remains near the recovery timescale (or unity in our chosen units).

Now that we have an estimate for crossover times, we can situate the inferred RSV contact-rate fluctuations and determine what regime they fall into. By plugging in the median and quartile inferred parameter values given in Sec. II A into Eq. (19), and using Eqs. (7) and (16), we find that the ratio of the noise correlation time to the crossover time, τ/τ_c, falls between $0.1-0.2$. As the τ estimates are substantially smaller than the crossover times, we expect the outbreak-size statistics to be well approximated by the white-noise theory. Hence, our analytical results can be used to make quantitative estimates for future RSV outbreak size probabilities, assuming parameters remain relatively similar to the 2019-2020 epidemic.

The second crossover that we consider is that of finite system size, namely: how large does a population have to be before demographic noise becomes irrelevant compared to reaction-rate noise? For this exploration, we perform a discrete time stochastic simulation (with small time steps) of the discrete state reactions defined for the SIR model in Sec. II [above Eqs. (1)], while the reaction rates fluctuate according to the OU processes in Eqs. (3). In Fig. 5(b), we plot the outbreak-size variance as a function of N^{-1} for three values of the correlation time $\tau=10,\ 1$, and 0.1 (from top to bottom). The curves are the expected total white-noise variance, $\sigma_{w,\mathrm{tot}}^2$, which is a sum of reaction-rate and demographic noise

$$\sigma_{w,\mathrm{tot}}^2=\frac{(\sigma_1^2+\sigma_2^2)(x_0-1)R_0x_0^2[2-R_0(1+x_0)]}{2(1-R_0x_0)^2}+\frac{x_0(1-x_0)(R_0^2x_0+1)}{N(R_0x_0-1)^2}. \tag{20}$$

The demographic term [second line of Eq. (20)] was calculated in Ref. [15]. Note that here we have assumed that the total variance is the sum of the variances from the independent noise sources [62] and that all of the noise amplitudes are relatively small: $\sigma_1^2,\sigma_2^2,N^{-1}\ll 1$.

In Fig. 5(b), we can see that for large system sizes the variance becomes flat with respect to changes in N and approaches approximately the expected white-noise limit, Eq. (16)—especially for the two smaller values of τ where the white-noise approximation is more appropriate. On the other hand, the crossover can occur for quite large system sizes, e.g., $N\sim 10^5$ for $\tau=0.1$ and $N\sim 10^4$ for $\tau=1$, meaning that demographic noise tends to persist if the reaction-rate noise is fast, but disappears quickly with N if the noise is slow; notice that the top series with $\tau=10$ has almost no N dependence.

VI. CONCLUSIONS

Temporal fluctuations in the parameters that control contagion dynamics are inevitable, and have been shown in many epidemic data analyses. Motivated by this, we analyzed the effects of fluctuating infection and recovery rates on the outbreak-size distribution in the canonical SIR model. The SIR reaction rates were modeled with OU noise, allowing us to extract the outbreak statistics as a function of the noise standard deviations and correlation times. Our simple choice was demonstrated by performing a model inference of the 2019-2020 RSV season in the U.S., where we observed significant temporal fluctuations in infectious contact rates.

In terms of analytical results, we found solutions for the outbreak-size distribution in the adiabatic and white noise regimes and showed that the distributions can be highly skewed with significant probabilities for large fluctuations away from mean-field predictions. Interestingly, we discovered that the outbreak variance is generally maximized for a value of the basic reproductive number that depends on the correlation time of the noise, which in the white-noise limit is independent of where noise resides (infection or recovery). In addition, we compared the typical fluctuations emerging from demographic and reaction-rate noise and determined the population sizes, correlation times, and reproductive numbers that place noisy SIR systems in the various limiting regimes. Altogether then, our paper illustrated a rich interplay between noise and outbreak dynamics—depending sensitively on fundamental noise characteristics and population size.

We expect that our results can be used to estimate the uncertainty in future outbreak sizes (e.g., the variance) due to contact-rate noise in, for instance, seasonal diseases if the noise and outbreak timescales are short compared to the seasonal time scale. Many control theories and policies for epidemics are based on mean-field analyses. As such, our results could provide more accurate analytical tools for disease control that account for and leverage uncertainty. More broadly, we predict that our Hamiltonian approach can be applied to a wider range of application systems, since the underlying contagion dynamics of the SIR model are common to many processes that are based upon population dynamics. As in the epidemic context discussed in this paper, we conjecture that observables such as the transient peak power in class-b laser systems [8] or the prevalence of rumors in

complex social networks [10] can be described in terms of one-parameter families of outbreak distributions when the dynamics are subjected to relatively small noise.

Currently, the theory presented pertains to well-mixed populations in which individuals come into contact with a contagion with homogeneous rates. In actuality, the contact rates within a population can be highly heterogeneous and/or spatially distributed and, therefore, an important extension of our paper includes the generalization of the outbreak-size distribution to complex network topology. Another common assumption that we used, which is only an approximation, was the implicit exponential waiting times for both the infection and recovery processes. Future work should include generalization to gamma and other more realistic waiting-time distributions. Finally, our work has relied substantially on small-noise assumptions allowing us to focus on the dominant, exponential contribution to the outbreak-size distribution. Corrections to this approach, which would include next-order contributions for larger noise amplitudes, are an important avenue for future analysis.

The numerical code we have used is available upon request.

ACKNOWLEDGMENTS

J.H. and I.B.S. were supported by the U.S. Naval Research Laboratory funding (No. N0001419WX00055), and the Office of Naval Research (No. N0001419WX01166 and No. N0001419WX01322). M.A. was supported by the Israel Science Foundation Grant No. 531/20.

APPENDIX A: RSV DATA ANALYSIS

We implement a time-discrete version of an SIR model to carry out Bayesian parameter inference on the hospitalization data for the 2019-2020 season of RSV in the U.S. [44]. At day d, the daily model updates follow the dynamics

$$i_d = \beta_d\, S_{d-1}\, I_{d-1}, \tag{A1a}$$

$$S_d = S_{d-1} - i_{d-1}, \tag{A1b}$$

$$I_d = I_{d-1} + i_d - \gamma\, I_{d-1} \tag{A1c}$$

$$R_d = R_{d-1} + \gamma\, I_{d-1}, \tag{A1d}$$

where (S_d, I_d, R_d) are the susceptible, infected, and recovered fractions ($S_d + I_d + R_d = 1$), respectively, and i_d is the fractional incidence. The parameters β_d and γ refer to the daily contact and recovery rates, respectively.

In our inference model, the dynamics of β_d are given by a daily discretized version of the OU process,

$$\beta_1 = \mathcal{N}(\beta_0, \sigma_\beta), \tag{A2}$$

$$\beta_d = \beta_{d-1} - \alpha(\beta_{d-1} - \beta_0) + \mathcal{N}(0, 2\alpha\sigma_\beta), \quad d > 1, \tag{A3}$$

where $\mathcal{N}(\mu, \sigma)$ is a normal random variable with mean μ and variance σ^2.

The parameter inference process is done by tying the daily discretized statistical model to the number of daily hospitalizations contained in the data H_d by the Poisson observation process

$$H_d = \mathrm{Pois}(\eta \cdot N \cdot i_d), \tag{A4}$$

where η is the hospitalization rate and N denotes the total population. With this model, we do Bayesian parameter inference using the platform `Stan` via the `R` package `rstan` [46,47]; `R` code for the model inference is available upon request.

APPENDIX B: ACTION FOR WHITE-NOISE FLUCTUATIONS IN RECOVERY

In addition to small fluctuations in the outbreaks, we can gain further analytical insight into the outbreak distribution for white noise by looking at other limiting-case scenarios. One such scenario is when the recovery-noise dominates over infection noise, namely, $f \gg 1$ while $\sigma_\gamma \ll 1$. In this limit, the Hamiltonian reduces to

$$H = x_i[p_i(R_0 x_s - 1) - R_0 x_s p_s] + \tfrac{1}{2} f x_i^2 p_i^2. \tag{B1}$$

As in Sec. IV B, given the simpler Hamiltonian Eq. (B1), the $H = 0$ condition can be combined with Hamilton's equations for $\dot{x}_s$ and $\dot{p}_i$, to give p_i as an explicit function of x_s and the initial condition $\delta \equiv p_i(t = 0)$:

$$p_i(x_s) = \frac{2\delta}{2 - f\delta \ln(x_s)/R_0}. \tag{B2}$$

Similarly, if we substitute Eq. (B2) and the zero-energy condition into $\dot{x}_i$ and divide by $\dot{x}_s$, we get the following differential equation for x_i as a function of x_s:

$$\frac{dx_i}{dx_s} = -1 + \frac{1}{R_0 x_s} - \frac{2 f x_i \delta}{R_0 x_s\Big(2 - f\delta \ln(x_s)/R_0\Big)}. \tag{B3}$$

Next, we can solve for $x_i(x_s)$ by separating the fraction of the population infected into a product of two functions that depend on x_s, i.e., $x_i(x_s) = u(x_s)v(x_s)$. By substituting the product form into Eq. (B3), setting $u\, dv/dx_s = -1 + 1/(R_0 x_s)$ and conditioning $u(x_s = 1) = 1$ and $v(x_s = 1) = 0$, we find

$$u(x_s) = \frac{1}{4}\left(\frac{f\delta}{R_0}\ln(x_s) - 2\right)^2, \quad v(x_s) = \int_1^{x_s} \frac{1 - R_0 x_s'}{R_0 x_s' u(x_s')}\, dx_{s'}, \tag{B4}$$

where the integral in Eq. (B4) can be expressed in terms of incomplete exponential integrals, though the formula is cumbersome. As $x_i(t \to \infty) \to 0$, Eq. (B4) gives us a condition for the final outbreak $1 - x_s^*$, namely, given R_0 and δ, we can solve $v(x_s^*) = 0$ for the unique value of x_s^*.

The final step for calculating the action associated with a given outbreak in the limit of recovery-only fluctuations is to differentiate $x_i(x_s) = u(x_s)v(x_s)$ and substitute Eqs. (B2) and (B4) into Eq. (12). The result is the following limiting-case action:

$$S = \int_1^{x_s^*} \frac{2\delta\Big(-1 + \dfrac{1}{R_0 x_s}\Big)}{\Big(f\delta \ln(x_s)/R_0 - 2\Big)}\, dx_s, \tag{B5}$$

where, as mentioned, the boundary condition for the integral (the final outbreak size) can be determined numerically from the condition $v(x_s^*) = 0$.

[1] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, NJ, 2008).

[2] H. Andersson and T. Britton, *Stochastic Epidemic Models and Their Statistical Analysis* (Springer-Verlag, New York, 2000).

[3] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. **87**, 925 (2015).

[4] H. S. Rodrigues, Application of sir epidemiological model: New trends, Int. J. Appl. Math. Inform. **10**, 92 (2016).

[5] L. Billings, W. M. Spears, and I. B. Schwartz, A unified prediction of computer virus spread in connected networks, Phys. Lett. A **297**, 261 (2002).

[6] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Critical phenomena in complex networks, Rev. Mod. Phys. **80**, 1275 (2008).

[7] J. Hindes and M. Assaf, Degree dispersion increases the rate of rare events in population networks, Phys. Rev. Lett. **123**, 068301 (2019).

[8] G. Kozyreff and T. Erneux, Contagious photons: Laser dynamics and the susceptible-infected-recovered model, Phys. Rev. E **108**, 024218 (2023).

[9] A. Volkening, D. F. Linder, M. A. Porter, and G. A. Rempala, Forecasting elections using compartmental models of infection, SIAM Rev. **62**, 837 (2020).

[10] X. Chen and N. Wang, Rumor spreading model considering rumor credibility, correlation and crowd classification based on personality, Sci. Rep. **10**, 5887 (2020).

[11] IHME COVID-19 Forecasting Team, Modeling COVID-19 scenarios for the United States, Nat. Med. **27**, 94 (2021).

[12] M. Ghaemi, M. Shojafar, Z. Zabihinpour, and Y. Asgari, On the possibility of oscillating in the Ebola virus dynamics and investigating the effect of the lifetime of T lymphocytes, PLoS ONE **17**, e0265065 (2022).

[13] I. Rodriguez-Barraquer, L. Mier-y-Teran-Romero, I. B. Schwartz, D. S. Burke, and D. A. T. Cummings, Potential opportunities and perils of imperfect dengue vaccines, Vaccine **32**, 514 (2014).

[14] A. Catching, S. Capponi, M. T. Yeh, S. Bianco, and R. Andino, Examining face-mask usage as an effective strategy to control COVID-19 spread, medRxiv (2021), doi:10.1101/2020.08.12.20173047.

[15] J. Hindes, M. Assaf, and I. B. Schwartz, Outbreak size distribution in stochastic epidemic models, Phys. Rev. Lett. **128**, 078301 (2022).

[16] J. Hindes, S. Bianco, and I. B. Schwartz, Optimal periodic closure for minimizing risk in emerging disease outbreaks, PLOS ONE **16**, e0244706 (2021).

[17] F. Brauer, Mathematical epidemiology: Past, present, and future, Infect. Dis. Model. **2**, 113 (2017).

[18] E. L. Ray, N. Wattanachit, J. Niemi, A. H. Kanji, K. House, E. Y. Cramer, J. Bracher, A. Zheng, T. K. Yamana, X. Xiong *et al.*, Ensemble forecasts of coronavirus disease 2019 (COVID-19) in the U.S, medrxiv (2020), doi:10.1101/2020.08.19.20177493.

[19] T. Ganyani, C. Faes, and N. Hens, Simulation and analysis methods for stochastic compartmental epidemic models, Ann. Rev. Stat. Appl. **8**, 69 (2021).

[20] C. M. Batistela, D. P. F. Correa, Á. M. Bueno, and J. C. Piqueira, SIRSi compartmental model for COVID-19 pandemic with immunity loss, Chaos Solit. Fractals **142**, 110388 (2021).

[21] J. C. Miller, Distribution of outbreak sizes for sir disease in finite populations, arXiv:1907.05138.

[22] J. L. Aron and I. B. Schwartz, Seasonality and period-doubling bifurcations in an epidemic model, J. Theor. Biol. **110**, 665 (1984).

[23] L. J. S. Allen, A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis, Infect. Dis. Model. **2**, 128 (2017).

[24] T. House, J. V. Ross, and D. Sirl, How big is an outbreak likely to be? Methods for epidemic final-size calculation, Proc. R. Soc. A: Math. Phys. Eng. Sci. **469**, 20120436 (2013).

[25] F. Ball, A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models, Adv. Appl. Probab. **18**, 289 (1986).

[26] F. Ball and D. Clancy, The final size and severity of a generalised stochastic multitype epidemic model, Adv. Appl. Probab. **25**, 721 (1993).

[27] M. I. Dykman, E. Mori, J. Ross, and P. M. Hunt, Large fluctuations and optimal paths in chemical kinetics, J. Chem. Phys. **100**, 5735 (1994).

[28] H. G. Hong and Y. Li, Estimation of time-varying reproduction numbers underlying epidemiological processes: A new statistical tool for the COVID-19 pandemic, PLOS ONE **15**, 1 (2020).

[29] R. Kahn, C. M. Peak, J. Fernández-Gracia, A. Hill, A. Jambai, L. Ganda, M. C. Castro, and C. O. Buckee, Incubation periods impact the spatial predictability of cholera and Ebola outbreaks in Sierra Leone, Proc. Natl. Acad. Sci. USA **117**, 5067 (2020).

[30] H. Vossler, P. Akilimali, Y. Pan, W. R. KhudaBukhsh, E. Kenah, and G. A. Rempała, Analysis of individual-level data from 2018–2020 Ebola outbreak in democratic republic of the congo, Sci. Rep. **12**, 5534 (2022).

[31] S. Dharmaratne, S. Sudaraka, I. Abeyagunawardena, K. Manchanayake, M. Kothalawala, and W. Gunathunga, Estimation of the basic reproduction number (R0) for the novel coronavirus disease in Sri Lanka, Virol. J. **17**, 144 (2020).

[32] R. Bsat, H. Chemaitelly, P. Coyle, P. Tang, M. R. Hasan, Z. Al Kanaani, E. Al Kuwari, A. A. Butt, A. Jeremijenko, A. H. Kaleeckal, A. N. Latif, R. M. Shaik, G. K. Nasrallah, F. M. Benslimane, H. A. Al Khatib, H. M. Yassine, M. G. Al Kuwari, H. E. Al Romaihi, M. H. Al-Thani, A. Al Khal *et al.*, Characterizing the effective reproduction number during the COVID-19 pandemic: Insights from Qatar's experience, J. Glob. Health **12**, 05004 (2022).

[33] A. Spannaus, T. Papamarkou, S. Erwin, and J. B. Christian, Inferring the spread of COVID-19: The role of time-varying reporting rate in epidemiological modelling, Sci. Rep. **12**, 10761 (2022).

[34] J. Dureau, K. Kalogeropoulos, and M. Baguelin, Capturing the time-varying drivers of an epidemic using stochastic dynamical systems, Biostatistics **14**, 541 (2013).

[35] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday, N. Davies, A. Gimma, K. van Zandvoort, H. Gibbs, J. Hellewell, C. I. Jarvis, S. Clifford, B. J. Quilty, N. I. Bosse, S. Abbott *et al.*, Early dynamics of transmission and control of COVID-19: A mathematical modelling study, Lancet Infect. Dis. **20**, 553 (2020).

[36] W. London and J. Yorke, Recurrent outbreaks of measles, chickenpox and mumps, Am. J. Epidemiol. **98**, 453 (1973).

[37] J. Yorke, N. Nathanson, G. Pianigiani, and J. Martin, Seasonality and the requirements for perpetuation and

eradication of viruses in populations, Am. J. Epidemiol. **109**, 103 (1979).

[38] P. Fine and J. Clarkson, Measles in England and Wales–I: An analysis of factors underlying seasonal patterns, Int. J. Epidemiol. **11**, 5 (1982).

[39] B. Grenfell, O. Bjørnstad, and B. Finkenstädt, Dynamics of measles epidemics: Scaling noise, determinism, and predictability with the TSIR model, Ecol. Monogr. **72**, 185 (2002).

[40] I. B. Schwartz, Small amplitude, long period outbreaks in seasonally driven epidemics, J. Math. Biol. **30**, 473 (1992).

[41] M. Drolet, A. Godbout, M. Mondor, G. Béraud, L. Drolet-Roy, P. Lemieux-Mellouki, A. Bureau, É. Demers, M.-C. Boily, C. Sauvageau, G. De Serres, N. Hens, P. Beutels, B. Dervaux, and M. Brisson, Time trends in social contacts before and during the COVID-19 pandemic: The connect study, BMC Public Health **22**, 1032 (2022).

[42] Y.-C. Chen, P.-E. Lu, C.-S. Chang, and T.-H. Liu, A time-dependent SIR model for COVID-19 with undetectable infected persons, IEEE Trans. Netw. Sci. Eng. **7**, 3279 (2020).

[43] S. Setianto and D. Hidayat, Modeling the time-dependent transmission rate using Gaussian pulses for analyzing the COVID-19 outbreaks in the world, Sci. Rep. **13**, 4466 (2023).

[44] RSV-NET Respiratory Syncytial Virus Hospitalization Surveillance Network, Centers for Disease Control and Prevention, J. Infect. Dis. **208**(S3), S165 (2013).

[45] If reinfection happens on short timescales, then models such as SIRS may be more appropriate [1].

[46] Stan Development Team (2023), RStan: the R interface to Stan, R package v. 2.21.8, https://cran.r-project.org/web/packages/rstan/vignettes/rstan.html.

[47] I. Hawryluk, T. A. Mellan, H. Hoeltgebaum, S. Mishra, R. P. Schnekenberg, C. Whittaker, H. Zhu, A. Gandy, C. A. Donnelly, S. Flaxman, and S. Bhatt, Inference of COVID-19 epidemiological distributions from Brazilian hospital data, J. R. Soc. Interface. **17**, 20200596 (2020).

[48] E. B. Rose, A. Wheatley, G. Langley, S. Gerber, and A. Haynes, Respiratory syncytial virus seasonality—United States, 2014-2017, MMWR Morb. Mortal. Wkly. Rep. **67**, 71 (2018).

[49] The Bayesian framework estimates the initial state for the SIR model that is used for prediction in a single year. The state includes the fraction of the population recovered from previous years. Since the model does not include birth, death, reinfection, periodicity, etc., it is not appropriate for multiple-year predictions [51], but is reasonable for a single season.

[50] Code available upon request.

[51] L. J. White, J. N. Mandl, M. G. M. Gomes, A. T. Bodley-Tickell, P. A. Cane, P. Perez-Brena, J. C. Aguilar, M. M. Siqueira, S. A. Portes, S. M. Straliotto, M. Waris, D. J. Nokes, and G. F. Medley, Understanding the transmission dynamics of respiratory syncytial virus using multiple time series and nested models, Math. Biosci. **209**, 222 (2007).

[52] M. Assaf, A. Kamenev, and B. Meerson, Population extinction in a time-modulated environment, Phys. Rev. E **78**, 041123 (2008).

[53] E. Roberts, S. Be'er, C. Bohrer, R. Sharma, and M. Assaf, Dynamics of simple gene-network motifs subject to extrinsic fluctuations, Phys. Rev. E **92**, 062717 (2015).

[54] We must assume that, e.g., $\sigma_\beta/\beta_0 \lesssim 1/3$ so that over 99% of the distribution is in the physical domain.

[55] M. I. Dykman, Large fluctuations and fluctuational transitions in systems driven by colored Gaussian noise: A high-frequency noise, Phys. Rev. A **42**, 2020 (1990).

[56] M. I. Freidlin, J. Szücs, and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Grundlehren der mathematischen Wissenschaften (Springer-Verlag, Berlin, 2012).

[57] M. Assaf and B. Meerson, WKB theory of large deviations in stochastic populations, J. Phys. A: Math. Theor. **50**, 263001 (2017).

[58] E. Forgoston and R. O. Moore, A primer on noise-induced transitions in applied dynamical systems, SIAM Rev. **60**, 969 (2018).

[59] We note that the general Hamiltonian approach will be similar for more complex noise distributions. Namely, one writes down a probability-flux equation for any state variables in the model that change in time, including both the deterministic and random parts. Then, one takes a small-noise limit, which results generally in a Hamilton-Jacobi equation according to large-deviations theory [55–58,63].

[60] M. Assaf, M. Mobilia, and E. Roberts, Cooperation dilemma in finite populations under fluctuating environments, Phys. Rev. Lett. **111**, 238101 (2013).

[61] A. Taitelbaum, R. West, M. Assaf, and M. Mobilia, Population dynamics in a changing environment: Random versus periodic switching, Phys. Rev. Lett. **125**, 048105 (2020).

[62] M. Thattai and A. van Oudenaarden, Intrinsic noise in gene regulatory networks, Proc. Natl. Acad. Sci. USA **98**, 8614 (2001).

[63] J. Hindes, P. Jacquod, and I. B. Schwartz, Network desynchronization by non-Gaussian fluctuations, Phys. Rev. E **100**, 052314 (2019).