The Limits to Learn a Diffusion Model: Online Appendix

Appendices EC.1, EC.2 and EC.3 contain the proofs of Theorems 2, 3 and 4 respectively. Appendix EC.5 contains the proofs of Propositions 1, 2, 3 and 4, each in their own subsections. Section EC.6 provides details on the datasets used in Section 5, and Section EC.7 contains a detailed description of the COVID-19 forecasting model from Section 5.4.

EC.1. Proof of Theorem 2

We finish the sections of the proof that were not included in the main paper. This includes the proof of Lemma 1, Lemma 3, calcuations for Lemma 2, and details regarding the final step of the proof.

We define $\lambda(N,k-1,C_{k-1}) = \left(\frac{\beta(N-C_{k-1})}{N}+\gamma\right) I_{k-1}$ and $\eta(N,C_{k-1}) = \frac{\beta(N-C_{k-1})}{\beta(N-C_{k-1})+N\gamma}$. Thus, for $k \le \tau$, $\lambda(N,k-1,C_{k-1})$ is the mean of the k-th inter-arrival time and $\eta(N,C_{k-1})$ is the probability that the arrival in the k-th instance is a new infection rather than a recovery.

EC.1.1. Proof of Lemma 1

Proof. Suppose $k<\tau$ i.e $E_k=1$. Then, k is equal to total number of jumps that have occurred so far (the number of movements from S to I and from I to R). The number of individuals that have moved from S to I is $C_k - I_0 - R_0$, and the number of movements from I to R is $C_k - I_k - R_0$. Therefore, $k = 2C_k - I_0 - I_k - 2R_0$. Since $I_k > 0$, $C_k > r_k$.

Suppose $k \ge \tau$ i.e $E_k = 0$. Then, k is greater than or equal to the total number of jumps, which is still equal to $2C_k - I_0 - I_k - 2R_0$. Hence $C_k \le r_k$ in this case.

□

EC.1.2. Proof of Lemma 3

Proof. Let $X_k \overset{iid}{\sim} \text{Bern}(p)$ for $k = 1,2,\ldots$. Let $\{A_k : k \ge 0\}$ be a stochastic process defined by:

$$A_k = \begin{cases} C_0 & \text{if } k=0 \\ C_0 + X_1 + \cdots + X_k & \text{if } A_i > r_i \; \forall i<k \\ A_{k-1} & \text{otherwise.} \end{cases}$$

Let $\tau_A = \min\{k : A_k \le r_k\}$ be the “stopping time” of this process.

Claim EC.1. $\Pr(\tau \le m) \le \Pr(\tau_A \le m)$.

The proof of this claim involves showing the process $\{A_k\}$ is stochastically less than $\{C_k\}$; the proof can be found in Section EC.1.2.1. We now upper bound $\Pr(\tau_A \le m)$. $\tau_A \le m$ if and only if $A_k \le r_k$ for some $k \le m$. Before this happens, $A_k = C_0 + X_1 + \cdots + X_k$. Therefore, if $\tau_A \le m$, it must be that $C_0 + X_1 + \cdots + X_k \le \frac{k+I_0+2R_0}{2}$ for some $k \le m$.

$$\Pr(\tau_A \le m) \le \sum_{k=1}^{m} \Pr\left(C_0 + X_1 + \cdots + X_k \le \frac{k+I_0+2R_0}{2}\right)$$

$$= \sum_{k=1}^{m} \Pr\left(X_1 + \cdots + X_k < pk\left(1 - \frac{2pk - k + I_0}{2pk}\right)\right).$$

Since $\mathbb{E}[X_1 + \cdots + X_k] = pk$, using the Chernoff bound (multiplicative form: $\Pr(\sum_{i=1}^{k} X_i \leq (1-\delta)\mu) \leq \exp(-\delta^2\mu/2)$) gives

$$\begin{aligned}
\Pr(\tau_A \leq m) &\leq \sum_{k=1}^{m} \exp\left(-\frac{pk}{2}\left(\left(1 - \frac{1}{2p}\right) + \frac{I_0}{2pk}\right)^2\right) \\
&= \sum_{k=1}^{m} \exp\left(-\frac{pk}{2}\left(1 - \frac{1}{2p}\right)^2 - \frac{I_0}{2}\left(1 - \frac{1}{2p}\right) - \frac{I_0^2}{8pk}\right) \\
&\leq \sum_{k=1}^{m} \exp\left(-\frac{pk}{2}\left(1 - \frac{1}{2p}\right)^2 - \frac{I_0}{2}\left(1 - \frac{1}{2p}\right)\right) \\
&\leq \exp\left(-\left(\frac{1}{2} - \frac{1}{4p}\right) I_0\right) \sum_{k=1}^{m} \exp\left(-\frac{pk}{2}\left(1 - \frac{1}{2p}\right)^2\right) \\
&\leq C_1 \exp(-C_2 I_0), \qquad \text{(EC.1)}
\end{aligned}$$

for constants $C_1 = \sum_{k=1}^{\infty} \exp\left(-\frac{pk}{2}\left(1 - \frac{1}{2p}\right)^2\right), C_2 = \frac{1}{2} - \frac{1}{4p} > 0$. ($C_1$ is a constant since it is a geometric series with a ratio smaller than 1, since $p > 1/2$.) Let D be the solution to $C_1 \exp(-C_2 D) = \frac{1}{2}$. Then, if $I_0 \geq D$, $\Pr(E_m) = 1 - \Pr(\tau \leq m) \geq 1 - \Pr(\tau_A \leq m) \geq \frac{1}{2}$.

□

EC.1.2.1. Proof of Claim EC.1.

Definition EC.1. For scalar random variables X, Y, we say that X is *stochastically less than* Y (written $X \leq_{st} Y$) if for all $t \in \mathbb{R}$,

$$\Pr(X > t) \leq \Pr(Y > t).$$

For random vectors $X, Y \in \mathbb{R}^n$ we say that $X \leq_{st} Y$ if for all increasing functions $\phi : \mathbb{R}^n \to \mathbb{R}$,

$$\phi(X_1, \ldots, X_n) \leq_{st} \phi(Y_1, \ldots, Y_n).$$

We make use of the following known result for establishing stochastic order for stochastic processes.

Theorem EC.1 **(Veinott 1965)**. *Suppose $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ are random variables such that $X_1 \leq_{st} Y_1$ and for any $x \leq y$,*

$$(X_k | X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) \leq_{st} (Y_k | Y_1 = y_1, \ldots, Y_{k-1} = y_{k-1})$$

for every $2 \leq k \leq n$. Then, $(X_1, \ldots, X_n) \leq_{st} (Y_1, \ldots, Y_n)$.

Proof of Claim EC.1. Because of the condition $\frac{\beta(N-m-C_0)}{\beta(N-m-C_0)+N\gamma} > p$, for $k \le m$ and $k \le \tau$, $C_k - C_{k-1} \sim \text{Bern}(q)$ for $q > p$. First, we show $(A_0, A_1, \ldots, A_m) \le_{st} (C_0, C_1, \ldots, C_m)$ using Theorem EC.1. $C_0 \le_{st} A_0$ since $C_0 = A_0 = I_0$. We condition on $A_{k-1} = x$ and $C_{k-1} = y$ for $x \le y$, and we must show $A_k \le_{st} C_k$. (We do not need to condition on all past variables since the both processes are Markov.) If $x \le r_{k-1}$, then $A_k = A_{k-1} = x \le y = C_{k-1} \le C_k$. Otherwise, the process A_k has not stopped, and neither has C_k since $y \ge x$. Then, $A_k \sim x + \text{Bern}(p)$ and $C_k \sim y + \text{Bern}(q)$ for some $q \ge p$. Clearly, $A_k \le_{st} C_k$ in this case. We apply Theorem EC.1, which implies $A_m \le_{st} C_m$.

Define the function $u : \mathbb{R}^{m+1} \to \{0,1\}$, $u(x_0, x_1, \ldots, x_m) = \mathbb{1}\{\cup_{k=1}^{m} \{x_k \le r_k\}\}$. Then, $u(A_0, A_1, \ldots, A_m) = 1$ if and only if $\tau_A \le m$, and $u(C_0, C_1, \ldots, C_m) = 1$ if and only if $\tau \le m$. u is a decreasing function. Therefore, $u(A_0, A_1, \ldots, A_m) \ge_{st} u(C_0, C_1, \ldots, C_m)$. Then, $\Pr(\tau \le m) = \Pr(u(C_0, C_1, \ldots, C_m) \ge 1) \le \Pr(u(A_0, A_1, \ldots, A_m) \ge 1) = \Pr(\tau_A \le m)$ as desired. □

EC.1.3. Calculations for Lemma 2

We define $\lambda(N, k-1, C_{k-1}) = \left(\frac{\beta(N-C_{k-1})}{N} + \gamma\right) I_{k-1}$ and $\eta(N, C_{k-1}) = \frac{\beta(N-C_{k-1})}{\beta(N-C_{k-1})+N\gamma}$. Thus, for $k \le \tau$, $\lambda(N, k-1, C_{k-1})$ is the mean of the k-th inter-arrival time and $\eta(N, C_{k-1})$ is the probability that the arrival in the k-th instance is a new infection rather than a recovery.

Derivation of $\mathbb{E}_{C_k}[g_{C_k|C_{k-1}}(C_k, C_{k-1}, N) | E_{k-1} = 1]$. When $E_{k-1} = 1$, we have $C_k \sim C_{k-1} + \text{Bern}(\eta(N, C_{k-1}))$. Therefore, $\mathbb{E}_{C_k}[g_{C_k|C_{k-1}}(C_k, C_{k-1}, N) | E_{k-1} = 1] = \mathcal{J}_{C_k \sim \text{Bern}(\eta(N, C_{k-1}))}(N)$. We reparameterize to write the Fisher information as:

$$\begin{aligned}
\mathbb{E}_{C_k}[g_{C_k|C_{k-1}}(C_k, C_{k-1}, N) | E_{k-1} = 1] &= \mathcal{J}_{C_k \sim \text{Bern}(\eta)}(\eta) \left(\frac{\partial}{\partial N} \eta(N, C_{k-1})\right)^2 \\
&= \frac{1}{\eta(1-\eta)} \left(\frac{\partial}{\partial N} \eta(N, C_{k-1})\right)^2.
\end{aligned}$$

Use $\eta(N, C_{k-1}) = \frac{\beta(N-C_{k-1})}{\beta(N-C_{k-1})+N\gamma}$ to derive

$$\begin{aligned}
\frac{\partial}{\partial N} \eta(N, C_{k-1}) &= \frac{\beta(\beta(N-C_{k-1}) + \gamma N) - \beta(N-C_{k-1})(\beta+\gamma)}{(\beta(N-C_{k-1}) + \gamma N)^2} \\
&= \frac{\beta\gamma C_{k-1}}{(\beta(N-C_{k-1}) + \gamma N)^2}.
\end{aligned}$$

Also, $\frac{1}{\eta(1-\eta)} = \frac{(\beta(N-C_{k-1})+N\gamma)^2}{(N-C_{k-1})\beta N\gamma}$.
Substituting,

$$\begin{aligned}
\mathbb{E}_{C_k}[g_{C_k|C_{k-1}}(C_k, C_{k-1}, N) | E_{k-1} = 1] &= \frac{(\beta(N-C_{k-1}) + N\gamma)^2}{(N-C_{k-1})\beta N\gamma} \left(\frac{\beta\gamma C_{k-1}}{(\beta(N-C_{k-1}) + \gamma N)^2}\right)^2 \\
&= \frac{\beta\gamma C_{k-1}^2}{(N-C_{k-1}) N (\beta(N-C_{k-1}) + \gamma N)^2}
\end{aligned}$$

Derivation of $\mathbb{E}_{T_k}[g_{T_k|C_{k-1}}(T_k, C_{k-1}, N)|E_{k-1} = 1]$. Similarly, conditioned on $E_{k-1} = 1, T_k \sim \text{Exp}(\lambda(N, k-1, C_{k-1}))$. Therefore, $\mathbb{E}_{T_k}[g_{T_k|C_{k-1}}(T_k, C_{k-1}, N)] = \mathcal{J}_{T_k \sim \text{Exp}(\lambda(N,k-1,C_{k-1}))}(N)$. We reparameterize to write

$$\begin{aligned}
\mathbb{E}_{T_k}[g_{T_k|C_{k-1}}(T_k, C_{k-1}, N)] &= \mathcal{J}_{T_k \sim \text{Exp}(\lambda)}(\lambda) \left(\frac{\partial}{\partial N} \lambda(N, k-1, C_{k-1}) \right)^2 \\
&= \frac{1}{\lambda^2} \left(\frac{\partial}{\partial N} \lambda(N, k-1, C_{k-1}) \right)^2 .
\end{aligned}$$

Use $\lambda(N, k-1, C_{k-1}) = (\frac{\beta(N-C_{k-1})}{N} + \gamma)(2C_{k-1} - (k-1) - I_0 - 2R_0)$ to derive

$$\begin{aligned}
\frac{\partial}{\partial N} \lambda(N, k-1, C_{k-1}) &= \frac{\beta C_{k-1}(2C_{k-1} - (k-1) - I_0 - 2R_0)}{N^2} \\
\frac{1}{\lambda(N, k-1, C_{k-1})} &= \frac{N}{(\beta(N - C_{k-1}) + \gamma N)(2C_{k-1} - (k-1) - I_0 - 2R_0)}.
\end{aligned}$$

Substituting,

$$\mathbb{E}_{T_k}[g_{T_k|C_{k-1}}(T_k, C_{k-1}, N)] = \left(\frac{\beta C_{k-1}}{N(\beta(N - C_{k-1}) + \gamma N)} \right)^2$$

Derivation of $\mathcal{J}_{O_m}(N)$. Using the expressions derived above for $\mathbb{E}_{C_k}[g_{C_k|C_{k-1}}(C_k, C_{k-1}, N)|E_{k-1} = 1]$ and

$\mathbb{E}_{T_k}[g_{T_k|C_{k-1}}(T_k, C_{k-1}, N)]$, we get

$$\begin{aligned}
&\mathbb{E}_{C_k}[g_{C_k|C_{k-1}}(C_k, C_{k-1}, N)|E_{k-1} = 1] + \mathbb{E}_{T_k}[g_{T_k|C_{k-1}}(T_k, C_{k-1}, N)] \\
&= \frac{\beta\gamma C_{k-1}^2}{(N - C_{k-1})N(\beta(N - C_{k-1}) + \gamma N)^2} + \left(\frac{\beta C_{k-1}}{N(\beta(N - C_{k-1}) + \gamma N)} \right)^2 \\
&= \frac{C_{k-1}^2}{(N - C_{k-1})N^2(N - C_{k-1} + \frac{\gamma}{\beta}N)}
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathcal{J}_{O_m}(N) &= \sum_{k=1}^{m} \mathbb{E}[g_{C_k|C_{k-1}}(C_k, C_{k-1}, N) + g_{T_k|C_{k-1}}(T_k, C_{k-1}, N)|E_{k-1} = 1] \Pr(E_{k-1} = 1) \\
&= \sum_{k=1}^{m} \mathbb{E}\left[\frac{C_{k-1}^2}{(N - C_{k-1})N^2(N - C_{k-1} + \frac{\gamma}{\beta}N)} \,\middle|\, E_{k-1} = 1 \right] \Pr(E_{k-1} = 1).
\end{aligned}$$

EC.1.4. Details of Final Step of Theorem 2

Define $p \triangleq \frac{1}{2}(\frac{\beta}{\beta+\gamma} + \frac{1}{2}) > \frac{1}{2}$ as in Lemma 3. Assume N is large enough so that $m + C_0 \leq \frac{N}{2}$ and $\frac{\beta(N-m-C_0)}{\beta(N-m-C_0)+P\gamma} > p$ (this is possible since $\frac{\beta}{\beta+\gamma} > p$ and $m = o(N)$).

For the upper bound, we have that $C_k \leq k + I_0 + R_0$ by definition. Since $I_0, R_0 \leq m$ by assumption, $C_k \leq 3m$. Moreover, by assumption, $C_k \leq m + C_0 \leq \frac{N}{2}$. Plugging these into (11) results in

$$\mathcal{J}_{O_m}(N) \leq \sum_{k=0}^{m-1} \Pr(E_{k-1} = 1) \frac{(3m)^2}{N^2(N - \frac{1}{2}N)((N - \frac{1}{2}N) + \frac{\gamma}{\beta}N)} \leq H_1 \frac{m^3}{N^4},$$

for a constant H_1.

Then, similarly to the upper bound, $\mathcal{J}_{O_m}(N) \geq H_2 \frac{m^3}{N^4}$ follows from using $\Pr(E_m = 1) \geq \frac{1}{2}$ and the fact that $C_k \geq \frac{k+I_0+2R_0}{2} \geq \frac{k}{2}$ when $E_k = 1$ (Lemma 1):

$$\mathcal{J}_{O_m}(N) \geq \sum_{k=0}^{m-1} \frac{1}{2} \frac{\left(\frac{k}{2}\right)^2}{N^4} \geq H_2 \frac{m^3}{N^4},$$

Combining the upper and lower bounds finish the proof.

EC.1.5. Generalization of Theorem 1: a Finite-sample Result

Note that the proof of Theorem 2 provides exact formulas for the Fisher information, where quantities such as N and m are finite. This leads to the following result of which Theorem 1 is a special case, and it holds for any initial conditions I_0 and R_0.

THEOREM EC.2. *Consider any observation $(T_0, I_0, R_0, T_1, I_1, R_1, \ldots, T_m, I_m, R_m)$ from either a Bass model or an SIR model (with any initial I_0 and R_0). Suppose $\hat{N}$ is an un-biased estimator for N. Let $I_{\max} = \max_{1\leq i\leq m} I_i$ with $I_{\max} \leq c(p,\gamma,\beta)N$. Then*

$$\mathbb{E}\left[\frac{(\hat{N}-N)^2}{N^2}\right] \geq C(p,\gamma,\beta)\frac{N^2}{I_{\max}^3}$$

where $C(p,\gamma,\beta)$ and $c(p,\gamma,\beta)$ are constants that are explicit functions of p,γ,β.

EC.2. Proof of Theorem 3

Proof. We construct the estimators $\hat{a}$ for a and $\hat{\beta}$ for β as the following. To begin, let

$$\hat{A} := \frac{\sum_{i=1}^{m/2} \min(T_i, T_{m-i})}{m/2}$$
$$\hat{B} := \frac{\sum_{i=1}^{m/4} \min(T_i, T_{m/2-i})}{m/4}.$$

We will show momentarily that $\hat{A}$ approximates $\frac{1}{2a+m\beta}$ and $\hat{B}$ approximates $\frac{1}{2a+(m/2)\beta}$. We then construct $\hat{a}$ and $\hat{\beta}$:

$$\hat{\beta} := \left(\frac{1}{\hat{A}} - \frac{1}{\hat{B}}\right)\frac{2}{m}$$
$$\hat{a} := \left(\frac{2}{\hat{B}} - \frac{1}{\hat{A}}\right)\frac{1}{2}.$$

To start the proof, let us analyze $\hat{A}$. Let $A_i = \min(T_i, T_{m-i}), 1 \leq i \leq \lfloor m/2 \rfloor$. By the property of independent exponential random variables, we have $A_i \sim \mathrm{Exp}(l_i)$ where

$$l_i := (2a+\beta m) - (a+i\beta)\frac{i}{N} - (a+(m-i)\beta)\frac{m-i}{N}.$$

Note that when $N \gg m$, we shall have $l_i \approx 2a + \beta m$, which is independent from i. This inspires us to use $\hat{A} := \frac{\sum_i A_i}{m/2}$ as an estimator for $\frac{1}{2a+\beta m}$.

More specifically, Let $\mu = \frac{\sum_i E[A_i]}{m/2} = \frac{1}{m/2}\sum_i \frac{1}{l_i}$. Note that

$$\frac{N-m}{N}(2a+\beta m) \le l_i \le 2a+\beta m.$$

Then, this implies that μ is close to $\frac{1}{2a+\beta m}$:

$$\frac{1}{2a+m\beta} \le \mu \le \frac{N}{N-m}\frac{1}{2a+m\beta} \tag{EC.2}$$

On the other hand, we can invoke the multiplicative Bernstein inequality (Janson (2018)) to obtain, with probability $1 - O(1/N^2)$,

$$(1-\delta)\mu \le \hat{A} \le \mu(1+\delta) \tag{EC.3}$$

where $\delta := O(\sqrt{\log(N)/m})$. Combining Eq. (EC.2) and Eq. (EC.3), we then have

$$\frac{1}{2a+m\beta}(1-\delta) \le \hat{A} \le \frac{N}{N-m}(1+\delta)\frac{1}{2a+m\beta}.$$

This further implies desired bounds for using $\frac{1}{\hat{A}}$ to estimate $2a+m\beta$:

$$\begin{aligned}\left|\frac{1}{\hat{A}} - (2a+m\beta)\right| &\lesssim (\delta + \frac{m}{N})(2a+m\beta)\\ &\lesssim \sqrt{\frac{\log(N)}{m}}(2a+m\beta)\end{aligned}$$

where the last inequality uses $m = O(N^{2/3}\log^{1/3}(N))$ (hence $m/N \lesssim \delta$).

A similar analysis can be conducted for $\hat{B}$, which implies that

$$\begin{aligned}\left|\frac{1}{\hat{B}} - (2a+(m/2)\beta)\right| &\lesssim (\delta + \frac{m}{N})(2a+(m/2)\beta)\\ &\lesssim \sqrt{\frac{\log(N)}{m}}(2a+(m/2)\beta)\end{aligned}$$

Combining the bounds of $\hat{A}$ and $\hat{B}$, we then obtain the bounds for $\hat{a}$ and $\hat{\beta}$, which completes the proof[7]. $\square$

[7] A further refinement can be performed for analyzing $\hat{a}$ by considering a set of estimators $\hat{S}_k = \sum_{i=1}^{k/2}\min(T_i, T_{k-i}$ that generalize $\hat{A}$ and $\hat{B}$. We omit the details for simplicity.

EC.3. Proof of Theorem 4

EC.3.1. Construction of Estimators

Our construction of estimators $\hat{\beta}$ for β and $\hat{\gamma}$ for γ is the following. To begin, let

$$\hat{A} := \frac{C_m - C_0}{m}$$
$$\hat{B} := \frac{\sum_{k=1}^{\min(m,\tau)} I_{k-1}T_k}{m}.$$

We will show momentarily that $\hat{A}$ can be viewed as an estimator for $\frac{\beta}{\beta+\gamma}$ and $\hat{B}$ an estimator for $\frac{1}{\beta+\gamma}$. Then given $\hat{A}$ and $\hat{B}$, we construct

$$\hat{\beta} := \frac{\hat{A}}{\hat{B}}$$
$$\hat{\gamma} := \frac{1}{\hat{B}} - \hat{\beta}.$$

This construction leads to the guarantees stated in Theorem 4.

The proof is based on a series of lemmas stated below. The first lemma bounds the probability that the epidemic diminishes before m samples, which follows from (EC.1) of the proof of Lemma 3.

Lemma EC.1. *If $\frac{\beta}{\beta+\gamma}\frac{N-m-C_0}{N} > p$, $\Pr(\tau < m) \le B_1 e^{-B_2 I_0}$, where $B_1, B_2 > 0$ are constant that depend only on β and γ.*

The next two lemmas give a high probability confidence bound for estimators $\hat{A}$ and $\hat{B}$.

Lemma EC.2. *For any m, I_0 where $\frac{\beta}{\beta+\gamma}\frac{N-m-C_0}{N} > \frac{1}{2}$, for any $\delta > 0$,*

$$\Pr\left(\frac{C_m - C_0}{m} \notin \left[\frac{\beta}{\beta+\gamma}(1-\delta)\frac{N-m-C_0}{N}, \frac{\beta}{\beta+\gamma}(1+\delta)\right], \tau \ge m\right) \le 2\exp(-m\delta^2/(4+2\delta)).$$

Lemma EC.3. *Let $\tilde{S}_m = \sum_{k=1}^{\min(m,\tau)} I_{k-1}T_k$. Then*

$$\Pr\left(\frac{\tilde{S}_m}{m} \notin [\frac{(1-\delta)}{\beta+\gamma}, \frac{(1+\delta)}{\beta+\gamma}\frac{N}{N-m-C_0}], \tau \ge m\right) \le 2e^{-m\frac{N-m-C_0}{N}(\delta - \ln(1+\delta))}.$$

The next proposition combines the two estimators from the above lemmas and into estimators $\hat{\beta}$ and $\hat{\gamma}$.

Proposition EC.1. *Assume $\beta > \gamma > 0$. Let $I_0 \le m < N$ such that $\frac{\beta}{\beta+\gamma}\frac{N-m-C_0}{N} > p$. Let $z = \frac{N-m-C_0}{N}$. Then, for any $0 < \delta < 1$, with probability $1 - 4e^{-m(\delta - \ln(1+\delta))} - 4e^{-m\delta^2/(4+2\delta)} - 2B_1e^{-B_2 I_0}$,*

$$\hat{\beta} \in \left[\beta\frac{(1-\delta)z^2}{1+\delta}, \beta\frac{1+\delta}{1-\delta}\right] \tag{EC.4}$$
$$\hat{\gamma} \in \left[\gamma\frac{z}{1+\delta} + \beta\frac{(1-\delta)z - (1+\delta)^2}{(1+\delta)(1-\delta)}, \gamma\frac{1}{1-\delta} + \beta\frac{1+\delta-(1-\delta)^2z^2}{(1-\delta)(1+\delta)}\right], \tag{EC.5}$$

where $B_1, B_2 > 0$ are constants that depend on β and γ.

We first show Theorem 4 using these results. We then prove Lemma EC.2, Lemma EC.3, and Proposition EC.1 in Section EC.3.3.

EC.3.2. Proof of Theorem 4

Proof. Let $\delta = \sqrt{\frac{5\log m}{m}}$. First, we claim that the probability in Proposition EC.1 is greater than $1 - \frac{8}{m} - 2B_1 e^{-B_2 I_0}$. Note that $\ln(1+\delta) \le \delta - \frac{\delta^2}{2} + \delta^3$, implying $\delta - \ln(1+\delta) \ge \delta^2(\frac{1}{2} - \delta)$. Since $\delta \le \frac{1}{4}$,

$$4e^{-m(\delta - \ln(1+\delta))} \le 4e^{-m\frac{\delta^2}{4}} \le \frac{4}{m}.$$

Using $\delta \le \frac{1}{4}$ again,

$$4e^{-m\delta^2/(4+2\delta)} \le 4e^{-m\frac{\delta^2}{5}} = \frac{4}{m}.$$

Hence, the bound in EC.1 holds with probability greater than $1 - \frac{8}{m} - 2B_1 e^{-B_2 I_0}$.

Since we assume $m(m + C_0) \le N$ and $z = 1 - \frac{m + C_0}{N}$,

$$1 - z \le \frac{1}{m}. \tag{EC.6}$$

From here on, assume the confidence bounds (EC.4)-(EC.5) hold. Note that $\frac{1+\delta}{1-\delta} \le 1 + 3\delta$ and $\frac{1-\delta}{1+\delta} \ge 1 - 3\delta$ for $\delta < \frac{1}{4}$. Then,

$$\begin{aligned}
(\hat\beta - \beta)^2 &\le \beta^2 \left(1 + 3\delta - (1 - 3\delta)z^2\right)^2 \\
&\le \beta^2 \left((1 - z) + 3\delta(1 + z)\right)^2 \\
&\le \beta^2 \left(\frac{1}{m} + 6\sqrt{\frac{5\log m}{m}}\right)^2 \\
&\le \beta^2 M_3 \frac{\log m}{m}
\end{aligned}$$

for an absolute constant $M_3 > 0$. The second last step uses (EC.6) and $1 + z \le 2$. Therefore, $\mathrm{RelError}(\hat\beta, \beta) \le M_1 \frac{\log m}{m}$.

Similarly,

$$(\hat\gamma - \gamma)^2 \le \left(\gamma\left(\frac{1}{1-\delta} - \frac{z}{1+\delta}\right) + \beta\left(\frac{1 + \delta - (1-\delta)^2 z^2}{(1-\delta)(1+\delta)} - \frac{(1-\delta)z - (1+\delta)^2}{(1+\delta)(1-\delta)}\right)\right)^2. \tag{EC.7}$$

Using the fact that $(1-\delta)(1+\delta) \ge \frac{1}{2}$,

$$\frac{1}{1-\delta} - \frac{z}{1+\delta} \le 2((1 - z) + \delta(1 + z)) \le 2\left(\frac{1}{m} + 2\sqrt{\frac{5\log m}{m}}\right).$$

$$\begin{aligned}
\frac{1 + \delta - (1-\delta)^2 z^2}{(1-\delta)(1+\delta)} - \frac{(1-\delta)z - (1+\delta)^2}{(1+\delta)(1-\delta)} &= \frac{(1+\delta) - (1-\delta)z + (1+\delta)^2 - (1-\delta)^2 z^2}{1 - \delta^2} \\
&\le 2(1 - z) + 4\delta(1 + z) + \frac{1+\delta}{1-\delta} - \frac{1-\delta}{1+\delta} z^2 \\
&\le 2(1 - z) + 8\delta + (1 + 3\delta) - (1 - 3\delta)z^2 \\
&\le 2(1 - z) + 8\delta + (1 - z^2) + 6\delta(1 + z^2) \\
&\le (1 - z)(3 + z) + \delta(8 + 6(1 + z^2)) \\
&\le \frac{4}{m} + 20\sqrt{\frac{5\log m}{m}}.
\end{aligned}$$

Substituting back into (EC.7) results in

$$(\hat{\gamma}-\gamma)^2 \leq \left(\gamma\left(\frac{2}{m}+4\sqrt{\frac{5\log m}{m}}\right)+\beta\left(\frac{4}{m}+20\sqrt{\frac{5\log m}{m}}\right)\right)^2$$
$$\leq M_2\beta^2\frac{\log m}{m},$$

for an absolute constant M_2, since $\beta>\gamma$. This implies the desired result. □

EC.3.3. Proofs of Intermediate Results

EC.3.3.1. Proof of Lemma EC.2.

Proof. Fix m, let $z:=\frac{N-m-C_0}{N}, p=\frac{\beta}{\beta+\gamma}z$. Then $p>\frac{1}{2}$. Define three stochastic processes $\{A_k: k\geq 0\}$, $\{B_k: k\geq 0\}$, $\{\tilde{C}_k: k\geq 0\}$:

$$A_k=\begin{cases}C_0 & \text{if } k=0\\ A_{k-1}+\text{Bern}(p) & \text{otherwise.}\end{cases}$$
$$B_k=\begin{cases}C_0 & \text{if } k=0\\ B_{k-1}+\text{Bern}(p/z) & \text{otherwise.}\end{cases}$$
$$\tilde{C}_k=\begin{cases}C_0 & \text{if } k=0\\ \tilde{C}_{k-1}+\text{Bern}\left\{\frac{\beta(N-\tilde{C}_{k-1})}{\beta(N-\tilde{C}_{k-1})+N\gamma}\right\} & \text{otherwise.}\end{cases}$$

Note that $\tilde{C}_k$ is a modified version of C_k where $\tilde{C}_k$ still evolves after the stopping time.

CLAIM EC.2. *A_m is stochastically less than $\tilde{C}_m$ ($A_m\leq_{st}\tilde{C}_m$); $\tilde{C}_m$ is stochastically less than B_m ($\tilde{C}_m\leq_{st}B_m$); that is, for any $\ell\in\mathbb{R}$,*

$$\Pr(B_m\leq\ell)\leq\Pr(\tilde{C}_m\leq\ell)\leq\Pr(A_m\leq\ell).$$

This claim follows from Theorem EC.1, using a similar argument to Claim EC.1.

Let $A_k=C_0+X_1+X_2+\ldots X_k$ where $X_i\sim\text{Bern}(p)$ are independent. We provide the left tail bound for C_m. Note that when $\tau\geq m$, $C_m\overset{d}{=}\tilde{C}_m$. Hence,

$$\begin{aligned}\Pr(C_m\leq mp(1-\delta)+C_0,\tau\geq m)&=\Pr(\tilde{C}_m\leq mp(1-\delta)+C_0,\tau\geq m)\\&\leq\Pr(\tilde{C}_m\leq mp(1-\delta)+C_0)\\&\leq\Pr(A_m\leq mp(1-\delta)+C_0).\end{aligned}\tag{EC.8}$$

Using the Chernoff bound gives,

$$\begin{aligned}\Pr(A_m\leq mp(1-\delta)+C_0)&=\Pr(C_0+X_1+\cdots+X_m\leq pm(1-\delta)+C_0)\\&=\Pr\left(X_1+\cdots+X_m\leq mp\left(1-\delta\right)\right)\\&\leq\exp\left(-\frac{mp}{2}\delta^2\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\Pr\left(\frac{C_m - C_0}{m} \le \frac{p}{z}(1-\delta)z, \tau \ge m\right) &= \Pr\left(C_m \le mp(1-\delta) + C_0, \tau \ge m\right)\\ &\le \exp\left(-\frac{mp}{2}\delta^2\right) \le \exp(-m\delta^2/4).\end{aligned}$$

Let $B_k = C_0 + Y_1 + \ldots + Y_k$ where $Y_i \sim \mathrm{Bern}(p/z)$ are independent. Similarly, for the upper tail bound, we have

$$\begin{aligned}\Pr\left(\frac{C_m - C_0}{m} \ge \frac{p}{z}(1+\delta), \tau \ge m\right) &= \Pr(C_m \ge mp/z(1+\delta) + C_0, \tau \ge m)\\ &\le \Pr(B_m \ge mp/z(1+\delta) + C_0)\\ &\le \Pr(C_0 + Y_1 + \cdots + Y_m \ge mp/z(1+\delta) + C_0)\\ &\le \exp(-\frac{mp/z}{2+\delta}\delta^2) \le \exp(-m\delta^2/(4+2\delta))\end{aligned}$$

due to the multiplicative Chernoff bound $\Pr(Z \ge E[Z](1+\delta)) \le e^{-\frac{Z}{2+\delta}\delta^2}$ where Z is the sum of i.i.d Bernoulli random variables.

Combine upper and lower tail bounds and note that $p/z = \frac{\beta}{\beta+\gamma}$. Then, we can conclude, for any $\delta > 0$,

$$\Pr\left(\frac{C_m - C_0}{m} \notin [\frac{\beta}{\beta+\gamma}(1-\delta)z, \frac{\beta}{\beta+\gamma}(1+\delta)], \tau \ge m\right) \le 2\exp(-m\delta^2/(4+2\delta)).$$

□

EC.3.3.2. Proof of Lemma EC.3.

Proof. Conditioned on $(I_0, C_0, I_1, C_1, \ldots, I_{m-1}, C_{m-1})$ with $\tau \ge m$, we have

$$I_{k-1}T_k \sim \mathrm{Exp}\left(\beta\frac{N - C_{k-1}}{N} + \gamma\right)$$

are independent exponential random variables.

Theorem 5.1 in Janson (2018) gives us a tail bound for the sum of independent exponential random variables: let $X = \sum_{i=1}^n X_i$ with $X_i \sim \mathrm{Exp}(a_i)$ independent, then for $\delta > 0$,

$$\Pr(X \ge (1+\delta)\mu) \le \frac{1}{1+\delta}e^{-a_*\mu(\delta - \ln(1+\delta))} \le e^{-a_*\mu(\delta-\ln(1+\delta))} \tag{EC.9}$$

$$\Pr(X \le (1-\delta)\mu) \le e^{-a_*\mu(\delta-\ln(1+\delta))} \tag{EC.10}$$

where $\mu = E[X], a_* = \min_{1\le i\le n} a_i$.

Let $\tilde{S}_{m|\vec{C},\vec{I}}$ be $\tilde{S}_m$ conditioned on $(I_0, C_0, I_1, C_1, \ldots, I_{m-1}, C_{m-1})$ with $\tau \ge m$. Let $\mu = E[\tilde{S}_{m|\vec{C},\vec{I}}] = \sum_{k=1}^m \frac{1}{\beta(N-C_{k-1})/N+\gamma}$, $a_* = \min_{1\le k\le m}\beta(N - C_{k-1})/N + \gamma$. It is easy to verify the following facts

$$\mu a_* \ge \sum_{k=1}^m \frac{a_*}{(\beta+\gamma)} \ge m\frac{N - m - C_0}{N}$$

$$\frac{1}{\beta+\gamma} \le \frac{\mu}{m} \le \frac{1}{\beta+\gamma}\frac{N}{N-m-C_0}.$$

Combining these with Eqs. (EC.9) and (EC.10), we have

$$\Pr\left(\frac{\tilde{S}_{m|\vec{C},\vec{I}}}{m} \notin \left[\frac{(1-\delta)}{\beta+\gamma}, \frac{(1+\delta)}{\beta+\gamma}\frac{N}{N-m-C_0}\right]\right) \le \Pr\left(\frac{\tilde{S}_{m|\vec{C},\vec{I}}}{m} \notin \left[\frac{\mu(1-\delta)}{m}, \frac{\mu(1+\delta)}{m}\right]\right)$$
$$\le 2e^{-m\frac{N-m-C_0}{N}(\delta-\ln(1+\delta))}.$$

Therefore,

$$\begin{aligned}\Pr\left(\frac{\tilde{S}_m}{m} \notin I, \tau \ge m\right) &= \int_{\vec{C},\vec{I}|\tau\ge m} \Pr\left(\frac{\tilde{S}_m}{m} \notin I \mid \vec{C},\vec{I},\tau \ge m\right) f(\vec{C},\vec{I}|\tau \ge m)\Pr(\tau \ge m)\\ &\le 2e^{-m\frac{N-m-C_0}{N}(\delta-\ln(1+\delta))}\Pr(\tau \ge m)\\ &\le 2e^{-m\frac{N-m-C_0}{N}(\delta-\ln(1+\delta))}.\end{aligned}$$

□

EC.3.3.3. Proof of Proposition EC.1.

Proof. Let $\hat{\beta} = \frac{C_m - C_0}{\tilde{S}_m}, z = \frac{N-C_0-m}{N}$. Suppose $x \in \frac{\beta}{\beta+\gamma}[(1-\delta)z, 1+\delta], y \in \frac{1}{\beta+\gamma}[1-\delta,(1+\delta)1/z]$. Then,

$$\frac{x}{y} \in \left[\beta\frac{(1-\delta)z^2}{1+\delta}, \beta\frac{1+\delta}{1-\delta}\right] \tag{EC.11}$$

Similarly, let $\hat{\gamma} = \frac{m}{\tilde{S}_m} - \hat{\beta}$. Suppose $a \in (\beta+\gamma)[\frac{z}{1+\delta}, \frac{1}{1-\delta}], b \in \beta[\frac{(1-\delta)z^2}{1+\delta}, \frac{1+\delta}{1-\delta}]$. Then

$$a - b \in \left[\gamma\frac{z}{1+\delta} + \beta\frac{(1-\delta)z-(1+\delta)^2}{(1+\delta)(1-\delta)}, \gamma\frac{1}{1-\delta} + \beta\frac{1+\delta-(1-\delta)^2z^2}{(1-\delta)(1+\delta)}\right]. \tag{EC.12}$$

Then, for any sets U_1, U_2,

$$\begin{aligned}\Pr(\hat{\beta} \in U_1, \hat{\gamma} \in U_2) &\ge 1 - \Pr(\hat{\beta} \notin U_1) - \Pr(\hat{\gamma} \notin U_2)\\ &\ge 1 - \Pr(\hat{\beta} \notin U_1, \tau > m) - \Pr(\hat{\beta} \notin U_2, \tau > m) - 2\Pr(\tau < m)\\ &\ge 1 - 4e^{-m(\delta-\ln(1+\delta))} - 4e^{-m\delta^2/(4+2\delta)} - 2B_1e^{-B_2I_0},\end{aligned}$$

where the last step uses Lemma EC.1, Lemma EC.2 and Lemma EC.3, using the intervals (EC.11) and (EC.12) for U_1 and U_2 respectively.

□

EC.4. Proof of Proposition 6

Proof. For a given N, let $\beta = 1, \gamma = 1/2, f_0 = N, f_1 = c, f_2 = 0.$[8] We construct the following two instances: $\mathcal{M}_1 = (f_0, f_1, f_2, \beta, \gamma, N_1), \mathcal{M}_2 = (f_0, f_1, f_2, \beta, \gamma, N_2)$ where $N_1 = N + N^{2/3}, N_2 = N - N^{2/3}$.

[8] We select f_1 in a way such that $f_1 c_\infty = N$, taking into account that c_∞ is linear in N in the deterministic SIR model.

Intuitively, we need at least $m = N^{2/3}$ samples to distinguish between $\mathcal{M}_1$ and $\mathcal{M}_2$, which is precisely why a lower bound on the regret will be incurred. To be precise, let the policy π_0 be the policy that chooses to implement the drastic intervention at the beginning ($m = 0$) since any intervention after $m = 0$ will be worse. Then, we have

$$\mathrm{cost}^{\pi_0}(\mathcal{M}_1) := N, \quad \mathrm{cost}^{\pi_0}(\mathcal{M}_2) := N.$$

Let the policy π_1 be the policy that does not implement the intervention at all. By our construction, we have

$$\mathrm{cost}^{\pi_1}(\mathcal{M}_1) := N + N^{2/3}, \quad \mathrm{cost}^{\pi_1}(\mathcal{M}_2) := N - N^{2/3}.$$

The optimal cost for these two problem instances is given by

$$\begin{aligned}\mathrm{cost}^{*}(\mathcal{M}_1) &:= N\\ \mathrm{cost}^{*}(\mathcal{M}_2) &:= N - N^{2/3}.\end{aligned}$$

On the other hand, for any policy π, consider the probability of choosing to implement the drastic intervention given $m = N^{2/3}$ observations. Let

$$\begin{aligned}p_1 &:= \mathrm{Prob}(\pi(O_m) = \text{using drastic intervention}), \quad O_m \sim \mathcal{M}_1\\ p_2 &:= \mathrm{Prob}(\pi(O_m) = \text{using drastic intervention}), \quad O_m \sim \mathcal{M}_2.\end{aligned}$$

It is easy to verify that $|p_1 - p_2| \le D_{\mathrm{TV}}(O_m^1, O_m^2)$, where $O_m^1 := O_m$ with $O_m \sim \mathcal{M}_1$ and $O_m^2 := O_m$ with $O_m \sim \mathcal{M}_2$ and D_{TV} is the total variation distance. By Pinsker's inequality,

$$D_{\mathrm{TV}}(O_m^1, O_m^2)^2 \le \frac{1}{2} D_{\mathrm{KL}}(O_m^1, O_m^2).$$

Further, by the first-order approximation of KL divergence using Fisher information, we have

$$D_{\mathrm{KL}}(O_m^1, O_m^2) \lesssim (N_1 - N_2)^2 J_{O_m}(N) = (N^{2/3})^2 \frac{m^3}{N^4} = \left(\frac{1}{N^{1/3}}\right)^2.$$

This then implies

$$|p_1 - p_2| = O(1/N^{1/3}).$$

On the other hand, it is clear that in order to make the regret of $\mathcal{M}_1$ and $\mathcal{M}_2$ both less than $o(N^{2/3})$, p_1 must be close to 1 and p_2 must be close to 0. However, this contradicts the fact that $|p_1 - p_2| = O(1/N^{1/3})$. Therefore, for any policy π, we have:

$$\sup_{\mathcal{M} \in \{\mathcal{M}_1, \mathcal{M}_2\}} \mathrm{regret}^{\pi}(\mathcal{M}) = \Omega(N^{2/3}).$$

This completes the proof. □

Decision Problem with Surveillance Testing. In this section, we introduce a variant of the decision problem that incorporates optional surveillance testing. The testing incurs a cost of $c \cdot K$, where K represents the number of individuals tested, and c is a constant. Given the same setup as presented in Proposition 6, the total cost of a policy is now defined as $f_0 + f_1 C_m + c \cdot K$, where a drastic intervention is conducted at step m and surveillance testing is performed with a size of K. We establish the following lower bound on regret.

PROPOSITION EC.2. *There exists a set of problem instances $\mathcal{S}_N$ parameterized by N. For any policy π, the following result holds:*

$$\sup_{\mathcal{M} \in \mathcal{S}_N} \text{regret}^{\pi}(\mathcal{M}) = \Omega(N^{1/2}).$$

In other words, the lower bound on the regret for any policy π is $\Omega(N^{1/2})$.

Proof. The proof follows a similar framework as Proposition 6. We use the same counterexample as constructed in Proposition 6, with the only difference being $N_1 = N + N^{1/2}$ and $N_2 = N - N^{1/2}$. Suppose there exists a policy π that can achieve a regret of $o(N^{1/2})$ for both N_1 and N_2. Then the testing size of the policy, whether for N_1 or N_2, must satisfy $K = O(N^{1/2})$. Given this, we consider the following probabilities:

$$p_1 := \text{Prob(using a drastic intervention before } m = O(N^{1/2})) \text{ for } N = N_1,$$
$$p_2 := \text{Prob(using a drastic intervention before } m = O(N^{1/2})) \text{ for } N = N_2.$$

Applying Pinsker's inequality as in Proposition 6, we have:

$$|p_1 - p_2|^2 \lesssim (N_1 - N_2)^2 \left(\frac{m^3}{N^4} + \frac{Km}{N^3} \right) \lesssim (N^{1/2})^2 \frac{1}{N^2}) = O(1/N).$$

This implies that p_1 and p_2 must be close enough, which contradicts the requirement that the regret for both N_1 and N_2 should be $O(N^{1/2})$. This completes the proof. □

By way of comparison, the regret for the diffusion scenario alone, without the inclusion of surveillance testing, amounts to $\Omega(N^{2/3})$. This illustrates the positive impact of testing, reducing the regret from $N^{2/3}$ to $N^{1/2}$. Furthermore, this lower bound underscores the essential cost, specifically $N^{1/2}$, even when surveillance testing is incorporated. It underscores the challenges associated with efficiently learning all diffusion model parameters (N, β, γ) during the early stages of an epidemic.

As a side note, we also notice that the rate of convergence in Proposition 6 and EC.2 is not optimized (e.g., $|p_1 - p_2|$ is far from being $O(1)$) and can likely be further improved, however we believe the primary takeaway is evident and leave the task of optimizing these bounds for future research.

EC.5. Proofs of Propositions

EC.5.1. Proof of Proposition 1

Proof. As in Miller (2017, 2012), the solution $\{(s'(t), i'(t), r'(t)) : t \geq 0\}$ can be written as:

$$
\begin{aligned}
s'(t) &= s'(0)e^{-\xi'(t)} \\
i'(t) &= N' - s'(t) - r'(t) \\
r'(t) &= r(0) + \frac{\gamma' N'}{\beta'}\xi'(t) \\
\xi'(t) &= \frac{\beta'}{N'}\int_0^t i'(t^*)\, dt^*
\end{aligned}
$$

Making the appropriate substitutions yields the following equivalent system:

$$
i'(t) = N' - s'(0)\exp\left(-\frac{\beta'}{N'}\xi(t)\right) - r(0) - \frac{\gamma' N'}{\beta'}\xi'(t) \tag{EC.13}
$$

$$
\xi'(t) = \frac{\beta'}{N'}\int_0^t i'(t^*)\, dt^*. \tag{EC.14}
$$

Therefore, it remains to show that for $\eta > 0$, $\{(s(t), i(t), r(t)) : t \geq 0\} \triangleq \{(\eta s'(t), \eta i'(t), \eta r'(t)) : t \geq 0\}$ is a solution for (EC.13) and (EC.14) where N' is replaced with $\eta N'$. Starting with (EC.13),

$$
\begin{aligned}
i'(t) &= N' - s'(0)\exp\left(-\xi'(t)\right) - r'(0) - \frac{\gamma' N'}{\beta'}\xi'(t) \\
\eta i'(t) &= \eta\left(N' - s'(0)\exp\left(-\xi'(t)\right) - r'(0) - \frac{\gamma' N'}{\beta'}\xi'(t)\right) \\
&= \eta N' - \alpha s'(0)\exp\left(-\xi(t)\right) - \eta r(0) - \frac{\gamma' \eta N'}{\beta'}\xi(t)
\end{aligned}
$$

where $\xi(t) = \xi'(t) = \frac{\beta'}{N'\eta}\int_0^t \eta i'(t^*)\, dt^*$. Noting that $\xi'(t) = \xi(t)$ and substituting $i(t) = \eta i'(t)$ yields the equations below, clearly showing that $\{(s(t), i(t), r(t)) : t \geq 0\}$ satisfy (EC.13) and (EC.14):

$$
\begin{aligned}
i(t) &= \eta N' - s(0)\exp\left(-\xi(t)\right) - r(0) - \frac{\gamma' \eta N'}{\beta'}\xi(t) \\
\xi(t) &= \frac{\beta'}{N'\eta}\int_0^t i(t^*)\, dt^*.
\end{aligned}
$$

$\square$

EC.5.2. Proof of Proposition 2

EC.5.2.1. SIR Model

Proof. Consider initial conditions $(s(0), i(0), 0)$, as in Miller (2017, 2012), the analytical solution is given by

$$
\begin{aligned}
s(t) &= s(0)e^{-\xi(t)}, \\
i(t) &= N - s(t) - r(t), \\
r(t) &= \frac{\gamma N}{\beta}\xi(t), \\
\xi(t) &= \frac{\beta}{N}\int_0^t i(t')dt'.
\end{aligned}
$$

Consider two SIR models with parameters (N, β, γ) and (N', β', γ'), and initial conditions $(s_0, i_0, 0)$ and $(s_0', i_0', 0)$ respectively. We claim that infection trajectories $i(t)$ and $i'(t)$ being identical on an open set $[0, T)$ implies the parameters and initial conditions are identical as well.

Assume $i(t) = i'(t)$ for all $t \in [0, T)$; then, given the exact solution above it follows that

$$
N - s_0 e^{-\frac{\beta}{N}x} - \gamma x = N' - s_0' e^{-\frac{\beta'}{N'}x} - \gamma' x, \qquad \text{for all } x \in \left[0, \int_0^T i(t)dt\right]
$$

As functions of x, both the RHS and LHS in the equality above are holomorphic, and hence, using the identity theorem, we then have for all $x \in \mathbb{R}$, there is $N - s_0 e^{-\frac{\beta}{N}x} - \gamma x = N' - s_0' e^{-\frac{\beta'}{N'}x} - \gamma' x$.

Then the following implies $\gamma = \gamma'$:

$$
-\gamma = \lim_{x \to +\infty} \frac{N - s_0 e^{-\frac{\beta}{N}x} - \gamma x}{x} = \lim_{x \to +\infty} = \frac{N' - s_0' e^{-\frac{\beta'}{N'}x} - \gamma' x}{x} = -\gamma'.
$$

Hence for all $x \in \mathbb{R}$, $N - s_0 e^{-\frac{\beta}{N}x} = N' - s_0' e^{-\frac{\beta'}{N'}x}$. Again, by taking x to infinity, we can conclude $N = N'$ by the following

$$
N = \lim_{x \to +\infty} \left(N - s_0 e^{-\frac{\beta}{N}x}\right) = \lim_{x \to +\infty} \left(N' - s_0' e^{-\frac{\beta'}{N'}x}\right) = N'.
$$

Furthermore, by taking $x = 0$, we can also get $s_0 = s_0'$ and then $\beta = \beta'$ follows. This completes the proof. □

EC.5.2.2. Bass Model

Proof. Consider the initial condition $i(0) = 0$. By the analytic solution given by Bass (1969), we have

$$
i(t) = N \frac{1 - e^{-(p+\beta)t}}{\frac{\beta}{p}e^{-(p+\beta)t} + 1}.
$$

Consider two bass models with parameters (N, β, p) and (N', β', p') and initial conditions $i(0) = 0, i'(0) = 0$ respectively. We claim that trajectories $i(t)$ and $i'(t)$ being identical on an open set $[0, T)$ implies the parameters are identical as well.

Assume $i(t) = i'(t)$ for all $t \in [0,T)$; then, given the exact solution above it follows that

$$N\frac{1-e^{-(p+\beta)t}}{\frac{\beta}{p}e^{-(p+\beta)t}+1} = N'\frac{1-e^{-(p'+\beta')t}}{\frac{\beta'}{p'}e^{-(p'+\beta')t}+1}, \qquad \text{for all } t \in [0,T) \tag{EC.15}$$

As functions of t, both the RHS and LHS in the equality above are holomorphic, and hence, using the identity theorem, we then have Eq. (EC.15) holds for all $t \in \mathbb{R}$.

By taking t to infinity, we can easily obtain $N = N'$. Furthermore, taking the derivative for t on both sides of Eq. (EC.15), one can obtain

$$\frac{(p+\beta)^2}{p}\frac{e^{-(p+\beta)t}}{(\beta/p\cdot e^{-(p+\beta)t}+1)^2} = \frac{(p'+\beta')^2}{p'}\frac{e^{-(p'+\beta')t}}{(\beta'/p'\cdot e^{-(p'+\beta')t}+1)^2}. \tag{EC.16}$$

By taking $t=0$ on both sides of Eq. (EC.16), one can verify that $p = p'$. Furthermore, let $g(t) = \frac{(p+\beta)^2}{p}\frac{e^{-(p+\beta)t}}{(\beta/p\cdot e^{-(p+\beta)t}+1)^2}$ and $g'(t) = \frac{(p'+\beta')^2}{p'}\frac{e^{-(p'+\beta')t}}{(\beta'/p'\cdot e^{-(p'+\beta')t}+1)^2}$.

Note that

$$-(p+\beta) = \lim_{t\to+\infty}\frac{\ln(g(t))}{t} = \lim_{t\to+\infty}\frac{\ln(g'(t))}{t} = -(p'+\beta').$$

We then can conclude $\beta = \beta'$. This completes the proof.

□

EC.5.3. Proof of Proposition 3

Proof. Note that $\mathbb{E}[T_i] = \frac{N}{pN(N-i)+\beta i(N-i)}$. Then

$$\mathbb{E}[t_{k^{\mathrm{CR}}}] = \mathbb{E}\left[\sum_{i=1}^{N^{2/3}-1} T_i\right] = \sum_{i=1}^{N^{2/3}-1}\frac{N}{pN(N-i)+\beta i(N-i)}.$$

Let $f(x) = \frac{N}{pN(N-x)+\beta x(N-x)}$, we use $f(x)$ as a proxy to bound $\mathbb{E}[t_{k^{\mathrm{CR}}}]$. Easy to verify that $f(x)$ is decreasing when $x \in (0, \mathring{r}]$ where $\mathring{r} = (1-p/\beta)N/2$. Note that $p/\beta < c$ for some constant c since $p/\beta = \Theta(N^{-\alpha})$ for $\alpha > 0$. Hence when $N \to \infty$, we have $\mathring{r} \gg N^{2/3}$ and

$$\begin{aligned}
\sum_{i=1}^{N^{2/3}-1}\frac{N}{pN(N-i)+\beta i(N-i)} &\geq \int_{x=1}^{N^{2/3}} f(x)dx \\
&= \frac{\ln(\beta x+Np)-\ln(N-x)}{p+\beta}\Big|_{x=1}^{N^{2/3}} \\
&= \frac{\ln(\beta N^{2/3}+Np)-\ln(\beta+Np)+\ln(N-1)-\ln(N-N^{2/3})}{p+\beta} \\
&\geq \frac{\ln(\beta N^{2/3}+Np)-\ln(\beta+Np)}{p+\beta}.
\end{aligned}$$

Similarly, for t_{k^*}, we have

$$\begin{aligned}
E[t_{k^*}] &= \sum_{i=1}^{\mathring{r}-1} \frac{N}{pN(N-i)+\beta i(N-i)} \\
&\le f(1) + \int_{x=1}^{\mathring{r}} f(x)dx \\
&\le f(1) + \frac{\ln(\beta N + Np) - \ln(pN+\beta) + \ln\frac{1}{1-c}}{p+\beta} \\
&\le \frac{\ln(\beta N + Np) - \ln(pN+\beta) + c'}{p+\beta}
\end{aligned}$$

for some absolute constant c'.

Let $\frac{\beta}{p} = C\cdot N^\alpha$ for some constant C. We then have

$$\begin{aligned}
\frac{E[t_{k^{\mathrm{CR}}}]}{E[t_{k^*}]} &\ge \frac{\ln(\beta N^{2/3}+Np) - \ln(\beta+Np)}{\ln(\beta N + pN) - \ln(pN+\beta)+c'} \\
&\ge \frac{\ln\left(\frac{CN^{2/3+\alpha}+N}{CN^\alpha+N}\right)}{\ln\left(\frac{CN^{1+\alpha}+N}{CN^\alpha+N}\right)+c'} =: k_N.
\end{aligned}$$

Then, it is easy to verify that when $\frac{1}{3} < \alpha \le 1$, $\lim_{N\to\infty} k_N = \frac{\alpha-1/3}{\alpha}$. When $\alpha > 1$, $\lim_{N\to\infty} k_N = \frac{2}{3}$.

Note that we also have $\mathbb{E}[t_{k^{\mathrm{CR}}}] \le f(1) + \int_{x=1}^{N^{2/3}} f(x)dx$ and $\mathbb{E}[t_{k^*}] \ge \int_{x=1}^{\mathring{r}} f(x)dx$. Similarly, one can verify that

$$\limsup_{N\to\infty} \frac{\mathbb{E}[t_{k^{\mathrm{CR}}}]}{\mathbb{E}[t_{k^*}]} \le \begin{cases} 0 & \alpha \le \frac{1}{3} \\ \frac{\alpha-\frac{1}{3}}{\alpha} & \frac{1}{3} < \alpha \le 1 \\ \frac{2}{3} & \alpha > 1 \end{cases}.$$

This completes the proof. □

EC.5.4. Proof of Proposition 4

Let $t_*^d = \inf\{t : \beta(s)t/N < \gamma\}$ be the time when the number of infections is at its peak. It is easy to show that $t_2^d \le t_*^d$. We show the analog of Proposition 4 with the peak defined instead as t_*^d — i.e. we show $\liminf_{N\to\infty} \frac{t_{\mathrm{CR}}^d}{t_*^d} \ge \frac{2}{3}$. Then, the desired result follows since $t_2^d \le t_*^d$.

First, we prove $t_2^d \le t_*^d$. We can write $\frac{d^2s}{dt^2}$ as

$$\begin{aligned}
\frac{d^2 s}{dt^2} &= \frac{-\beta}{N}\left(\frac{ds}{dt}i + \frac{di}{dt}s\right) \\
&= \frac{-\beta}{N}\left(\frac{-\beta s}{N}i^2 + \left(\frac{\beta s}{N}-\gamma\right)is\right) \\
&= \frac{\beta^2 is}{N^2}\left(i - s + \frac{\gamma}{\beta}N\right).
\end{aligned} \tag{EC.17}$$

From (EC.17), we see that $\frac{d^2s}{dt^2} > 0$ if and only if

$$s < \frac{\gamma}{\beta}N + i.$$

By definition, t_*^d occurs at a time when

$$s < \frac{\gamma}{\beta} N.$$

Since s is decreasing and i is non-negative, clearly t_2^d occurs before t_*^d.

Next, we prove $\liminf_{N\to\infty} \frac{t_{\mathrm{CR}}^d}{t_*^d} \geq \frac{2}{3}$. The crux of the problem is summarised in two smaller results, bounding t_{CR}^d and t_*^d respectively. Let $\rho_1 = 1 - \frac{1}{\log\log N}$ and $\rho_2 = \frac{\gamma}{\beta}$.

Proposition EC.3. *There exists a constant ν_1 that only depends on γ, β such that*

$$t_{\mathrm{CR}}^d \geq \frac{1}{\beta-\gamma}\left(\frac{2}{3}\log\frac{\nu_1 N}{c(0)^{3/2}} + \log\frac{\nu_1^{2/3}}{c(0)}\left(1 - \frac{c(0)}{N^{2/3}}\right)\right).$$

Proposition EC.4. *There exists a constant ν_2 that only depends on γ, β and a constant $C = O(1)$, such that*

$$t_*^d \leq \frac{1}{\beta\rho_1 - \gamma}\log\frac{\nu_2 N}{i(0)} + \frac{C}{1-\rho_1}.$$

The argument follows directly by taking the limit of the bounds we provide in Propositions EC.3-EC.4. Specifically, using that the constants ν_1, ν_2 do not depend on N, we arrive at

$$\begin{aligned}
\limsup_{N\to\infty} \frac{t_*^d}{t_{\mathrm{CR}}^d} &\leq \limsup_{N\to\infty} \frac{\frac{1}{\beta\rho_1-\gamma}\log\frac{\nu_2 N}{i(0)} + \frac{C}{(1-\rho_1)}}{\frac{1}{\beta-\gamma}\left(\frac{2}{3}\log\frac{\nu_1 N}{c(0)^{3/2}} + \log\frac{\nu_1^{2/3}}{c(0)}\left(1-\frac{c(0)}{N^{2/3}}\right)\right)} \\
&= \limsup_{N\to\infty} \frac{\beta-\gamma}{\beta\rho_1-\gamma} \cdot \frac{\log N + \log\nu_2 - \log i(0)}{\frac{2}{3}\log N + \frac{4}{3}\log\nu_1 - 2\log c(0) + \log\left(1-\frac{c(0)}{N^{2/3}}\right)} \\
&\quad + \limsup_{N\to\infty} \frac{(\beta-\gamma)C\log\log N}{\frac{2}{3}\log N + \frac{4}{3}\log\nu_1 - 2\log c(0) + \log\left(1-\frac{c(0)}{N^{2/3}}\right)}
\end{aligned}$$

$\rho_1 \to 1$ as $N\to\infty$, so $\frac{\beta-\gamma}{\beta\rho_1-\gamma} \to 1$. Since $c(0) = O(\log(N))$ by assumption (and $i(0) \leq c(0)$), and $C = O(1)$ by Proposition EC.4, the limits of the two summands above are $3/2$ and 0 respectively, which concludes the proof.

EC.5.4.1. Proof of Proposition EC.3.

Proof of Proposition EC.3. Define $\tilde{i}(t)$ such that $\tilde{i}(0) = i(0)$ and $\frac{d\tilde{i}}{dt} = (\beta-\gamma)\tilde{i}$, implying

$$\tilde{i}(t) = i(0)\exp\{(\beta-\gamma)t\}.$$

Since $\frac{d\tilde{i}}{dt} \geq \frac{di}{dt}$ for all t, $\tilde{i}(t) \geq i(t)$ for all t. Then, for all t,

$$\frac{ds}{dt} = -\beta\frac{s}{N}i \geq -\beta i \geq -\beta\tilde{i}.$$

Hence we can write

$$\begin{aligned} s(t) &\geq s(0) + \int_0^t -\beta \tilde{i}(t')dt' \\ &= s(0) - \beta i(0) \int_0^t \exp\{(\beta-\gamma)t'\}dt' \\ &= s(0) - \frac{\beta i(0)}{\beta-\gamma}(\exp\{(\beta-\gamma)t\} - 1) \end{aligned}$$

Since $s(0) - s(t^d_{\mathrm{CR}}) = N^{2/3} - c(0)$, setting $t = t^d_{\mathrm{CR}}$ and solving for t^d_{CR} in the inequality above results in

$$\begin{aligned} t^d_{\mathrm{CR}} &\geq \frac{1}{\beta-\gamma} \log\left(\frac{\beta-\gamma}{\beta i(0)}(N^{2/3} - c(0))\right) \\ &\geq \frac{1}{\beta-\gamma} \log\left(\frac{\beta-\gamma}{\beta c(0)}(N^{2/3} - c(0))\right) \\ &= \frac{1}{\beta-\gamma} \left(\log \frac{\beta-\gamma}{\beta c(0)}(N^{2/3}) + \log \frac{\beta-\gamma}{\beta c(0)}\left(1 - \frac{c(0)}{N^{2/3}}\right)\right) \\ &= \frac{1}{\beta-\gamma} \left(\frac{2}{3}\log \frac{\nu_1 N}{c(0)^{3/2}} + \log \frac{\nu_1^{2/3}}{c(0)}\left(1 - \frac{c(0)}{N^{2/3}}\right)\right) \end{aligned}$$

for $\nu_1 = \left(\frac{\beta-\gamma}{\beta}\right)^{3/2}$ as desired. □

EC.5.4.2. Proof of Proposition EC.4. For $\rho \in [0, \frac{\gamma}{\beta}]$, let t_ρ be the time t when $\frac{s(t)}{N} = \rho$. ρ will represent the fraction of the total population that is susceptible. Since $\rho \leq \frac{\gamma}{\beta}$, i is increasing for the time period of interest.

Let $\beta > \gamma$, N be fixed. Let $\rho_1 = 1 - \frac{1}{\log\log N}$ and $\rho_2 = \frac{\gamma}{\beta}$. We assume N is large enough that $\rho_1 > \rho_2$, hence $t_{\rho_1} < t_{\rho_2}$. $t^d_* = t_{\rho_2}$.

Lemma EC.4. *For any* $\rho \in [0, \frac{\gamma}{\beta}]$, $i(t_\rho) \geq N(1-\rho)\frac{\beta\rho-\gamma}{\beta\rho} - \frac{c(0)}{2}$.

Proof of Lemma EC.4. Fix ρ. At time t_ρ, the total number of people infected is $c(t_\rho) = i(t_\rho) + r(t_\rho) = N(1-\rho)$, by definition. At any time $t \leq t_\rho$, the rate of increase in i is $\frac{\beta\frac{s(t)}{N}-\gamma}{\beta\frac{s(t)}{N}} \geq \frac{\beta\rho-\gamma}{\beta\rho}$ of the rate of increase in c. Therefore, $i(t_\rho) - i(0) \geq \left(\frac{\beta\rho-\gamma}{\beta\rho}\right)\left(c(t_\rho) - c(0)\right)$ and $i(t_\rho) \geq \left(\frac{\beta\rho-\gamma}{\beta\rho}\right)N(1-\rho) - \frac{\beta\rho-\gamma}{\beta\rho}c(0) + i(0)$. Using the fact that $i(0) \geq \frac{c(0)}{2}$ and rearranging terms gives the desired result. □

Lemma EC.5. *For* $t \in [t_{\rho_1}, t_{\rho_2}]$, *where* $\rho_2 > \rho_1$ *for* $\rho_1, \rho_2 \in [0, \frac{\gamma}{\beta}]$, $t_{\rho_2} - t_{\rho_1} \leq \frac{N(\rho_1-\rho_2)}{\beta\rho_2 i(t_{\rho_1})}$.

Proof of Lemma EC.5. The difference in s between t_{ρ_1} and t_{ρ_2} is $s(t_{\rho_1}) - s(t_{\rho_2}) = N(\rho_1 - \rho_2)$. As a consequence of the mean value theorem, $\frac{s(t_{\rho_2})-s(t_{\rho_1})}{t_{\rho_2}-t_{\rho_1}} \leq \max_{t\in[t_{\rho_1},t_{\rho_2}]}\{\frac{ds}{dt}\}$. Using these two expressions,

$$\frac{N(\rho_1-\rho_2)}{t_{\rho_2}-t_{\rho_1}} \geq \min\left\{-\frac{ds}{dt}\right\} = \min\left\{\beta\frac{s(t)}{N}i(t) : t \in [t_{\rho_1}, t_{\rho_2}]\right\} \geq \beta\rho_2 i(t_{\rho_1})$$

The desired expression follows from rearranging terms. □

LEMMA EC.6. *For any* $\rho \le \min\{\frac{\gamma}{\beta}, 1/2\}$, $t_\rho \le \frac{1}{\beta\rho-\gamma}\log\frac{\nu_2}{i(0)}N$, *for* $\nu_2 = \frac{2(\beta-\gamma)}{\beta}$.

The proof of this lemma follows the exact same procedure as the proof of Proposition EC.3.

Proof of Lemma EC.6. We proceed in the same way as the proof of Proposition EC.3 except in this case we will lower bound $s(0)-s(t)$. We achieve this by letting $\tilde{i}$ be defined to grow slower than i, so it is used as a lower bound. Define $\tilde{i}(t)$ such that $\tilde{i}(0)=i(0)$ and $\frac{d\tilde{i}}{dt} = (\beta\rho-\gamma)\tilde{i}$, implying

$$\tilde{i}(t) = i(0)\exp\{(\beta\rho-\gamma)t\}.$$

Since $\frac{d\tilde{i}}{dt} \le \frac{di}{dt}$ when , $\tilde{i}(t)\le i(t)$ for all $t<t_{\rho_2}$. In addition, when $t<t_{\rho_2}$, $\frac{s}{N}\ge\frac{\gamma}{\beta}\ge\rho$. Then, for $t<t_{\rho_2}$,

$$\frac{ds}{dt} = -\beta\frac{s}{N}i \le -\beta\rho\tilde{i}.$$

Hence we can write

$$\begin{aligned}
s(t) &\le s(0) + \int_0^t -\beta\rho\tilde{i}(t')dt' \\
&= s(0) - \beta\rho i(0)\int_0^t \exp\{(\beta\rho-\gamma)t'\}dt' \\
&= s(0) - \frac{\beta\rho i(0)}{\beta\rho-\gamma}(\exp\{(\beta\rho-\gamma)t\}-1)
\end{aligned}$$

Since $s(t_\rho)=\rho N$,

$$\rho N \le s(0) - \frac{\beta\rho i(0)}{\beta\rho-\gamma}(\exp\{(\beta\rho-\gamma)t_\rho\}-1).$$

Solving for t_ρ results in

$$t_\rho \le \frac{\log\left(\frac{\beta\rho-\gamma}{\beta\rho i(0)}(s(0)-\rho N)+1\right)}{\beta\rho-\gamma} \le \frac{1}{\beta\rho-\gamma}\log\left(\frac{\nu_2}{i(0)}N\right)$$

where $\nu_2 = \frac{2(\beta-\gamma)}{\beta}$, using the fact that $\rho\le 1/2$. □

Proof of Proposition EC.4. Using the results from Lemmas EC.4-EC.6,

$$\begin{aligned}
t_{\rho_2} &= t_{\rho_1} + (t_{\rho_2}-t_{\rho_1}) \\
&\le \frac{1}{\beta\rho_1-\gamma}\log\left(\frac{\nu_2}{i(0)}N\right) + \frac{N(\rho_1-\rho_2)}{\beta\rho_2 i(t_{\rho_1})} \\
&\le \frac{1}{\beta\rho_1-\gamma}\log\left(\frac{\nu_2}{i(0)}N\right) + \frac{N(\rho_1-\rho_2)}{\frac{\rho_2}{\rho_1}N(1-\rho_1)(\beta\rho_1-\gamma)-\frac{\beta\rho_2}{2}c(0)} \\
&= \frac{1}{\beta\rho_1-\gamma}\log\left(\frac{\nu_2}{i(0)}N\right) + \frac{C}{1-\rho_1},
\end{aligned}$$

where $C = \frac{\rho_1-\rho_2}{\frac{\rho_2}{\rho_1}(\beta\rho_1-\gamma)-\frac{\beta\rho_2}{2}\frac{c(0)}{N(1-\rho_1)}}$. Note that, as required in the statement, $C = O(1)$. Indeed,

$$C = \frac{(\rho_1-\rho_2)}{\frac{\rho_2}{\rho_1}(\beta\rho_1-\gamma)-\frac{\beta\rho_2}{2}\frac{c(0)}{N(1-\rho_1)}} = \frac{1-\rho_2-\frac{1}{\log\log N}}{\beta\rho_2-\frac{\gamma\rho_2}{1-\frac{1}{\log\log N}}-\frac{\beta\rho_2}{2}\frac{c(0)\log\log N}{N}},$$

and so, as N grows large, C tends to $(1-\rho_2)/\rho_2(\beta-\gamma)$ (recall that $c(0) = O(\log\log N)$).

□

EC.5.5. Proof of Proposition 7

Note that, conditioned on N, O_m and $\tilde{O}_m$ are independent. Thus,

$$J_{O_m\cup\tilde{O}_m}(N) = J_{O_m}(N) + J_{\tilde{O}_m}(N).$$

Note that $J_{O_m}(N) = \Theta\left(\frac{m^3}{N^4}\right)$ has been calculated in the main theorem. It is sufficient to consider $J_{\tilde{O}_m}(N)$, which is

$$J_{\tilde{O}_m}(N) = KJ_{\mathrm{Ber}(\kappa_m)}(N)$$

since X_k are independent from each other. Note that for any function $\eta(N)$, we have

$$J_{\mathrm{Ber}(\eta)}(\eta(N)) = \frac{1}{\eta(1-\eta)}\left(\frac{d\eta(N)}{dN}\right)^2.$$

Using $\eta(N) = E[C_m]/N$, we have

$$\begin{aligned} J_{\mathrm{Ber}(\eta)}(\eta(N)) &= \frac{N^2}{E[C_m](N-E[C_m])}\frac{E[C_m]^2}{N^4} \\ &= \frac{E[C_m]}{N^2(N-E[C_m])}. \end{aligned}$$

Note that $E[C_m] = \Theta(m)$ and $m = o(N)$. Therefore,

$$J_{\tilde{O}_m}(N) = K\cdot J_{\mathrm{Ber}(\kappa_m)}(N) = \Theta\left(\frac{Km}{N^3}\right)$$

which completes the proof.

EC.6. Datasets

Here we provide details on the datasets used in Section 5.

EC.6.1. Amazon product reviews

For the Bass model, we use the Amazon product dataset of Ni et al. (2019), which contains product reviews for Amazon products over more than twenty years. We take these reviews as a proxy for sales. Products in Amazon's electronics category typically have review trajectories well-approximated by the Bass model, marked by slow initial adoption and a long tail of sales towards the end of the product lifecycle – see Figure EC.1 for examples of such trajectories. For our experiments, we randomly selected 100 products with over four years of reviews, and over 100 reviews by the fourth year. Review counts are taken at a weekly granularity. Here we use $N_{\max} = 1e5$ – an order of magnitude larger than any of the true product sales numbers in the dataset.

Figure EC.2 below shows the cumulative distribution (over products) of the values of N estimated using this data set.

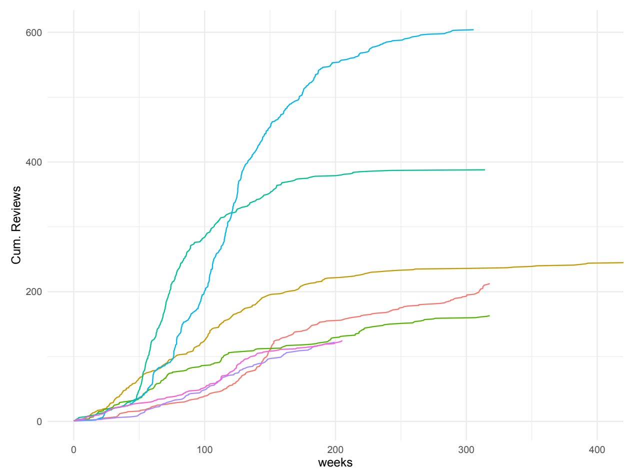

Figure EC.1 Cumulative weekly product reviews for randomly selected products from our subset of the Amazon dataset.

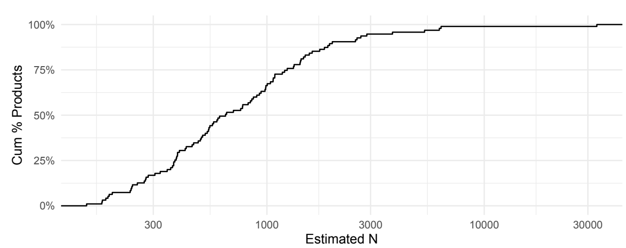

Figure EC.2 Cumulative distribution of estimated N over products in the Amazon data set. The mean is 1357.

EC.6.2. CDC ILINet influenza database

For the SIR model, we use the CDC's ILINet database of patient visits for flu-like illnesses in the United States, broken down by Department of Health and Human Services region. Each instance in the dataset consists of weekly patient visits in a given region, over the course of one year. Each year starts in September, at the low point of the flu season. We use data from 2010 through 2019 for each of 10 regions, for 100 instances total. As the dataset only includes cumulative infections $C_i[t]$, rather than observations of infection and recoveries $I_i[t], R_i[t]$, we simulate these based on the dynamics (13).

Here, we take $\gamma = 0.24$ as in Chowell et al. (2008), and a is assumed to be 0. We take $N_{\max}$ to be the total patient population (including for non-flu illnesses) in the dataset.

Figure EC.3 below shows the cumulative distribution (over region-years) of the values of N estimated using this data set.

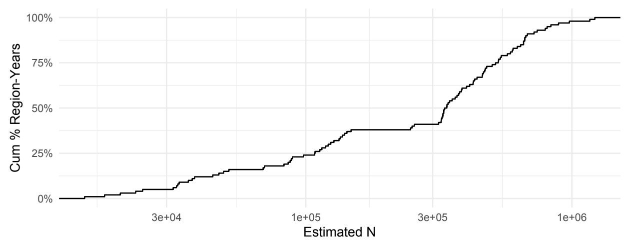

Figure EC.3 Cumulative distribution of the estimated N over region-years in the ILINet dataset. The mean is 349,314.

EC.6.3. COVID-19 Datasets

For observed COVID-19 cases, we use publicly available case data from the ongoing COVID-19 epidemic provided by Dong et al. (2020). We aggregate data into sub-state regions, corresponding broadly to public health service areas. The median state has seven regions. Here we take $\gamma = 1/4$.

The dataset contains static demographic covariates and time-varying mobility features that affect the disease transmission rate. The dynamic covariates proxy mobility by estimating the daily fraction of people staying at home relative to a region-specific benchmark of activity in early March before social distancing measures were put in place. We also include a regional binary indicator of the days when the fraction of people staying home exceeds the benchmark by 0.2 or more.

These data are provided by Safegraph, a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places. To enhance privacy, SafeGraph excludes census block group information if fewer than five devices visited an establishment in a month from a given census block group. Documentation can be found at Saf (2020).

The static covariates capture standard demographic features of a region that influence variation in infection rates. These features fall into several categories:

- Fraction of individuals that live in close proximity or provide personal care to relatives in other generations. These covariates are reported by age group by state from survey responses conducted by UMi (2020).
- Family size from U.S. Census data, aggregated and cleaned by Cla (2020).
- Fraction of the population living in group quarters, including colleges, group homes, military quarters, and nursing homes (U.S. Census via Cla (2020)).

- Population-weighted urban status (US Census via Cla (2020))
- Prevalence of comorbidities, such as cardiovascular disease and hypertension (CDC (2020a))
- Measures of social vulnerability and poverty (U.S. Census via Cla (2020); CDC (2020b))
- Age, race and occupation distributions (U.S. Census via Cla (2020))

EC.7. Detailed description of the COVID-19 model

EC.7.1. Approximating the arrival process with latent state

Recall the stochastic SIR process, $(S(t), I(t), R(t)) : t \geq 0$, a multi-variate counting process determined by parameters (N, β, γ). We now allow β to be time-varying, yielding a counting process with jumps $C_k - C_{k-1} \sim \text{Bern}\{\beta S_{k-1}/(\beta_k S_{k-1} + \gamma N I(t))\}$.

We obtain discrete-time diffusion processes, $\{(S_i[t], I_i[t], R_i[t]) : t \in \mathbb{N}\}$ for instances $i \in \mathcal{I}$ by considering the Euler-approximation to the stochastic diffusion process (3) (e.g. Jacod et al. (2005)). Specifically, let $\Delta I[t] = I[t] - I[t-1]$, and define $\Delta S[t]$ and $\Delta R[t]$ analogously. A discrete-time approximation to the SIR process is then given by:

$$\begin{aligned}
\Delta S_i[t+1] &= -\beta_i[t](S_i[t]/N_i)I_i[t] + \nu_{i,t}^S \\
\Delta I_i[t+1] &= \beta_i[t](S_i[t]/N_i)I_i[t] - \gamma I_i[t] + \nu_{i,t}^I \\
\Delta R_i[t+1] &= \gamma I_i[t] + \nu_{i,t}^R
\end{aligned} \tag{EC.18}$$

where $\{\nu_{i,t}^S\}, \{\nu_{i,t}^I\}, \{\nu_{i,t}^R\}$ are appropriately defined martingale difference sequences.

In the real world, the SIR model is a latent process – we never directly observe any of the state variables $S_i[t], I_i[t], R_i[t]$. Instead, we observe $C_i[t] = I_i[t] + R_i[t] = N_i - S_i[t]$. The MLE problem for parameters (N, β) is simply $\max_{(\beta,N)} \sum_{i,t} \log \mathbb{P}(C_i[t] | \beta, N)$.

This is a difficult non-linear filtering problem (and an interesting direction for research). We therefore consider an approximation: Denote by $\{(s_i[t], i_i[t], r_i[t]) : t \in \mathbb{N}\}$ the deterministic process obtained by ignoring the martingale difference terms in the definition of the discrete time SIR process. We consider the approximation $C_i[t] = N_i - S_i[t] \sim (N_i - s_i[t])\omega_i[t]$, where $\omega_i[t]$ is log-normally distributed with mean 1 and variance $\exp(\sigma^2) - 1$.

Under this approximation, we have the log likelihood function

$$\log p(C_i[t] | N, \beta) = (\log C_i[t] - \log(N_i - s_i[t]))^2 \tag{EC.19}$$

EC.7.2. Two-Stage Estimation of the SIR model

We parameterize our estimates of N as $\hat{N}_i(\phi, \delta) = \exp(\phi^\top Z_i + \delta_i)P_i$, where Z_i are non-time-varying, region-specific covariates, P_i is the population of region i, ϕ is a vector of fixed effects, and $\delta_i \sim \mathcal{N}(0, \sigma_\delta^2)$ are region-specific random effects.

Demographic and mobility factors also influence the reproduction rate of the disease. To model these effects, we estimate $\beta_i[t]$ as a mixed effects model incorporating covariates $\beta_i[t] = \exp(X_i[t]^\top \theta) + \epsilon_i$, where θ is a vector of fixed effects, and $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is a vector of random effects.

Given observations up to time T, we then estimate the model parameters $(\theta, \phi, \delta, \epsilon)$ in two stages:

1. Estimate the peak parameters $\hat{\phi}, \hat{\delta}$ via MLE, for the regions $i \in Q[t]$:

$$\hat{\phi}, \hat{\delta} = \arg\max_{\phi,\delta} \left\{ \max_{\theta,\epsilon} \left\{ \sum_{i\in Q[t]} \sum_{t\in[T]} \log p\left(C_i[t] \,\middle|\, \beta_i(\theta,\epsilon), \hat{N}_i(\phi,\delta)\right) + \log p(\epsilon,\delta) \right\} \right\}$$

where p is the likelihood defined in (EC.19). We let $\hat{\delta}_i = 0$ for $i \notin Q[t]$.

2. Estimate the remaining parameters over all regions $i \in \mathcal{I}$:

$$\hat{\theta}, \hat{\epsilon} = \arg\max_{\theta,\epsilon} \left\{ \sum_{i\in\mathcal{I}} \sum_{t\in[T]} \log p\left(C_i[t] \,\middle|\, \beta_i(\theta,\epsilon), \hat{N}_i(\hat{\phi},\hat{\delta}),\right) + \log p(\epsilon,\delta) \right\} \tag{EC.20}$$

We note that (EC.20) is differentiable with respect to the parameters $(\theta, \epsilon, \phi, \delta)$, and we solve it (or a weighted version) using Adam (Kingma and Ba 2014).[9]

To identify the set $Q[t]$ of regions for which the variance of $\hat{N}$ may be small, we simply look for regions that have passed their peak rate of new infections. Concretely, we define $Q[t]$ as:

$$Q[t] = \{i \in \mathcal{I} : C_i[t] - C_i[t-1] \le \gamma_1 \max_{\tau\le t} (C_i[\tau] - C_i[\tau-1])\}, \tag{EC.21}$$

where $\gamma_1 \in (0,1)$ is a hyperparameter.

EC.7.3. Performance relative to other models

To contextualize the quality of the Two-Stage model, we compare our analyzed models to the widely used IHME model ihm (2020). We note that there exist comparable models that may serve as stronger baselines; we include these results merely to demonstrate that the Two-Stage model yields high-quality predictions, comparable to widely-cited models in the literature.

Figure EC.4 compares state-level[10] WMAPE for MLE, Two-Stage and IHME models, for vintages stretching back 28 days. The IHME model up to this date is, in effect, an SI model with carefully tuned parameters. We report published IHME forecasts; 10 vintages of that model were reported between April 21 and May 21. *Two Stage* dominates IHME across all model vintages.

[9] Adam was run for 20k iterations, with learning rate tuned over a coarse grid. A weighted version of the loss function in (EC.20) with weights for (i,t)th observation set to $C_i[t]$ worked well.

[10] Due to IHME only providing state-level predictions. Additionally IHME only offers deaths predictions for these vintages; we show WMAPE on deaths for IHME and WMAPE on infections for MLE and Two-Stage.

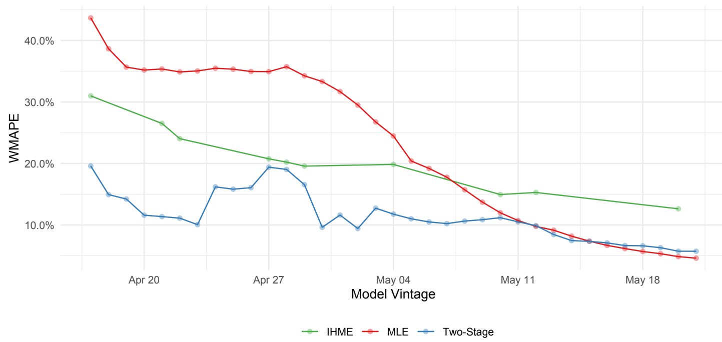

Figure EC.4 WMAPE for predicting state-level cumulative cases on May 21, 2020, comparing MLE and the Two-Stage approach against IHME.

References

(2020a) *CDC Interactive Atlas of Heart Disease and Stroke.* https://www.cdc.gov/dhdsp/maps/atlas/index.htm.

(2020b) *CDC Social Vulnerability Index.* https://svi.cdc.gov/.

(2020) *Claritas.* https://www.claritas.com/.

(2020) *IHME.* http://www.healthdata.org/sites/default/files/files/Projects/COVID/RA_COVID-forecasting-USA-EEA_042120.pdf.

(2020) *Safegraph Social Distancing Metrics.* https://docs.safegraph.com/docs/social-distancing-metrics.

(2020) *University of Michigan Health and Retirement Study.* https://hrs.isr.umich.edu/data-products.

Bass FM (1969) A new product growth for model consumer durables. *Management science* 15(5):215–227.

Chowell G, Miller M, Viboud C (2008) Seasonal influenza in the united states, france, and australia: transmission and prospects for control. *Epidemiology & Infection* 136(6):852–864.

Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases* 20(5):533–534.

Jacod J, Kurtz TG, Méléard S, Protter P (2005) The approximate euler method for lévy driven stochastic differential equations. *Annales de l'IHP Probabilités et statistiques*, volume 41, 523–558.

Janson S (2018) Tail bounds for sums of geometric and exponential variables. *Statistics & Probability Letters* 135:1–6.

Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

Miller JC (2012) A note on the derivation of epidemic final sizes. *Bulletin of mathematical biology* 74(9):2125–2141.

Miller JC (2017) Mathematical models of sir disease spread with combined non-sexual and sexual transmission routes. *Infectious Disease Modelling* 2(1):35–55.

Ni J, Li J, McAuley J (2019) Justifying recommendations using distantly-labeled reviews and fine-grained aspects. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 188–197.