Inferring Density-Dependent Population Dynamics Mechanisms through Rate Disambiguation for Logistic Birth-Death Processes

Linh Huynh[1,5,*], **Jacob G. Scott**[2,3], **and Peter J. Thomas**[1,4]

[1]Department of Mathematics, Applied Mathematics, and Statistics, Case Western Reserve University, Cleveland, OH, 44106 USA
[2]Department of Translational Hematology and Oncology Research, Cleveland Clinic, Cleveland, OH, 44106 USA
[3]Department of Systems Biology and Bioinformatics, School of Medicine, Case Western Reserve University, Cleveland, OH, 44106 USA
[4]Department of Biology, Case Western Reserve University, Cleveland, OH, 44106 USA
[5]Department of Mathematics, University of Utah, Salt Lake City, UT, 84112 USA
[*]Corresponding author: linhhuynh@math.utah.edu

ABSTRACT

Density dependence is important in the ecology and evolution of microbial and cancer cells. Typically, we can only measure net growth rates, but the underlying density-dependent mechanisms that give rise to the observed dynamics can manifest in birth processes, death processes, or both. Therefore, we utilize the mean and variance of cell number fluctuations to separately identify birth and death rates from time series that follow stochastic birth-death processes with logistic growth. Our nonparametric method provides a novel perspective on stochastic parameter identifiability, which we validate by analyzing the accuracy in terms of the discretization bin size. We apply our method to the scenario where a homogeneous cell population goes through three stages: (1) grows naturally to its carrying capacity, (2) is treated with a drug that reduces its carrying capacity, and (3) overcomes the drug effect to restore its original carrying capacity. In each stage, we disambiguate whether the dynamics occur through the birth process, death process, or some combination of the two, which contributes to understanding drug resistance mechanisms. In the case of limited sample sizes, we provide an alternative method based on maximum likelihood and solve a constrained nonlinear optimization problem to identify the most likely density dependence parameter for a given cell number time series. Our methods can be applied to other biological systems at different scales to disambiguate density-dependent mechanisms underlying the same net growth rate.

Keywords Parameter identifiability · Uncertainty quantification · Stochastic discretization error analysis · Stochastic processes · Density-dependent ecological modeling · Drug resistance

Mathematics Subject Classifications 60J27 · 92D25 · 62M10 · 60J25

Acknowledgements This work was made possible in part by NSF grant DMS-2052109, by research support from the Oberlin College Department of Mathematics, the National Institutes of Health (5R37CA244613-02), and the American Cancer Society Research Scholar Grant (RSG-20-096-01). The authors thank Dr. Vishhvaan Gopalakrishnan, Mina Dinh, and Dr. Kyle Card for discussing their preliminary experimental data from the EVE system (EVolutionary biorEactor).

Statements and Declarations The authors have declared no competing interests. No experimental or clinical datasets were generated or analysed during the current study. Code (for the methods and simulated data) is available at https://github.com/lhuynhm/birthdeathdisambiguation.

1 Introduction

Density dependence, a phenomenon in which a population's *per capita* growth rate changes with population density[1], plays an important role in the ecology and evolution of microbial and cancer cell populations, especially under drug treatments. For example, Karslake et al. 2016[2] shows experimentally that changes in *E.coli.* cell density can either increase or decrease the efficacy of antibiotics. Existing work such as[3–7] show that interactions between drug sensitive and resistant cancer cells can shape the population's evolution of drug resistance. The effects of these interactions have been modeled using a quadratic growth-suppression term; see[8] for example, reflecting their density dependence.

To analyze the role of density dependence, especially in drug resistance, we begin by considering a classical mathematical model of density-dependent population dynamics, Verhulst's logistic growth model[9]. Later, we will also discuss more general examples, including piecewise-specified density-dependent growth rates, see Appendix C. Verhulst's model describes the dynamics of a homogeneous population in terms of its net growth rate:

$$\frac{d\phi}{dt} = r\left(1-\frac{\phi}{K}\right)\phi = r\phi - \frac{r}{K}\phi^2. \tag{1}$$

In Equation (1), ϕ denotes population size, r denotes intrinsic *per capita* net growth rate, and K denotes carrying capacity. The density dependence term $\frac{r}{K}\phi^2$ describes the direct or indirect interactions between individuals in the population. The minus $(-)$ sign indicates that the interactions have a negative *net* effect on the population–in particular, reducing the population size. This term reflects negative density dependence, resulting from situations such as crowding, competition, or predation, which can hinder the growth of population size through either the birth process, the death process, or some combination of the two. However, the formulation in Equation (1) leaves the underlying mechanisms of negative density dependence unclear. To disambiguate birth-related versus death-related mechanisms, we rewrite the density dependence term with the parameter γ as follows:

$$\frac{r}{K}\phi^2 = \gamma\frac{r}{K}\phi^2 + (1-\gamma)\frac{r}{K}\phi^2, \quad 0 \le \gamma \le 1. \tag{2}$$

We interpret the term $\gamma\frac{r}{K}\phi^2$ as the reduction in the population's growth rate due to competition-regulated mechanisms affecting the birth process, and $(1-\gamma)\frac{r}{K}\phi^2$ as the population's competition-regulated mechanisms affecting the death process. For example, in ecology, one distinguishes exploitative competition, where limited resources hinder the reproduction of the populations, from interference competition, where individuals fight against one another causing death[10].

For completeness, we also disentangle the intrinsic net growth rate $r\phi$ into birth and death as follows:

$$r\phi = b_0\phi - (b_0 - r)\phi = b_0\phi - d_0\phi, \quad \text{with } b_0 \ge r > 0 \text{ and } d_0 := b_0 - r \ge 0. \tag{3}$$

We interpret b_0 as the population's intrinsic (low-density) *per capita* birth rate and d_0 as the population's intrinsic (low-density) *per capita* death rate. Hence, we parameterize Equation (1) with γ, b_0, and d_0 as follows:

$$\frac{d\phi}{dt} = \underbrace{\left(b_0 - \gamma\frac{r}{K}\phi\right)\phi}_{\text{birth}} - \underbrace{\left(d_0 + (1-\gamma)\frac{r}{K}\phi\right)\phi}_{\text{death}}, \quad 0 \le \gamma \le 1, \quad r = b_0 - d_0. \tag{4}$$

For fixed K, b_0, and d_0 (or r), while different values of γ in Equation (4) result in equations that are algebraically equivalent to Equation (1), they can describe different biological processes. This motivates us to ask the following question:

[Q]: Do birth- or death-related mechanisms determine the changes in density-dependent population dynamics?

The significance of the answer to question [Q] can also be seen in other contexts. In this paper, we restrict attention to negative density dependence, which can be interpreted as competitive dynamics. However, density dependence can also be positive. For example, the Allee effect[11] of density-dependent dynamics (which is defined as a positive correlation between population density and *per capita* net growth rate) provides another example. Although the Allee effect is typically modeled with cubic growth[12] instead of logistic growth, answering question [Q] would contribute to understanding the mechanisms that give rise to the effect. Increasing *per capita* net growth rates with increased population density could result from increased cooperation or mating among individuals (increased birth rates) or from reduced fighting due to habitat amelioration (decreased death rates)[13]. This distinction is important because populations that experience the Allee effect can become extinct if the population sizes fall below the Allee threshold[14]. Extinction problems are of interest because, for example, we hope to eventually eradicate tumors and harmful bacteria within individual hosts. Clinically, bactericidal drugs such as penicillin promote cell death, while bacteriostatic drugs such as chloramphenicol, clindamycin, and linezolid inhibit cell division[15].[16] shows that bactericidal and bacteriostatic drugs affect cellular metabolism differently, and the metabolic state of bacteria in turn influences drug efficacy. Hence, identifying "-cidal" versus "- static" drug effects may help contribute to developing more efficacious drug treatments. From an evolutionary perspective,[17] shows that assuming a zero death rate leads to overestimating bacterial mutation rates under stress, which in turn can lead to incorrect conclusions about the evolution of bacteria under drug treatments. The authors point out that it is important to separately identify birth and death rates. Theoretically, one may compute probability of extinction/escape and mean first-passage time to extinction/escape for cell populations under certain drug treatments such as[18–20]. If we model evolution with a birth-death process as in[21], computing the probability and mean first-passage time involves knowing separate birth and death rates[22–24]. That means cell populations with the same net growth rates–but different birth and death rates–can have different extinction/escape probabilities and mean first-passage times. Indeed,[21] points out that defining "fitness" as net growth rate (difference between birth rate and death rate) loses evolutionary information; instead, we should use separate birth and death rates to measure "survival of the fittest." Therefore, the significance of disambiguating birth and death rates underlying a given net growth rate is clear across multiple biological contexts on different scales.

We aim to answer question [Q] by extracting birth and death rates from observations of density-dependent population dynamics. One type of population dynamics information that is typically available is population size. For fixed volume, cell density and cell number are proportional. Deterministic modeling does not allow us to disentangle birth rate $b_{\hat{\phi}}$ and death rate $d_{\hat{\phi}}$ from net growth rate $(b_{\hat{\phi}} - d_{\hat{\phi}})$, because for any arbitrary rate $a_{\hat{\phi}}$, the transformation $b_{\hat{\phi}} \to b_{\hat{\phi}} + a_{\hat{\phi}}$ and $d_{\hat{\phi}} \to d_{\hat{\phi}} + a_{\hat{\phi}}$ leaves the net growth rate unchanged and results in exactly the same cell number trajectory. On the other hand, in stochastic models, cell number trajectories generated from different pairs of birth and death rates with the same net growth rate will have different noise levels, which makes it possible to distinguish different scenarios. Moreover, at a fundamen-

tal level, population growth is driven by the birth/division[1] and death of individual cells. At this level, cell birth and death are discrete rather than continuous processes, and may involve stochastic elements such as molecular fluctuations in the reactions within individual cells[25]. Therefore, although the tractability of deterministic population equations has made them attractive as a framework for modeling the growth of pathogenic populations and their responses to therapeutic agents[26–28], a stochastic modeling framework is more appropriate for the research question we consider, which is to disambiguate birth-related and death-related dynamics. We describe density dependence with logistic growth, because it is one of the simplest form of density dependence and still captures some realistic cell population dynamics such as the dynamics of cancer cells[29].

In this paper, we develop a nonparametric approach to modeling population growth that leverages stochastic fluctuations in order to estimate both the birth and death rates from population time series data (without requiring *a priori* assumption of a specific model such as logistic growth). Although we use a Verhulst-inspired logistic model for illustrating and validating our method, and investigating negative density dependence, our nonparametric estimation method applies equally well to general birth-death models that do not have a simply parametrized birth and death rate functions. When net growth rates are *not* density-dependent, the underlying population model is linear. Linear models have been used in the context of serial dilution experiments, a protocol that keeps the population at a low density, masking density-dependent effects (see[30] for example). When growth rates are allowed to be density-dependent, a population that could range up to K individuals (the carrying capacity) has in general $(K+1)$ net growth rate parameters, i.e. $\{r_n : 0 \leq n \leq K\}$, and hence $2(K+1)$ birth and death rates in total. When the underlying model has a large number of unknown parameters, maximum likelihood methods become intractable. While advancement in technology may allow experimental estimation of birth and death rates[31], our direct estimation method enjoys simple, commonly available data: population size.

We describe the details of our direct estimation method in Section 3, and demonstrate the method on the Verhulst-inspired logistic birth-death process model described in Section 2, where the birth and death rates are defined with the parameter set $\{b_0, d_0, \gamma, K\}$ as in Equation (4). We mention the model parameters here for data simulation and error analysis purposes; however, our estimation method is nonparametric and does not require knowing the forms of birth and death rates. We illustrate this advantage of our method by demonstrating it on a birth-death process with piecewise density-dependent *per capita* birth and death rates in Appendix C. In Section 7, we compare our method with related existing methods[32–34]. To deal with the sample size issue, we describe a discretization approach in Section 3.3 and analyze this approach in detail in Appendix A. In the extreme case where there is only one cell number sample trajectory available, we supplement our method with a maximum likelihood approach in Section 6. To study the role of density dependence in population dynamics and drug resistance, in Section 5, we apply our direct estimation method to disambiguate underlying mechanisms for a homogeneous cell population that goes through three stages: (1) auto-regulated growth, (2) drug treatment, and (3) defense against the treatment, which we define as drug resistance.

2 Verhulst-Inspired Stochastic Logistic Birth-Death Process Model

We consider systems of homogeneous cells described by a birth-death process, that is, a discrete-state continuous-time Markov chain tracking the number of individual cells $N(t)$ in the system over time t, with

[1] Although cells do not give birth to offspring in the biological sense, for the rest of the manuscript, we refer to cell division as birth to be consistent with the birth-death process model we use.

state transitions comprising either "birth" ($N \to N+1$) or "death" ($N \to N-1$), as shown in Figure 1. In linear birth-death processes, *per capita* birth and death rates are constants that do not depend on N. In contrast, here we consider birth-death processes whose *per capita* birth and death rates depend on N in order to incorporate density-dependent population dynamics. Specifically, motivated by Equation (4), we define the *per capita* birth rate b_N and death rate d_N in our model as follows:

$$b_N = \max\left\{b_0 - \gamma\frac{r}{K}N, 0\right\}, \tag{5}$$

$$d_N = d_0 + (1-\gamma)\frac{r}{K}N, \tag{6}$$

where $b_0 > 0$ and $d_0 \geq 0$ are intrinsic (low-density) *per capita* birth and death rates respectively, $r = b_0 - d_0 \geq 0$ is the intrinsic (low-density) *per capita* net growth rate, $K > 0$ is the population's carrying capacity, and $\gamma \in [0,1]$ determines the extent to which the nonlinear or density-dependent dynamics arises from the *per capita* birth versus death rates. When $\gamma = 0$, the birth process is density-independent; all density dependence lies in the death process. Conversely, when $\gamma = 1$, the density-dependent dynamics is fully contained in the birth process. When $0 < \gamma < 1$, the density-dependent dynamics is split between birth and death. We use the max function in Definition (5) to ensure b_N is nonnegative. The total birth and death rates of the population are $b_N N$ and $d_N N$.

Figure 1 **(B)** illustrates the need for separating birth and death rates, as the net growth rate may be the same under two different drugs, but the underlying drug effects are different.

For a single-species birth-death process of this form, with $d_1 > 0$ (d_1 is the death rate when $N = 1$) and no immigration, it is well known that the unique stationary probability distribution gives $N(t) \to 0$ as $t \to \infty$ with probability one[35]. Rather than concern ourselves with the long-term behavior, here we are interested in answering question [Q] by estimating b_N and d_N. Therefore, we will focus on the analysis of transient population behavior rather than long-time, asymptotic behavior.

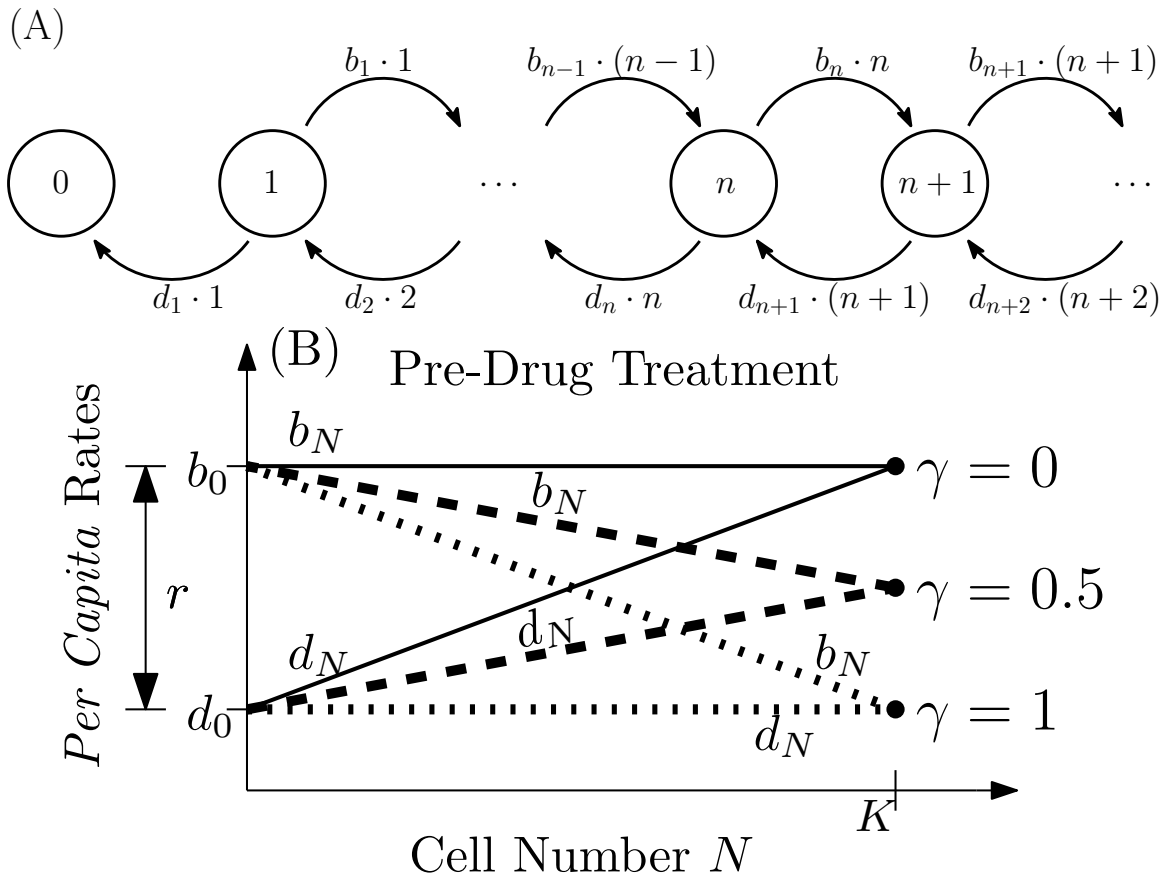

Figure 1. Schematic representation of our logistic birth-death process model. **(A)** State $N = n$ transitions to state $N = n+1$ at rate $b_n \cdot n$ and transitions to state $N = n-1$ at rate $d_n \cdot n$. At state $N = 0$, the system cannot transition to state $N = 1$, because there is no individual to give birth. **(B)** Graphical illustration of density-dependent *per capita* birth and death rates defined in Equations (5) and (6) for three cases (I) $\gamma = 0$ (solid lines), (II) $\gamma = 0.5$ (dashed lines), and (III) $\gamma = 1$ (dotted lines). The *net* growth rate curve $(b_N - d_N)$ is the same for each case.

.

3 Nonparametric Direct Estimation Method

3.1 Mathematical Derivation

Let $N(t) \geq 0$ be an integer-valued random variable representing the number of cells at time t. We consider a small time increment Δt, within which each cell can either divide (i.e. one cell is replaced by two cells), die (i.e. one cell disappears and is not replaced), or stay the same (i.e. there is still one cell). Focusing on a single timestep, let $(\Delta N_+ | N, \Delta t)$ and $(\Delta N_- | N, \Delta t)$ be two random variables representing the numbers of cells gained and lost, respectively, from an initial population of N cells, after a period of time Δt. The number of cells that neither die nor divide is thus equal to $(N - \Delta N_+ - \Delta N_-)$. Although the two random variables $(\Delta N_+ | N, \Delta t)$ and $(\Delta N_- | N, \Delta t)$ are not strictly independent (as one cell cannot both die and reproduce at the same time), we assume the correlation between them is small enough to be neglected and the birth and death processes are independent. In Section 7, we discuss the case where we do not have this independence assumption. Throughout the paper, we assume that the density-dependent birth rates b_n and death rates d_n are not fluctuating in time; if environmental fluctuations influenced these rates

this could introduce additional statistical dependencies between the birth and death processes, which we choose to neglect here.

Thus, among N cells, ΔN_+ cells are "chosen" to divide and ΔN_- cells are "chosen" to die. On a time interval of length Δt, the probabilities that a cell divides and dies are $b_N\Delta t + o(\Delta t)$ and $d_N\Delta t + o(\Delta t)$ respectively.[2] For convenience, we will omit the $o(\Delta t)$ correction where possible without introducing inaccuracies. The random variables $(\Delta N_+|N,\Delta t)$ and $(\Delta N_-|N,\Delta t)$ are binomially distributed. In particular,

$$(\Delta N_+|N,\Delta t) \sim \text{ Binomial}(N, b_N\Delta t) \text{ with mean } Nb_N\Delta t \text{ and variance } Nb_N\Delta t(1-b_N\Delta t), \tag{7}$$
$$(\Delta N_-|N,\Delta t) \sim \text{ Binomial}(N, d_N\Delta t) \text{ with mean } Nd_N\Delta t \text{ and variance } Nd_N\Delta t(1-d_N\Delta t). \tag{8}$$

Define a random variable $(\Delta N|N,\Delta t)$ to be the net change in population size from N cells after a period of time Δt, i.e. $(\Delta N|N,\Delta t) = (\Delta N_+|N,\Delta t) - (\Delta N_-|N,\Delta t)$. Typically, experimental or clinical measurements reflect only the net change $(\Delta N|N,\Delta t)$ rather than the increase $(\Delta N_+|N,\Delta t)$ or decrease $(\Delta N_-|N,\Delta t)$ separately. Because $(\Delta N_+|N,\Delta t)$ and $(\Delta N_-|N,\Delta t)$ are approximately independent, for sufficiently small Δt, we have

$$\mathbb{E}[\Delta N|N,\Delta t] = \mathbb{E}[\Delta N_+|N,\Delta t] - \mathbb{E}[\Delta N_-|N,\Delta t] = Nb_N\Delta t - Nd_N\Delta t = (b_N-d_N)N\Delta t, \tag{9}$$
$$\text{Var}[\Delta N|N,\Delta t] \approx \text{Var}[\Delta N_+|N,\Delta t] + \text{Var}[\Delta N_-|N,\Delta t] = Nb_N\Delta t(1-b_N\Delta t) + Nd_N\Delta t(1-d_N\Delta t) \tag{10}$$
$$= Nb_N\Delta t + Nd_N\Delta t + O(\Delta t^2) \tag{11}$$
$$\approx (b_N+d_N)N\Delta t. \tag{12}$$

Therefore, to estimate the two parameters birth and death rates, we solve the following linear system consisting of two equations:

$$(b_N-d_N)N = \frac{\mathbb{E}[\Delta N|N,\Delta t]}{\Delta t} \text{ and } (b_N+d_N)N = \frac{\text{Var}[\Delta N|N,\Delta t]}{\Delta t}, \tag{13}$$

and obtain:

$$b_N N = \frac{\mathbb{E}[\Delta N|N,\Delta t] + \text{Var}[\Delta N|N,\Delta t]}{\Delta t} \tag{14}$$
$$d_N N = \frac{\text{Var}[\Delta N|N,\Delta t] - \mathbb{E}[\Delta N|N,\Delta t]}{\Delta t} \tag{15}$$

In theory, we can divide the total birth rate $b_N N$ and total death rate $d_N N$ by N to obtain *per capita* birth and death rates. However, when we implement this on a computer, dividing by small N leads to numerical blowups. Hence, to visualize the estimated rates better, we present total rates $b_N N$ and $d_N N$ in our results instead of *per capita* rates. When the total rates are nonlinear functions of cell number, that means the *per capita* rates are density-dependent. In Section 3.3, we discuss how to approximate $\mathbb{E}[\Delta N|N,\Delta t]$ and $\text{Var}[\Delta N|N,\Delta t]$ from data.

3.2 Stochastic Data Simulation

To validate our method, we apply it to simulated *in silico* data. While our underlying model is time-continuous, in experimental and clinical settings, one can only observe cell numbers at discrete time points. In order to efficiently generate an ensemble of trajectories of the birth-death process, we construct

[2] We adopt the standard convention $\frac{o(\Delta t)}{\Delta t} \to 0$ as $\Delta t \to 0$.

a τ-leaping approximation[36] as follows.

Given $N(t)$ individuals at time t, we approximate the number of individuals after a short time interval Δt as

$$N(t+\Delta t) \approx N(t) + \Delta N_+(t) - \Delta N_-(t), \tag{16}$$

where $\Delta N_+ \sim \text{Binomial}\Big(N(t), b_{N(t)}\Delta t\Big)$ and $\Delta N_- \sim \text{Binomial}\Big(N(t), d_{N(t)}\Delta t\Big)$ representing the number of cells added to and lost from the system after a period of time Δt. We approximate ΔN_+ and ΔN_- as if they were independent random variables; see discussion in Section 3.1. When $N(t)$ is sufficiently large, we approximate the binomial distributions with Gaussian distributions that have the same means and variances as the binomial distributions. Our discrete-state process in Section 2 is now approximated with a continuous-state process:

$$\Delta N(t)_+ \sim \text{Normal}\left(N(t)b_{N(t)}\Delta t, N(t)b_{N(t)}\Delta t\Big(1-b_{N(t)}\Delta t\Big)\right), \tag{17}$$

$$\Delta N(t)_- \sim \text{Normal}\left(N(t)d_{N(t)}\Delta t, N(t)d_{N(t)}\Delta t\Big(1-d_{N(t)}\Delta t\Big)\right). \tag{18}$$

Thus, the net change in the number of cells after a timestep Δt is

$$\begin{aligned} N(t+\Delta t) - N(t) \approx & N(t)b_{N(t)}\Delta t + \Delta W_+(t)\sqrt{N(t)b_{N(t)}\Big(1-b_{N(t)}\Delta t\Big)} \\ & - N(t)d_{N(t)}\Delta t - \Delta W_-(t)\sqrt{N(t)d_{N(t)}\Big(1-d_{N(t)}\Delta t\Big)} \end{aligned} \tag{19}$$

$$\approx \Big(b_{N(t)} - d_{N(t)}\Big)N(t)\Delta t + \sqrt{\Big(b_{N(t)} + d_{N(t)}\Big)N(t)}\Delta W(t). \tag{20}$$

Here $\Delta W_\pm$ are independent Wiener process increments (with mean 0 and variance Δt), and ΔW is a Wiener process increment derived from a linear combination of the $\Delta W_\pm$. (Note that an arbitrary linear combination of two independent Wiener process increments $Z = \alpha\Delta W_1 + \beta\Delta W_2$ yields a Gaussian process with variance $E[Z^2] = (\alpha^2+\beta^2)\Delta t$.) Equation (20) is the τ-leaping approximation used in our data simulation, which is analogous to the forward Euler algorithm in the deterministic setting. Taking the limit $\Delta t \to dt$, we obtain a version of our population model as a continuous-time Langevin stochastic differential equation:

$$dN(t) = \Big(b_{N(t)} - d_{N(t)}\Big)N(t)dt + \sqrt{\Big(b_{N(t)} + d_{N(t)}\Big)N(t)}\,dW(t). \tag{21}$$

where $dW(t)$ is delta-correlated white noise satisfying $\langle dW(t)dW(t')\rangle = \delta(t-t')$. We use Equation (21) under the Ito interpretation.

3.3 Estimation from Data

Equations (14) and (15) require the mean $\mathbb{E}[\Delta N|N,\Delta t]$ and variance $\text{Var}[\Delta N|N,\Delta t]$ to compute the birth and death rates corresponding to each population size N. We estimate these two statistics directly from data. The higher the sample size, the closer the estimated statistics are to the theoretical ones. However,

conducting multiple experiments costs time and resources. To deal with this sample size issue, one approach is to discretize all the N values in the given dataset into bins of equal size to increase the number of samples corresponding to each cell number. Roughly speaking, in each bin, we compute the change in cell number ΔN from each data point after time period Δt, and then compute the mean and variance of those increments. We use the computed mean and variance to estimate the birth and death rates corresponding to the midpoint of the bin. In this approach, not all the computed cell number increments correspond to the midpoint of the bin–in fact, many of them are not cell number increments from the midpoint. However, this approximation is good enough for an optimal binsize, which we will discuss in Section 4 and Appendix A.

The τ-leaping simulation described in Section 3.2 provides a convenient format of the dataset, where all the cell number trajectory vectors have the same length and the ith entries in all the vectors correspond to the same time point t_i. For example, as an illustration, the orange points labeled 2, 3, 4, 5 in Figure 2 represent cell numbers at the same time for different trajectories/experiments; the same applies to the orange points labeled 6 and 7. More precisely, denote S as the number of cell trajectories in a given dataset $\mathscr{D}$. In general, $\mathscr{D}$ has the following form:

$$\mathscr{D} = \Big\{ \underbrace{[N_0^1; \ldots; N_i^1; \ldots; N_{T_1}^1]}_{\text{trajectory vector 1}}, \ldots, \underbrace{[N_0^s; \ldots; N_i^s; \ldots; N_{T_s}^s]}_{\text{trajectory vector s}}, \ldots, \underbrace{[N_0^S; \ldots; N_i^S; \ldots; N_{T_S}^S]}_{\text{trajectory vector } S} \Big\}, \tag{22}$$

where the superscript s of N (with $s = 1, \ldots, S$) indicates the sth trajectory/experiment and the ith subscripts of N (with $i = 0, \ldots, T_s$) indicate the time points. Each of the S trajectory vectors has length $(T_s + 1)$, meaning the cell number data from experiment s are collected at $T_s + 1$ time points. As discussed above, with the τ-leaping simulation, we set all the T_s (with $s = 1, \ldots, S$) to be equal to each other. Denote $T := T_s$, $\forall s$. The time period between the ith and $(i+1)$th entries in each trajectory vector s is equal to Δt, $\forall i$, with $i = 0, \ldots, T$, and $\forall s$, with $s = 1, \ldots, S$. This format of our simulated data is consistent with the actual experimental data produced by the EVolutionary biorEactor (EVE) in our lab[37]. Figure 2 helps visualize the dataset. In this figure, the vertical dashed lines indicate the time points. The data points in $\mathscr{D}$ are the intersections between these lines and the cell number trajectories such as the orange points in the figure. We carry out the steps in Algorithm 1 to estimate $\mathbb{E}[\Delta N|N, \Delta t]$ and variance $\mathrm{Var}[\Delta N|N, \Delta t]$ from a given dataset $\mathscr{D}$ with the properties we just described.

Algorithm 1 Estimation of Mean and Variance of Cell Number Increments from Dataset $\mathscr{D}$

Step 1: Find the minimum and maximum cell numbers $N_{\min}$ and $N_{\max}$ across the whole dataset $\mathscr{D}$.

Step 2: Discretize the interval $[N_{\min}, N_{\max}]$ into bins of equal size η. The total number of bins is equal to $k_{\max} = \left\lceil \frac{N_{\max} - N_{\min}}{\eta} \right\rceil \in \mathbb{Z}^+$, where $\lceil \cdot \rceil$ is the ceiling function that represents the smallest integer that is greater than or equal to $\frac{N_{\max} - N_{\min}}{\eta}$.

Step 3: Put all the data points across $\mathscr{D}$ into appropriate bins. The data points in each kth bin $[N_k, N_k + \eta)$ are greater than or equal to N_k and less than $(N_k + \eta)$, as graphically illustrated by Figures 2 and 3 **(A)**. The orange points in these figures are the data points in bin k.

Step 4: In each bin, compute the cell number increments after Δt from the data points in the bin. That means: suppose a data point N_i^s is the ith entry in trajectory vector s, then find the $(i+1)$th entry in the same vector, denoted as N_{i+1}^s, and compute $\Delta N_i^s := N_{i+1}^s - N_i^s$. This also mean: we cannot use the last entries of the trajectory vectors because there is no information about the cell number after Δt for those data points. Figure 3 serves as a "cartoon" graphical illustration for this step. Consider the orange data points labeled 1, 2, 3, 4, 5, 6, 7, 8 in Figure 3 **(A)**. Then, the corresponding cell numbers after Δt are the purple data points labeled 1, 2, 3, 4, 5, 6, 7, 8 in Figure 3 **(B)**. The eight ΔN samples for bin k are obtained by having the cell number values of the purple data points in Figure 3 **(B)** minus the cell number values of the orange data points Figure 3 **(A)** in the respective order 1, 2, 3, 4, 5, 6, 7, 8.

Step 5: Compute the mean and variance of the cell number increment samples from **Step 4**, and use these statistics to estimate the birth and death rates corresponding to the midpoints of the bins. The midpoint of bin $[N_k, N_k + \eta]$ is $\left(N_k + \frac{\eta}{2}\right)$. That means: we estimate $b_{(N_k + \eta/2)}$ and $d_{(N_k + \eta/2)}$ for $k = 1, \ldots, k_{\max}$.

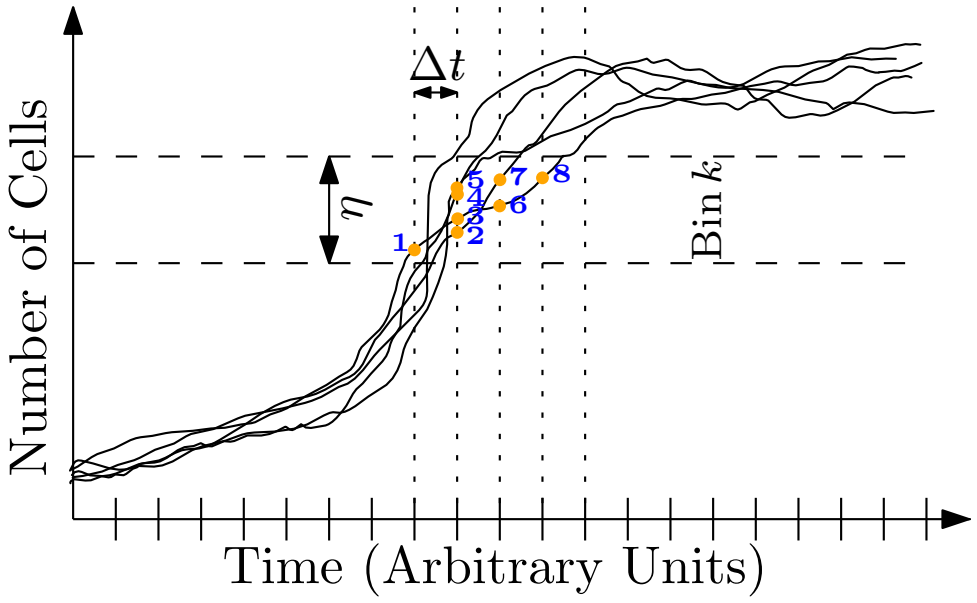

Figure 2. **Graphical illustration of the overall discretization and Step 3 in Algorithm 1.** The solid black curves represents stochastic cell number trajectories. The cell number data are collected after time period Δt. We group all the data points across the whole dataset into bins along the vertical axis. Each bin is of size η. The orange points labeled 1, 2, 3, 4, 5, 6, 7, 8 are the data points in an arbitrary bin k.

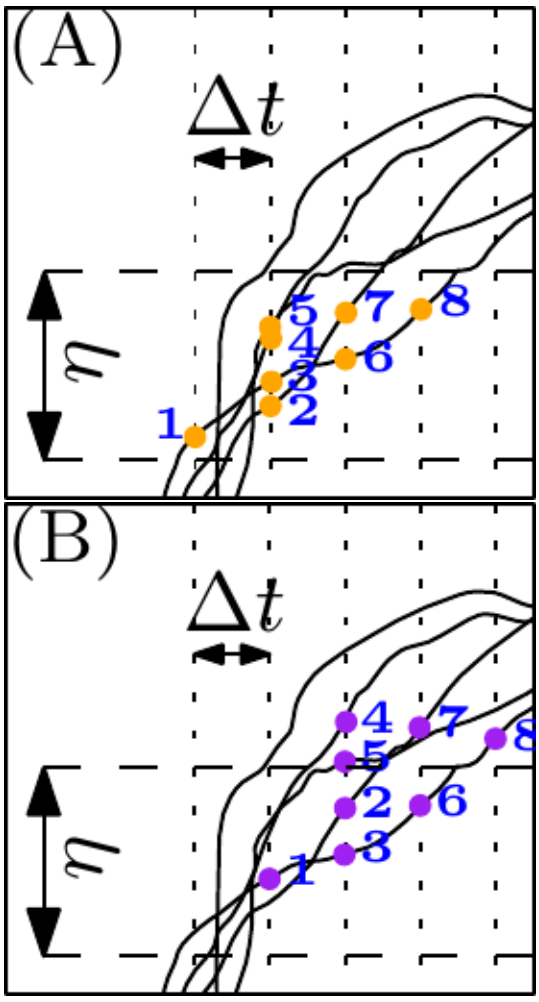

Figure 3. Graphical illustration of Step 4 in Algorithm 1. Figure **(A)** is a zoom-in of Figure 2, where the orange points labeled 1, 2, 3, 4, 5, 6, 7, 8 are the data points in an arbitrary bin k. In both **(A)** and **(B)**, Δt represents time period and η represents bin size. The cell number trajectories in **(B)** are the same ones in **(A)**. The purple points labeled 1, 2, 3, 4, 5, 6, 7, 8 in **(B)** represent the cell numbers after Δt from the orange points in **(A)** respectively. For example, after Δt, the orange point labeled 1 in **(A)** changes to the purple point labeled 1 in **(B)**. Hence, the cell number value of the purple point labeled 1 minus the cell number value of the orange point labeled 1 gives the cell number increment ΔN from the orange point labeled 1 after Δt. This increment is used approximating the mean and variance of cell number increment corresponding to the midpoint of bin k. The same applies to the other points.

Our Algorithm 1 requires a sufficiently large bin size so that there are at least two samples in each bin in order to compute the variances of the cell number increments. On the other hand, if the bin size is too large, then the cell number increments from the data points in each bin do not accurately represent the increment from the midpoint of the bin. That leads to the question: what is an optimal bin size η? We answer this question with a summary in Section 4 and rigorous analytical details in Appendix A.

4 Validation and Error Analysis

We validate our method by comparing the estimated rates with the true rates that are used to generate the simulated data. We simulate an ensemble of cell number trajectories, using a numerically efficient

τ-leaping approximation described in Section 3.2, and estimate birth and death rates using Equations (14) and (15) and the method described in Section 3.3. Figure 4 shows that the estimated and true rates are well-aligned. Denote $\hat{S}_k$ as the sample size in bin k. Using the discretization described in Section 3.3, we estimate birth and death rates via the empirical mean $\left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle$ and empirical variance $\sigma^2 \left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right]$ obtained from an ensemble of $S = 100$ simulated trajectories. To quantify the accuracy of our method, we define the error $\mathscr{E}_{k\text{birth}}$ in estimating the birth rate corresponding to population size $N = N_k + \frac{\eta}{2}$, and the error $\mathscr{E}_{k\text{death}}$ in estimating the death rate corresponding to $N = N_k + \frac{\eta}{2}$ as follows:

$$\mathscr{E}_{k\text{birth}} := \frac{\mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right] + \text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]}{2\Delta t} \tag{23}$$

$$- \frac{\left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle + \sigma^2 \left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right]}{2\Delta t}, \tag{24}$$

$$\mathscr{E}_{k\text{death}} := \frac{\text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right] - \mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]}{2\Delta t} \tag{25}$$

$$- \frac{\sigma^2 \left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right] - \left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle}{2\Delta t}. \tag{26}$$

$$\tag{27}$$

Under the assumption that the samples η_i are iid uniformly distributed on $[0, \eta)$, the theoretical means and variances of the errors $\mathscr{E}_{k\text{birth}}$ and $\mathscr{E}_{k\text{death}}$ are equal to:

$$\mathbb{E}\left[\mathscr{E}_{k\text{birth}}\right] = \frac{\mathbb{E}\left[\mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]\right] + \mathbb{E}\left[\text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]\right]}{2\Delta t} \tag{28}$$

$$- \frac{\mathbb{E}\left[\left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle\right] + \mathbb{E}\left[\sigma^2 \left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right]\right]}{2\Delta t} \tag{29}$$

$$= \frac{\mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right] - \mathbb{E}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0, \eta)\right]}{2\Delta t} \tag{30}$$

$$+ \frac{\text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right] - \text{Var}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0, \eta)\right]}{2\Delta t}, \tag{31}$$

$$\mathbb{E}\left[\mathscr{E}_{k\text{death}}\right] = \frac{\mathbb{E}\left[\text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]\right] - \mathbb{E}\left[\mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]\right]}{2\Delta t} \tag{32}$$

$$- \frac{\mathbb{E}\left[\left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle\right] + \mathbb{E}\left[\sigma^2\left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k\right]\right]}{2\Delta t} \tag{33}$$

$$= \frac{\mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right] - \mathbb{E}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right]}{2\Delta t} \tag{34}$$

$$+ \frac{\text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right] - \text{Var}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right]}{2\Delta t}. \tag{35}$$

Similarly,

$$\text{Var}\left[\mathscr{E}_{k\text{birth}}\right] = \text{Var}\left[\mathscr{E}_{k\text{death}}\right] \tag{36}$$

$$= \frac{\text{Var}\left[\left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle\right] + \text{Var}\left[\sigma^2\left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k\right]\right]}{4\Delta t^2} \tag{37}$$

$$= \frac{\frac{\text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]}{\hat{S}_k} + \frac{2\left(\text{Var}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right]\right)^2}{\hat{S}_k - 1}}{4\Delta t^2}. \tag{38}$$

We compute $\mathbb{E}\left[\mathscr{E}_{k\text{birth}}\right]$, $\mathbb{E}\left[\mathscr{E}_{k\text{death}}\right]$, $\text{Var}\left[\mathscr{E}_{k\text{birth}}\right]$, and $\mathbb{E}\left[\mathscr{E}_{k\text{death}}\right]$ as functions of bin size in Appendix A. We then compute the 2-norm of the theoretical means and standard deviations (i.e. square roots of the variances) over all k to obtain the plots in Figure 5. We observe that as the bin size η increases, the expected errors increase, the theoretical variances (or standard deviations) of the errors decreases, and the empirical errors (computed using data from a simulation of $S = 100$ cell number trajectories) balance between the expected values and variances (or standard deviations), as shown in Figure 5. The expected values of errors reflect the differences between ΔN at the midpoint $\left(N = N_k + \frac{\eta}{2}\right)$ and ΔN at multiple points $\left(N = N_k + \eta_i, 0 \le \eta_i < \eta\right)$. The smaller the bin size, the closer multiple points are to the midpoint, so the error is smaller. However, if the bin is too small, then there are too few samples to accurately estimate theoretical statistics with empirical statistics. The theoretical variances of errors involves sample sizes; the bigger the bin size, the more samples we have. These two competing effects of bin size result in the empirical errors being intermediate values between the two theoretical statistics (expected values and variances) of the estimation errors. This "Goldilocks principle" is an example of the bias/variance tradeoff common in many estimation problems.

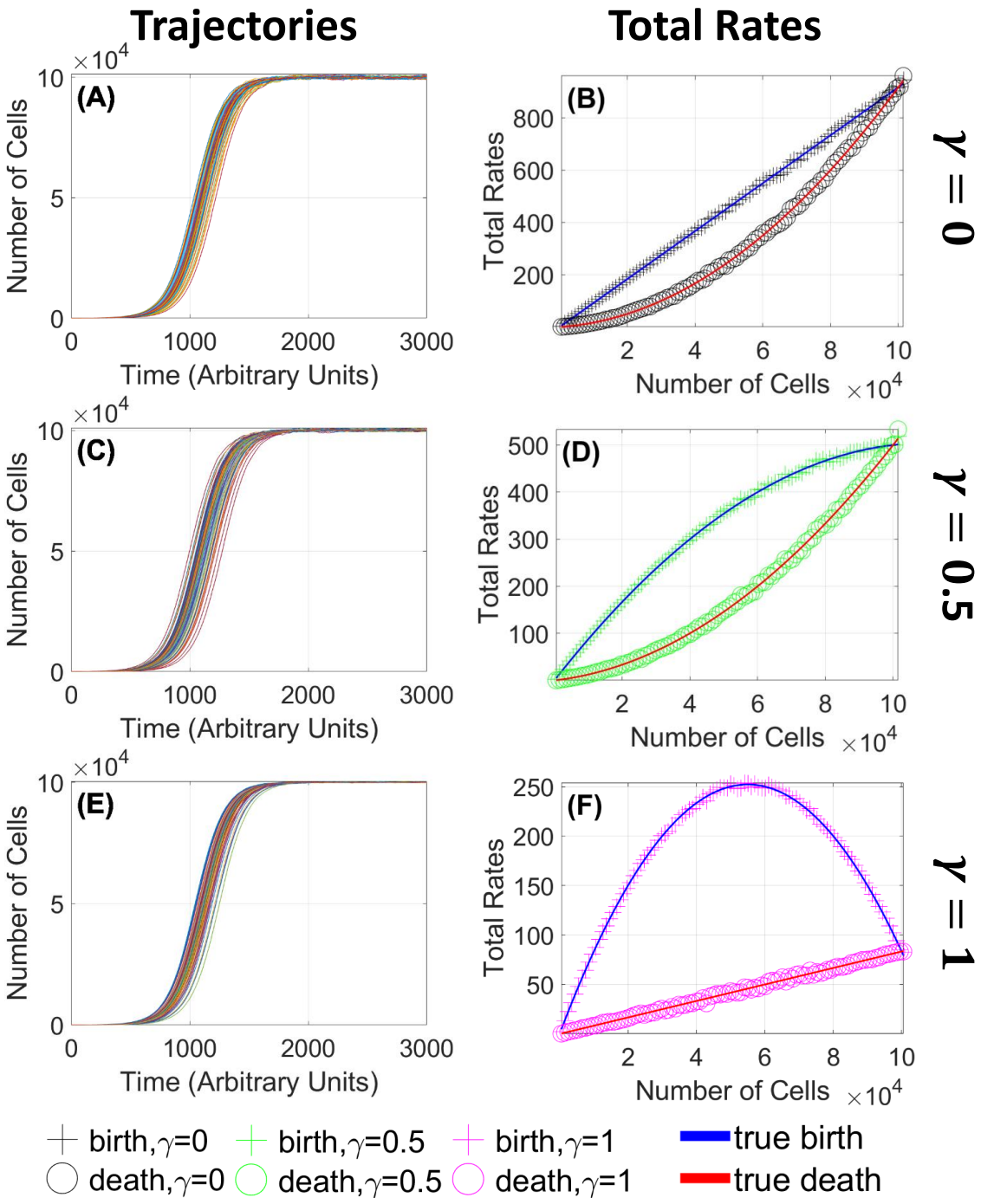

Figure 4. Agreement of estimated and true birth and death rates validates the direct estimation method. (A, C, E): Time series ensembles simulated using the τ-leaping approximation for the cases $\gamma = 0$ **(A)**, $\gamma = 0.5$ **(C)**, and $\gamma = 1$ **(E)** respectively. Each figure shows $S = 100$ trials. The estimated rates are computed using a bin size of $\eta = 10^3$. Carrying capacity $K = 10^5$ cells; low-density rates $b_0 = 1.1/120$ and $d_0 = 0.1/120$ (arbitrary time units). **(B, D, F)**: Estimated and true birth and death rates, as functions of population size. Blue line: true birth rate. Red line: true death rate. Plus signs $(+)$ denote estimated birth rates; circles $(\circ)$ denote estimated death rates. Throughout the paper we will use distinct colors to denote values of γ. **(B)** Black: $\gamma = 0$; **(D)** Green: $\gamma = 0.5$; **(E)** Magenta: $\gamma = 1.0$. We observe that the estimated birth and death rates are well-aligned with the true birth and death rates used to simulate the trajectories in **(A)**, **(C)**, and **(E)**.

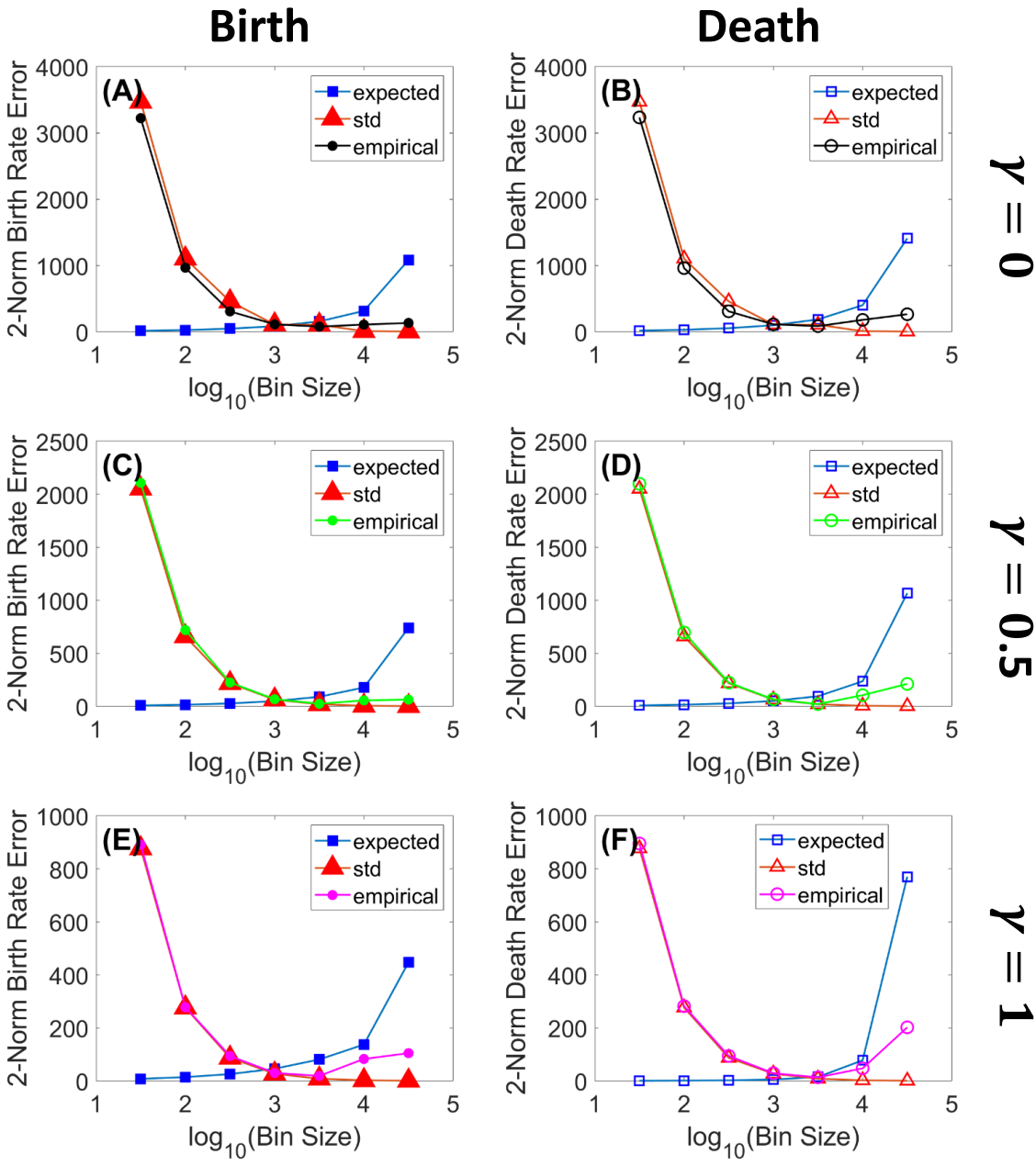

Figure 5. Intermediate bin sizes give optimal estimation performance. We plot the l_2-norm (over all bins) errors in estimating birth rate (left column) and death rate (right column) as functions of bin size η for carrying capacity $K = 10^5$. Squares (□) denote expected values of errors; triangles (△) denote standard deviations of errors; circles (○) denote empirical errors using data from a simulation of $S = 100$ cell number trajectories. **(A, C, E)**: errors in estimating birth rates. **(B, E, F)**: errors in estimating death rates. **(A, B)**: $\gamma = 0$ (black color); **(C, D)**: $\gamma = 0.5$ (green color); **(E, F)**: $\gamma = 0.5$ (magenta color). We observe that as the bin size η increases, the expected errors increase, the theoretical variances/standard deviations of the errors decreases, and the sample errors balance between the expected values and variances and have convex quadratic shapes.

5 Applications: Disambiguating Underlying Mechanisms of Autoregulation, Drug Effects, and Drug Resistance

In this section, we apply our direct estimation method (Section 3) to a homogeneous cell population that goes through three stages: (1) autoregulation, in which the cell population's growth saturates at a given carrying capacity, (2) drug treatment, in which an applied drug reduces the cell population's carrying capacity, and finally (3) drug resistance, in which the cell population regains its original carrying capacity (for a discussion on definitions of drug resistance, see Section 7). In each stage, we disambiguate whether the density-dependent mechanisms are related to the birth process, death process, or some combination of the two. While the mechanisms are different, they result in the same mean-field cell number time series.

5.1 General Analysis

Recall that *per capita* birth and death rates are:

$$b_N = \max\{b_0 - \gamma\frac{r}{K}N, 0\} \quad \text{and} \quad d_N = d_0 + (1-\gamma)\frac{r}{K}N \text{ respectively.} \tag{39}$$

The terms $\gamma\frac{r}{K}N$ and $(1-\gamma)\frac{r}{K}N$ indicate the rates' density dependence (i.e. the *per capita* rates are linear functions of cell number N). In this paper, we consider only negative density dependence (i.e. competitive interactions), so we set the density dependence parameter γ to be in $[0,1]$, meaning density dependence suppresses the cell population's growth by decreasing birth rates or increasing death rates. For the rest of this section, we omit $\max\{\cdot,0\}$ in the *per capita* birth rate for convenience in writing, but it is understood that the rate is the max of the formula and 0.

(1) In the autoregulation stage, different competition (negative density-dependent) mechanisms that cause the population to stop growing around the carrying capacity K are distinguished by different values of γ, all of which can manifest in the same mean-field cell number trajectories. When $\gamma = 0$, density dependence is fully in the death rate. Such a scenario could arise, for example, through interference (i.e. direct) competition such as predation or other conspecific lethal interactions. In contrast, when $\gamma = 1$, density dependence is fully in the birth rate. Such a scenario could arise, for example, through exploitation (i.e. indirect) competition mediated by accumulation of waste products or limited food resources that slow cell division. The scenario where $\gamma \in (0,1)$ represents a combination of those two mechanisms. We illustrate this stage in Figure 4 **(B)**.

(2) In the drug treatment stage, we consider a drug that reduces the population's carrying capacity by a factor $m > 1$ (i.e. $K \to K/m$). In general, parsing out birth and death rates allows us to know whether the drug affect the birth process, death process, or a combination of the two. Since this paper focuses on density-dependent mechanisms, we consider the situation where the drug does not change the intrinsic *per capita* rates (i.e. the rates remain b_0 and d_0); only the density dependence parameter changes (i.e. $\gamma \to ?$). Our estimation method is significant in disambiguating two particular clinically relevant drug effects: -cidal (increasing death rates and keeping birth rates the same) versus -static (decreasing birth rates and keeping death rates the same), both of which can manifest in the same mean-field cell number trajectories. Note that "increase" and "decrease" in this context are with respect to cell number, meaning for a given N, the new rates under drug treatment are either higher or lower than the original rates. For a detailed discussion on bactericidal and bacteriostatic antibiotics, see Section 7.

(2.1) For the -cidal drug effect, in order for the new carrying capacity to be K/m while the *per capita* birth rate is still $b_0 - \gamma\frac{r}{K}N$, the density dependence parameter has to become $\frac{\gamma}{m}$, because $b_0 + \frac{\gamma}{m}\frac{r}{K/m}N =$

$b_0 - \gamma\frac{r}{K}N$. Then, the *per capita* death rate under -cidal drug treatment is $d_0 + \left(1 - \frac{\gamma}{m}\right)\frac{r}{K/m}N$, which is indeed larger than the pre-treatment *per capita* death rate $d_0 + (1-\gamma)\frac{r}{K}N$, because $d_0 + \left(1 - \frac{\gamma}{m}\right)\frac{r}{K/m}N = d_0 + (m-\gamma)\frac{r}{K}N > d_0 + (1-\gamma)\frac{r}{K}N$, as $m > 1$.

(2.2) For the -static drug effect, in order for the new carrying capacity to be K/m while the *per capita* death rate is still $d_0 + (1-\gamma)\frac{r}{K}N$, the density dependence parameter has to become $1 - \frac{1-\gamma}{m}$, because $d_0 + \left(1 - 1 + \frac{1-\gamma}{m}\right)\frac{r}{K/m}N = d_0 + (1-\gamma)\frac{r}{K}N$. Then, the *per capita* birth rate under -static drug treatment is $b_0 - \left(1 - \frac{1-\gamma}{m}\right)\frac{r}{K/m}N$, which is indeed smaller than the pre-treatment *per capita* birth rate $b_0 - \gamma\frac{r}{K}N$, because $b_0 - \left(1 - \frac{1-\gamma}{m}\right)\frac{r}{K/m}N = b_0 - (\gamma + m - 1)\frac{r}{K}N < b_0 - \gamma\frac{r}{K}N$, as $(m-1) > 0$.

While the two drug mechanisms (-cidal versus -static) in (2.1) and (2.2) are different, they result in the same *net* growth rates (birth rates minus death rates). Our direct estimation method allows us to disambiguate the underlying mechanisms, despite the same mean-field cell number time series.

We illustrate the drug treatment stage in Figure 6.

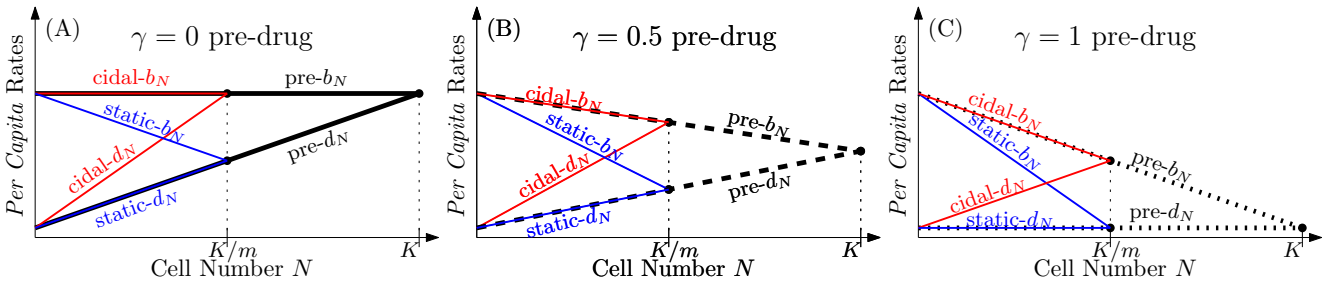

Figure 6. Schematic illustration of the -cidal versus -static drug effects. (A, B, C): density dependence parameter before drug treatment is $\gamma = 0$ (solid lines), 0.5 (dashed), 1 (dotted) respectively. K denotes the cell population's carrying capacity before treatment, and K/m denotes the cell population's carrying capacity after treatment. For illustration, we set $m = 2$. Pre-b_N and pre-d_N are the *per capita* birth and death rates before treatment (black curves); cidal-b_N and cidal-d_N are the *per capita* birth and death rates treatment with a -cidal drug (death-promoting; red curves); static-b_N and static-d_N are the *per capita* birth and death rate after treatment with a -static drug (birth-inhibiting; blue curves).

.

(3) In the drug resistance stage, we consider the situation where the cell population defends against the drug and gains back its original carrying capacity (i.e. $K/m \to K$). Similar to the drug treatment stage, in general, by estimating birth and death rates, we know whether the resistance happens through the birth process, death process, or both. Our estimation method is significant in disambiguating two particular mechanisms: reduced-mortality resistance (decreasing death rates and keeping the birth rates the same) and enhanced fecundity (increasing birth rates and keeping the death rates the same). There are many different ways the *per capita* birth and death rates can change to satisfy the criteria of these two mechanisms. However, in this paper, we are interested in density-dependent mechanisms, so we only consider the

situation where the intrinsic *per capita* rates are still b_0 and d_0, and analyze changes in density dependence parameter:

(3.1) For the reduced-mortality resistance mechanism to -cidal drugs, the density dependence parameter changes back to the original γ. This effect could occur through a mutation in the protein by which the drug gains entry to the cell, blocking the mechanism of drug action.

(3.2) For the enhanced-fecundity resistance mechanism to -cidal drugs, we require that (i) the intrinsic *per capita* birth and death rates are still b_0 and d_0, (ii) the population's carrying capacity changes to K (from K/m), and (iii) the *per capita* death rate remains $d_0 + (m-\gamma)\frac{r}{K}N$, so the density dependence parameter becomes $(1-m+\gamma)$ and the resistant *per capita* birth rate changes to $b_0 - (1-m+\gamma)\frac{r}{K}N$, which is always greater than the -cidal *per capita* birth rate $b_0 - \gamma\frac{r}{K}N$ because $(1-m) < 0$. Since we only consider negative density dependence, we also require $0 \leq 1-m+\gamma \leq 1$, which implies $m \leq \gamma+1 \leq 1+1 = 2$. Hence, for $m > \gamma+1$, it is not possible to have enhanced-fecundity resistance to -cidal drug through only competitive interactions. In order for the resistance to happen for the case $m > \gamma+1$ and the intrinsic *per capita* birth and death rates to remain b_0 and d_0, we have to allow $(1-m+\gamma)$ to be negative, reflecting cooperative interactions that increase the birth rate. This observation is consistent with the results in[7,8], which show that cooperation between cells contributes to drug resistance.[3]

While the drug resistance mechanisms in (3.1) and (3.2) are different, they result in the same *net* growth rates (birth rates minus death rates).Our direct estimation method allows us to disambiguate the underlying mechanisms, despite the same mean-field cell number time series.

[3] If we still require only competitive interactions (no cooperation), then the intrinsic *per capita* birth rate has to become larger than b_0, suggesting a mutation-related mechanism. In particular, if we rewrite the *per capita* death rate as $d_0 + \left(1-\frac{\gamma}{m}\right)\frac{(mr)}{K}N$ and interpret the density dependence parameter as $\frac{\gamma}{m}$, then the intrinsic *per capita* birth rate has to be $b_0 + (m-1)r$.

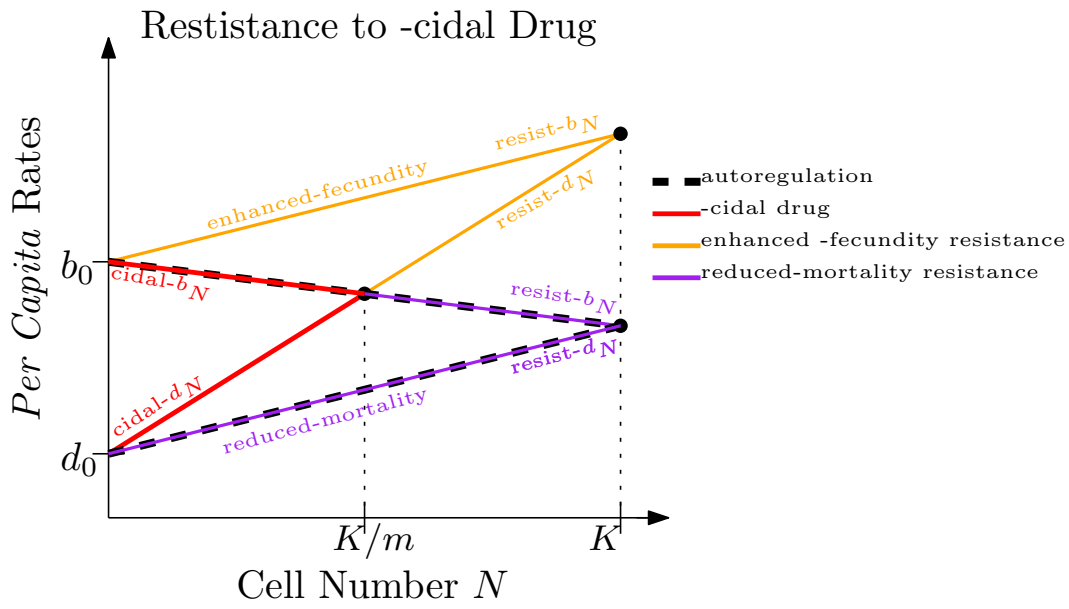

Figure 7. Schematic illustration of resistance to -cidal drug. -Cidal-b_N and -cidal-d_N are *per capita* birth and death rates under a -cidal drug; resist-b_N and resist-d_N are resistant *per capita* birth and death rates.

.

The same applies to resistance mechanisms to -static drugs:

(3.3) For enhanced-fecundity resistance mechanism to -static drugs, the density dependence parameter changes back to the original γ. Like a reduced-mortality resistance strategy versus a -cidal drug, this effect could occur through a mutation in the protein by which the drug gains entry to the cell, blocking the mechanism of drug action.

(3.4) For reduced-mortality mechanism resistance to -static drugs, since we require that (i) the intrinsic *per capita* birth and death rates are still b_0 and d_0, (ii) the population's carrying capacity changes to K (from K/m), and (iii) the *per capita* birth rate remains $b_0-(\gamma+m-1)\frac{r}{K}N$, the density dependence parameter becomes $(\gamma+m-1)$ and hence the resistance *per capita* death rate is $d_0+(1-\gamma-m+1)\frac{r}{K}N$, which is always lower than the -static *per capita* death rate $d_0+(1-\gamma)\frac{r}{K}N$ because $m>1$. Since we only consider competitive interactions (i.e. negative density dependence), we require $0\leq(\gamma+m-1)\leq 1$, which implies $1\leq\gamma+m\leq 2$, which means $m\leq 2-\gamma<2$. Hence, for $m>2-\gamma$, it is not possible to have reduced-mortality resistance to -static drugs through only competitive interactions. In order for the resistance to happen for the case $m\geq 2$ and the intrinsic *per capita* birth and death rates remains b_0 and d_0, we have to allow $\Big(1-(\gamma+m-1)\Big)$ to be negative, reflecting cooperative interactions that reduce the death rate.[4] Note that if $(1-\gamma-m+1)<0$,

[4] If we still require only competitive interactions (no cooperation), then we have to decrease the intrinsic *per capita* death rate to lower than d_0. In particular, we rewrite the *per capita* birth rate as $b_0-\Big(\frac{\gamma+m-1}{m}\Big)\frac{(mr)}{K}$ and interpret the density dependence parameter as $\frac{\gamma+m-1}{m}$, then the intrinsic *per capita* death rate becomes $d_0-(m-1)r$. Unlike the enhanced-fecundity resistance to -cidal case, the new intrinsic *per capita* rate $d_0-(m-1)r$ can become negative when $m>\frac{b_0}{b_0-d_0}$, which is not biologically meaningful. Therefore, this d_0-related mechanism seems unlikely to be found in real populations.

then the *per capita* death rate $d_0 + (1-\gamma-m+1)\frac{r}{K}N$ can potentially get negative, which is not biologically meaningful.

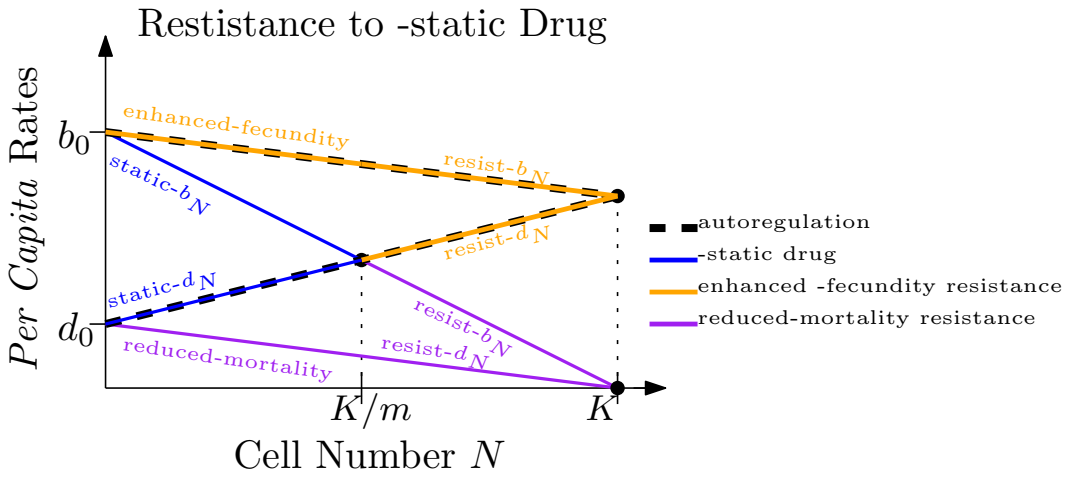

Figure 8. Schematic illustration of resistance to -static drug. -Static-b_N and -static-d_N are *per capita* birth and death rates under a -static drug; resist-b_N and resist-d_N are resistant *per capita* birth and death rates.

.

From the analysis, we observe that:

- While drug treatments can affect competitive interactions alone, in some cases, cell populations have to utilize other mechanisms such as cooperative interactions or mutation to defend against drug treatments.

- Resistance by decreasing intrinsic *per capita* death rates is unlikely, as discussed in (3.4), suggesting that mutation-induced effects are more likely to accomplish resistance through changes in birth rates.

Table 1. Summary of Dynamics and Mechanisms. In the table, b_0, d_0, K, γ denote the original parameters during the autoregulation stage before drug treatment. The carrying capacity under drug treatment is K/m, with $m > 1$. For the drug treatment and resistance stages, the density dependence parameters and intrinsic *per capita* birth and death rates are expressed in terms of the original parameters b_0, d_0, γ, K and the reduction factor m. For example, the intrinsic *per capita* birth rate for the enhanced-fecundity resistance mechanism to -cidal drugs is $\left(b_0 + (m-1)r\right)$ for $m > (1+\gamma)$ and density dependence parameter $\frac{\gamma}{m}$.

Dynamics	Other Parameters	Density Dependence
Autoregulation	K, b_0, d_0	$\gamma \in [0,1]$ (competition)
-Cidal Drug Effect	$K/m, b_0, d_0$	$\frac{\gamma}{m} \in \left[0, \frac{1}{m}\right], m > 1$ (more competition in death)
-Static Drug Effect	$K/m, b_0, d_0$	$\left(1 - \frac{1-\gamma}{m}\right) \in \left[1 - \frac{1}{m}, 1\right], m > 1$ (more competition in birth)
Reduced-Mortality Resistance (to -Cidal)	K, b_0, d_0	$\gamma \in [0,1]$ (less competition in death)
Enhanced-Fecundity Resistance (to -Cidal)	K, b_0, d_0	$(\gamma + 1 - m) \in [0, \gamma), m \in (1, \gamma+1]$ (less competition in birth)
	K, b_0, d_0	$(\gamma + 1 - m) \in [1-m, 0), m > \gamma + 1$ (increased birth due to cooperation)
	$K, b_0 + (m-1)r, d_0$	$\frac{\gamma}{m} \in \left[0, \frac{\gamma}{\gamma+1}\right), m > \gamma + 1$ (increased intrinsic birth)
Enhanced-Fecundity Resistance (to -Static)	K, b_0, d_0	$\gamma \in [0,1]$ (less competition in birth)
Reduced-Mortality Resistance (to -Static)	K, b_0, d_0	$(\gamma + m - 1) \in (\gamma, 1], m \in (1, 2-\gamma]$ (less competition in death)
	K, b_0, d_0	$(\gamma + m - 1) \in (0, \gamma), m > (2-\gamma)$ (decreased death due to cooperation)
	$K, b_0, d_0 - (m-1)r$	$\frac{\gamma + m - 1}{m} \in \left(0, \frac{\gamma}{m}\right)$, $(2-\gamma) < m \leq \frac{b_0}{b_0 - d_0}$ (decreased intrinsic death)

5.2 Illustration of Specific Scenarios Using Direct Estimation Method

In order to make the preceding analysis concrete, and to further validate our direct estimation method, we present detailed case studies of each of the scenarios summarized above, using specific parameter values. Figure 9 illustrates the autoregulation stage; Figure 10 illustrates the drug treatment stage; Figures 11 and 12 illustrate the resistance to -cidal and -static drug stage. For all of these figures, we simulate 100 trajectories of the cell population with an initial population size $N(t_0) = 10$ and the same parameter

values as in Section 4 and Table 3. In addition to using $K = 10^5$ for carrying capacity, we also simulate the population growth with a carrying capacity $K = 10^2$ for Figure 10 and $K = 10^3$ for Figures 11 and 12 to better visualize the noise levels. The method works robustly for larger values of K as well. After simulating an ensemble of cell number trajectories, we estimate birth and death rates from that ensemble of trajectories using the method given in section 3.3. The birth rate $b_{N(t)}$ and death rate $d_{N(t)}$ as functions of time are computed by treating the rates as composite functions of the cell number $N(t)$, and finding the rates that correspond to the (randomly) selected cell number time series.

For the autoregulation stage, Figure 9 shows how our direct estimation method can disambiguate the three autoregulation scenarios: (I) $\gamma = 0$, (II) $\gamma = 0.5$, and (III) $\gamma = 1$. We demonstrate that when the carrying capacity is small (e.g. 10^2), it is easier to see the noise levels than when the carrying capacity is large (e.g. 10^5), as seen in Figure 9 **(C)** and **(D)**, because the fluctuations are larger relative to the mean population. When the same net growth rate arises from different density-dependent mechanisms, at the level of birth and death rates, the birth and death rates as functions of *time* can appear markedly different. For example, while in scenarios (I) and (II), the birth and death rates show monotonically increasing, sigmoidal shapes throughout time, in scenario (III), the birth rate has the shape of a concave-down quadratic function as shown by the "+" magenta curves in Figure 9 **(E)** and **(F)**.

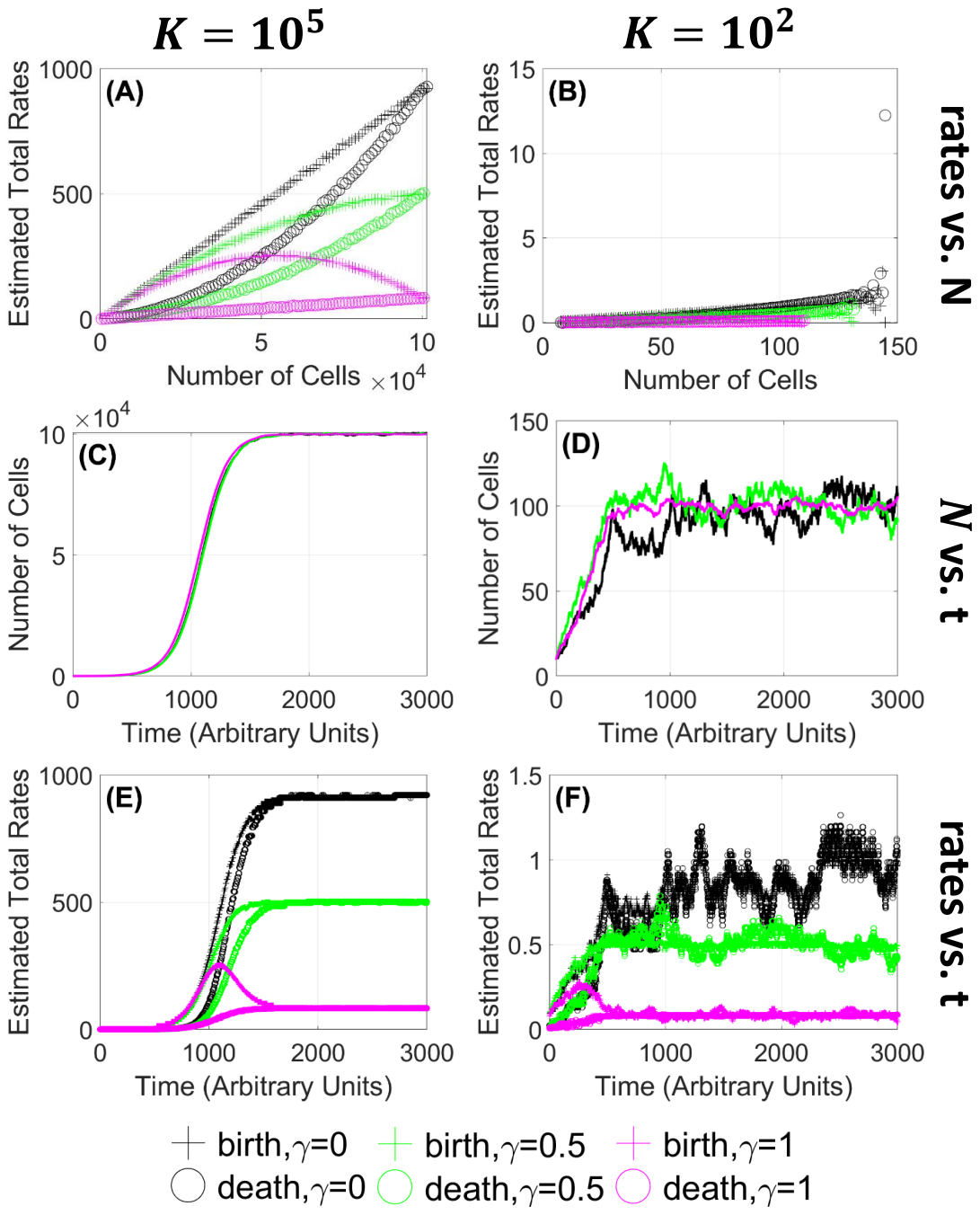

Figure 9. Underlying autoregulation mechanisms are distinguished by separately identified birth and death rates, not necessarily by net changes in total population size. Plots showing that cell populations with the same net growth rate and carrying capacity grow to the carrying capacity under different density-dependent mechanisms, although the observed dynamics (shown in (C) and (D)) look indistinguishable. Noise levels are more visible for smaller carrying capacities due to smaller scales. **(A, C, E)**: logistic birth-death processes with carrying capacity $K = 10^5$. **(B, D, F)**: logistic birth-death processes with carrying capacity $K = 10^2$. **(A-F)**: black curves correspond to the scenario $\gamma = 0$; green curves correspond to the scenario $\gamma = 0.5$; magenta curves correspond to the scenario $\gamma = 1$. **(A, B)**: estimated birth and death rates for three scenarios using the direct estimation method with an ensemble of 100 trajectories. **(C, D)**: one selected trajectory for each scenario. **(E, F)**: estimated birth and death rates throughout time corresponding to the trajectories in **(C)** and **(D)**. Plus signs $(+)$ denote estimated birth rates; circles $(\circ)$ denote estimated death rates.

For the drug treatment stage, Figure 10 illustrates the effects of -cidal versus -static drugs in scenario (I) in the first column (panels **A, D, G**), scenario (II) in the second column (panels **B, E, H**), and scenario (III) in the third column (panels **C, F, I**). As is evident in Figure 10 **(D, E, F)**, the observed cell number dynamics can be very similar in each scenario. However, Figure 10 panels **(A, B, C)** and **(G, F, H)** show that the underlying birth and death processes that give rise to the dynamics can be very different. Specifically, in **(D, E, F)**, we see that the red and blue curves are almost indistinguishable. Thus, these scenarios could not easily be distinguished from the general shape of the growth curve alone. However, to obtain the red curves, we keep the *per capita* birth rates the same and increase the *per capita* death rates, and to obtain the blue curves, it is the other way around–as illustrated in panels **(A, B, C)**. The time-dependent birth and death rates in panels **(G, H, I)** also show significant differences. In particular, the *per capita* birth rates under the "-static" drug treatment (blue curves) are monotonically increasing in scenario (I) (density-dependent death rate, as shown in **(G)**), but show a pronounced increase and then decrease in scenario (III), as shown in **(I)**. Thus, by extracting birth and death rates separately from cell number time series, we are able to disambiguate underlying drug mechanisms.

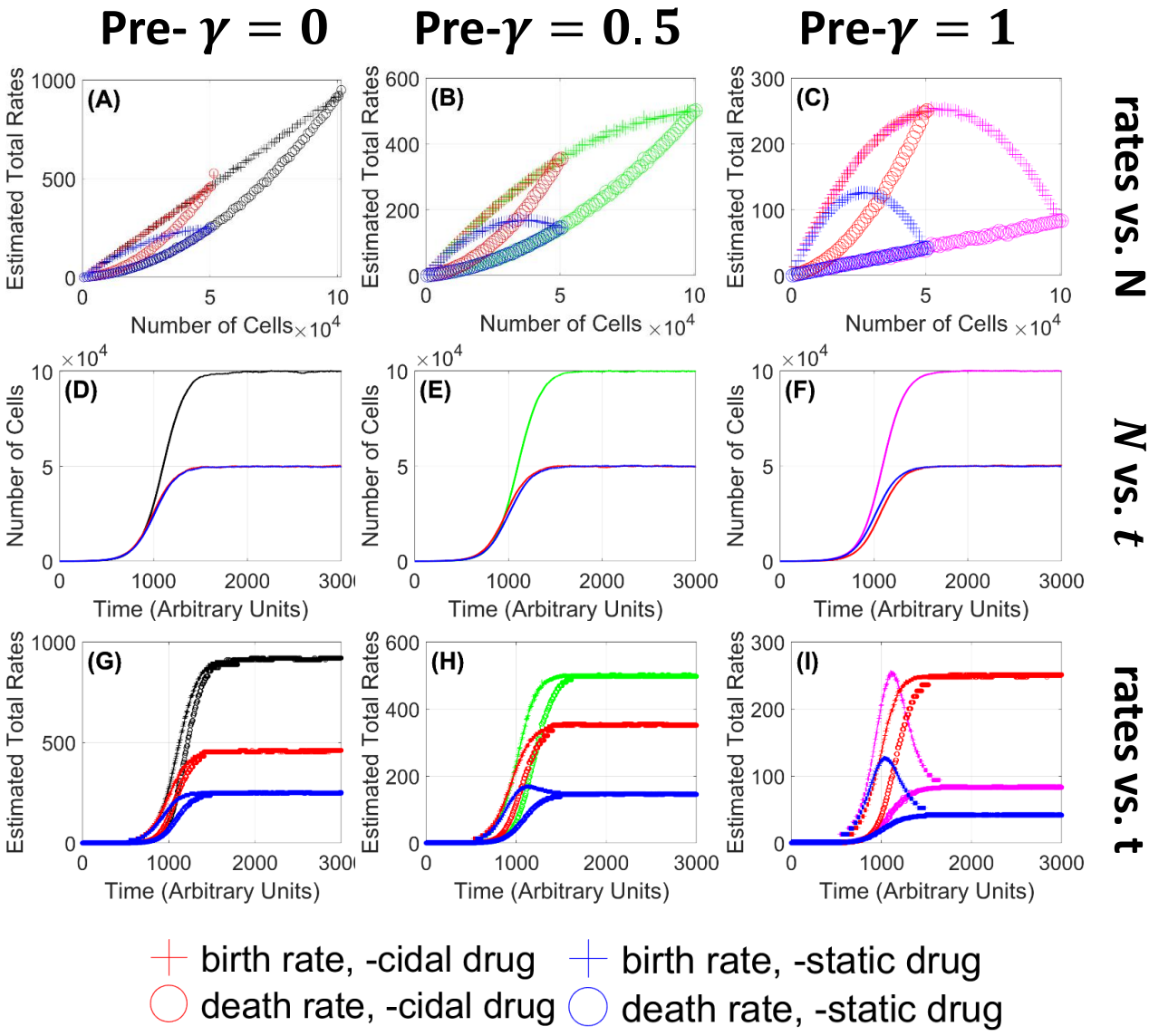

Figure 10. Separating birth and death rates distinguishes the underlying -cidal versus -static action of drugs. In each of the density-dependent cases (I), (II), (III), two different drugs reduce the cell population's carrying capacity to the same level (shown in red and blues curves in (D), (E), (F)), but the underlying mechanisms are different: increasing death rates (red curves) or decreasing birth rates (blue curves). Black, green, and magenta curves represent scenarios (I), (II), and (III) without drugs. Red curves represent the scenarios under a "-cidal" drug, and blue curves represent the scenarios under a "-static" drug. Plus signs $(+)$ denote estimated birth rates; circles $(\circ)$ denote estimated death rates. **(A, D, G)**: scenario (I) with $\gamma = 0$ (black curves), **(B, E, H)**: scenario (II) with $\gamma = 0.5$ (green curves), **(C, F, I)**: scenario (III) with $\gamma = 1.0$ (magenta curves). **(A, B, C)**: birth and death rates estimated from 100-trajectory ensembles. **(D, E, F)**: a representative trajectory without drug and two representative trajectories treated with drugs. The red and curves trajectories have the same mean-field behavior but the drug mechanisms are different. **(G, H, I)**: estimated birth and death rates throughout time corresponding to the trajectories in **(D, E, F)**.

For the drug resistance stage, Figure 11 shows that having been treated with a "-cidal" drug, the cell population can develop resistance either by decreasing its death rate and reverting to its original dynamics (the red curves change back to the black, green, and magenta curves for scenarios (I), (II), and (III) respectively), or by increasing its *per capita* birth rate as illustrated by the cyan curves. Without computing the birth and death rates explicitly, we observe from cell number time series that if the resistant cell population (cyan curves) reaches its original carrying capacity earlier than the wild-type population (black, green, magenta curves) as in Figure 11 **(B, D)** or if the typical fluctuations around the mean population size are visibly larger than the fluctuations of the wild-type as in Figure 11 **(F)**, we may hypothesize that the population has developed drug resistance through the enhanced-fecundity mechanism.

Figure 12 shows that having been treated with a "-static" drug, the cell population can develop resistance either by increasing its birth rate and reverting to its original dynamics (the blue curves change back to the black, green, and magenta curves for scenarios (I), (II), and (III) respectively), or by decreasing its *per capita* death rate as illustrated by the cyan curves. We note that the decreased *per capita* death rate can become algebraically negative and not biologically meaningful, which is consistent with the fact that drug resistance has previously been considered mainly for "-cidal" drugs, not "-static" drugs, in the literature, see[38]. However, in contrast to some recent literature[38], in this paper, we propose the possibility of mechanisms through which cell populations can overcome the "-static" effect (birth inhibition) of drugs–that is, increasing the *per capita* birth rates back to the original rates, as seen in the black, green, and magenta curves in Figure 12. For instance if, through preexisting genetic variation, the cell population contained a mutant with an alternative sequence for the protein by which the drug targets the cell, then as this variant propagated in favor of the principal variant, the cell line could develop resistance to the "-static" drug. It is interesting to observe in Figure 12 **(G)** that even after being with a "-static" drug that inhibits birth, the cell population can develop resistance by reducing birth rates *throughout time*–as we can see the cyan curves are lower than the blue curves as time increases. We note that for scenario (I) where pre-$\gamma = 0$, we observe a second possible drug resistance mechanism, in which the cell population decreases its *per capita* death rate without making it negative. In this scenario, the cell population also changes its density dependence parameter from pre-$\gamma = 0$ to pre-$\gamma = 1$ as it becomes resistant to the "-static" drug.

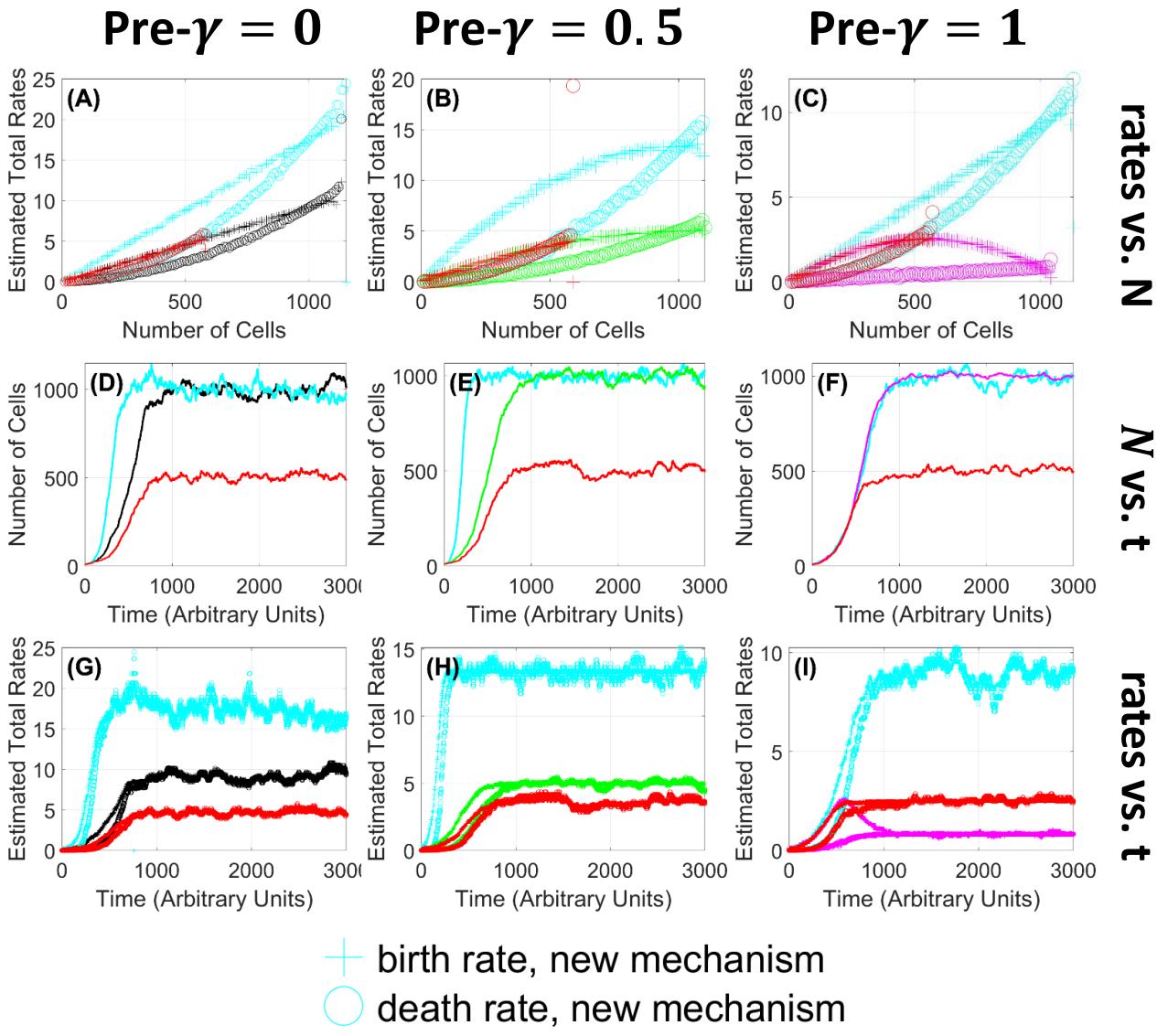

Figure 11. Resolving separate birth versus death rates distinguishes different underlying mechanisms of resistance to -cidal drugs. In each of the three density-dependent scenarios (I), (II), (III), a cell population can restore its carrying capacity after a "-cidal" drug treatment via different mechanisms: decreasing death rate to return to the original dynamics (shown in the black, green, magenta curves) or increasing birth rate (shown in the cyan curves). **(A, B, C)**: estimated birth and death rates using an ensemble of 100 cell number trajectories. Plus signs $(+)$ denote estimated birth rates; circles $(\circ)$ denote estimated death rates. **(D, E, F)**: selected cell number trajectories. **(G, H, I)**: estimated birth and death rates corresponding to the cell number trajectories in **(D, E, F)**. **(A, D, G)**: scenario (I) where the density dependence pre-$\gamma = 0$. **(B, E, H)**: scenario (II) where the density dependence pre-$\gamma = 0.5$. **(C, F, I)**: scenario (III) where the density dependence pre-$\gamma = 1$. The red curves represent the case where the cell population has been treated with a "-cidal" drug. The black, green, and magenta curves represent the case where the cell population develops resistance by decreasing its *per capita* death rate and returns to the original dynamics for the scenarios (I), (II), and (III) introduced earlier. The cyan curves represent the case in which the cell population develops resistance by increasing its *per capita* birth rate.

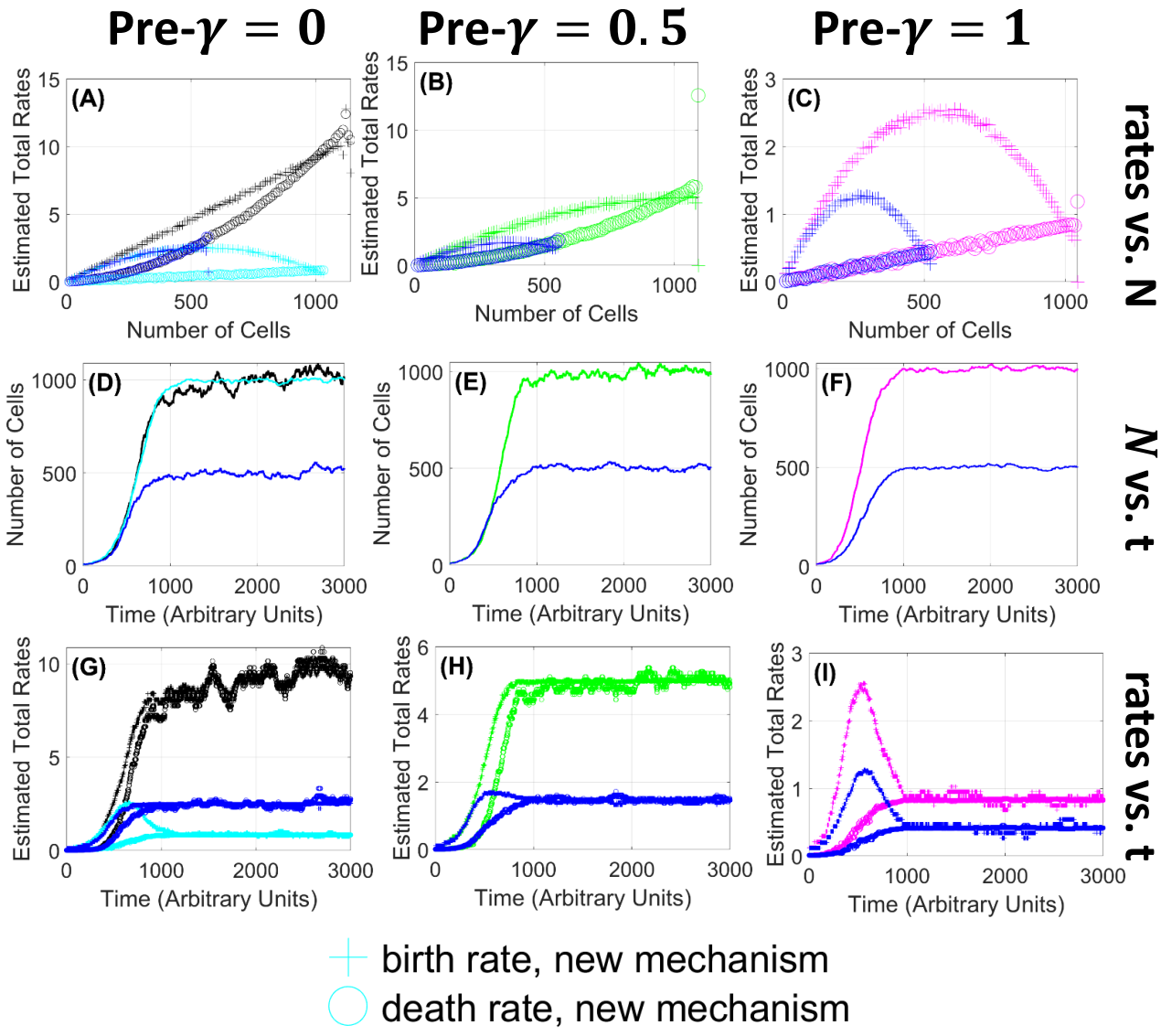

Figure 12. Resolving separate birth versus death rates distinguishes different underlying mechanisms of resistance to -static drugs. A cell population can restore its carrying capacity after a "-static" drug treatment via different mechanisms: increasing birth rate to return to the original dynamics (shown in the black, green, magenta curves) or decreasing death rate (shown in the cyan curves). The latter can happen only for scenario (I) where originally, the density dependence is fully in the death rate. **(A, B, C)**: estimated birth and death rates using an ensemble of 100 cell number trajectories. Plus signs $(+)$ denote estimated birth rates; circles $(\circ)$ denote estimated death rates. **(D, E, F)**: selected cell number trajectories. **(G, H, I)**: estimated birth and death rates corresponding to the cell number trajectories in **(D, E, F)**: selected cell number trajectories. **(A, D, G)**: scenario (I) where the density dependence pre-$\gamma = 0$. **(B, E, H)**: scenario (II) where the density dependence pre-$\gamma = 0.5$. **(C, F, I)**: scenario (III) where the density dependence pre-$\gamma = 1$. The blue curves represent the case where the cell population has been treated with a "-static" drug. The black, green, and magenta curves represent the case where the cell population develops resistance by decreasing its *per capita* death rate and returns to the original dynamics for the scenarios (I), (II), and (III) introduced earlier. The cyan curves represent the case in which the cell population develops resistance by increasing its *per capita* birth rate.

6 Likelihood-Based Statistical Inference

In the instance that there is only a single cell number trajectory, we would like to be able to assert how likely it is that the time series belongs to one of several scenarios, parameterized by the density dependence parameter γ. This question leads us to consider a maximum likelihood approach.

Let a cell number time series $\boldsymbol{X}_{\text{data}} = [x_0, x_1, \ldots, x_T]$ be a realization for the random variable $\boldsymbol{N} = [N(t_0), N(t_1), \ldots, N(t_T)]$. For clarity, we denote x_j as $x_{j,\text{data}}$, $j = 0, \ldots, T$. In Section 3.2, we approximate that $\boldsymbol{N}$ follows a normal distribution [5] characterized by the parameter set $\Theta_{\text{data}} = \{b_{0,\text{data}}, d_{0,\text{data}}, \gamma_{\text{data}}, K_{\text{data}}\}$. These parameters determine the birth and death rates of the birth-death process from which the time series is generated. In particular, the *per capita* birth and death rates are defined as follows:

$$b_{x_{j,\text{data}}} = \max\left\{b_{0,\text{data}} - \gamma_{\text{data}}\frac{r_{\text{data}}}{K_{\text{data}}}x_{j,\text{data}}, 0\right\}, \tag{40}$$

$$d_{x_{j,\text{data}}} = d_{0,\text{data}} + (1-\gamma_{\text{data}})\frac{r_{\text{data}}}{K_{\text{data}}}x_{j,\text{data}}, \tag{41}$$

where $r_{\text{data}} = b_{0,\text{data}} - d_{0,\text{data}}$. Recall that the mean and variance of the cell number increment $(N_{k+1} - N_k | N_k, \Delta t)$ are $(b_{N_k} - d_{N_k})N_k\Delta t$ and $(b_{N_k} + d_{N_k})N_k\Delta t$ respectively.

To test whether a given time series $\boldsymbol{X}_{\text{data}}$ belongs to a scenario characterized by the parameter set $\Theta_{\text{test}} = \{b_{0,\text{test}}, d_{0,\text{test}}, \gamma_{\text{test}}, K_{\text{test}}\}$, we evaluate the log likelihood function at the time series:

$$\mathscr{L}(\boldsymbol{X}_{\text{data}}|\Theta_{\text{test}}) = \underbrace{\ln\left(P(x_{0,\text{data}})\right)}_{=\ln(1)} + \sum_{j=1}^{T-1}\ln\left(P(x_{j+1,\text{data}}|x_{j,\text{data}}, \Theta_{\text{test}})\right) \tag{42}$$

$$= \sum_{j=1}^{T-1}\frac{1}{2}\ln\left(\frac{1}{2\pi x_{j,\text{data}}(b_{\text{test},x_{j,\text{data}}} + d_{\text{test},x_{j,\text{data}}})\Delta t}\right) \tag{43}$$

$$- \frac{1}{2}\frac{\left(x_{j+1,\text{data}} - x_{j,\text{data}} - x_{j,\text{data}}(b_{\text{test},x_{j,\text{data}}} - d_{\text{test},x_{j,\text{data}}})\Delta t\right)^2}{x_{j,\text{data}}(b_{\text{test},x_{j,\text{data}}} + d_{\text{test},x_{j,\text{data}}})\Delta t}, \tag{44}$$

where

$$b_{\text{test},x_{j,\text{data}}} = \max\left\{b_{0,\text{test}} - \gamma_{\text{test}}\frac{r_{\text{test}}}{K_{\text{test}}}x_{j,\text{data}}, 0\right\}, \tag{45}$$

$$d_{\text{test},x_{j,\text{data}}} = d_{0,\text{test}} + (1-\gamma_{\text{test}})\frac{r_{\text{test}}}{K_{\text{test}}}x_{j,\text{data}}, \tag{46}$$

$$r_{\text{test}} = b_{0,\text{test}} - d_{0,\text{test}}. \tag{47}$$

Suppose we know $b_{0,\text{data}}, d_{0,\text{data}}$, and K_{data}. That is, suppose that $b_{0,\text{test}} = b_{0,\text{data}}$, $d_{0,\text{test}} = d_{0,\text{data}}$, and $K_{\text{test}} = K_{\text{data}}$. Given one cell number time series, to infer which density dependence scenario the data mostly likely belongs, we treat the log-likelihood function as a function f of $\gamma := \gamma_{\text{test}}$, and find $\gamma \in [0,1]$ that maximizes $f(\gamma)$. We thus formulate a constrained nonlinear optimization problem as follows:

$$\max_{\gamma} f(\gamma) = \mathscr{L}(\boldsymbol{X}_{\text{data}}|\Theta_{\text{test}}) \quad \text{subject to} \quad 0 \le \gamma \le 1. \tag{48}$$

[5] This approximation requires sufficiently large summed birth and death rates.

For shorter notation, here we denote $b_{0,\text{test}}$ and $b_{0,\text{data}}$ as b_0, $d_{0,\text{test}}$ and $d_{0,\text{data}}$ as d_0, K_{test} and K_{data} as K, and $x_{j,\text{data}}$ as x_j. We calculate the first derivative $df/d\gamma$ in Appendix B and find critical points by solving $\frac{df}{d\gamma} = 0, \gamma \in [0,1]$ numerically using the Bisection method on the interval $[0-\Delta\gamma, 1+\Delta\gamma], \Delta\gamma = 0.5 > 0$. For a discussion on the maximality of the critical points, please refer to Appendix B.

Given multiple samples of cell number time series (e.g. from multiple experimental trials), we obtain an empirical distribution of solutions γ to the optimization problem (48). In Figure 13, for each of the three scenarios (I) $\gamma_{\text{data}} = 0$, (II) $\gamma_{\text{data}} = 0.5$, and (III) $\gamma_{\text{data}} = 1$, we plot the results upon solving the optimization problem 100 times for 100 independent time series, and obtain a distribution of estimated γ parameters. In addition we obtain a distribution of the estimation error, defined as the absolute difference $(\gamma_{\text{data}} - \gamma_{\text{estimated}})$, where $\gamma_{\text{estimated}}$ is the numerical solution to the optimization problem (48). The values of the parameters b_0, d_0, and K used in data simulation are the same as in Section 4. The empirical means and variances of the estimated γ values and estimation errors for the three scenarios (I), (II), and (III) are listed in Table 2.

Table 2. Numerical solution to the optimization problem (48) for a time series of length $T = 90,000$ points (timestep $\Delta t = 1/30$, total time 3000 arbitrary units).

True γ Value	Mean Estimated γ	Variance of Estimated γ	Mean Error	Error Variance
0	0.0010	1.4858×10^{-4}	-0.0010	1.4858×10^{-4}
0.5	0.4996	8.4605×10^{-5}	2.7218×10^{-5}	8.4605×10^{-5}
1	1.0000	3.6839×10^{-5}	-2.8675×10^{-5}	3.6839×10^{-5}

We note that the mean values of γ for the three scenarios (I), (II), and (III) are separated by margins that are an order of magnitude larger than the standard errors of the estimates. Thus, for the data generated by our birth/death simulations, the distribution the density-dependent effects can clearly be distinguished in terms of fully a birth-rate effect, fully a death-rate effect, or an evenly mixed effect.

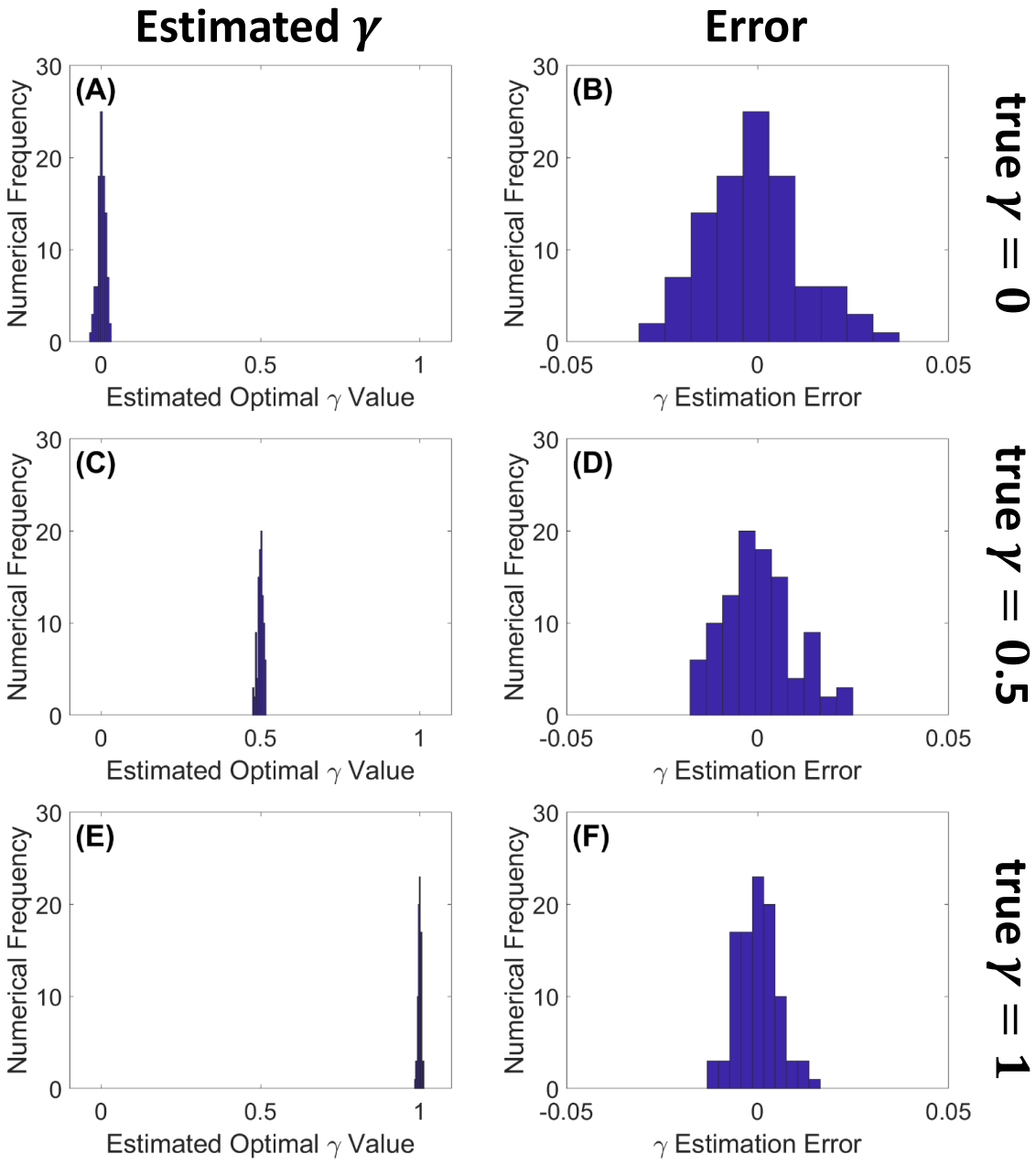

Figure 13. Numerical solutions to the optimization problem for different density dependence scenarios are clearly separated. We plot empirical distributions of estimated γ values, and corresponding errors, for an ensemble of individual cell number time series. **(A, B)**: cell number time series simulated with $\gamma = 0$; **(C, D)**: cell number time series simulated with $\gamma = 0.5$; **(E, F)**: cell number time series simulated with $\gamma = 1$. **(A, C, E)**: distributions of estimated γ; **(B, D, F)**: distributions of the corresponding estimation errors, defined as $(\gamma_{\text{data}} - \gamma_{\text{estimated}})$. We observe that the distributions of the estimated γ have small variances, and the errors are approximately normally distributed. Here we used time series of length $T = 90,000$ points (timestep $\Delta t = 1/30$, total time 3000).

7 Conclusion and Discussion

In order to infer density-dependent population dynamics mechanisms from data, we separately identify density-dependent *per capita* birth and death rates from net growth rates using the method described in Section 3 and infer whether density dependence is manifest in the birth process, death process, or some combination of the two. Our method involves directly estimating the mean and variance of cell number increments, as functions of population size, and expressing birth and death rates in terms of these two statistics. In order to obtain the mean and variance with tolerable accuracy, we compute them from an ensemble of cell number time series (e.g. multiple experiments). We analyze the accuracy of this method and derive analytical expressions for the theoretical expected errors and variance of errors in estimating birth and death rates as functions of the bin size (details are in Appendix A). We discover that small bin sizes do not necessarily result in small errors in estimating birth and death rates, due to small sample sizes. In fact, we find that intermediate bin sizes are optimal. Our error analysis also shows that if the product $r\Delta t$ is large relative to the carrying capacity K, then the expected error in estimating the mean cell number increment is high, as shown in Equation (95), which suggests that the estimation is not as good for fast-producing cell types.

Our method is distinct from other methods in the literature. It provides a novel perspective on the problem of stochastic parameter identification. Existing methods typically require numerical solution of a high-dimensional optimization problem, e.g. in a Bayesian inverse problem setting[39] or a likelihood function maximization framework. Crawford et al. 2014[32] constructs an expectation-maximization algorithm to identify birth and death rates for general birth-death processes. This method enjoys fast convergence and benefits from an elegant formulation of conditional expectations in terms of convolutions of transition probabilities. Their approach results from solving a maximum likelihood problem and is designed for large time periods Δt. In contrast, we suggest a nonparametric direct estimation approach that accurately extracts birth and death rates from the conditional first and second moments of the cell number time series data sampled at sufficiently small time periods. Our method does not require *a priori* assumption of specific forms of the rates. Aside from[32], to the best of our knowledge, other work addressing disambiguation of birth and death rates has been confined to *linear* birth-death processes. For example,[33] uses a Bayesian approach to parameter estimation for linear birth-death models in order to quantify the effects of changing drug concentrations. Here, we also consider different drug treatment scenarios, but in the context of nonlinear, logistic population models rather than linear growth models.[34,40] estimate birth and death rates as functions of time for a continuous-time branching process. Their method applies to multi-type cell populations and is illustrated with density-independent *per capita* birth and death rates. In contrast, our framework encompasses density-dependent *per capita* rates.

Our direct estimation method based on Equations (14) and (15) requires an ensemble of cell number time series. To deal with the sample size issue, we provide two solutions. The first solution is to utilize discretization as described in Section 3.3. We provide a rigorous error analysis for this approach. As an alternative solution, given only one cell number trajectory, we also present a maximum likelihood approach, in which we evaluate the log-likelihood function and maximize it over the density dependence parameter $\gamma \in [0,1]$. This approach, which involves solving a one-dimensional constrained nonlinear optimization problem, is limited to the assumption that the other system parameters are known.

The significance of both approaches is the application to studying treatments of pathogens and their resistance to the treatments. Specifically, in Section 5, we consider the scenario where a homogeneous

cell population goes through three stages: (1) grows naturally to its carrying capacity, (2) is treated with a drug that reduces its carrying capacity, and (3) overcomes the drug effect to gain back its carrying capacity. Our method allows us to identify whether each stage happens through the birth process, death process, or some combination of the two. Our analysis contributes to disambiguating underlying mechanisms such as exploitation versus interference competition in ecology, bacteriostatic versus bactericidal antibiotics in clinical treatments, and enhanced-fecundity versus reduced-mortality in pathogens' defense against drug treatments, which we may define as drug resistance. The mechanisms shown in this paper can help explain biological phenomena and may suggest novel approaches for engineering synthetic biological systems. More microscopic mechanisms within the birth process or death process, such as inactivating mutations of the gene for p53 protein[41], are beyond the scope of the model in this paper.

In the derivation of the relations between birth/death rates and the mean/variance of cell number increments in Section 3, we assume that the birth and death processes are approximately independent. However, the birth and death processes cannot be truly independent, as a cell cannot both divide and die at the same time. Suppose we do not assume independence in data, which means our data simulations would include the nonzero covariance between the number of cells going through division and the number of cells going through death. The covariance is computed as follows. Let $I_+^{(p)}$ denote a binary random variable indicating whether the pth cell divides after a time period Δt: $I_+^{(p)} = 1$ in case the cell divides, otherwise $I_+^{(p)} = 0$. Similarly, let $I_-^{(q)}$ denote a binary random variable indicating whether the qth cell dies after a time period Δt. Then,

$$\Delta N_+ = \sum_{p=1}^{N} I_+^{(p)} \quad \text{and} \quad \Delta N_- = \sum_{q=1}^{N} I_-^{(q)}, \tag{49}$$

and (50)

$$\mathrm{Cov}\left(\Delta N_+, \Delta N_- \middle| N\right) = \mathrm{Cov}\left(\sum_{p=1}^{N} I_+^{(p)}, \sum_{q=1}^{N} I_-^{(q)}\right) \tag{51}$$

$$= \sum_{p=1}^{N}\sum_{q=1}^{N} \mathrm{Cov}\left(I_+^{(p)}, I_-^{(q)}\right) \tag{52}$$

$$= \sum_{q=1}^{N} \mathrm{Cov}\left(I_+^{(q)}, I_-^{(p)}\right) + \sum_{p=1}^{N}\sum_{q\neq p} \mathrm{Cov}\left(I_+^{(p)}, I_-^{(q)}\right) \tag{53}$$

$$= \sum_{q=1}^{N} \mathrm{Cov}\left(I_+^{(q)}, I_-^{(q)}\right), \text{ as the transitions of different cells } p \neq q \text{ are independent.} \tag{54}$$

$$= \sum_{q=1}^{N} \mathbb{E}\left[I_+^{(q)} I_-^{(q)}\right] - \sum_{q=1}^{N} \mathbb{E}\left[I_+^{(q)}\right]\mathbb{E}\left[I_-^{(q)}\right] \tag{55}$$

$$= 0 - \sum_{q=1}^{N} \mathbb{E}\left[I_+^{(q)}\right]\mathbb{E}\left[I_-^{(q)}\right], \text{ as } I_+^{(q)} I_-^{(q)} = 0, \text{ since a cell cannot simultaneously divide and die.} \tag{56}$$

$$= -N\left(b_N \Delta t\right)\left(d_N \Delta t\right) \tag{57}$$

$$= -N b_N d_N \Delta t^2, \tag{58}$$

which is the covariance of two multinomial random variables on three outcomes (birth, death, and neither), with parameter N and probability vector $(b_N\Delta t, d_N\Delta t, 1-(b_N+d_N)\Delta t)$. We could include the covariance term in our data simulation using Gaussian approximation and τ-leaping simulation as follows:

$$N(t+\Delta t) - N(t) = \Big(b_{N(t)} - d_{N(t)}\Big)N(t)\Delta t + \sqrt{\sigma^2}\,\Delta W(t), \tag{59}$$

$$\text{where } \sigma^2\Delta t = \Big(b_{N(t)} + d_{N(t)}\Big)N(t)\Delta t \underbrace{-N(t)b_{N(t)}d_{N(t)}\Delta t^2}_{\text{from nonzero covariance}} \text{ is the variance of } \Delta N(t). \tag{60}$$

We simulate cell number trajectories using Equation (59) and estimate birth and death rates with the independence assumption, as in Equations (13) with the approximation $\text{Var}[\Delta N|N] \approx \Big(b_N + d_N\Big)N\Delta t$. Although the data do not assume independence between the birth and death processes and our estimation does, we see that the estimated and true birth and death rates are well-aligned, showing our method described in Section 3 with the independence assumption is robust, as shown in Figure 14.

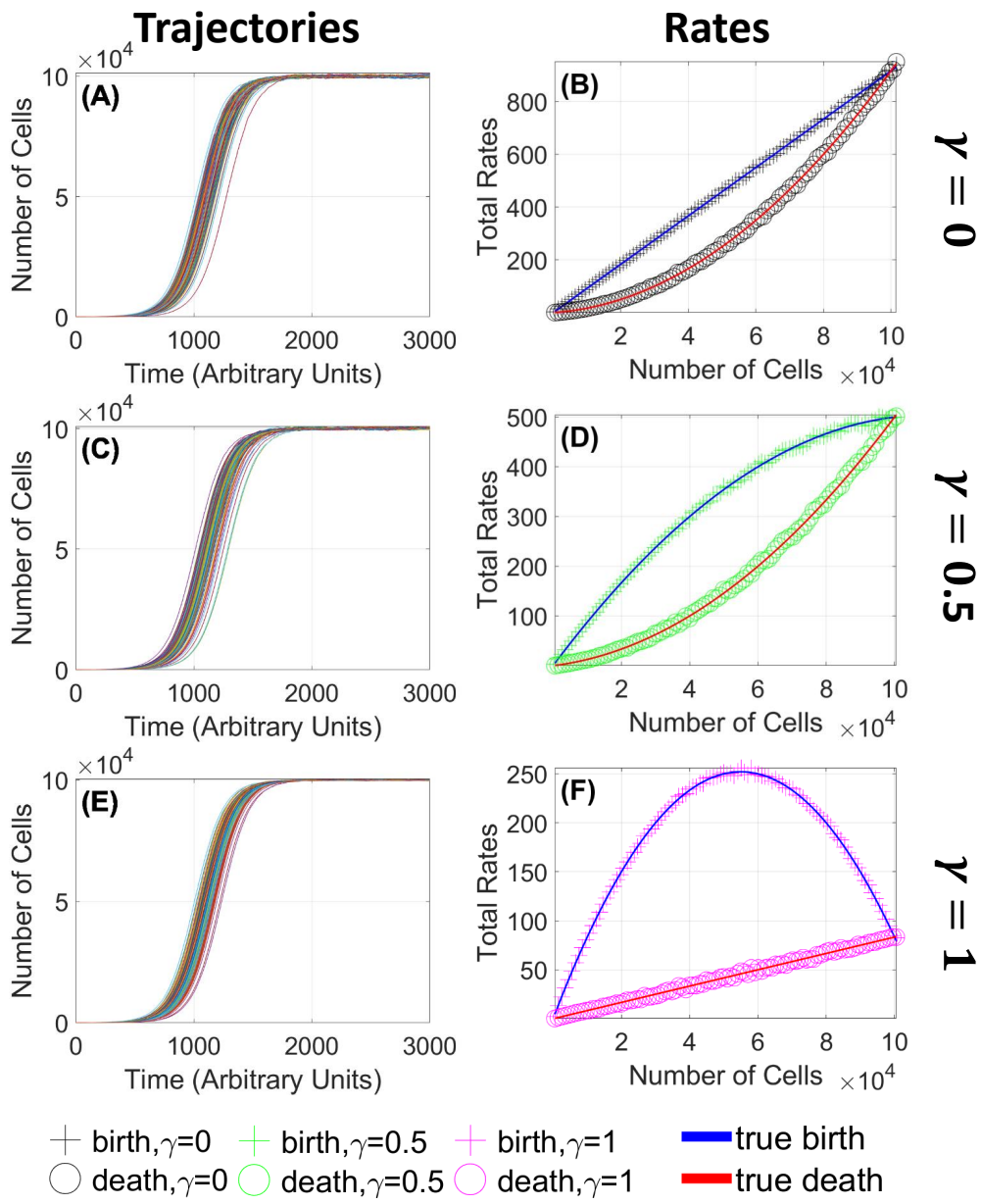

Figure 14. The estimation method works well even without the independence assumption in the simulated data. (A, C, E): Time series ensembles simulated using the τ-leaping approximation for the cases $\gamma = 0$ **(A)**, $\gamma = 0.5$ **(C)**, and $\gamma = 1$ **(E)** respectively. In the simulation, we do not assume independence between the birth and death processes, and include the nonzero covariance between the two processes. Each figure shows $S = 100$ trials. The estimated rates are computed using a bin size of $\eta = 10^3$. Carrying capacity $K = 10^5$ cells; low-density rates $b_0 = 1.1/120$ and $d_0 = 0$ (arbitrary time units); the other parameters are the same as in Section 4. **(B, D, F)**: Estimated and true birth and death rates, as functions of population size. Blue line: true birth rate. Red line: true death rate. Plus signs $(+)$ denote estimated birth rates; circles $(\circ)$ denote estimated death rates. Throughout the we will use distinct colors to denote values of γ. **(B)** Black: $\gamma = 0$; **(D)** Green: $\gamma = 0.5$; **(E)** Magenta: $\gamma = 1.0$. We observe that the estimated birth and death rates are well-aligned with the true birth and death rates used to simulate the trajectories in **(A)**, **(C)**, and **(E)**.

In Section 5, we show how to apply our method to distinguish the action of "-static" (birth-inhibiting) versus "-cidal" (death-promoting) drugs. However, the classification of drugs as being "-static" or "-cidal" is complicated by potentially stochastic factors such as external growth conditions[15]. For bacterial infections in a clinical setting, the "-static/-cidal" distinction is defined in terms of drug concentrations–specifically in terms of the ratio between Minimum Inhibitory Concentration (MIC) and Minimum Bactericidal Concentration (MBC). The Minimum Inhibitory Concentration (MIC) is defined as the lowest drug concentration that prevents visible growth, and the Minimum Bactericidal Concentration (MBC) is defined as the lowest drug concentration that results in a 99.99% decrease in the initial population size over a fixed period of time. Bacteriostatic drugs have been defined as those for which the ratio of the MBC to the MIC is larger than 4. Bactericidal drugs are those for which the ratio is ≤ 4[42]. Including the differential effects of drugs at larger or smaller concentrations will be an interesting direction for expanding our birth/death rate analysis in future work.

In Section 5, we use our direct estimation method to disambiguate different drug resistance mechanisms. In our paper, we define "drug resistance" as the cell population's ability to overcome the drug effect and gain back its original carrying capacity. However, the term "drug resistance" is used to mean different things in the research literature. For example, in Davison et al. 2000[43], drug resistance is defined in terms of the drug concentration needed to inhibit growth or kill the pathogen. Brauner et al. 2016[38] quantify cell populations overcoming drug effects in terms of MIC and the minimum time needed to kill the pathogens (MDK). Based on these two measures, MIC and MDK, the pathogens' defense against the drug can be called drug tolerance, persistence, or resistance. For future work, we will look into different definitions of "drug resistance".

For the present study, we confine our investigation to simulated data because of several factors. First, generating large ensembles of cell population trajectories is expensive, although high-throughput methods continue to accelerate the pace of data generation[37]. In a typical bioreactor, the data available are optical density time series, rather than direct cell number measurements. In theory, the relation between optical density and cell count is expected to be linear. Unfortunately, that is not always the case. McClure et al. 1993[44] show that it can be second order and Stephens et al. 1997[45] show that it can be third order. Moreover, Stevenson et al. 2016[46] show that the relation between cell count and optical density varies for different cell sizes and shapes, as well as other properties such as the index of refraction of the media. Some experimental calibration techniques have been developed to overcome these discrepancies, such as Francois et al. 2005[47] and Beal et al. 2020[48]. Finally, experimental data may include measurement noise that obscures finite population driven density fluctuations. Swain et al. 2016[49] attempts to estimate net growth rates from optical density data using a Gaussian process framework. In contrast, we would like disambiguate net growth rate into separate birth and death rates. Extending our method to take into account the mapping from cell number to noisy optical density measurements is an interesting subject for future work.

As mentioned in the Introduction (Section 1), throughout the paper, we interpret the density dependence term (interaction between individuals) as competition, which either reduces birth rates or increases death rates. However, in some situations, interactions among individuals can be cooperative, and increase the birth rate or reduce death rate with increasing population size[50]. To address this possibility, in future

work one might introduce to a cooperation parameter $c \geq 0$:

$$\frac{d\phi}{dt} = r\phi - \frac{r}{K}\phi^2 = r\phi + \underbrace{c\frac{r}{K}\phi^2}_{\text{cooperation}} - \underbrace{(1+c)\frac{r}{K}\phi^2}_{\text{competition}}. \tag{61}$$

One may interpret the cooperation term $c\frac{r}{K}\phi^2$ as a positive interaction between individuals that increases cell population growth. One could parameterize this term with parameter γ_c, to quantify how much of the cooperation increases birth and how much of the cooperation decreases death. Similarly, one may interpret the competition term $(1+c)\frac{r}{K}\phi^2$ as a negative interaction between individuals that reduces cell population growth. One could parameterize the competition term with parameter $\gamma_{\sim c}$ to quantify how much of the competition decreases birth and how much of the competition increases death:

$$\frac{d\phi}{dt} = r\phi + \underbrace{\gamma_c c\frac{r}{K}\phi^2}_{\text{cooperation}} + \underbrace{(1-\gamma_c)c\frac{r}{K}\phi^2}_{\text{cooperation}} - \underbrace{\gamma_{\sim c}(1+c)\frac{r}{K}\phi^2}_{\text{competition}} - \underbrace{(1-\gamma_{\sim c})(1+c)\frac{r}{K}\phi^2}_{\text{competition}} \tag{62}$$

$$= \underbrace{\left(b_0\phi + \gamma_c c\frac{r}{K}\phi^2 - \gamma_{\sim c}(1+c)\frac{r}{K}\phi^2\right)}_{\text{birth}} \tag{63}$$

$$- \underbrace{\left(d_0\phi - (1-\gamma_c)c\frac{r}{K}\phi^2 + (1-\gamma_{\sim c})(1+c)\frac{r}{K}\phi^2\right)}_{\text{death}}. \tag{64}$$

This study would provide a new perspective on modeling and analyzing the Allee effect and help disentangle positive and negative density dependence. Exploring these and other extensions provide interesting directions for future investigation.

References

1. Hixon, M. A. & Johnson, D. W. *Density Dependence and Independence. In: Encyclopedia of Life Sciences (ELS)* (John Wiley and Sons, Ltd: Chichester, 2009).
2. Karslake, J., Maltas, J., Brumm, P. & Wood, K. B. Population density modulates drug inhibition and gives rise to potential bistability of treatment outcomes for bacterial infections. *PLoS computational biology* **12**, e1005098 (2016).
3. Kaznatcheev, A., Peacock, J., Basanta, D., Marusyk, A. & Scott, J. G. Fibroblasts and alectinib switch the evolutionary games played by non-small cell lung cancer. *Nat. ecology & evolution* **3**, 450–456 (2019).
4. Paczkowski, M. *et al.* Reciprocal interactions between tumour cell populations enhance growth and reduce radiation sensitivity in prostate cancer. *Commun. Biol.* **4**, 1–13 (2021).
5. Bottery, M. J. *et al.* Inter-species interactions alter antibiotic efficacy in bacterial communities. *The ISME journal* **16**, 812–821 (2022).
6. Farrokhian, N. *et al.* Measuring competitive exclusion in non–small cell lung cancer. *Sci. Adv.* **8**, eabm7212 (2022).
7. Gutwillig, A. *et al.* Transient cell-in-cell formation underlies tumor relapse and resistance to immunotherapy. *Elife* **11**, e80315 (2022).
8. Emond, R. *et al.* Ecological interactions in breast cancer: Cell facilitation promotes growth and survival under drug pressure. *bioRxiv* (2021).
9. Verhulst, P. Notice sur la loi que la population suit dans son accroissement. *Corresp. mathematique et physique* **10**, 113–121 (1838).
10. Jesen, A. Simple models for exploitive and inference competition. *Ecol. Model.* **35**, 113–121 (1987).
11. Allee, W. & Bowen, E. S. Studies in animal aggregations: mass protection against colloidal silver among goldfishes. *J. Exp. Zool.* **61**, 185–207 (1932).
12. Kanarek, A. & Webb, C. Allee effects, adaptive evolution, and invasion success. *Evol. Appl.* **3**, 122–135, DOI: https://doi.org/10.1111/j.1752-4571.2009.00112.x (2010).
13. Drake, J. M. & Kramer, A. M. Allee effects. *Nat. Educ. Knowl.* **3**, 2 (2011).
14. Strang, A., Abbott, K. & Thomas, P. How to avoid an extinction time paradox. *Theor. Ecol.* **12**, 467–487 (2019).
15. Pankey, G. & Sabath, L. Clinical relevance of bacteriostatic versus bactericidal mechanisms of action in the treatment of gram-positive bacterial infections. *Clin. Infect. Dis.* **38**, 864–870, DOI: https://doi.org/10.1086/381972 (2004).
16. Lobritz, M. *et al.* Antibiotic efficacy is linked to bacterial cellular respiration. *Proc. Natl. Acad. Sci. United States Am.* **112**, 8173–8180, DOI: https://doi.org/10.1073/pnas.1509743112 (2015).
17. Frenoy, A. & Bonhoeffer, S. Death and population dynamics affect mutation rate estimates and evolvability under stress in bacteria. *Plos Biol.* **16**, DOI: https://doi.org/10.1371/journal.pbio.2005056 (2018).
18. Iwasa, Y., Michor, F. & Nowak, M. A. Evolutionary dynamics of escape from biomedical intervention. *Proc. Royal Soc. London. Ser. B: Biol. Sci.* **270**, 2573–2578 (2003).

19. Komarova, N. Stochastic modeling of drug resistance in cancer. *J. Theor. Biol.* **239**, 351–366, DOI: 10.1016/j.jtbi.2005.08.003 (2006).

20. Foo, J. & Michor, F. Evolution of resistance to anti-cancer therapy during general dosing schedules. *J. theoretical biology* **263**, 179–188 (2010).

21. Doebeli, M., Ispolatov, Y. & Simon, B. Towards a mechanistic foundation of evolutionary theory. elife 6 (february) (2017).

22. Bailey, N. T. *The elements of stochastic processes with applications to the natural sciences*, vol. 25 (John Wiley & Sons, 1991).

23. Norris, J. R. *Markov chains*. 2 (Cambridge university press, 1998).

24. Gardiner, C. *Stochastic methods*, vol. 4 (Springer Berlin, 2009).

25. Lei, X., Tian, W., Zhu, H., Chen, T. & Ao, P. Biological sources of intrinsic and extrinsic noise in ci expression of lysogenic phage lambda. *Sci. Reports* **5**, 1–12, DOI: 10.1038/srep13597 (2015).

26. Yoon, N., Veld, R. V., Marusyk, A. & Scott, J. Optimal therapy scheduling based on a pair of collaterally sensitive drugs. *Bull. Math. Biol.* **80**, DOI: https://doi.org/10.1007/s11538-018-0434-2 (2018).

27. Yoon, N., Krishnan, N. & Scott, J. Theoretical modeling of collaterally sensitive drug cycles: shaping heterogeneity to allow adaptive therapy. *J. Math. Biol.* **83**, 1–29 (2021).

28. Scarborough, J., Tom, M., Kattan, M. & Scott, J. Revisiting a null hypothesis: Exploring the parameters of oligometastasis treatment. *Int. J. Radiat. Oncol. Biol. Phys.* 1–11, DOI: https://doi.org/10.1016/j.ijrobp.2020.12.044 (2021).

29. Gerlee, P. The model muddle: in search of tumor growth laws. *Cancer research* **73**, 2407–2411 (2013).

30. Maltas, J. & Wood, K. B. Pervasive and diverse collateral sensitivity profiles inform optimal strategies to limit antibiotic resistance. *PloS Biol.* **17**, e3000515, DOI: https://doi.org/10.1371/journal.pbio.3000515 (2019).

31. Schwartz, H. R. *et al.* Drug grade: an integrated analysis of population growth and cell death reveals drug-specific and cancer subtype-specific response profiles. *Cell reports* **31**, 107800 (2020).

32. Crawford, F. W., Minin, V. N. & Suchard, M. A. Estimation for general birth-death processes. *J. Am. Stat. Assoc.* **109**, 730–747, DOI: https://doi.org/10.1080/01621459.2013.866565 (2014).

33. Liu, Y. & Crawford, F. W. Estimating dose-specific cell division and apoptosis rates from chemo-sensitivity experiments. *Sci. Reports* **8**, 2705, DOI: https://doi.org/10.1038/s41598-018-21017-5 (2018).

34. Ferlic, J. *Quantitative Approaches to Cancer and Cellular Differentiation*. Ph.D. thesis, Harvard University, Graduate School of Arts and Sciences (2019).

35. Allen, L. *An introduction to stochastic processes with applications to biology* (CRC Press, 2010).

36. Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *The J. chemical physics* **115**, 1716–1733 (2001).

37. Gopalakrishnan, V. *et al.* A low-cost, open-source evolutionary bioreactor and its educational use. *Elife* **11**, e83067 (2022).

38. Brauner, A., Fridman, O., Gefen, O. & Balaban, N. Q. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nat. Rev. Microbiol.* **14**, 320–330 (2016).

39. Calvetti, D. & Somersalo, E. *An introduction to Bayesian scientific computing: ten lectures on subjective computing*, vol. 2 (Springer Science & Business Media, 2007).

40. Roney, J. P., Ferlic, J., Michor, F. & McDonald, T. O. Estipop: a computational tool to simulate and estimate parameters for continuous-time markov branching processes. *Bioinformatics* **36**, 4372–4373 (2020).

41. Baguley, B. C. Multiple drug resistance mechanisms in cancer. *Mol. biotechnology* **46**, 308–316 (2010).

42. Wald-Dickler, N., Paul Holtom, P. & Spellberg, B. Busting the myth of "static vs cidal": A systemic literature review. *Clin. Infect. Dis.* **66**, 1470–1474, DOI: https://doi.org/10.1093/cid/cix1127 (2018).

43. Davison, H., Woolhouse, M. & Low, J. What is antibiotic resistance and how can we measure it? *Trends microbiology* **8**, 554–559 (2000).

44. McClure, P., Cole, B., Davies, K. & Anderson, W. The use of automated turbidimetric data for the construction of kinetic models. *J. Ind. Microbiol.* **12**, 277–285 (1993).

45. Stephens, P. *et al.* The use of an automated growth analyser to measure recovery times of single heat-injured salmonella cells. *J. Appl. Microbiol.* **83**, 445–455 (1997).

46. Stevenson, K., Alexander F. McVey, A. F., Clark, I., Swain, P. & Pilizota, T. General calibration of microbial growth in microplate readers. *Sci. Reports* **6**, DOI: 10.3934/proc.2011.2011.1279 (2016).

47. Francois, K. *et al.* Environmental factors influencing the relationship between optical density and cell count for listeria monocytogenes. *J. Appl. Microbiol.* **99**, 503–1515 (2005).

48. Beal, J. *et al.* Robust estimation of bacterial cell count from optical density. *Commun. Biol.* **3**, DOI: https://doi.org/10.1038/s42003-020-01127-5 (2020).

49. Swain, P. S. *et al.* Inferring time derivatives including cell growth rates using gaussian processes. *Nat. Commun.* DOI: DOI:10.1038/ncomms13766 (2016).

50. Bhowmick, A. R., Saha, B., Chattopadhyay, J., Ray, S. & Bhattacharya, S. Cooperation in species: Interplay of population regulation and extinction through global population dynamics database. *Ecol. Model.* **312**, 150–165 (2015).

51. Reding-Roman, C. *et al.* The unconstrained evolution of fast and efficient antibiotic-resistant bacterial genomes. *Nat. ecology & evolution* **1**, 1–11 (2017).

A Error Analysis of the Direct Estimation Method

As described in Section 3.3, we discretize all the values of cell number across the whole ensemble of trajectories into bins. Denote the bin size as η. The left end point N_k of the kth bin $[N_k, N_k+\eta)$ with $k = 1,2,\ldots,k_{\max}$ is equal to $N_k := N_{\min} + (k-1)\eta$, where $N_{\min}$ is the smallest value of cell number across the whole ensemble of trajectories. In many instances, $N_{\min} = N(t_0)$, the initial population size. The total number of bins $k_{\max} \in \mathbb{Z}^+$ is equal to $\left\lceil \frac{N_{\max} - N_{\min}}{\eta} \right\rceil$, where $N_{\max}$ is the largest value of cell number across the whole ensemble of trajectories, and $\lceil n \rceil$ is the smallest integer not less than n. The ith cell number element to have landed in the kth bin $[N_k, N_k+\eta)$ is equal to $N_k + \eta_i$. For simplicity, we make the approximation that for each bin, the random variables η_i are i.i.d. and uniformly distributed on $[0,\eta)$. We expect this approximation to be reasonably accurate when the bin size η is small enough that a given trajectory is unlikely to land in any particular bin twice in succession; the approximation may become inaccurate for excessively large bin sizes. In light of this uniform distribution assumption, we use the midpoint $N_k + \frac{\eta}{2}$ to represent the kth bin $[N_k, N_k+\eta)$.

We approximate the theoretical mean $\mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]$ with the empirical mean $\left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle$ and the theoretical variance $\mathrm{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]$ the empirical variance $\sigma^2\left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k\right]$ obtained from simulation of S cell number trajectories. Recall that $\hat{S}_k$ denotes the number of population size $N_k + \eta_i$ landing in bin k. These sample sizes $\hat{S}_k, k = 1,2,\ldots,k_{\max}$, are not necessarily equal to each other or equal to the number of cell number trajectories S, which is pre-determined and independent of the bin size η. Different bin sizes η result in different sets of $\hat{S}_k, k = 1,2,\ldots,k_{\max}$. With the same bin size η, different simulations may also result in different sets of cell number values and hence different sets of $\hat{S}_k, k = 1,2,\ldots,k_{\max}$. It is well-known that as the larger the sample size $\hat{S}_k$, the smaller the estimation errors[34].

In this section, we analyze how the bin size influences distributions of estimation errors of birth and death rates. In particular, we compute the theoretical means and variances of errors as functions of bin size η. We use the notation N for cell number to be consistent with the mathematical model discussed in Section 2. A summary of notations can be found in Section A.3.

A.1 Theoretical Mean and Variance of Cell Number Increment as Functions of Bin Size

As mentioned above, our estimation of the birth and death rates corresponding to $N = N_k + \frac{\eta}{2}$ uses the empirical mean $\left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle$ and empirical variance $\sigma^2\left[\Delta N | N = N_k + \eta_i] \middle| 0 \le \eta_i < \eta, \hat{S}_k\right]$. The theoretical means and variances of the estimation errors involves the theoretical mean $\mathbb{E}\left[\Delta N \middle| N = N_k + U, U \sim \mathrm{Unif}[0,\eta)\right]$ and theoretical variance $\mathrm{Var}\left[\Delta N \middle| N = N_k + U, U \sim \mathrm{Unif}[0,\eta)\right]$, as shown in Section 4. In this subsection, we analyze how the bin size η influences these theoretical mean and variance. We present the analysis for nonnegative birth rates, that is, in which we can drop the max function in Equation (5), as the birth rates are always positive in our simulated datasets.

Theoretical mean:

$$\mathbb{E}\Big[\Delta N\Big|N=N_k+U,U\sim\text{Unif}[0,\eta)\Big]=\mathbb{E}\Big[\mathbb{E}[\Delta N|N=N_k+U]\Big|U\sim\text{Unif}[0,\eta)\Big] \tag{65}$$

$$=\mathbb{E}\Big[(b_{N_k+U}-d_{N_k+U})(N_k+U)\Delta t\Big|U\sim\text{Unif}[0,\eta)\Big] \tag{66}$$

$$=\mathbb{E}\Big[(r-\frac{r}{K}N_k-\frac{r}{K}U)(N_k+U)\Delta t\Big|U\sim\text{Unif}[0,\eta)\Big] \tag{67}$$

$$=\mathbb{E}\Big[(r-\frac{r}{K}N_k)N_k\Delta t\Big|U\sim\text{Unif}[0,\eta)\Big]-\frac{r}{K}N_k\Delta t\mathbb{E}\Big[U\Big|U\sim\text{Unif}[0,\eta)\Big] \tag{68}$$

$$+(r-\frac{r}{K}N_k)\Delta t\mathbb{E}\Big[U\Big|U\sim\text{Unif}[0,\eta)\Big]-\frac{r}{K}\Delta t\mathbb{E}\Big[U^2\Big|U\sim\text{Unif}[0,\eta)\Big]$$

$$=\mathbb{E}[\Delta N\,|\,N=N_k]+(r-2\frac{r}{K}N_k)\Delta t\frac{\eta}{2}-\frac{r}{K}\Delta t\frac{\eta^2}{3}. \tag{69}$$

Theoretical variance:

$$\text{Var}\Big[\Delta N\Big|N=N_k+U,U\sim\text{Unif}[0,\eta)\Big]$$

$$=\mathbb{E}\Big[\Delta N^2\Big|N=N_k+U,U\sim\text{Unif}[0,\eta)\Big]-\Big(\mathbb{E}\Big[\Delta N\Big|N=N_k+U,U\sim\text{Unif}[0,\eta)\Big]\Big)^2, \tag{70}$$

where

$$\mathbb{E}\Big[\mathbb{E}[\Delta N^2|N=N_k+U]\Big|U\sim\text{Unif}[0,\eta)\Big]$$

$$=\mathbb{E}\left[\text{Var}[\Delta N|N=N_k+U]+\Big(\mathbb{E}[\Delta N|N=N_k+U]\Big)^2\Big|U\sim\text{Unif}[0,\eta)\right] \tag{71}$$

$$=\mathbb{E}\Big[\text{Var}[\Delta N|N=N_k+U]\Big|U\sim\text{Unif}[0,\eta)\Big]+\mathbb{E}\Big[\Big(\mathbb{E}[\Delta N|N=N_k+U]\Big)^2\Big|U\sim\text{Unif}[0,\eta)\Big], \tag{72}$$

and

$$\mathbb{E}\Big[\text{Var}[\Delta N|N=N_k+U]\Big|U\sim\text{Unif}[0,\eta]\Big]$$

$$=\mathbb{E}\Big[(b_{N_k+U}+d_{N_k+U})(N_k+U)\Delta t\Big|U\sim\text{Unif}[0,\eta)\Big] \tag{73}$$

$$=\mathbb{E}\Big[(b_0+d_0+(1-2\gamma)\frac{r}{K}N_k+(1-2\gamma)\frac{r}{K}U)(N_k+U)\Delta t\Big|U\sim\text{Unif}[0,\eta)\Big] \tag{74}$$

$$=\mathbb{E}\Big[(b_0+d_0+(1-2\gamma)\frac{r}{K}N_k)N_k\Delta t\Big|U\sim\text{Unif}[0,\eta)\Big]+\mathbb{E}\Big[(b_0+d_0+(1-2\gamma)\frac{r}{K}N_k)U\Delta t\Big|U\sim\text{Unif}[0,\eta)\Big] \tag{75}$$

$$+\mathbb{E}\Big[(1-2\gamma)\frac{r}{K}N_kU\Delta t\Big|U\sim\text{Unif}[0,\eta)\Big]+\mathbb{E}\Big[(1-2\gamma)\frac{r}{K}U^2\Delta t\Big|U\sim\text{Unif}[0,\eta)\Big]$$

$$=\text{Var}[\Delta N_k]+\Big(b_0+d_0+2(1-2\gamma)\frac{r}{K}N_k\Big)\Delta t\mathbb{E}[U|U\sim\text{Unif}[0,\eta)]+(1-2\gamma)\frac{r}{K}\Delta t\mathbb{E}[U^2|U\sim\text{Unif}[0,\eta)] \tag{76}$$

$$=\text{Var}[\Delta N_k]+\Big(b_0+d_0+2(1-2\gamma)\frac{r}{K}N_k\Big)\Delta t\frac{\eta}{2}+(1-2\gamma)\frac{r}{K}\Delta t\frac{\eta^2}{3}, \tag{77}$$

and

$$\mathbb{E}\Big[(b_{N_k+U}-d_{N_k+U})^2(N_k+U)^2\Delta t^2\Big|U\sim \text{Unif}[0,\eta)\Big]$$
$$=\Delta t^2\mathbb{E}\Big[\Big((b_0-d_0)(N_k+U)-\frac{r}{K}(N_k+U)^2\Big)^2\Big|U\sim \text{Unif}[0,\eta)\Big] \tag{78}$$
$$=\Delta t^2\mathbb{E}\Big[r^2(N_k+U)^2-2\frac{r^2}{K}(N_k+U)^3+\frac{r^2}{K^2}(N_k+U)^4\Big|U\sim \text{Unif}[0,\eta)\Big] \tag{79}$$
$$=\Delta t^2r^2\mathbb{E}\Big[(N_k+U)^2\Big|U\sim \text{Unif}[0,\eta)\Big]-2\frac{r^2}{K}\Delta t^2\mathbb{E}\Big[(N_k+U)^3\Big|U\sim \text{Unif}[0,\eta)\Big] \tag{80}$$
$$+\frac{r^2}{K^2}\Delta t^2\mathbb{E}\Big[(N_k+U)^4\Big|U\sim \text{Unif}[0,\eta)\Big]$$
$$=\Delta t^2r^2\frac{(N_k+\eta)^3-N_k^3}{3\eta}-2\frac{r^2}{K}\Delta t^2\frac{(N_k+\eta)^4-N_k^4}{4\eta}+\frac{r^2}{K^2}\Delta t^2\frac{(N_k+\eta)^5-N_k^5}{5\eta}. \tag{81}$$

Therefore,

$$\text{Var}\Big[\Delta N\Big|N=N_k+U,U\sim \text{Unif}[0,\eta)\Big]$$
$$=\text{Var}[\Delta N_k]+\Big(b_0+d_0+2(1-2\gamma)\frac{r}{K}N_k\Big)\Delta t\frac{\eta}{2}+(1-2\gamma)\frac{r}{K}\Delta t\frac{\eta^2}{3} \tag{82}$$
$$+\Delta t^2r^2\frac{(N_k+\eta)^3-N_k^3}{3\eta}-2\frac{r^2}{K}\Delta t^2\frac{(N_k+\eta)^4-N_k^4}{4\eta}+\frac{r^2}{K^2}\Delta t^2\frac{(N_k+\eta)^5-N_k^5}{5\eta}$$
$$-\left(\mathbb{E}[\Delta N\,|\,N=N_k]+(r-2\frac{r}{K}N_k)\Delta t\frac{\eta}{2}-\frac{r}{K}\Delta t\frac{\eta^2}{3}\right)^2.$$

In Figure 15, we compare the theoretical mean $\mathbb{E}\Big[\Delta N\Big|N=N_k+U,U\sim \text{Unif}[0,\eta)\Big]$ that we just computed with the theoretical mean $\mathbb{E}\Big[\Delta N\Big|N=N_k+\frac{\eta}{2}\Big]$ and the empirical mean $\Big\langle\Delta N\Big|N=N_k+U,U\sim \text{Unif}[0,\eta)\Big\rangle$ using data from a simulation of $S=100$ cell number trajectories. Similarly, we also compare the population variance $\text{Var}\Big[\Delta N\Big|N=N_k+U,U\sim \text{Unif}[0,\eta)\Big]$ that we just computed with the theoretical variance $\text{Var}\Big[\Delta N\Big|N=N_k+\frac{\eta}{2}\Big]$ and the empirical variance $\sigma^2\Big[\Delta N\Big|N=N_k+\eta_i,0\le\eta_i<\eta\Big]$ using data from a simulation of $S=100$ cell number trajectories.

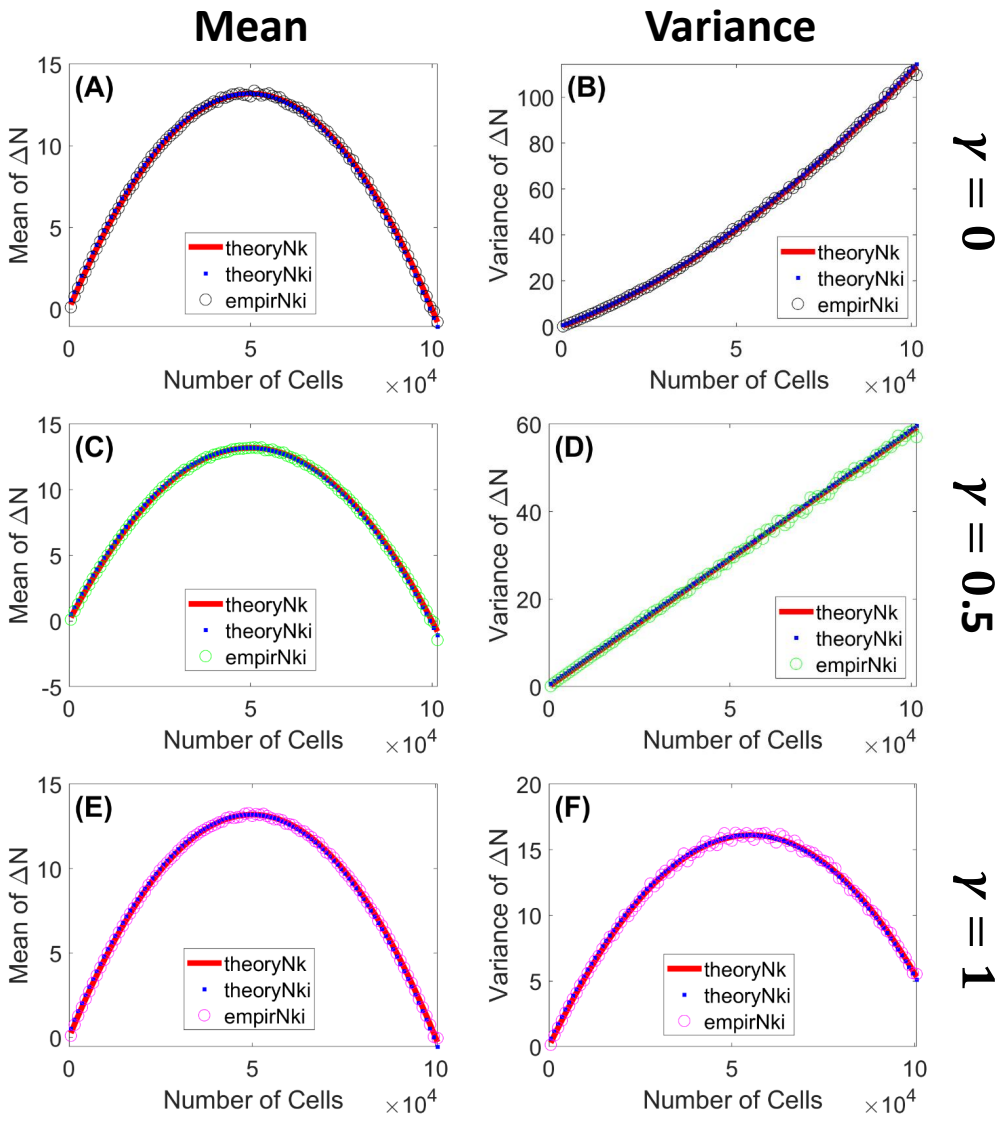

Figure 15. Theoretical mean and variance of cell number increments ΔN as functions of population size are well-aligned with empirical mean and variance. The statistics are computed using carrying capacity $K = 10^5$ and bin size $\eta = 10^3$. In **(A, C, E)**, we compare the theoretical mean $\mathbb{E}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right]$ with the theoretical mean $\mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]$ and the empirical mean $\left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta \right\rangle$ using data from a simulation of $S = 100$ cell number trajectories. In **(B, E, F)**, we compare the theoretical variance $\text{Var}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right]$ with theoretical variance $\text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]$ and the empirical variance $\sigma^2\left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta\right]$ using data from a simulation of $S = 100$ cell number trajectories. **(A, B)**: $\gamma = 0$ (black color); **(C, D)**: $\gamma = 0.5$ (green color); **(E, F)**: $\gamma = 0.5$ (magenta color). Red lines (-) denote theoretical statistics (i.e. mean and variance) for $N = N_k + \frac{\eta}{2}$; blue squares (□) denote theoretical statistics for $N = N_k + U, U \sim \text{Unif}[0,\eta)$; circles (○) denote empirical statistics for $N = N_k + \eta_i$ with $i = 1, \ldots, \hat{S}_k$.

A.2 Errors of Birth and Death Rate Estimation as Functions of Bin Size

In this section we consider the effect of bin size on the accuracy with which we can estimate the birth and death rates. Thus we compare the theoretical mean and variance of the population increment, given that a point of the trajectory lies within a given bin, versus the empirical mean and variance obtained from simulation with a finite sample size. We use $\mathscr{E}$ to represent expected differences in these errors. Define

$$\mathscr{E}_{k\text{mean}} := \mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right] - \left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle, \tag{83}$$

$$\mathscr{E}_{k\text{var}} := \text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right] - \sigma^2\left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k\right]. \tag{84}$$

The errors in estimating the birth and death rates corresponding to $N = N_k + \dfrac{\eta}{2}$ are

$$\mathscr{E}_{k\text{birth}} = \frac{\mathscr{E}_{k\text{var}} + \mathscr{E}_{k\text{mean}}}{2\Delta t} \quad \text{and} \quad \mathscr{E}_{k\text{death}} = \frac{\mathscr{E}_{k\text{var}} - \mathscr{E}_{k\text{mean}}}{2\Delta t}. \tag{85}$$

The theoretical means of the errors over *all* realizations η_i of the iid uniform random variable $U \sim \text{Unif}[0, \eta)$ are

$$\mathbb{E}\left[\mathscr{E}_{k\text{birth}}\right] = \frac{\mathbb{E}\left[\mathscr{E}_{k\text{var}}\right] + \mathbb{E}\left[\mathscr{E}_{k\text{mean}}\right]}{2\Delta t} \quad \text{and} \quad \mathbb{E}\left[\mathscr{E}_{k\text{death}}\right] = \frac{\mathbb{E}\left[\mathscr{E}_{k\text{var}}\right] - \mathbb{E}\left[\mathscr{E}_{k\text{mean}}\right]}{2\Delta t}. \tag{86}$$

The theoretical variances of the errors over *all* realizations η_i of U are

$$\text{Var}\left[\mathscr{E}_{k\text{birth}}\right] = \text{Var}\left[\mathscr{E}_{k\text{death}}\right] = \frac{\text{Var}\left[\mathscr{E}_{k\text{var}}\right] + \text{Var}\left[\mathscr{E}_{k\text{mean}}\right]}{4\Delta t^2}. \tag{87}$$

We analyze how the bin size η influences these analytical expected values and variances of errors $\mathbb{E}\left[\mathscr{E}_{k\text{mean}}\right]$, $\mathbb{E}\left[\mathscr{E}_{k\text{var}}\right]$, $\text{Var}\left[\mathscr{E}_{k\text{var}}\right]$, and $\text{Var}\left[\mathscr{E}_{k\text{mean}}\right]$.

Treating the samples of $\left(\Delta N \middle| N = N_k + U, U \sim \text{Unif}[), \eta)\right)$ as if they were identically and independently distributed, the expected value of the sample mean is equal to the theoretical mean. Therefore,

$$\mathbb{E}\left[\mathscr{E}_{k\text{mean}}\right] = \mathbb{E}\left[\underbrace{\mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]}_{\text{independent of } \eta_i}\right] - \mathbb{E}\left[\left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle\right] \tag{88}$$

$$= \mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right] - \mathbb{E}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0, \eta)\right], \tag{89}$$

where

$$\mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]$$

$$= \left(b_{N_k + (\eta/2)} - d_{N_k + (\eta/2)}\right)\left(N_k + \frac{\eta}{2}\right)\Delta t \tag{90}$$

$$= \left(r - \frac{r}{K}N_k - \frac{r}{K}\frac{\eta}{2}\right)\left(N_k + \frac{\eta}{2}\right)\Delta t \tag{91}$$

$$= \left(r - \frac{r}{K}N_k\right)N_k\Delta t + \left(r - \frac{r}{K}N_k\right)\Delta t\frac{\eta}{2} - \frac{r}{K}N_k\Delta t\frac{\eta}{2} - \frac{r}{K}\frac{\eta^2}{4}\Delta t \tag{92}$$

$$= \mathbb{E}[\Delta N | N = N_k] + \left(r - 2\frac{r}{K}N_k\right)\Delta t\frac{\eta}{2} - \frac{r}{K}\Delta t\frac{\eta^2}{4}. \tag{93}$$

Hence,

$$\mathbb{E}\left[\mathscr{E}_{k\text{mean}}\right] = \mathbb{E}[\Delta N | N = N_k] + \left(r - 2\frac{r}{K}N_k\right)\Delta t\frac{\eta}{2} - \frac{r}{K}\Delta t\frac{\eta^2}{4} \tag{94}$$

$$- \mathbb{E}[\Delta N \,|\, N = N_k] - (r - 2\frac{r}{K}N_k)\Delta t\frac{\eta}{2} + \frac{r}{K}\Delta t\frac{\eta^2}{3}$$

$$= \frac{1}{12}\frac{r}{K}\Delta t\eta^2. \tag{95}$$

We observe that the expected error $\mathbb{E}\left[\mathscr{E}_{k\text{mean}}\right]$ in approximating the true mean $\mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]$ for each bin k is independent of k and is increasing quadratically for $\eta > 0$. If we write the expected error $\mathbb{E}\left[\mathscr{E}_{k\text{mean}}\right]$ as $\frac{1}{12}\left(r\Delta t\right)\left(\frac{\eta}{K}\right)\eta$, then we see that the expected error depends on the ratio $\left(\frac{\eta}{K}\right)$, which shows how big the bin size is relative to the system size (i.e. carrying capacity K), and also depends on the product $r\Delta t$, which can be interpreted roughly as the *per capita* change in cell number $\left(\frac{\Delta N}{N}\right)$ after Δt. The higher these ratios are, the higher expected error is. Looking from a different angle, the expected error $\mathbb{E}\left[\mathscr{E}_{k\text{mean}}\right]$ can be written as $\left(\frac{r\Delta t}{K}\right)\left(\frac{1}{12}\eta^2\right)$. This shows that the expected error depends on $\left(\frac{1}{12}\eta^2\right)$, which is the variance of the random variable η_i, and how big the *per capita* change in cell number $r\Delta t$ after Δt is relative to the system size K. This observation suggests that it may be harder to estimate the cell number increments with high accuracy for fast-reproducing cell types. Further analysis on the relation between r and K would be interesting for future work, since existing work such as[51] shows that the product rK can influence the evolution of antibiotic-resistant bacterial genomes.

We assume that the samples of $\left(\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right)$ are independently and identically distributed, so the expected value of the sample variance is equal to the population variance. Therefore,

$$\mathbb{E}\left[\mathscr{E}_{k\text{var}}\right] = \mathbb{E}\left[\underbrace{\text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]}_{\text{independent of } \eta_i}\right] - \mathbb{E}\left[\sigma^2\left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k\right]\right] \tag{96}$$

$$= \text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right] - \text{Var}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right], \tag{97}$$

where

$$\text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]$$

$$= \left(b_0 + d_0 + (1-2\gamma)\frac{r}{K}N_k + (1-2\gamma)\frac{r}{K}\frac{\eta}{2}\right)\left(N_k + \frac{\eta}{2}\right)\Delta t \tag{98}$$

$$= \text{Var}[\Delta N_k] + \left(b_0 + d_0 + 2(1-2\gamma)\frac{r}{K}N_k\right)\Delta t\frac{\eta}{2} + (1-2\gamma)\frac{r}{K}\Delta t\frac{\eta^2}{4}. \tag{99}$$

Therefore,

$$\mathbb{E}\left[\mathscr{E}_{k\text{var}}\right] = -(1-2\gamma)\frac{r}{K}\Delta t\frac{\eta^2}{12} \tag{100}$$

$$+ \Delta t^2 r^2\frac{(N_k+\eta)^3 - N_k^3}{3\eta} - 2\frac{r^2}{K}\Delta t^2\frac{(N_k+\eta)^4 - N_k^4}{4\eta} + \frac{r^2}{K^2}\Delta t^2\frac{(N_k+\eta)^5 - N_k^5}{5\eta}$$

$$- \left(\mathbb{E}[\Delta N \,|\, N = N_k] + (r - 2\frac{r}{K}N_k)\Delta t\frac{\eta}{2} - \frac{r}{K}\Delta t\frac{\eta^2}{3}\right)^2.$$

Now, we compute the theoretical variances $\text{Var}\left[\mathscr{E}_{k\text{mean}}\right]$ and $\text{Var}\left[\mathscr{E}_{k\text{var}}\right]$ over *all* realizations of η_i. We assume the samples of $\left(\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right)$ are identically distributed, so the variance of the sample mean is equal to the population variance divided by the sample size. Therefore,

$$\text{Var}\left[\mathscr{E}_{k\text{mean}}\right] = \text{Var}\left[\underbrace{\mathbb{E}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]}_{\text{independent of } \eta_i}\right] + \text{Var}\left[\left\langle \Delta N \middle| N = N_k + U, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle\right] \tag{101}$$

$$= \text{Var}\left[\left\langle \Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k \right\rangle\right] \tag{102}$$

$$= \frac{\text{Var}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right]}{\hat{S}_k}. \tag{103}$$

As mentioned above, the samples of $\left(\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right)$ are independently and identically distributed. For computation convenience here, we approximate the binomial distribution of these samples with the Gaussian distribution with the empirical mean and variance as discussed in Section 3.2. We still use the notation N instead of X here to be consistent with the other statistics computed above. With this approximation, the theoretical variance of the empirical variance is equal to two times the theoretical variance squared divided by the sample size minus one. Therefore,

$$\text{Var}\left[\mathscr{E}_{k\text{var}}\right] = \text{Var}\left[\underbrace{\text{Var}\left[\Delta N \middle| N = N_k + \frac{\eta}{2}\right]}_{\text{independent of } \eta_i}\right] + \text{Var}\left[\sigma^2\left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k\right]\right] \tag{104}$$

$$= \text{Var}\left[\sigma^2\left[\Delta N \middle| N = N_k + \eta_i, 0 \le \eta_i < \eta, \hat{S}_k\right]\right] \tag{105}$$

$$= \frac{2\left(\text{Var}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right]\right)^2}{\hat{S}_k - 1}. \tag{106}$$

The theoretical variance $\text{Var}\left[\Delta N \middle| N = N_k + U, U \sim \text{Unif}[0,\eta)\right]$ is given by Equation (82).

Using the $\mathbb{E}\left[\mathscr{E}_{k\text{mean}}\right]$, $\mathbb{E}\left[\mathscr{E}_{k\text{var}}\right]$, $\text{Var}\left[\mathscr{E}_{k\text{mean}}\right]$, and $\text{Var}\left[\mathscr{E}_{k\text{var}}\right]$ that we just computed, we obtain the theoretical means and variances of the errors in estimating birth and death rates corresponding to $N = N_k + \frac{\eta}{2}$ for all $k = 1, 2, \ldots, k_{\max}$ using Equations (85) and (86).

In Figure 5, we compare the l_2-norm of the theoretical means and variances of the errors and compare them with the l_2-norm of the empirical errors (i.e. realizations of the error random variables) computed using data from a simulation of $S = 100$ cell number trajectories. To computed the theoretical variances of the errors shown in Figure 5, we use the empirical sample sizes $\hat{S}_k, k = 1, 2, \ldots, k_{\max}$, from the same data simulation.

We observe that as the bin size η increases, the theoretical means of the errors increase, the theoretical variances (or standard deviations) of the errors decreases, and the empirical errors balance between the

theoretical means and variances (or standard deviations) and have convex quadratic shapes. The theoretical means of the errors reflect the differences between ΔN at one point $\left(N = N_k + \frac{\eta}{2}\right)$ and ΔN at multiple points $\left(N = N_k + \eta_i, 0 \leq \eta_i < \eta\right)$; the smaller the bin size, the closer multiple points are to one point, so the error is smaller (for example, Equation (95) shows that the expected errors in estimating the mean of cell number increments are $(r\Delta t/12K)\eta^2$). However, if the bin is too small, then there are not enough samples to estimate theoretical statistics with empirical statistics with accuracy. The theoretical variances of errors involves sample sizes; the bigger the bin size, the more samples we have. These two competing effects of bin size result in the empirical errors being intermediate values between the two theoretical statistics (means and variances) of the estimation errors. The optimal bin size reflects a balancing of these two effects. When the bin size is smaller than the optimal bin size, the sample error coincides with the sum of the expected error and the standard deviation of the error. When the bin size is bigger than the optimal bin size, this relationship breaks down, which may reflect growing inaccuracy of our approximation that the trajectory points are uniformly and i.i.d. within each bin.

A.3 Notation

N	denotes	cell number random variable
t_0 and t_T	denotes	deterministic initial and final times respectively
η	denotes	deterministic bin size
k	denotes	bin index, $k = 1, 2, \ldots, k_{\max}$
U	denotes	uniformly distributed random variable such that $\left(N = N_k + U\right) \in [N_k, N_k + \eta)$
η_i	denotes	realization of the random variable U
S	denotes	number of cell number trajectories/time series
$\hat{S}_k$	denotes	number of samples of $\Delta N := N(t + \Delta t) - N(t)$ in bin $[N_k, N_k + \eta)$
$\mathbb{E}[\cdot]$	denotes	theoretical mean
$\langle \cdot \rangle$	denotes	empirical mean
$\mathrm{Var}[\cdot]$	denotes	theoretical variance
$\sigma^2[\cdot]$	denotes	empirical variance
$\mathscr{E}_{[\cdot]}$	denotes	error

B Analysis of Log-Likelihood Function

We calculate the first and second derivatives of the log-likelihood function $f(\gamma)$ (44) for a single trajectory as a function of the density dependence parameter γ. Let $\Delta x_j = x_{j+1} - x_j$.

$$f(\gamma) = \sum_{j=1}^{T-1} \frac{1}{2} \ln\left(\frac{1}{2\pi \mathrm{Var}[\Delta x_j]}\right) - \frac{1}{2} \frac{\left(\Delta x_j - \mathbb{E}[\Delta x_j]\right)^2}{\mathrm{Var}[\Delta x_j]} \tag{107}$$

$$= \sum_{j=1}^{T-1} -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln\left(\mathrm{Var}[\Delta x_j]\right) - \frac{1}{2} \frac{\left(\Delta x_j - \mathbb{E}[\Delta x_j]\right)^2}{\mathrm{Var}[\Delta x_j]}, \tag{108}$$

where

$$\mathbb{E}[\Delta x_j] = x_j \Delta t (b_{x_j} - d_{x_j}) = x_j \Delta t \left(\frac{b_0 - \gamma(r/K)x_j + \left| b_0 - \gamma(r/K)x_j \right|}{2} - d_0 - (1-\gamma)(r/K)x_j \right), \tag{109}$$

$$\mathrm{Var}[\Delta x_j] = x_j \Delta t (b_{x_j} + d_{x_j}) = 2x_j \Delta t \left(\frac{b_0 - \gamma(r/K)x_j + \left| b_0 - \gamma(r/K)x_j \right|}{2} + d_0 + (1-\gamma)(r/K)x_j \right). \tag{110}$$

We observe that $\mathbb{E}[\Delta x_j]$ is a piecewise linear function of γ, i.e. $\mathbb{E}[\Delta x_j]$ has the form $c_1^j + c_2^j \gamma$, where

$$c_1^j = \begin{cases} x_j \Delta t \left(b_0 - d_0 - \frac{r}{K} x_j \right), & \text{for } \left(b_0 - \gamma \frac{r}{K} x_j \right) > 0, \\ -x_j \Delta t \left(d_0 + \frac{r}{K} x_j \right), & \text{for } \left(b_0 - \gamma \frac{r}{K} x_j^2 \right) = 0, \end{cases} \tag{111}$$

and

$$c_2^j = \begin{cases} 0, & \text{for } \left(b_0 - \gamma \frac{r}{K} x_j \right) > 0, \\ \Delta t \frac{r}{K} x_j, & \text{for } \left(b_0 - \gamma \frac{r}{K} x_j^2 \right) = 0. \end{cases} \tag{112}$$

The variance $\mathrm{Var}[\Delta x_j]$ is also a linear function of γ, i.e. $\mathrm{Var}[\Delta x_j]$ has the form $c_3^j - c_4^j \gamma$ with

$$c_3^j = \begin{cases} x_j \Delta t \left(b_0 + d_0 + \frac{r}{K} x_j \right), & \text{for } \left(b_0 - \gamma \frac{r}{K} x_j \right) > 0, \\ x_j \Delta t \left(d_0 + \frac{r}{K} x_j \right), & \text{for } \left(b_0 - \gamma \frac{r}{K} x_j^2 \right) = 0, \end{cases} \tag{113}$$

and

$$c_4^j = \begin{cases} 2\Delta t \frac{r}{K} x_j^2, & \text{for } \left(b_0 - \gamma \frac{r}{K} x_j \right) > 0, \\ \Delta t \frac{r}{K} x_j^2, & \text{for } \left(b_0 - \gamma \frac{r}{K} x_j^2 \right) = 0, \end{cases} \tag{114}$$

Therefore,

$$f(\gamma) = \sum_{j=1}^{T-1} -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \left(c_3^j - c_4^j \gamma \right) - \frac{1}{2} \frac{\left(\Delta x_j - c_1^j - c_2^j \gamma \right)^2}{c_3^j - c_4^j \gamma}. \tag{115}$$

Denote $v_j = \dfrac{1}{c_3^j - c_4^j \gamma} \Rightarrow v_j > 0$ and $\dfrac{dv_j}{d\gamma} = \dfrac{c_4^j}{(c_3^j - c_4^j \gamma)^2} = c_4^j v_j^2$. We have

$$f(\gamma) = \sum_{j=1}^{T-1} -\frac{1}{2} \ln(2\pi) + \frac{1}{2} \ln(v_j) - \frac{1}{2} \left(\Delta x_j - c_1^j - c_2^j \gamma \right)^2 v_j. \tag{116}$$

If $b_0 - \gamma(r/K)x_j > 0$, then $c_2^j = 0$ and $\mathbb{E}[\Delta x_j] = c_1^j$ is independent of γ. Hence,

$$\frac{df}{d\gamma} = \sum_{j=1}^{T-1} \frac{1}{2}\frac{1}{v_j}\frac{dv_j}{d\gamma} - \frac{1}{2}\left(\Delta x_j - c_1^j\right)^2 \frac{dv_j}{d\gamma} = \sum_{j=1}^{T-1} \frac{1}{2}c_4^j v_j - \frac{1}{2}\left(\Delta x_j - c_1^j\right)^2 c_4^j v_j^2 \tag{117}$$

$$\Rightarrow \frac{d^2 f}{d\gamma^2} = \sum_{j=1}^{T-1} \frac{1}{2}c_4^j \frac{dv_j}{d\gamma} - \frac{1}{2}\left(\Delta x_j - c_1^j\right)^2 c_4^j 2v_j \frac{dv_j}{d\gamma} \tag{118}$$

$$= \sum_{j=1}^{T-1} \frac{1}{2}(c_4^j)^2 v_j^2 - \left(\Delta x_j - c_1^j\right)^2 (c_4^j)^2 v_j^3 \tag{119}$$

$$= \sum_{j=1}^{T-1} (c_4^j)^2 v_j^2 \left(\frac{1}{2} - \left(\Delta x_j - c_1^j\right)^2 v_j\right) \tag{120}$$

$$= \sum_{j=1}^{T-1} (c_4^j)^2 v_j^2 \left(\frac{1}{2} - \frac{\left(\Delta x_j - \mathbb{E}[\Delta x_j]\right)^2}{\mathrm{Var}[\Delta x_j]}\right). \tag{121}$$

In general,

$$\frac{df}{d\gamma} = \sum_{j=1}^{T-1} \frac{1}{2}\frac{1}{v_j}\frac{dv_j}{d\gamma} - \frac{1}{2}\left(\Delta x_j - c_1^j - c_2^j\gamma\right)^2 \frac{dv_j}{d\gamma} + c_2^j\left(\Delta x_j - c_1^j - c_2^j\gamma\right)v_j \tag{122}$$

$$= \sum_{j=1}^{T-1} \frac{1}{2}c_4^j v_j - \frac{1}{2}\left(\Delta x_j - c_1^j - c_2^j\gamma\right)^2 c_4^j v_j^2 + c_2^j\left(\Delta x_j - c_1^j - c_2^j\gamma\right)v_j \tag{123}$$

$$\Rightarrow \frac{d^2 f}{d\gamma^2} = \sum_{j=1}^{T-1} \frac{1}{2}c_4^j \frac{dv_j}{d\gamma} + \frac{1}{2}2c_2^j\left(\Delta x_j - c_1^j - c_2^j\gamma\right)c_4^j v_j^2 - \frac{1}{2}\left(\Delta x_j - c_1^j - c_2^j\gamma\right)^2 c_4^j 2v_j \frac{dv_j}{d\gamma} \tag{124}$$

$$+ \sum_{j=1}^{T-1} c_2^j\left(\Delta x_j - c_1^j - c_2^j\gamma\right)\frac{dv_j}{d\gamma} \tag{125}$$

$$= \sum_{j=1}^{T-1} \frac{1}{2}(c_4^j)^2 v_j^2 + 2c_2^j c_4^j\left(\Delta x_j - c_1^j - c_2^j\gamma\right)v_j^2 - \left(\Delta x_j - c_1^j - c_2^j\gamma\right)^2 (c_4^j)^2 v_j^3 \tag{126}$$

$$= \sum_{j=1}^{T-1} c_4^j v_j^2 \left(\frac{1}{2}c_4^j + 2c_2^j\left(\Delta x_j - c_1^j - c_2^j\gamma\right) - \left(\Delta x_j - c_1^j - c_2^j\gamma\right)^2 c_4^j v_j\right). \tag{127}$$

Figure 16 shows the histogram of $\frac{d^2 f}{d\gamma^2}$ evaluated at the numerical root of $\frac{df}{d\gamma}$ on $[0,1]$. The second derivative is negative among for each of 100 instances of solving the optimization problem (48). We observe that the second derivatives are negative for all of the cases, which implies that the numerical root is reasonably presumed to be a maximum.

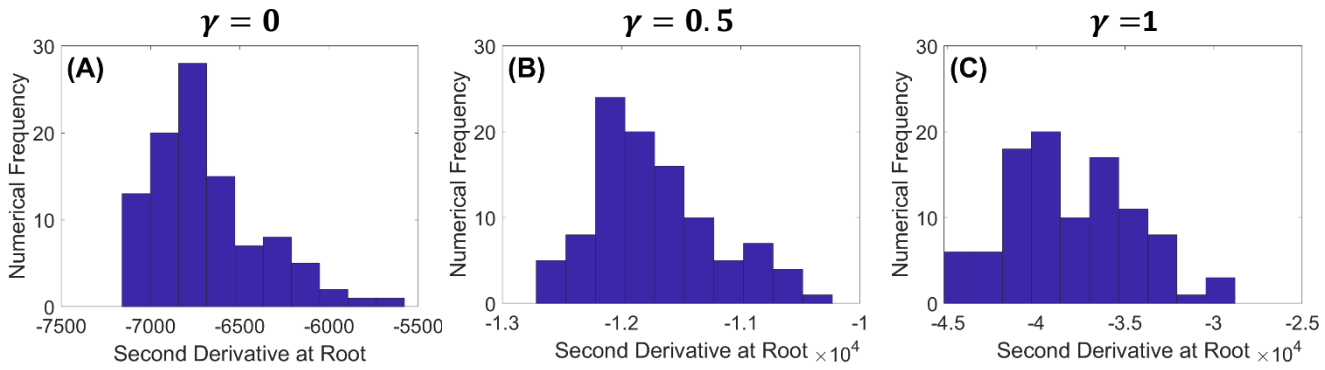

Figure 16. The second derivative of the log likelihood is confirmed empirically to be negative at the numerical solution of the optimization problem. We plot empirical distributions of the second derivative of the log-likelihood function $f(\gamma)$ evaluated at the numerical root on $[0,1]$ of the first derivative of $f(\gamma)$ for 100 cell number trajectories. **(A), (B), (C)** correspond to three different scenarios of density dependence $\gamma = 0$, $\gamma = 0.5$, $\gamma = 1$ respectively. The distributions are obtained from maximizing the log-likelihood function $f(\gamma)$ 100 times for each of the three γ scenarios. We observe that the second derivatives are negative for all of the cases.

We explicitly calculate the first derivative of $f(\gamma)$ below to find critical points:

$$\frac{df}{d\gamma} = \sum_{j=1}^{T-1} -\frac{1}{2}\frac{1}{x_j(b_{x_j}+d_{x_j})} x_j \frac{d}{d\gamma}(b_{x_j}+d_{x_j}) \tag{128}$$

$$-\sum_{j=1}^{T-1} \frac{1}{2} 2\Big(\Delta x_j - x_j(b_{x_j}-d_{x_j})\Delta t\Big)\Delta t(-x_j)\frac{d}{d\gamma}(b_{x_j}-d_{x_j})\frac{1}{x_j(b_{x_j}+d_{x_j})}\frac{1}{\Delta t} \tag{129}$$

$$-\sum_{j=1}^{T-1} \frac{1}{2}(\Delta x_j - x_j(b_{X_j}-d_{x_j})\Delta t)^2 \frac{-1}{\Delta t x_j^2 (b_{x_j}+d_{x_j})^2} x_j \frac{d}{d\gamma}(b_{x_j}+d_{x_j}) \tag{130}$$

$$= \sum_{j=1}^{T-1} -\frac{1}{2}\frac{1}{(b_{x_j}+d_{x_j})}\frac{d}{d\gamma}(b_{x_j}+d_{x_j}) \tag{131}$$

$$+\sum_{j=1}^{T-1} \Big(\Delta x_j - x_j(b_{x_j}-d_{x_j})\Delta t\Big)\frac{1}{(b_{x_j}+d_{x_j})}\frac{d}{d\gamma}(b_{x_j}-d_{x_j}) \tag{132}$$

$$+\sum_{j=1}^{T-1} \frac{1}{2}(\Delta x_j - x_j(b_{X_j}-d_{x_j})\Delta t)^2 \frac{1}{\Delta t x_j (b_{x_j}+d_{x_j})^2}\frac{d}{d\gamma}(b_{x_j}+d_{x_j}), \tag{133}$$

where

$$b_{x_j}+d_{x_j} = \frac{b_0 - \gamma(r/K)x_j + \left|b_0 - \gamma(r/K)x_j\right|}{2} + d_0 + (1-\gamma)(r/K)x_j \tag{134}$$

$$\Rightarrow \frac{d}{d\gamma}\Big(b_{x_j}+d_{x_j}\Big) = -\frac{(r/K)x_j}{2} - (r/K)x_j - \frac{1}{2}(r/K)x_j\frac{\left|b_0-\gamma(r/K)x_j\right|}{b_0-\gamma(r/K)x_j} \tag{135}$$

$$= -\frac{3}{2}(r/K)x_j - \frac{1}{2}(r/K)x_j\frac{\left|b_0-\gamma(r/K)x_j\right|}{b_0-\gamma(r/K)x_j}, \tag{136}$$

and

$$b_{x_j} - d_{x_j} = \frac{b_0 - \gamma(r/K)x_j + \left|b_0 - \gamma(r/K)x_j\right|}{2} - d_0 - (1-\gamma)(r/K)x_j \tag{137}$$

$$\Rightarrow \frac{d}{d\gamma}\left(b_{x_j} - d_{x_j}\right) = -\frac{(r/K)x_j}{2} + (r/K)x_j - \frac{1}{2}(r/K)x_j \frac{\left|b_0 - \gamma(r/K)x_j\right|}{b_0 - \gamma(r/K)x_j} \tag{138}$$

$$= \frac{1}{2}(r/K)x_j - \frac{1}{2}(r/K)x_j \frac{\left|b_0 - \gamma(r/K)x_j\right|}{b_0 - \gamma(r/K)x_j}. \tag{139}$$

Using these expressions, we numerically obtain the root of the first derivative on the interval $[0,1]$.

C Demonstration of the Nonparametric Property of Our Direct Estimation Method

Our birth-rate/death-rate disambiguation method applies to a broader class of models than the Verhulst-inspired logistic model used for illustration in other parts of the paper. To demonstrate the method's utility, we apply the method here to a model that does not have a simple low-dimensional parametric description, and is thus intractable via maximum likelihood methods. We consider a stochastic birth-death process for which the density-dependent *per capita* birth and death rates are defined piecewise as functions of the cell number N as follows:

$$b_N = \max\left\{\frac{1}{60}, \frac{1}{120} - \frac{1}{15K}N\right\}, \tag{140}$$

$$d_N = \max\left\{\frac{1}{15K}N - \frac{1}{20}, \frac{1}{105K}N\right\}, \tag{141}$$

$K = 10^5$ is the carrying capacity.

This birth-death process has cell number trajectories as in Figure 17.

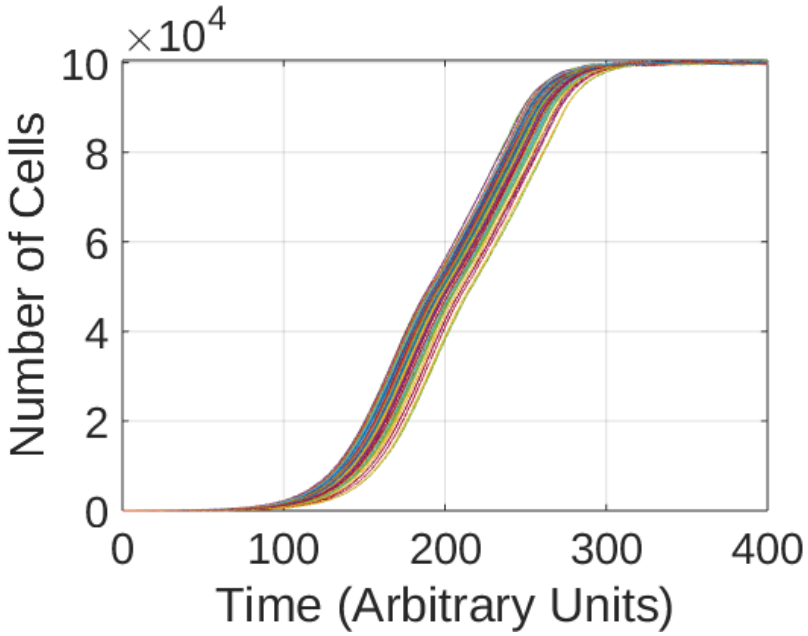

Figure 17. Cell number trajectories for the birth-death process with the *per capita* birth and death rates defined by Equations (140) and (141).

Suppose we are given only cell number trajectories like those in Figure 17; we do not know the forms of the birth and death rates (note: we defined b_N and d_N in Equations (140) and (141) above for data simulation and validation only). Using only the given ensemble of cell number trajectories, we use our direct estimation method discussed in Section 3.3 and are able to infer the birth and death rates, as shown in Figure 18.

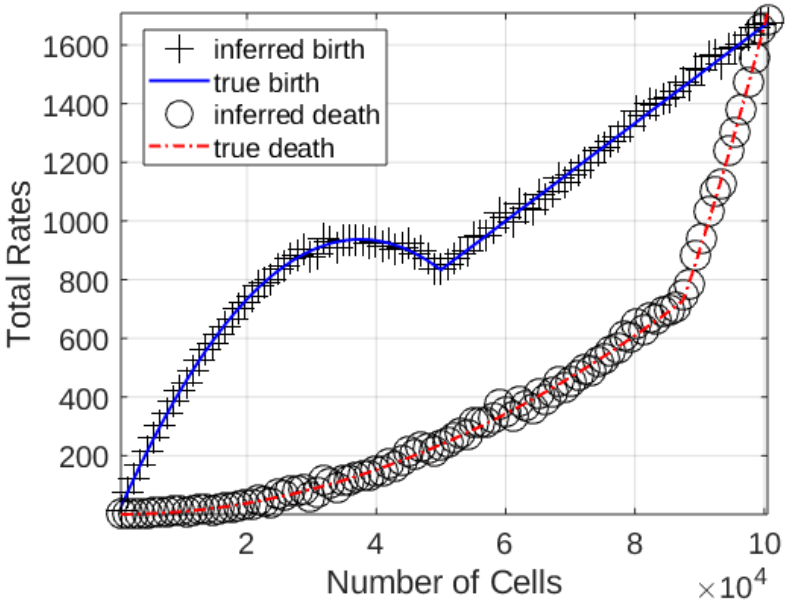

Figure 18. Inferred birth and death rates for the birth-death process with the *per capita* birth and death rates defined by Equations (140) and (141) are well-aligned with the true rates.

Direct inspection of the cell number trajectories in Figure 17 would not readily lead one to suppose that the birth and death rates have the shapes as in Figure 18; moreover, we cannot take a maximum likelihood approach to make inferences. By utilizing the mean and variance of the cell number stochastic fluctuations, our direct estimation method allows us to see the shapes of the rates and their values corresponding to certain cell numbers. In this particular example, the *per capita* birth rate is density-dependent for $N \leq \frac{K}{2}$, but not for $N > \frac{K}{2}$. In addition to not knowing the forms of the rates, the piecewise-defined functions are not differentiable at the transition points (between the pieces), which may make it difficult to solve nonlinear optimization problems.

D Model Parameters Used in Simulation

Table 3. Model Parameters Used in Simulation

Parameter	Value	Unit
b_0	1.1/120	1/time
d_0	0.1/120	1/time
r	1/120	1/time
K	10^5	Dimensionless