

# Likelihood-based inference for the Gompertz model with Poisson errors

Paolo Onorati\*, Sofia Ruiz-Suarez†, and Radu V Craiu‡

## Abstract

Population dynamics models play an important role in a number of fields, such as actuarial science, demography, and ecology. Statistical inference for these models can be difficult when, in addition to the process' inherent stochasticity, one also needs to account for sampling error. Ignoring the latter can lead to biases in the estimation, which in turn can produce erroneous conclusions about the system's behavior. The Gompertz model is widely used to infer population size dynamics, but a full likelihood approach can be computationally prohibitive when sampling error is accounted for. We close this gap by developing efficient computational tools for statistical inference in the Gompertz model based on the full likelihood. The approach is illustrated in both the Bayesian and frequentist paradigms. Performance is illustrated with simulations and data analysis.

*Keywords:* Bayesian inference, EM algorithm, MCMC, Sampling error.

## 1 Introduction

Population dynamics models are crucial in both applied and theoretical ecology, because they help to explain past population fluctuations and project future population abundances [Newman et al., 2014]. These models are used to manage the conservation of species, understand the dynamics of biological invasions, and assess the response of certain species to changes in their environment which can range from those produced by human developments to those triggered by climate change. The accuracy and reliability of these models are often affected by two major components of uncertainty due to: i) the process' stochasticity connected to, say, demographic and environmental factors, and ii) the sampling error.

Ecologists have long recognized the importance of separating observation from process error in ecological modeling, and significant progress has been achieved through the use of state-space models to analyze time series of population fluctuations [Dennis et al., 2006, Hostetler and Chandler, 2015, Auger-Méthé et al., 2021]. However, including both process noise and observation error in the model remains challenging, as it produces significant computational difficulties. To illustrate, Staples et al.

---

\*University of Padua, Italy

†University of Toronto, Canada

‡University of Toronto, Canada

[2004] note that the sampling error adds to the variability in the data leading to positively biased estimators of process variation and propose a restricted maximum likelihood estimation of the latter and of the error variance. Subsequently, [Lindén and Knape, 2009] show empirically that observation errors can change the autocorrelation structure of a time series with potential biases resulting from models that ignore such errors.

The Gompertz model [Gompertz, 1825] is widely used to describe and characterize population dynamics. Its use permeates multiple disciplines, including actuarial science [e.g., Butt and Haberman, 2004, Baione and Levantesi, 2018], demography [e.g., Alexander et al., 2017, Tai and Noymer, 2018], as well as in life sciences where it can be used to model the growth patterns of animals, plants, tumors and the volume of bacteria [Winsor, 1932, Tjørve and Tjørve, 2017]. Most importantly, the Gompertz model can be modified to account for sampling variability and, in some clearly structured problems, it is possible to explicitly formulate the likelihood function in the presence of sampling error. For example, if the observation noise is assumed to follow a log-normal distribution, it is feasible to write the likelihood function [Staples et al., 2004] and the model can be transformed into a linear Gaussian state-space model which can be fitted using the Kalman filter [Dennis et al., 2006]. However, if the sampling error is not log-normal, the expression of the likelihood becomes more complex. Relevant for this paper is the study of Lele [2006] which shows that likelihood-based inference for a stationary Gompertz model with Poisson sampling errors requires the calculation of a high-dimensional integral that brings on a prohibitive computational burden. He proposes a composite likelihood approach that reduces the dimension of the integral and makes the computation manageable for longer time series. However, replacing the full likelihood with a composite one comes with an inferential cost because the resulting confidence, or credible, intervals will not have nominal coverage.

The main contribution of this paper is the construction of algorithms that allow for a full likelihood-based analysis of the Gompertz model with Poisson errors. Central to our approach is an efficient algorithm for computing the likelihood, which allows us to conduct both frequentist and Bayesian inference. The developed methods are available in a new R package published on GitHub. The paper is organized as follows. In Section 2 we introduce the notation and the statistical model. Section 3 contains the estimation procedure for a frequentist approach based on a simulation-aided EM algorithm [Dempster et al., 1977, Wei and Tanner, 1990]. The Bayesian analysis relies on a new MCMC sampler which is described in Section 4. The paper continues with a Section 5 of numerical experiments that includes simulations and a data analysis. A discussion of open questions and future research directions closes the paper in Section 6.

## 2 Statistical Model

Consider a population observed at times  $t = 1, 2, \dots, T$ . The Gompertz model with stochastic errors establishes a probabilistic model for the evolution of  $N_t$ , the population size at time  $t$ , for  $1 \leq t \leq T$ . Specifically, assume that for all  $1 \leq t \leq T - 1$ ,

$$N_{t+1} = N_t \exp(a + b \log(N_t) + \varepsilon_{t+1}), \quad (1)$$

where  $a$  is an individual intrinsic growth rate parameter,  $b$  links the current measure with the past one, and  $\varepsilon_t$  are independent errors normally distributed:  $\varepsilon_t \sim N(0, \sigma^2)$ . The logarithmic transformation of  $N_t$  in (1) leads to

$$Z_{t+1} = a + (1 + b)Z_t + \varepsilon_{t+1} \quad (2)$$

where  $Z_t = \log(N_t)$ . Equation (2) implies that the Gompertz model corresponds, on the logarithmic scale, to an autoregressive process of order 1. If we assume that  $|1 + b| < 1$ , then for any subset of  $n$  observations, the  $n$ -dimensional marginal distribution of the log transformed population size,  $Z_t$ , has a multivariate normal distribution  $(Z_{i_1}, Z_{i_2}, \dots, Z_{i_n})' \sim N_n(\boldsymbol{\mu}, \Sigma)$  where

$$\boldsymbol{\mu} = (\theta_1, \theta_1, \dots, \theta_1)', \Sigma = \theta_2 B, \text{ and } B_{jk} = (1 + b)^{|i_j - i_k|} \quad (3)$$

and  $\theta_1 = -a/b$  and  $\theta_2 = -\sigma^2/[b(2 + b)]$ . In what follows, we denote  $\boldsymbol{\theta} = (\theta_1, \theta_2, b)$ .

In this paper, we consider the case where the exact population size  $N_t$  is unknown. Instead, we assume that at time  $t \in \{1, \dots, T\}$ ,  $N_t^*$  is the observed population size, whose conditional probability mass function is denoted  $\pi(N_t^*|N_t, \phi)$  and is indexed by the parameter  $\phi$ . Furthermore, we assume that the random variables  $\{N_t^*|1 \leq t \leq T\}$  are independent conditionally on  $\{N_t : 1 \leq t \leq T\}$ . The observed-data likelihood is obtained by taking advantage of the conditional distribution of  $N_t^*|N_t$  and averaging the missing true population values  $\{N_1, \dots, N_T\}$ .

$$L(\boldsymbol{\theta}, \phi | \mathbf{N}^*) = \int \int \dots \int \pi(N_1^*|N_1, \phi) \pi(N_2^*|N_2, \phi) \dots \pi(N_T^*|N_T, \phi) \pi(\mathbf{N}|\boldsymbol{\theta}) dN_1 dN_2 \dots dN_T, \quad (4)$$

where  $\pi(\mathbf{N}|\boldsymbol{\theta})$  indicates the joint distribution of the unobserved time series of exact population sizes.

If the sampling error distribution  $\pi(N_t^*|N_t, \phi)$  is assumed to be log-normal, then it is possible to explicitly formulate the complete likelihood function as in Dennis et al. [2006]. However, for other sampling error distributions, the likelihood (4) cannot be expressed in closed form using elementary functions. Here we focus on the case of a Poisson sampling error distribution, i.e.  $N_t^* \sim \text{Poisson}(N_t)$ , for all  $1 \leq t \leq T$ . In this case, no additional parameters,  $\phi$ , are associated with the sampling error distribution. However, the use of a Poisson distribution leads to computational challenges because the likelihood is not available in closed form, as discussed by [Lele, 2006]. In Section 3 we present a method for directly addressing the expression (4) to perform parameter inference. Our approach is very general and can also be implemented in a Bayesian framework, as we show in Section 4. The performance of the proposed algorithms is examined in Section 5.1.

### 3 Maximum likelihood based inference

We describe a frequentist inferential method for the Gompertz model with Poisson observation errors. In particular, we are interested in computing the maximum likelihood estimator (MLE) for  $\boldsymbol{\theta}$  and producing confidence intervals. The proposed algorithm takes advantage of the data augmentation strategy implied by (4). It is apparent that if the population sizes  $\mathbf{N}$  were available, then the

inference would be straightforward, since it would rely on the analysis of a first-order autoregressive model for  $Z_t = \log N_t$ . However, since population sizes  $N_t, t \in \{1, 2, \dots, T\}$  are unobserved latent variables, we adopt a data augmentation approach.

The expectation-maximization (EM) algorithm is a powerful iterative procedure that is used to estimate the parameters of models with missing data. In our case,  $\mathbf{N}^*$  is the vector of observed data,  $\mathbf{Z} = \log \mathbf{N}$  is the vector of missing data, and  $\boldsymbol{\theta}^{(k)}$  is the value of the parameter estimate in the  $k$ -th iteration. The EM algorithm [Dempster et al., 1977] iteratively computes  $\boldsymbol{\theta}^{(k+1)} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$  where the  $Q$  function is

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \mathbb{E}[\log \pi(\mathbf{N}^*, \mathbf{Z}|\boldsymbol{\theta})|\mathbf{N}^*, \boldsymbol{\theta}^{(k)}].$$

The calculation of the conditional expectation of the complete data log-likelihood,  $\log \pi(\mathbf{N}^*, \mathbf{Z}|\boldsymbol{\theta})$ , represents the E-step and its optimization as a function of  $\boldsymbol{\theta}$  is the M-step. The algorithm cycles between the E- and M-steps until the change in  $\boldsymbol{\theta}^{(k)}$  becomes negligible.

In the specific case of the Gompertz model with Poisson sampling error, the complete data log-likelihood can be separated as

$$\log \pi(\mathbf{N}^*, \mathbf{Z}|\boldsymbol{\theta}) = \log \pi(\mathbf{N}^*|\mathbf{Z}) + \log \pi(\mathbf{Z}|\boldsymbol{\theta}),$$

and only the second term depends on  $\boldsymbol{\theta}$ . Nonetheless, we cannot compute the expectation required in the E-step in closed form, so a Monte Carlo strategy must be used. Specifically, we approximate the expectation with a Monte Carlo average

$$\mathbb{E}[\log \pi(\mathbf{Z}|\boldsymbol{\theta})|\mathbf{N}^*, \boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}] \approx \frac{1}{J} \sum_{j=1}^J \log \pi(\tilde{\mathbf{Z}}_j|\boldsymbol{\theta}), \quad (5)$$

where  $\tilde{\mathbf{Z}}_j$  are sampled from  $\mathbf{Z}|\mathbf{N}^*, \boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$ . Replacing the expectation with a Monte Carlo estimate is the idea behind the so-called Monte Carlo EM (MCEM) introduced by Wei and Tanner [1990] and further expanded by McCulloch [1994], Chan and Ledolter [1995], Booth and Hobert [1999], and Caffo et al. [2005].

We start the MCEM algorithm with a value of  $J = 10^3$  in (5) and, as prescribed by Booth and Hobert [1999] and Caffo et al. [2005], we increase the value of  $J$  with the number of iterations in order to achieve higher precision in the estimate of the expected complete data log-likelihood. Specifically, we use the method proposed in Caffo et al. [2005] to determine whether an increase in  $J$  is necessary or not. If so, we double the current value of  $J$  up to a maximum value of  $2 \times 10^4$ . The conditional distribution  $\mathbf{Z}|\mathbf{N}^*, \boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$  is nonstandard, so we must customize a Gibbs sampler to obtain the necessary draws. In the following, we omit the superscript  $(k)$  for ease of notation.

According to the Gibbs sampler design, we need to obtain samples from the full conditional distributions of each latent variable, given the observed data, the current parameter values, and the remaining unobserved population sizes. For brevity, we denote  $\mathbf{Z}_{-t}$  the vector  $\mathbf{Z}$  from which the  $t$

-th component,  $Z_t$  is excluded. Thus, we need to sample from

$$Z_t | \mathbf{Z}_{-t}, \mathbf{N}^*, \boldsymbol{\theta} = \begin{cases} Z_1 | Z_2, N_1^*, \boldsymbol{\theta} & \text{if } t = 1 \\ Z_t | Z_{t-1}, Z_{t+1}, N_t^*, \boldsymbol{\theta} & \text{if } 1 < t < T \\ Z_T | Z_{T-1}, N_T^*, \boldsymbol{\theta} & \text{if } t = T \end{cases} . \quad (6)$$

The expression above exploits the independence of the components of  $\mathbf{N}^*$  conditionally on  $\mathbf{Z}$  and the sparse structure of the inverse of the correlation matrix  $B$ . These facts dramatically improve the mixing of the Gibbs sampler.

From expression (6), we obtain the unnormalized density for every full conditional distribution of  $Z_t$ , that is,

$$\pi(Z_t | \mathbf{Z}_{-t}, \mathbf{N}^*, \boldsymbol{\theta}) \propto \begin{cases} \pi(N_1^* | Z_1) \pi(Z_1 | Z_2, \boldsymbol{\theta}) & \text{if } t = 1 \\ \pi(N_t^* | Z_t) \pi(Z_t | Z_{t-1}, Z_{t+1}, \boldsymbol{\theta}) & \text{if } 1 < t < T \\ \pi(N_T^* | Z_T) \pi(Z_T | Z_{T-1}, \boldsymbol{\theta}) & \text{if } t = T \end{cases} , \quad (7)$$

and the second factor in the above expression is always a Gaussian density with mean  $\mu_t$  and variance  $\tau_t^2$  given by

$$\mu_t = \begin{cases} \frac{\theta_1 \sigma^2 + (Z_2 - a)(1+b)\theta_2}{\sigma^2 + (1+b)\theta_2} & \text{if } t = 1 \\ \frac{a + (1+b)(Z_{t+1} + Z_{t-1} - a)}{1 + (1+b)^2} & \text{if } 1 < t < T \\ a + (1+b)Z_{T-1} & \text{if } t = T \end{cases} , \quad \tau_t^2 = \begin{cases} \frac{\sigma^2 \theta_2}{\sigma^2 + (1+b)\theta_2} & \text{if } t = 1 \\ \frac{\sigma^2}{1 + (1+b)^2} & \text{if } 1 < t < T \\ \sigma^2 & \text{if } t = T \end{cases} , \quad (8)$$

Therefore, expression (7) is equal to the product between a Gaussian density and a Poisson mass probability, so that

$$\pi(Z_t | \mathbf{Z}_{-t}, \mathbf{N}^*, \boldsymbol{\theta}) \propto \exp \left( -e^{Z_t} - \frac{(Z_t - \mu_t)^2}{2\tau_t^2} \right) , \quad (9)$$

where  $\mu_t$  and  $\tau_t^2$  are given by (8) and channel the dependence between  $Z_t$  and  $\mathbf{Z}_{-t}$ .

We sample from the above density using an accept-reject algorithm with a  $N(\xi, \omega^2)$  serving as the proposal density. The following proposition provides the upper bound between the target and the proposal and, in the special case of  $\omega^2 = \tau_t^2$ , the best value for  $\xi$ .

**Proposition 1.** *Using the above notation the following hold:*

- i) *The ratio between expression (9) and the density of  $N(\xi, \omega^2)$  is unbounded if  $\omega^2 < \tau_t^2$ .*

ii) If  $\omega^2 \geq \tau_t^2$  then there exists an unique maximum equal to

$$\widehat{Z}_t = \begin{cases} \log\left(\frac{N_t^* \tau_t^2 - \xi + \mu_t}{\tau_t^2}\right) & \text{if } \omega^2 = \tau_t^2 \\ \frac{N_t^* \omega^2 \tau_t^2 - \xi \tau_t^2 + \mu \omega^2}{\omega^2 - \tau_t^2} W_0\left(\frac{\omega^2 \tau_t^2}{\omega^2 - \tau_t^2} \exp\left(\frac{N_t^* \omega^2 \tau_t^2 - \xi \tau_t^2 + \mu \omega^2}{\omega^2 - \tau_t^2}\right)\right) & \text{if } \omega^2 > \tau_t^2 \end{cases},$$

where  $W_0$  denotes the upper branch of the Lambert W function.

iii) When  $\omega^2 = \tau_t^2$  the value of  $\xi$  that minimizes the maximum  $\widehat{Z}_t$  is

$$\hat{\xi} = N_t^* \tau_t^2 + \mu_t - W_0(\tau_t^2 \exp(N_t^* \tau_t^2 + \mu_t)).$$

*Proof.* i) The ratio between the target and the proposal density is proportional to

$$h(Z_t) = \exp\left(Z_t N_t^* - e^{Z_t} - \frac{(Z_t - \mu_t)^2}{2\tau_t^2} + \frac{(Z_t - \xi)^2}{2\omega^2}\right). \quad (10)$$

It is easy to show that is unbounded if  $\omega^2 < \tau_t^2$ .

ii) We compute the first and second log-derivative and obtain

$$\begin{aligned} \frac{d \log h(Z_t)}{dZ_t} &= N_t^* + \frac{Z_t - \xi}{\omega^2} - \frac{Z_t - \mu_t}{\tau_t^2} - e^{Z_t}, \\ \frac{d^2 \log h(Z_t)}{dZ_t^2} &= \frac{1}{\omega^2} - \frac{1}{\tau_t^2} - e^{Z_t}. \end{aligned}$$

Therefore, the second derivative is always negative as long as  $\omega^2 \geq \tau_t^2$ ; in this case  $h(\cdot)$  is log-concave and the unique maximum occurs at the value  $Z_t = \widehat{Z}_t$  where the first derivative is equal to 0. When  $\omega^2 \geq \tau_t^2$ , setting the first derivative to 0 yields:

$$e^{\widehat{Z}_t} = \frac{N_t^* \omega^2 \tau_t^2 - \xi \tau_t^2 + \mu_t \omega^2}{\omega^2 \tau_t^2} - \widehat{Z}_t \frac{\omega^2 - \tau_t^2}{\omega^2 \tau_t^2}.$$

If we settle on  $\omega^2 > \tau_t^2$ , after some algebra we obtain

$$\widehat{Z}_t = \frac{N_t^* \omega^2 \tau_t^2 - \xi \tau_t^2 + \mu \omega^2}{\omega^2 - \tau_t^2} W_0\left(\frac{\omega^2 \tau_t^2}{\omega^2 - \tau_t^2} \exp\left(\frac{N_t^* \omega^2 \tau_t^2 - \xi \tau_t^2 + \mu \omega^2}{\omega^2 - \tau_t^2}\right)\right),$$

where  $W_0$  denotes the upper branch of the Lambert W function. However, this solution is cumbersome to work with and, more importantly, does not guarantee a more efficient solution.

iii) If we set  $\omega^2 = \tau_t^2$  we obtain

$$\widehat{Z}_t = \log\left(\frac{N_t^* \tau_t^2 - \xi + \mu_t}{\tau_t^2}\right).$$

Replacing  $\omega^2 = \tau_t^2$  in (10) we obtain

$$\log h(\widehat{Z}_t) = \widehat{Z}_t N_t^* + \frac{(\widehat{Z}_t - \xi)^2 - (\widehat{Z}_t - \mu_t)^2}{2\tau_t^2} - e^{\widehat{Z}_t}.$$

The value  $\hat{\xi}$  which minimizes  $\log h(\hat{Z}_t)$  is found using

$$\begin{aligned}\frac{d \log h(\hat{Z}_t)}{d \xi} &= \frac{d \log h(x)}{d x} \bigg|_{x=\hat{Z}_t} \frac{d \hat{Z}_t}{d \xi} + \frac{d \log h(x)}{d \xi} \bigg|_{x=\hat{Z}_t} \\ &= \frac{d \log h(x)}{d \xi} \bigg|_{x=\hat{Z}_t} \\ &= -\frac{\hat{Z}_t - \xi}{\tau_t^2}.\end{aligned}$$

Since the second derivative is always positive,  $\log h(\hat{Z}_t)$  is a log-convex function with an unique minimum at  $\xi = \hat{Z}_t$ . However, since  $\hat{Z}_t$  is itself a function of  $\xi$ , from the following condition

$$\xi = \log \left( \frac{N_t^* \tau_t^2 - \xi + \mu_t}{\tau_t^2} \right),$$

after some algebra, we obtain the final expression

$$\hat{\xi} = N_t^* \tau_t^2 + \mu_t - W_0(\tau_t^2 \exp(N_t^* \tau_t^2 + \mu_t)).$$

□

The proposed algorithm falls within the class of Markov chain Monte Carlo EM (MCMC-EM) since it is an MCEM optimization algorithm in which the draws required to complete the E-step are obtained using the Gibbs sampler. For the efficiency of the algorithm, it is important to choose carefully the initialization point. Regarding the starting point  $\boldsymbol{\theta}^{(0)}$ , we set it equal to the method of moments estimator that is provided by the following proposition.

**Proposition 2.** *In the Gompertz model with Poisson sampling error distribution, the mean, variance and covariances for  $\{N_t^*, 1 \leq t \leq T\}$  are:*

$$\begin{aligned}i) \mathbb{E}(N_t^*) &= \exp\left(\theta_1 + \frac{\theta_2}{2}\right), \\ ii) \text{Var}(N_t^*) &= \exp(2\theta_1 + \theta_2) \left( \exp(\theta_2) - 1 \right), \\ iii) \text{Cov}(N_t^*, N_{t+h}^*) &= \exp(2\theta_1 + \theta_2) \left( \exp(\theta_2(1+b)^h) - 1 \right).\end{aligned}$$

*Proof.* Equation (3) implies

$$\mathbb{E}(Z_t) = \theta_1, \text{Var}(Z_t) = \theta_2, \text{Cov}(Z_t, Z_{t+h}) = \theta_2(1+b)^{|h|}.$$

On the other hand,  $N_t = \exp(Z_t)$  and based on the from the expression for the moment generating function of a Gaussian random variable, we get

$$\mathbb{E}(\exp(cZ_t)) = \exp\left(c\theta_1 + \frac{c^2\theta_2}{2}\right),$$

and thus the following expressions hold

$$\begin{aligned}\mathbb{E}(N_t) &= \mathbb{E}(\exp(Z_t)) = \exp\left(\theta_1 + \frac{\theta_2}{2}\right), \\ \text{Var}(N_t) &= \mathbb{E}(\exp(2Z_t)) - \mathbb{E}^2(\exp(Z_t)) = \exp(2\theta_1 + \theta_2)(\exp(\theta_2) - 1), \\ \text{Cov}(N_t, N_{t+h}) &= \mathbb{E}(\exp(Z_t + Z_{t+h})) - \mathbb{E}^2(\exp(Z_t)) = \exp(2\theta_1 + \theta_2)(\exp(\theta_2(1+b)^h) - 1).\end{aligned}$$

The last expression is valid because  $Z_t + Z_{t+h} \sim N(2\theta_1, 2\theta_2(1 + (1+b)^h))$  and  $\mathbb{E}(\exp(Z_t)) = \mathbb{E}(\exp(Z_{t+h}))$ . Furthermore, since  $N_t^*|N_t \stackrel{ind}{\sim} \text{Po}(N_t)$  we have  $\mathbb{E}(N_t^*|N_t) = \text{Var}(N_t^*|N_t) = N_t$  and  $\text{Cov}(N_t^*, N_{t+h}^*|N_t, N_{t+h}) = 0$  due to conditional independence. Using the law of total expectation, the law of total variance and the law of total covariance is easy to obtain the first two moments for  $N_t^*$ , as

$$\begin{aligned}\mathbb{E}(N_t^*) &= \mathbb{E}(\mathbb{E}(N_t^*|N_t)) = \exp\left(\theta_1 + \frac{\theta_2}{2}\right), \\ \text{Var}(N_t^*) &= \mathbb{E}(\text{Var}(N_t^*|N_t)) + \text{Var}(\mathbb{E}(N_t^*|N_t)) = \exp(2\theta_1 + \theta_2)(\exp(\theta_2) - 1), \\ \text{Cov}(N_t^*, N_{t+h}^*) &= \mathbb{E}(\text{Cov}(N_t^*, N_{t+h}^*|N_t, N_{t+h})) + \text{Cov}(\mathbb{E}(N_t^*|N_t), \mathbb{E}(N_{t+h}^*|N_{t+h})) \\ &= \exp(2\theta_1 + \theta_2)(\exp(\theta_2(1+b)^h) - 1).\end{aligned}$$

□

We set the starting point  $\boldsymbol{\theta}^{(0)}$  in order to match the first two moments provided by the above proposition with their sample counterparts. Regarding the covariance, we set  $h = 1$  because the first lag is the most efficient to estimate.

We use a sampling importance resampling strategy for the initialization of  $\mathbf{Z}$ . The importance density is equal to

$$\begin{aligned}q(\mathbf{Z}) &= \pi(Z_1|N_1^*, Z_2 = \log(N_2^*))\pi(Z_2|N_2^*, Z_1, Z_3 = \log(N_3^*)) \dots \\ &\quad \times \pi(Z_{t-1}|N_{t-1}^*, Z_{t-2}, Z_t = \log(N_t^*))\pi(Z_t|N_t^*, Z_{t-1}).\end{aligned}$$

Thus, the importance density for  $Z_t$  is equal to its full conditional used in the Gibbs sampler; the full conditional is obtained setting  $Z_{t+1} = \log(N_{t+1}^*)$ . Therefore, we draw 10,000 values of  $\mathbf{Z}$  from  $q(\cdot)$  and compute the importance weights. The initial value is then selected with probability proportional to these weights.

The proposed MCMC-EM method can be used computing the asymptotic variance of the MLE too. As noted in Caffo et al. [2005], it is easy to use the output of the sampling algorithm to calculate the inverse of the Fisher information matrix using the method of Louis [1982]. This allows us to produce asymptotic variance estimates for  $\hat{\boldsymbol{\theta}}$ .

## 4 Bayesian Inference

The Bayesian paradigm offers a different probabilistic mechanism to estimate finite sample variances for the estimators of interest, allows principled ways to incorporate prior knowledge when it is available, and to integrate model uncertainty into the predictions via model averaging. The crux of the approach is the posterior distribution which encodes all the uncertainty after observing the data. In the Gompertz model with Poisson errors, the posterior distribution is analytically intractable, so we must study it using MCMC sampling. The data augmentation strategy presented in Section 3 also plays a central role in the design of the MCMC algorithm. Perhaps surprisingly, the numerical experiments show that the algorithm for sampling the Bayesian posterior is much more efficient than the MCMC-EM in terms of computation time.

To perform a fair comparison with the MLE, we propose using a weakly informative prior. We use an uniform prior for  $b$ , that is,  $b \sim U(-2, 0)$ , and normal-inverse gamma priors for  $\theta_1$  and  $\theta_2$ ,

$$\theta_2 \sim \text{Inv.Gamma}(\varphi_1, \varphi_2) \text{ and } \theta_1 | \theta_2 \sim N(\eta_1, \eta_2 \theta_2).$$

We also assume that  $\theta_1, \theta_2$  are a priori independent of  $b$ . We choose values for the hyperparameters  $\varphi_1, \varphi_2, \eta_1$ , and  $\eta_2$  that lead to a weakly informative prior. We set  $\varphi_1 = \varphi_2 = 0.1$ ,  $\eta_1 = 0$ , and  $\eta_2 = 100$ . The sampling algorithm's steps do not depend on the particular values we choose for the hyperparameters, so we describe them in terms of generic values.

The dependent draws from the conditional distribution of all parameters and augmented data  $\pi(\theta_1, \theta_2, b, \mathbf{Z}, | N^*)$  are obtained using a Gibbs sampler.

Let  $\theta_1^{(k)}, \theta_2^{(k)}, b^{(k)}, \mathbf{Z}^{(k)}$  be the sample values in the  $k$ -th iteration of the MCMC algorithm. The starting values  $\theta_1^{(0)}, \theta_2^{(0)}, b^{(0)}$  are set equal to the values produced by the moment estimator method and  $\mathbf{Z}^{(0)}$  is drawn using sampling importance resampling, as in the MCMC-EM initialization.

Thus, given  $\theta_1^{(k)}, \theta_2^{(k)}, b^{(k)}$ , and  $\mathbf{Z}^{(k)}$ , we obtain  $\theta_1^{(k+1)}, \theta_2^{(k+1)}, b^{(k+1)}$ , and  $\mathbf{Z}^{(k+1)}$  using the following update scheme:

1. For  $t = 1, 2, \dots, T$ , sample  $Z_t^{(k+1)} | N_t^*, \theta_1^{(k)}, \theta_2^{(k)}, b^{(k)}, \mathbf{Z}_{-t}^{(k)}$ .
2. Sample  $\theta_1^{(k+1)}, \theta_2^{(k+1)}, b^{(k+1)} | \mathbf{Z}^{(k+1)}$ .
  - (a) Sample  $b^{(k+1)} | \mathbf{Z}^{(k+1)}$ .
  - (b) Sample  $\theta_2^{(k+1)} | b^{(k+1)}, \mathbf{Z}^{(k+1)}$ .
  - (c) Sample  $\theta_1^{(k+1)} | \theta_2^{(k+1)}, b^{(k+1)}, \mathbf{Z}^{(k+1)}$ .

In step (1), we use the acceptance-rejection algorithm described in section 3. In steps (2b) and (2c) the full conditionals are standard, since

$$\begin{aligned} \theta_2 | b, \mathbf{Z} &\sim \text{Inv.Gamma} \left( \phi_1 + \frac{T}{2}, \phi_2 + \frac{1}{2}(\mathbf{Z} - \eta_1 \mathbf{1}_T)'(\eta_2 \mathbf{1}_T \mathbf{1}_T' + B)^{-1}(\mathbf{Z} - \eta_1 \mathbf{1}_T) \right), \\ \theta_1 | \theta_2, b, \mathbf{Z} &\sim N \left( \frac{\eta_1 + \eta_2 \mathbf{1}_T' B^{-1} \mathbf{Z}}{1 + \eta_2 \mathbf{1}_T' B^{-1} \mathbf{1}_T}, \frac{\eta_2 \theta_2}{1 + \eta_2 \mathbf{1}_T' B^{-1} \mathbf{1}_T} \right), \end{aligned}$$

where  $1_T$  denotes the  $T$ -dimensional vector of ones. In contrast, step (2a) is not standard. To sample from the conditional density of  $b$  given  $\mathbf{Z}$  we use another accept-reject algorithm. The proposal is the uniform prior itself; thus the upper bound between the target density and the proposal density is the maximum of  $\pi(\mathbf{Z}|b)$ . Since the prior of  $(\theta_1, \theta_2)$  is a normal-inverse gamma distribution, straightforward calculation yields  $Z|\theta_2, b \sim N_T(\eta_1 1_T, \theta_2(\eta_2 1_T 1'_T + B))$  and then

$$\begin{aligned}\pi(\mathbf{Z}|b) &= \int_{\theta_2} \int_{\theta_1} \pi(\mathbf{Z}|\theta_1, \theta_2, b) \pi(\theta_1, \theta_2) d\theta_1 d\theta_2 \\ &\propto \det^{-\frac{1}{2}}(\eta_2 1_T 1'_T + B) \int_0^{+\infty} \theta_2^{-\frac{2\phi_1+T}{2}-1} \exp\left(-\frac{1}{\theta_2}\left(\phi_2 + \frac{1}{2}(\mathbf{Z} - \eta_1 1_T)'(\eta_2 1_T 1'_T + B)^{-1}(\mathbf{Z} - \eta_1 1_T)\right)\right) d\theta_2 \\ &\propto \det^{-\frac{1}{2}}(\eta_2 1_T 1'_T + B) \left(1 + \frac{1}{2\phi_2}(\mathbf{Z} - \eta_1 1_T)'(\eta_2 1_T 1'_T + B)^{-1}(\mathbf{Z} - \eta_1 1_T)\right)^{-\frac{2\phi_1+T}{2}}.\end{aligned}\quad (11)$$

The expression above can be further simplified using the matrix determinant lemma, Sherman-Morrison formula, and the closed-form expression for  $B^{-1}$ . The latter is available because  $B$  is the correlation matrix of a stationary autoregressive process of order 1.

**Proposition 3.** *In the Gompertz model with Poisson sampling error distribution, the density of  $\mathbf{Z}|b$  is*

$$\begin{aligned}\pi(\mathbf{Z}|b) &\propto (1-r^2)^{1-\frac{T}{2}} \left((\eta_2(T-2)-1)r^2 - 2\eta_2(T-1)r + \eta_2T + 1\right)^{-\frac{1}{2}} \\ &\times \left(1 + \frac{r^2 \sum_{s=2}^{T-1} W_s^2 - 2r \sum_{h=1}^{T-1} W_h W_{h+1} + \sum_{t=1}^T W_t^2}{2\phi_2(1-r^2)} - \frac{\eta_2(1-r^2)}{(\eta_2(T-2)-1)r^2 - 2\eta_2(T-1)r + \eta_2T + 1}\right. \\ &\times \left. \frac{r^2 \left(\sum_{s=2}^{T-1} W_s\right)^2 - 2r \sum_{s=2}^{T-2} W_s \sum_{t=1}^T W_t + \left(\sum_{t=1}^T W_t\right)^2}{2\phi_2(1+r^2)}\right)^{-\phi_1-\frac{T}{2}},\end{aligned}\quad (12)$$

where  $W_t = Z_t - \eta_1$  and  $r = 1 + b$ .

*Proof.* Let  $r = 1 + b$  and  $W = Z - \eta_1 1_T$ . Then [Hamilton, 1994][Section 5.2]

$$\begin{aligned}B &= \begin{bmatrix} 1 & r & r^2 & \dots & r^{T-1} \\ r & 1 & r & \dots & r^{T-1} \\ \dots & \dots & \dots & \dots & \dots \\ r^{T-1} & r^{T-2} & r^{T-3} & r^{T-4} & 1 \end{bmatrix}, \det(B) = (1-r^2)^{T-1}, \\ B^{-1} &= \frac{1}{1-r^2} \begin{bmatrix} 1 & -r & & & \\ -r & 1+r^2 & -r & & \\ & \ddots & \ddots & \ddots & \\ & & -r & 1+r^2 & -r \\ 0 & & & -r & 1 \end{bmatrix}.\end{aligned}$$

Using the matrix determinant lemma and Sherman-Morrison formula, it is obtained

$$\begin{aligned}\det(\eta_2 \mathbf{1}_T \mathbf{1}_T' + B) &= \det(B)(1 + \eta_2 \mathbf{1}_T' B^{-1} \mathbf{1}_T), \\ (\eta_2 \mathbf{1}_T \mathbf{1}_T' + B)^{-1} &= B^{-1} - \eta_2 \frac{B^{-1} \mathbf{1}_T \mathbf{1}_T' B^{-1}}{1 + \eta_2 \mathbf{1}_T' B^{-1} \mathbf{1}_T}.\end{aligned}$$

This implies

$$\pi(Z|b) \propto \det^{-\frac{1}{2}}(B)(1 + \eta_2 \mathbf{1}_T' B^{-1} \mathbf{1}_T)^{-\frac{1}{2}} \left( 1 + \frac{W' B^{-1} W}{2\phi_2} - \frac{\eta_2}{(1 + \eta_2 \mathbf{1}_T' B^{-1} \mathbf{1}_T)} \frac{W' B^{-1} \mathbf{1}_T \mathbf{1}_T' B^{-1} W}{2\phi_2} \right). \quad (13)$$

Consider separately  $1 + \eta_2 \mathbf{1}_T' B^{-1} \mathbf{1}_T$ . It is straightforward that  $\mathbf{1}_T' B^{-1} \mathbf{1}_T$  is the sum of all components of the matrix  $B^{-1}$ ; in this matrix there are  $(T-2)$ ,  $2(T-1)$ , and two times  $1+r^2$ ,  $-r$ , and one respectively. Then

$$\begin{aligned}1 + \eta_2 \mathbf{1}_T' B^{-1} \mathbf{1}_T &= 1 + \frac{\eta_2}{1-r^2} (2 + (1+r^2)(T-2) - 2r(T-1)) \\ &= \frac{1-r^2 + \eta_2(r^2(T-2) - 2r(T-1) + T)}{1-r^2} \\ &= \frac{(\eta_2(T-2) - 1)r^2 - 2\eta_2(T-1)r + \eta_2 T + 1}{1-r^2}.\end{aligned} \quad (14)$$

We now focus our attention on  $W' B^{-1} W$ , which can be simplified as follows:

$$\begin{aligned}W' B^{-1} W &= \frac{1}{1-r^2} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_T \end{bmatrix}' \begin{bmatrix} 1 & -r & 0 & \cdots & 0 \\ -r & 1+r^2 & -r & \cdots & 0 \\ 0 & -r & 1+r^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -r \\ 0 & 0 & \cdots & -r & 1 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_T \end{bmatrix} \\ &= \frac{1}{1-r^2} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_T \end{bmatrix}' \begin{bmatrix} W_1 - rW_2 \\ (1+r^2)W_2 - r(W_1 + W_3) \\ (1+r^2)W_3 - r(W_2 + W_4) \\ \vdots \\ (1+r^2)W_t - r(W_{t-1} + W_{t+1}) \\ \vdots \\ (1+r^2)W_{T-1} - r(W_{T-2} + W_T) \\ W_T - rW_{T-1} \end{bmatrix} \\ &= \frac{1}{1-r^2} \left( W_1^2 - rW_1W_2 + W_T^2 - rW_{T-1}W_T + \sum_{s=2}^{T-1} \left[ (1+r^2)W_s^2 - rW_s(W_{s-1} + W_{s+1}) \right] \right) \\ &= \frac{1}{1-r^2} \left( \sum_{t=1}^T W_t^2 - rW_1W_2 - r \sum_{s=2}^{T-1} W_s W_{s+1} - rW_T W_{T-1} - r \sum_{s=2}^{T-1} W_s W_{s-1} + r^2 \sum_{s=2}^{T-1} W_s^2 \right)\end{aligned}$$

$$= \frac{1}{1-r^2} \left( r^2 \sum_{s=2}^{T-1} W_s^2 - 2r \sum_{s=2}^{T-1} W_s W_{s+1} + \sum_{t=1}^T W_t^2 \right). \quad (15)$$

Similarly,  $W'B^{-1}1_T$  yields

$$\begin{aligned} W'B^{-1}1_T &= \frac{1}{1-r^2} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_T \end{bmatrix}' \begin{bmatrix} 1 & -r & 0 & \cdots & 0 \\ -r & 1+r^2 & -r & \cdots & 0 \\ 0 & -r & 1+r^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -r \\ 0 & 0 & \cdots & -r & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \\ &= \frac{1}{1-r^2} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_T \end{bmatrix}' \begin{bmatrix} 1-r \\ r^2-2r+1 \\ \cdots \\ r^2-2r+1 \\ 1-r \end{bmatrix} \\ &= \frac{(W_1 + W_T)(1-r) + (1-r)^2 \sum_{s=2}^{T-1} W_s}{(1-r)(1+r)} \\ &= \frac{\sum_{t=1}^T W_t - r \sum_{s=2}^{T-1} W_s}{1+r}. \end{aligned}$$

Therefore,

$$\begin{aligned} W'B^{-1}1_T 1_T' B^{-1} W &= (W'B^{-1}1_T)^2 \\ &= \frac{\left( \sum_{t=1}^T W_t - r \sum_{s=2}^{T-1} W_s \right)^2}{(1+r)^2} \\ &= \frac{r^2 \left( \sum_{s=2}^{T-1} W_s \right)^2 - 2r \sum_{s=2}^{T-1} W_s \sum_{t=1}^T W_t + \left( \sum_{t=1}^T W_t \right)^2}{(1+r)^2} \end{aligned} \quad (16)$$

Finally, the form (12) is obtained by substituting into (13) the expressions (14), (15), and (16).  $\square$

It is important to note that equation (12) is inexpensive to evaluate, as it is just a combination of ratios and products of polynomials in  $b$ . This allows us to adopt the following two-step strategy. We first perform a grid search for the maximum of  $\pi(\mathbf{Z}|b)$ , as a function of  $b$ . To this end, we compute (12) over a dense sequence of values of  $b$ , say  $-1.99, -1.98, \dots, -0.01$ , and set

$$b_0 = \arg \max_{b=-1.99, -1.98, \dots, -0.01} \pi(\mathbf{Z}|b).$$

It is reasonable that the value of  $b_0$  is close to the global maximum of  $\pi(\mathbf{Z}|b)$ . Then, we use the

L-BFGS-B algorithm to find the global maximum starting from  $b_0$ . Although this optimization must be repeated at each iteration of the Gibbs sampler, our method is fast due to the simplicity of expression (12). In fact, as reported in Section 5.1, the Gibbs sampler is much faster than the MCMC-EM optimizer. Furthermore, as reported in the Supplementary Material, our proposed Gibbs sampler is also faster than the state-of-the-art represented by the STAN implementation.

## 5 Numerical Experiments

The purpose of running the numerical experiments in this section is two-fold. First, we examine the difference between the frequentist and Bayesian inferences, which can be informative when the data has a modest size and we want to gauge the influence of the prior. Second, we provide proof-of-concept for the algorithms proposed and examine their performance on data generated under different scenarios, including one in which the sampling error distribution is misspecified. We also consider a comparison between the Bayesian inference produced using the sampler we design here and the one using a generic STAN implementation. The latter can be considered at this point state of the art since, to our knowledge, no other MCMC sampler has been designed for this problem. Finally, we illustrate these methods by analyzing a time series of American Redstart counts from the Breeding Bird Survey data.

### 5.1 Simulations

#### 5.1.1 Simulation scenarios

The simulation study evaluates the performance of the proposed method when the sampling model is correct and when it is misspecified. Specifically, we generated time series of varying lengths under the Poisson noise model and under a model with sampling errors that have a negative binomial distribution. In addition, we varied the model parameters to examine the algorithms under different scenarios. One of the scenarios has parameter values close to the estimates obtained in the Redstart data analysis.

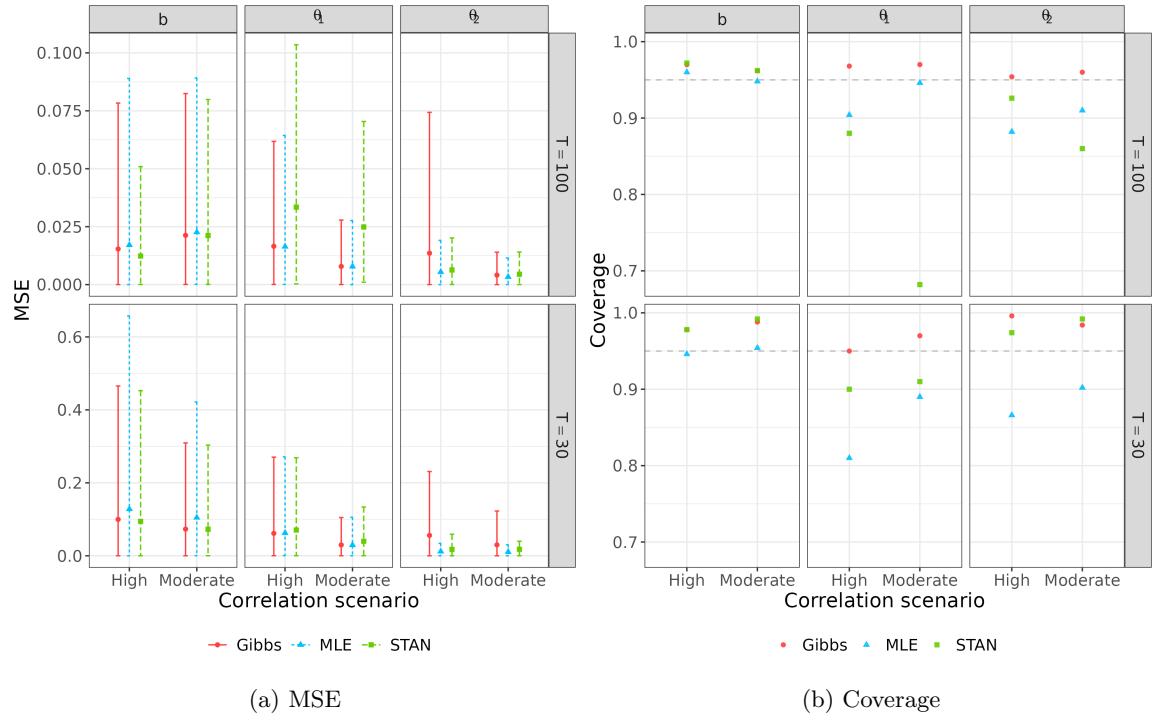
The data are fitted using the frequentist approach described in Section 3, and using the Bayesian methods described in Section 4. In addition, Bayesian inference is produced using an off-the-shelf STAN implementation.

We analyze data generated under eight simulation scenarios. Scenario **S1** yields a moderate level of serial correlation given by  $b^\dagger = -0.5$ , while setting **S2** exhibits high levels of correlation given by  $b^\dagger = -0.22$ . Based on the parameters estimates obtained from the Redstart data analysis, we fixed  $\theta_1^\dagger = 2$  and  $\theta_2^\dagger = 0.22$ , for both **S1** and **S2**. We simulated these settings considering time series of lengths  $T = 30$ . The scenarios **S3** and **S4** use the same parameters as, respectively, **S1** and **S2**, but have  $T = 100$ .

Scenarios **S5** - **S8** are produced using the same parameter values as, respectively, **S1** - **S4**, but use a misspecified sampling error model. Specifically, the sampling error was simulated using a negative binomial distribution with mean  $\exp(Z_t)$  and variance  $2\exp(Z_t)$ . These values were chosen to provide a realistic comparison with the Poisson model. Note that a negative binomial with the

mean and variance specified above corresponds to one with a success probability of 0.5 and the dispersion (size) parameter selected so that the mean equals  $\exp(Z_t)$ . As the probability of success  $p$  approaches zero, the negative binomial distribution converges to a Poisson distribution with the same mean. Thus, setting  $p = 0.5$  ensures that the model is sufficiently different to illustrate the impacts of misspecification. For each setting, we then computed the mean square error (MSE) and the coverage of the 95% confidence and credible intervals, for every parameter in each scenario.

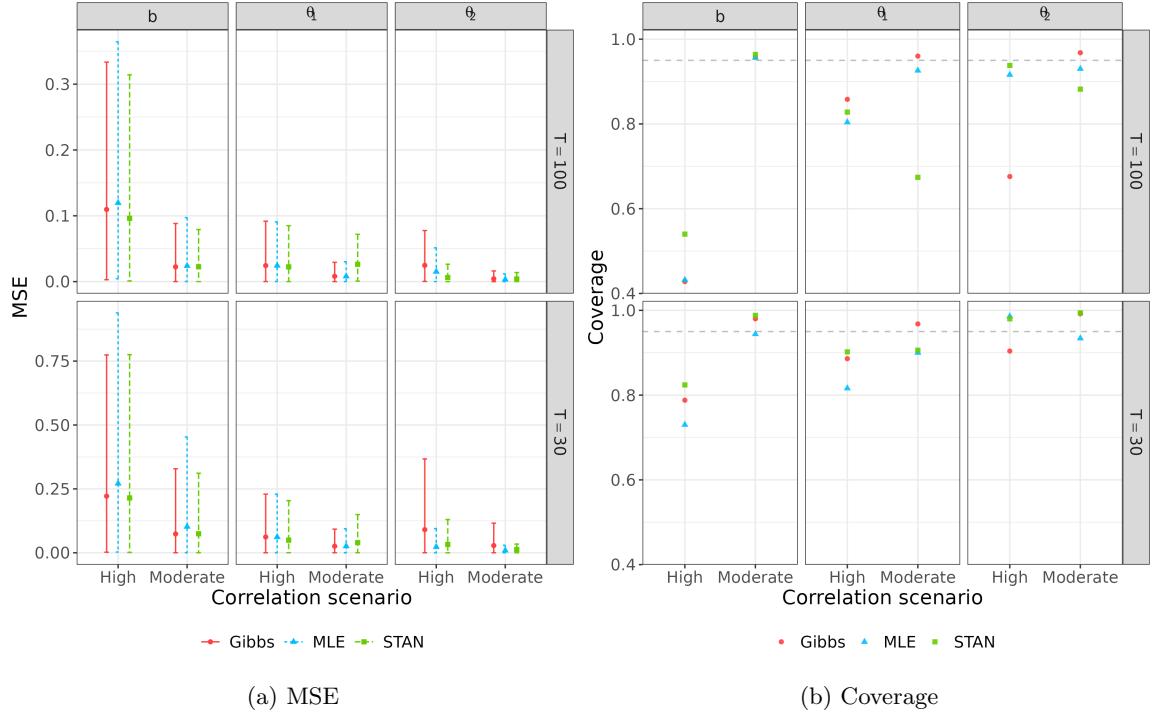
The simulation analysis is conducted with 500 independent replicates, and for the Bayesian methods, each replicate is based on an MCMC sample of size 10,000. In addition, we report the effective sample sizes obtained by our sampler and STAN. Finally, we compute and report summaries of computation times.



**Figure 1: Results for the correctly specified model with Poisson distributed errors:** Mean square error and 95% coverage by correlation setting for each parameter and time series length  $T$  for scenarios **S1 - S4**. Within each sub-figure, columns represent parameters and rows represent the different time series length  $T$ . Results from the Gibbs sampler are shown with points and solid lines, results for the MLE approach are indicated by triangles and fine dashed lines, and results for STAN are displayed with squares and bold dashed lines. In the MSE plots (left panels), the central points/triangles indicate the mean values, and the error bars represent the 5th and 95th percentiles. The coverage plots (right panels) show the coverage obtained using credible intervals approximated from Gibbs samples (red points), STAN samples (green squares), and using Wald confidence intervals obtained using the MLE and its asymptotic variance produced by MCMC-EM.

### 5.1.2 Simulation results

As expected, the analysis based on the three algorithms improves as the time series length increases. Figure 1 presents the results for the correctly specified model with Poisson distributed errors. In



**Figure 2: Results for the misspecified model with Negative binomial distributed errors:** Mean square error and 95% coverage by correlation scenario for each parameter and time series length  $T$  for scenarios **S5 - S8**. Within each sub-figure, columns represent parameters and rows represent the different time series length  $T$ . Results from the Gibbs sampler are shown with points and solid lines, results for the MLE approach are indicated by triangles and fine dashed lines, and results for STAN are displayed with squares and bold dashed lines. In the MSE plots (left panels), the central points/triangles indicate the mean values, and the error bars represent the 5th and 95th percentiles. The coverage plots (right panels) show the coverage obtained using credible intervals approximated from Gibbs samples (red points), STAN samples (green squares), and using Wald confidence intervals obtained using the MLE and its asymptotic variance produced by MCMC-EM.

most cases, the MSE values across the three inference methods are similar. The uncertainty is generally higher for the estimates of parameter  $b$ , lower for  $\theta_1$ , and even lower for  $\theta_2$ . An exception occurs with STAN when  $T = 100$ , where the uncertainty for  $\theta_1$  is greater than for  $b$ . Furthermore, for  $\theta_2$ , the Gibbs sampler produced estimates with greater uncertainty compared to the other two algorithms. In terms of coverage, all three inference methods yield values close to 0.95 for the parameter  $b$ . However, the MLE approach shows lower coverage, particularly when  $T = 30$  and the correlation is high. This is likely due to the sample size being too small for the asymptotic variance estimate to be reliable. When  $T = 100$ , the Gibbs sampler demonstrates the most accurate coverage across all parameters and correlation scenarios.

The results for the misspecified case are shown in Figure 2. In this setting, the performance of the algorithms differs noticeably between the high and moderate correlation scenarios. When the correlation is lower, the three algorithms exhibit improved performance, with lower MSE, reduced uncertainty, and more accurate coverage across all parameters and both sample sizes. Again, higher uncertainty is observed for the estimates of  $b$  and  $\theta_1$ , and lower uncertainty for  $\theta_2$ . Moreover, when

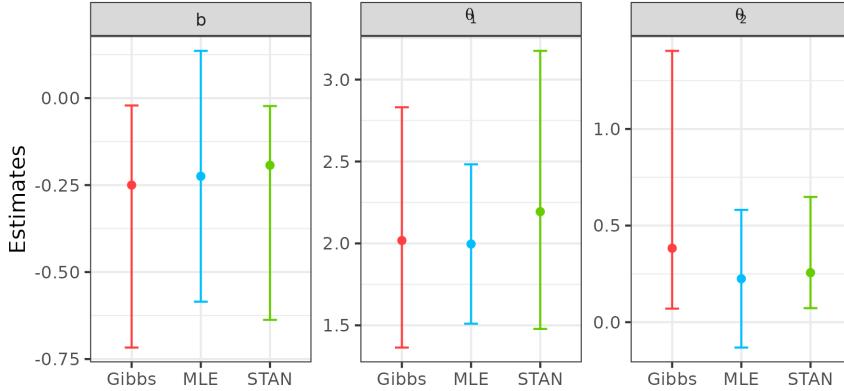


Figure 3: Point estimates for the three parameters obtained using each inference method. For the two Bayesian methods, the error bars represent the 95% credible intervals, while for the MLE, they indicate the Wald 95% confidence intervals .

$T = 100$ , the Gibbs sampler provides the most accurate coverage for all parameters.

	1st Quantile	Mean	Median	3rd Quantile
<b>Gibbs</b>	1.33	1.33	1.33	1.34
<b>MLE</b>	9.21	11.82	11.23	13.38
<b>STAN</b>	6.98	7.66	7.09	7.31

Table 1: Summary statistics of the computational times (in minutes) under scenario **S4**

The computational times for scenario **S4** are reported in Table 1. The MLE approach was the slowest, taking between two and nine times longer than the other two. The Gibbs sampler outperforms STAN in terms of speed, with a mean computation time that is approximately five times shorter. The other scenarios exhibited similar patterns and are presented in the Supplementary Material. Finally, a comparison of the effective sample sizes for the three parameters across the two Bayesian inference methods is also provided there. In all cases, the Gibbs sampler outperformed STAN, yielding effective sample sizes per second that were around 1.5 to 3.5 times larger. This is of course not surprising, since the algorithm presented here has been designed specifically for this problem, while STAN is a sampling algorithm which is able to tackle a wide variety of models.

## 5.2 Real data analysis

We analyze here the American Redstart counts dataset which was previously discussed by Lele [2006] and Dennis et al. [2006]. This data set is recorded with the number 0214332808636 in the North American Breeding Bird Survey [Peterjohn, 1994, Robbins et al., 1986] and contains the number of specimens observed from 1966 to 1995 at a survey location. We fit the Gompertz model with Poisson-distributed errors to this dataset using both the Gibbs sampler and the MLE approach proposed here. STAN was also used as a baseline for comparison within the Bayesian framework.

We computed point estimates for the three parameters, along with the 95% credible intervals for

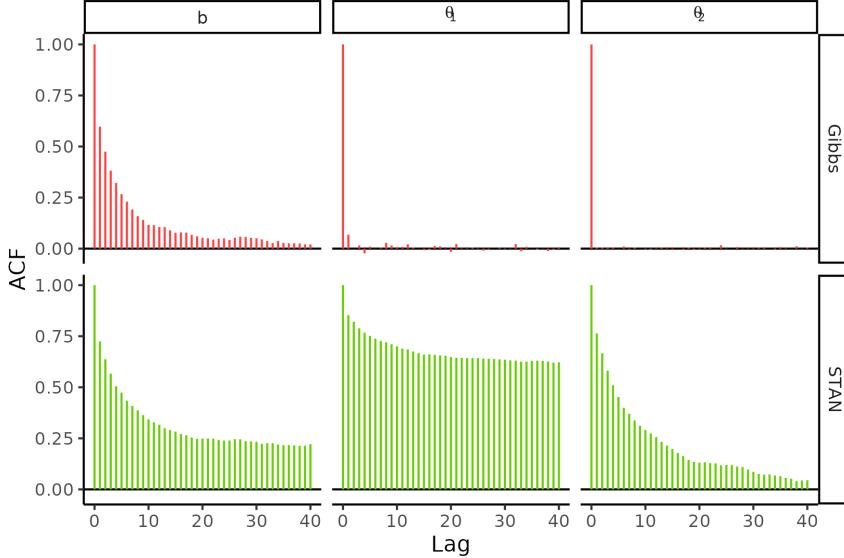


Figure 4: Autocorrelation functions of the posterior samples for the three parameters under both Bayesian inference methods.

	b	$\theta_1$	$\theta_2$
<b>Gibbs</b>	1106.8	8239.5	10000
<b>STAN</b>	209.4	30.3	583.4

Table 2: Effective sample sizes for the three parameters from the analysis of the American Redstart for both Bayesian inference methods

the Bayesian methods and the 95% confidence intervals based on Gaussian approximation for the MLE approach (Figure 3). For the two Bayesian methods, we also calculated the effective sample size and autocorrelation function (Figure 4) for each parameter (Table 2).

Although the Gibbs sampler shows greater uncertainty in the estimation of  $\theta_2$ , the point estimates across the three methods remain comparable. However, when examining the effective sample size and autocorrelation functions of the posterior distributions for the two Bayesian methods, the Gibbs sampler demonstrates superior performance. The effective sample size for  $\theta_2$  is nearly twice as large with the Gibbs sampler, six times larger for  $b$ , and increases for  $\theta_1$ , from only 30 in STAN to more than 8200 with the Gibbs sampler.

## 6 Discussion

We develop full-likelihood-based inference within the frequentist and Bayesian paradigms for the Gompertz model with Poisson sampling errors. The proposed approaches remove the need to consider pseudo-likelihood methods that mitigate computational challenges at the price of reducing the information provided by the data.

In our future work, we would like to investigate whether similar developments can be produced to modified versions of the model considered here. The latter can be created by modifying the

growth curve by adding parameters that allow curvature and long-term behavior [Asadi et al., 2023] or when the population dynamics is determined by a stochastic differential equation as in Donnet et al. [2010].

## Acknowledgements

The first author was supported by MUR - Prin 2022 - Grant no. 2022FJ3SLA, funded by the European Union – Next Generation EU. The third author was supported by NSERC of Canada discovery grant RGPIN-2024-04506. The authors thank Dr. Monica Alexander and Dr. Vianey Leos-Barajas for comments that have improved the paper.

## Code

The developed methods are available in a new R package: **gse**. The new R package and the scripts of Section 5 are available online on GitHub at the following link: <https://github.com/sofiar/GMLossF>

## References

M. Alexander, E. Zagheni, and M. Barbieri. A flexible Bayesian model for estimating subnational mortality. *Demography*, 54(6):2025–2041, 2017.

M. Asadi, A. Di Crescenzo, F. A. Sajadi, and S. Spina. A generalized Gompertz growth model with applications and related birth-death processes. *Ricerche di Matematica*, 72(2):1–36, 2023.

M. Auger-Méthé, K. Newman, D. Cole, F. Empacher, R. Gryba, A. A. King, V. Leos-Barajas, J. Mills Flemming, A. Nielsen, G. Petris, and L. Thomas. A guide to state-space modeling of ecological time series. *Ecological Monographs*, 91(4):e01470, 2021.

F. Baione and S. Levantesi. Pricing critical illness insurance from prevalence rates: Gompertz versus Weibull. *North American Actuarial Journal*, 22(2):270–288, 2018.

J. G. Booth and J. P. Hobert. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(1):265–285, 1999.

Z. Butt and S. Haberman. Application of frailty-based mortality models using generalized linear models. *ASTIN Bulletin: The Journal of the IAA*, 34(1):175–197, 2004.

B. Caffo, W. Jank, and G. Jones. Ascent-based Monte Carlo expectation–maximization. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):235–251, 2005.

K.-S. Chan and J. Ledolter. Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, 90(429):242–252, 1995.

A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22, 1977.

B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples. Estimating density dependence, process noise, and observation error. *Ecological Monographs*, 76(3):323–341, 2006.

S. Donnet, J.-L. Foulley, and A. Samson. Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics*, 66(3):733–741, 2010.

B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS. *Philosophical transactions of the Royal Society of London*, (115):513–583, 1825.

J. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.

J. A. Hostetler and R. B. Chandler. Improved state-space models for inference about spatial and temporal variation in abundance from count data. *Ecology*, 96(6):1713–1723, 2015.

S. R. Lele. Sampling variability and estimates of density dependence: A composite-likelihood approach. *Ecology*, 87(1):189–202, 2006.

A. Lindén and J. Knape. Estimating environmental effects on population dynamics: consequences of observation error. *Oikos*, 118(5):675–680, 2009.

T. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 44(2):226–233, 1982.

C. E. McCulloch. Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association*, 89(425):330–335, 1994.

K. Newman, S. Buckland, B. Morgan, R. King, D. Borchers, D. Cole, P. Besbeas, O. Gimenez, and L. Thomas. *Modelling Population Dynamics: Model Formulation, Fitting and Assessment using State-Space Methods*. Springer, 2014.

B. Peterjohn. The North American breeding bird survey. *Birding*, 26(6):386–398, 1994.

C. S. Robbins, D. Bystrak, and P. H. Geissler. The Breeding Bird Survey: its first fifteen years, 1965-1979. Technical report, US Fish and Wildlife Service, 1986.

D. F. Staples, M. L. Taper, and B. Dennis. Estimating population trend and process variation for pva in the presence of sampling error. *Ecology*, 85(4):923–929, 2004.

T. H. Tai and A. Noymer. Models for estimating empirical Gompertz mortality: With an application to evolution of the Gompertzian slope. *Population Ecology*, 60(1-2):171–184, 2018.

K. M. C. Tjørve and E. Tjørve. The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the unified-Richards family. *PLOS ONE*, 12(6):1–17, 06 2017.

G. C. Wei and M. A. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. Journal of the American statistical Association, 85(411):699–704, 1990.

C. P. Winsor. The Gompertz curve as a growth curve. Proc Natl Acad Sci U S A, 18(1):1–8, Jan. 1932.

## Supplement: Extra Figures and Tables

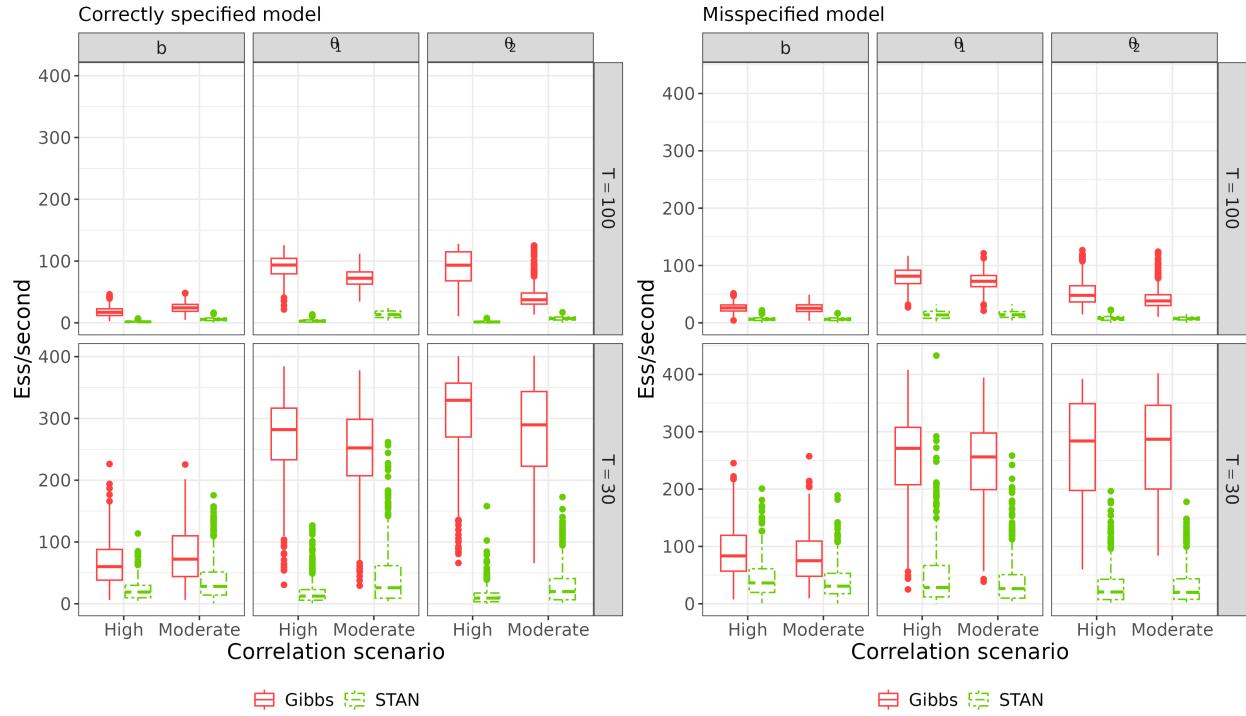


Figure 1: Effective sample size per second obtained from the simulation analysis for both Bayesian inference methods for each parameter and time series length  $T$ . Results from the Gibbs sampler are shown with solid red lines, while those from STAN are displayed with green dashed lines.

### Correctly specified model

	1st Quantile	Mean	Median	3rd Quantile	Scenario
<b>Gibbs</b>	0.44	0.45	0.45	0.46	<b>S1</b>
<b>MLE</b>	3.43	4.82	4.11	5.41	<b>S1</b>
<b>STAN</b>	0.57	0.97	0.76	0.97	<b>S1</b>
<b>Gibbs</b>	0.43	0.44	0.44	0.44	<b>S2</b>
<b>MLE</b>	3.04	4.16	3.48	4.55	<b>S2</b>
<b>STAN</b>	0.53	0.7	0.57	0.8	<b>S2</b>
<b>Gibbs</b>	1.32	1.34	1.33	1.37	<b>S3</b>
<b>MLE</b>	11.46	14.5	13.58	16.87	<b>S3</b>
<b>STAN</b>	8.77	12.7	11.82	14.87	<b>S3</b>

Table 1: Summary statistics of the computational times (in minutes) under scenarios **S1-S3**

Misspecified model					
	1st Quantile	Mean	Median	3rd Quantile	Scenario
<b>Gibbs</b>	0.44	0.45	0.45	0.46	<b>S1</b>
<b>MLE</b>	3.02	3.9	3.49	4.05	<b>S1</b>
<b>STAN</b>	0.53	0.67	0.58	0.72	<b>S1</b>
<b>Gibbs</b>	0.43	0.44	0.44	0.45	<b>S2</b>
<b>MLE</b>	3.04	3.9	3.59	4.36	<b>S2</b>
<b>STAN</b>	0.53	0.69	0.56	0.76	<b>S2</b>
<b>Gibbs</b>	1.33	1.36	1.37	1.39	<b>S3</b>
<b>MLE</b>	10.06	12.12	11.59	13.97	<b>S3</b>
<b>STAN</b>	7.04	7.71	7.15	3.37	<b>S3</b>
<b>Gibbs</b>	1.33	1.34	1.34	1.35	<b>S4</b>
<b>MLE</b>	9.49	11.83	11.24	13.29	<b>S4</b>
<b>STAN</b>	7	7.66	7.11	7.36	<b>S4</b>

Table 2: Summary statistics of the computational times (in minutes) under scenarios **S1-S4**