

The Penalized Likelihood for Detecting Mortality Deceleration

Silvio C Patricio*¹

¹Interdisciplinary Center on Population Dynamics, University of Southern Denmark

Abstract

This paper offers methodological advances that contribute to an ongoing scientific debate about aging patterns, especially detecting the mortality deceleration through heterogeneity. We address how to detect a deceleration in mortality using the gamma-Gompertz model and life tables. We propose penalizing the likelihood function, using its maximization to solve this problem. We compare the results with the widely used maximum likelihood estimate (MLE). The results are promising for parameter estimation, especially for the detection of heterogeneity.

Keywords: Gompertz model; gamma-Gompertz model, Mortality deceleration; Penalized likelihood function; Maximum a posteriori.

1 Introduction

The interest in mortality modeling started in the 17th century when attempts were made to develop a quantitative theory that would lead to aging, mortality, and life expectancy. However, as Haberman (1996) pointed out, a new era for mortality modeling began in 1825 with the mortality model proposed by British actuary Benjamin Gompertz. He postulated that human mortality increases exponentially with age (Gompertz, 1825).

As it is well-known, Gompertz's ideas can be adequately expressed in terms of what we now call the force of mortality and is one of the most persuasive proposals in the early days of survival modeling. Many contributions throughout the second half of the 19th century generalize the Gompertz model of mortality, or in some way, proceed from the ideas brought by him. A notable example is given by the mortality model proposed by another British actuary, William Makeham (Makeham, 1860, 1867, 1890), who observed that the Gompertz mortality model could be improved by adding an age-independent constant to the exponential growth.

However, due to improvements in medicine, nutrition, and public health, it has been observed that there is a reduction in mortality from infectious diseases and natural disasters. As a consequence, we have observed a constant increase in the human life record (Vaupel et al., 2021) and also in its life expectancy, which has increased by four decades over 160 years (Cutler et al., 2006; Oeppen and Vaupel, 2002)

*silca@sam.sdu.dk

These improvements in mortality also bring a significant improvement in the data quality from older people. Thus when the oldest-old data is analyzed, we can observe a downward deviation from the exponentially increasing force of mortality proposed by Gompertz (1825). This phenomenon is called mortality deceleration (Thatcher et al., 1998) and is not exclusive to humans. It is also observed in other species, such as bugs, worms, and yeasts (Vaupel et al., 1998). In some cases, such a deceleration may be enough to cause a "leveling off" or plateau in the force of mortality at older ages (Newman, 2018).

As well discussed by Vaupel et al. (1979) and Wienke (2010), this type of mortality deceleration results if hazards differ between individuals. A proportional hazard frailty model is the standard approach used to model such heterogeneity (Böhnstedt and Gampe, 2019).

First, let X be a non-negative continuous random variable that describes human life spans, and its hazard function at time x , also known as mortality function or force of mortality at age x (Dickson et al., 2019), is given by

$$\mu(x) = \lim_{\varepsilon \downarrow 0} \mathbb{P}(x \leq X < x + \varepsilon | X \geq x).$$

The simple and efficient idea brought up by Vaupel et al. (1979) was to introduce a non-negative continuous Z random variable to modulate the individual hazards. So we have a proportional hazard frailty model written as

$$\mu(x|Z = z) = z\mu_0(x),$$

with $\mu_0(x)$ being the baseline hazard, $\mu(x|Z = z)$ denotes the conditional hazard of an individual at age x given its frailty $Z = z$. When the baseline hazard is exponentially increasing, as described by Gompertz (1825), $\mu_0(x) = ae^{bx}$, with parameters $a > 0$ and $b > 0$, and the fragility Z has gamma distribution with $\mathbb{E}(Z) = 1$ and $\text{VAR}(Z) = \sigma^2$, we get the so-called gamma-Gompertz force of mortality. The frailty variance parameter σ^2 describes the heterogeneity in the risk of dying: i.e., individuals with higher frailty values tend to die earlier, while more robust individuals tend to survive (Böhnstedt and Gampe, 2019).

Despite the frailty component in the model, we can easily see that the individual hazard $\mu(x|Z = z)$ is still exponentially increasing. However, when we look at the marginal hazard

$$\mu(x) = \frac{ae^{bx}}{1 + \sigma^2 \frac{a}{b} (e^{bx} - 1)}$$

we can see that if $\sigma^2 > 0$: it shows a downward deviation from the exponential growth shown in the proposed force of mortality by Gompertz (1825). This deviation is noticed in the high ages. If $\sigma^2 = 0$: i.e., when there is no heterogeneity, the model reduces to the Gompertz (1825) model with the hazard function $\mu(x) = ae^{bx}$, which is widely used.

The change in human longevity and mortality has generated an important and pertinent debate among demographers, economists, and actuaries. This discussion revolves around the existence or not of mortality deceleration (Gavrilova and Gavrilov, 2015; Newman, 2018). For the gamma-Gompertz model this question reduces to whether $\sigma^2 > 0$ or $\sigma^2 = 0$.

However, when $\sigma^2 = 0$, the parameter σ^2 is on the boundary of the parametric space, which violates the standard assumptions underlying the asymptotic properties of likelihood-based inference (see, for example, Böhnstedt and Gampe (2019)). Furthermore, in this case, the parameter to be estimated lies on the boundary. Therefore, limiting the parameter's distribution may not be normal (since when we maximize the likelihood function, we do so over the parameter space and not outside it).

In this paper, we present the problem of identifying whether $\sigma^2 > 0$ or $\sigma^2 = 0$ as a model misspecification problem. To solve it, we consider adding a penalty function in the likelihood function and using the Mean Squared Error (MSE) to compare the estimates obtained by maximizing the likelihood function and the penalized likelihood function.

2 Methodology

To start with, let us draw the problem of identifying whether $\sigma^2 > 0$ or $\sigma^2 = 0$ as a model misspecification problem. First, let us see what is happening when we misspecify the model.

When dealing with statistical modeling, we can assume that the model we fit is the correct model (usually unknown). In this situation, a natural question to ask is, what are we estimating?

Let us suppose that \mathbf{X} is a random sample which have density $g(x)$. However, we fit the incorrect family of densities $\{f(x; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ to the data using the Maximum Likelihood Estimator (MLE) and estimate $\boldsymbol{\theta}$. The misspecified log-likelihood is

$$\ell(\boldsymbol{\theta}; \mathbf{X}) = \sum_{i=1}^n \log f(X_i; \boldsymbol{\theta}).$$

Then we can use the Law of Larger Numbers to obtain the limit of $\ell(\boldsymbol{\theta}; \mathbf{X})$ and understand what the MLE is estimating.

$$\frac{1}{n} \ell(\boldsymbol{\theta}; \mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \log f(X_i; \boldsymbol{\theta}) \xrightarrow{a.s.} \mathbb{E}_g(\log f(X_1; \boldsymbol{\theta})) = \int \log f(X_i; \boldsymbol{\theta}) dG(x).$$

Therefore, it is clear that $\hat{\boldsymbol{\theta}}_n = \arg \max \ell(\boldsymbol{\theta}; \mathbf{X})$ is an estimator of $\arg \max (\int \log f(X_i; \boldsymbol{\theta}) dG(x))$. Thus, when we estimate the wrong model, what we get is the estimate that comes closest to the behavior observed in the data.

Now let's bring up this example to identify whether $\sigma^2 > 0$ or $\sigma^2 = 0$. Let us assume there is no heterogeneity in the data ($\sigma^2 = 0$), and we will fit the gamma-Gompertz model. In other words, we observe an exponential increase in the data, but we will estimate a model that considers a downward deviation from the exponential increase. As shown above, we will estimate σ^2 close to zero but never equal zero.

In this sense, Böhnstedt and Gampe (2019) has done much work to derive a hypothesis test from detecting heterogeneity. However, hypothesis testing has been widely discussed and rethought in the statistical community, especially as the alpha level (almost always 0.1, 0.05, or 0.01) is arbitrary. Furthermore, the presentation of only test statistics, degrees of freedom, and P values limit the effectiveness of (future) meta-analyses (see, for example, Thompson (2001); Nester (1996)).

Because of this, some authors have proposed alternative methods to the use of hypothesis tests. An example is shrinkage estimation. Shrinkage is implicit in Bayesian inference and penalized likelihood inference; as an example, we can cite the Least Absolute Shrinkage and Selection Operator, known as LASSO regression, and also Ridge regression (see, for example, Tibshirani (1996)). These methods shrink the parameters to zero when they are not significant, thus avoiding a hypothesis test performance.

2.1 Inference

In the following, we shall discuss estimation of the parameters of the Gompertz-Makeham model by the maximum likelihood (ML) method and the penalized ML method for the gamma-Gompertz model and when the data present heterogeneity or lack thereof.

It is worth stressing that a complication that arises with a continuous distribution in the demographic context is that the variable of interest or the variable measured is discrete. For example, the age at death (a continuous variable) will be given in discrete form as age last birthday. Thus, in general, only discrete data aggregated over individuals is available. Hence, the unknown parameters of the gamma-Gompertz model cannot be directly estimated once it is a parametric model for continuous random variables.

On the other hand, if the number of deaths and the number of person-years exposed to the risk of dying can be observed, we can assume that the number of deaths in a given age interval follows a Poisson distribution, for example; that is, demographers use the standard assumption that a Poisson distribution generates death counts, and its mean parameter is related to the mortality function of the continuous distribution (see, for example, Brillinger (1986); Macdonald et al. (2018); Teimouri and Gupta (2012)).

Let D_x be the number of deaths in a given age interval $[x, x + 1)$ for $x = 0, \dots, m$. Also, let E_x denote the number of person-years at age x exposed to the risk of dying (see, for example Brillinger (1986); Macdonald et al. (2018)). Also, define $\mathbf{D} = (D_0, D_1, \dots, D_m)^\top$ and $\mathbf{E} = (E_0, E_1, \dots, E_m)^\top$. In addition, let $\boldsymbol{\theta} = (a, b, \sigma^2)^\top \in \Theta$ be the parameter vector that characterizes the force of mortality at age x of the gamma-Gompertz model, which is given by $\mu(x; \boldsymbol{\theta}) = \frac{ae^{bx}}{1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)}$. Finally, we assume that the number of deaths and the number of person-years exposed to the risk of dying can be observed.

Once we are considering the standard assumption that the number of deaths D_x are generated by a Poisson distribution with $\mathbb{E}(D_x) = \text{VAR}(D_x) = \mu(x; \boldsymbol{\theta})E_x$; that is, $D_x \sim \mathcal{P}(\mu(x; \boldsymbol{\theta})E_x)$ for $x = 0, \dots, m$ (see, for example Castellares et al. (2020a,b)). Under this assumption, the log-likelihood function for the parameter vector $\boldsymbol{\theta} = (a, b, \sigma^2)^\top$ is given by

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta} | \mathbf{D}, \mathbf{E}) = \sum_x [D_x \ln(\mu(x; \boldsymbol{\theta})E_x) - \mu(x; \boldsymbol{\theta})E_x].$$

The ML estimate $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{\sigma}^2)^\top$ of $\boldsymbol{\theta} = (a, b, \sigma^2)^\top$ is obtained by maximizing the log-likelihood function with respect to $\boldsymbol{\theta} = (a, b, \sigma^2)^\top$.

As mentioned above, the penalized likelihood function is an excellent alternative to using hypothesis tests to detect the presence or absence of heterogeneity. Therefore, we can define the penalized likelihood function as being

$$\ell_p(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - p(\boldsymbol{\theta}).$$

The function $\ell(\boldsymbol{\theta})$ denotes the likelihood function and $p(\boldsymbol{\theta})$ denotes the penalty function. A simplistic way to understand this function is as follows:

$$\begin{aligned}
\ell_p(\boldsymbol{\theta}) &= \ell(\boldsymbol{\theta}) - p(\boldsymbol{\theta}) \\
&= \sum_x \log \mathbb{P}(D_x = d_x | \boldsymbol{\theta}) - \log \exp(p(\boldsymbol{\theta})) \\
&= \log \left[\left(\prod_x \mathbb{P}(D_x = d_x | \boldsymbol{\theta}) \right) e^{-p(\boldsymbol{\theta})} \right] \\
&= \log \left(\mathbb{P}(\mathbf{D} = \mathbf{d} | \boldsymbol{\theta}) e^{-p(\boldsymbol{\theta})} \right) \\
&\propto \log \mathbb{P}(\boldsymbol{\theta} | \mathbf{D}, \mathbf{E}) = \log \left(\frac{\mathbb{P}(\mathbf{D} = \mathbf{d} | \boldsymbol{\theta}) \mathbb{P}(\boldsymbol{\theta})}{\int_{\Theta} \mathbb{P}(\mathbf{D} = \mathbf{d} | \boldsymbol{\theta}) \mathbb{P}(\boldsymbol{\theta}) d\boldsymbol{\theta}} \right)
\end{aligned}$$

Thus, it is easy to see that we can write the penalized likelihood function $\ell_p(\boldsymbol{\theta})$ as the parameter vector $\boldsymbol{\theta}$ posteriori probability distribution, where $e^{-p(\boldsymbol{\theta})}/C_p$ is taken as a prior of $\boldsymbol{\theta}$, where

$$C_p := \int_{\Theta} e^{-p(\boldsymbol{\theta})} d\boldsymbol{\theta}.$$

When we maximize $\ell_p(\boldsymbol{\theta})$, we get the mode of the subsequent distribution of $\boldsymbol{\theta}$. In Bayesian statistics, this estimation method is the maximum a posteriori probability (MAP) estimate. It is widely used in speech and audio processing (Gauvain and Lee, 1994; Benaroya et al., 2005), and image and video processing (Greig et al., 1989; Levitan and Herman, 1987; Belekos et al., 2010).

Therefore, we can define the penalty function as the following:

$$e^{-p(\boldsymbol{\theta})} = g(\sigma^2) = \left[(\sigma^2)^{-1/2} e^{-\sigma^2/2} \right]^{\frac{1}{n}}.$$

In this ways we can also interpret the penalty function as a gamma priori with the parameters $\alpha = 1 - \frac{1}{2n}$ and $\beta = \frac{1}{2n}$ as the σ^2 parameter distribution. Therefore, we can write the penalized versatility function as

$$\begin{aligned}
\ell_p(\boldsymbol{\theta}) &= \sum_x [D_x \ln(\mu(x; \boldsymbol{\theta}) E_x) - \mu(x; \boldsymbol{\theta}) E_x] + \log \left[(\sigma^2)^{-1/2} e^{-\sigma^2/2} \right]^{\frac{1}{n}} \\
&= \sum_x [D_x \ln(\mu(x; \boldsymbol{\theta}) E_x) - \mu(x; \boldsymbol{\theta}) E_x] - \frac{1}{2n} (\log \sigma^2 + \sigma^2).
\end{aligned}$$

Analogous to ML estimate, the MAP estimate $\widehat{\boldsymbol{\theta}}_p = (\widehat{a}_p, \widehat{b}_p, \widehat{\sigma}_p^2)^\top$ of $\boldsymbol{\theta} = (a, b, \sigma^2)^\top$ is obtained by maximizing the log-likelihood function with respect to $\boldsymbol{\theta} = (a, b, \sigma^2)^\top$.

2.2 Monte Carlo simulations

In this subsection, we report Monte Carlo simulation experiments to explore the performance of the MAP and ML method in estimating the parameters of the gamma-Gompertz model under discrete Poisson distribution. All simulations were performed using the R software (Team et al., 2021) with the optimization of the log-likelihood function and penalized log-likelihood function, obtained by using the `optim` function through the Nelder and Mead optimization method (Nelder and Mead, 1965).

The performance of the MAP and ML estimates was evaluated by considering the mean squared error (MSE), which is computed from $R = 5,000$ Monte Carlo replications; that is, if $\tilde{\omega}_r$ denotes an estimate of ω for $r = 1, \dots, R$, we have that

$$\text{MSE}(\tilde{\omega}) = \frac{1}{R} \sum_{r=1}^R (\tilde{\omega}_r - \omega)^2$$

where $\bar{\omega} = \frac{1}{R} \sum_{r=1}^R \tilde{\omega}_r$. To estimate the gamma-Gompertz model parameters, we generate 100,000 random samples from this model for some parameter values. From these samples, we generate life tables and have used them to estimate this model parameters under the discrete Poisson distribution. When there is heterogeneity the true parameter values are $a_1 = 0.0001$ and $a_2 = 0.00001$ for a , $b_1 = 0.1$ and $b_2 = 0.15$ for b , and $\sigma_1^2 = 0.1$ and $\sigma_2^2 = 0.003$ for σ^2 . Note that we choose small values for the parameters, the same presented in the Castellares et al. (2020a) simulation study. When there is no heterogeneity ($\sigma^2 = 0$) the true parameter values are $a_1 = 0.0001$, $a_2 = 0.0003$ and $a_3 = 0.0005$ for a , and $b_1 = 0.09$, $b_2 = 0.10$ and $b_3 = 0.11$ for b , we choose these values based on the estimated parameter presented by Castellares et al. (2020a).

Table 1: Simulation results: gamma-Gompertz model.

There is heterogeneity						
ML estimates			MAP estimates			
Parameter	MSE(\hat{a})	MSE(\hat{b})	MSE($\hat{\sigma}^2$)	MSE(\hat{a})	MSE(\hat{b})	MSE($\hat{\sigma}^2$)
(a_1, b_1, σ_1^2)	0.000078	0.001590	0.022866	0.000080	0.001597	0.023000
(a_1, b_1, σ_2^2)	0.000082	0.001833	0.036225	0.000083	0.001834	0.036142
(a_1, b_2, σ_1^2)	0.000103	0.002255	0.022462	0.000102	0.002260	0.022395
(a_1, b_2, σ_2^2)	0.000105	0.002633	0.035316	0.000106	0.002626	0.035249
(a_2, b_1, σ_1^2)	0.000011	0.001434	0.021706	0.000011	0.001462	0.022020
(a_2, b_1, σ_2^2)	0.000012	0.001675	0.034674	0.000012	0.001685	0.034917
(a_2, b_2, σ_1^2)	0.000013	0.002043	0.020782	0.000013	0.002062	0.020872
(a_2, b_2, σ_2^2)	0.000014	0.002385	0.033200	0.000014	0.002398	0.033338
There is no heterogeneity						
ML estimates			MAP estimates			
Parameter	MSE(\hat{a}) (10^{-10})	MSE(\hat{b}) (10^{-6})	MSE($\hat{\sigma}^2$)	MSE(\hat{a}) (10^{-10})	MSE(\hat{b}) (10^{-6})	MSE($\hat{\sigma}^2$) (10^{-52})
(a_1, b_1, σ^2)	0.317381	0.678322	0.000087	0.005284	0.009019	0.000663
(a_1, b_2, σ^2)	0.340420	0.882027	0.000099	0.115079	0.043352	0.031499
(a_1, b_3, σ^2)	0.375371	1.162997	0.000110	0.204270	0.122622	0.001694
(a_2, b_1, σ^2)	2.117426	0.784103	0.000098	0.319646	0.005873	0.000037
(a_2, b_2, σ^2)	2.355995	1.042755	0.000115	0.750419	0.027933	1.861002
(a_2, b_3, σ^2)	2.416706	1.267139	0.000127	1.078662	0.067992	0.193178
(a_3, b_1, σ^2)	5.064508	0.855126	0.000117	0.696732	0.006491	0.001526
(a_3, b_2, σ^2)	5.379506	1.065385	0.000122	0.878248	0.032037	0.000862
(a_3, b_3, σ^2)	5.514831	1.284763	0.000132	0.098827	0.756976	0.000769

The simulation results are presented in Table 1. When there is heterogeneity, it is evident that the performance of the ML method in estimating the model parameters are pretty good, and that the Nelder and Mead's optimization method works appropriately; that is, the ML estimates are close to the true parameter values.

Further, it is interesting to note when there is the heterogeneity the gamma-Gompertz model parameters estimates both methods provide similar results. Despite the ML estimates provide better results,

the difference between them is tiny; for the parameter a that difference is less than 3%, for parameters σ^2 and b that is less than 2%. These results suggest that both methods provide good results, with similar MSE.

Furthermore, when there is no heterogeneity, the MAP estimates method gives us better results, presenting lower MSE for all parameters, especially for the parameter σ^2 , which has a value close to zero ($\approx 10^{-30}$).

In short, the Monte Carlo simulation results are promising and reveal that the penalized likelihood can also be used to estimate the parameters of the gamma-Gompertz model. In other words, the penalized likelihood seems to be a powerful tool to detect heterogeneity in the gamma-Gompertz model, complementing the work developed by Böhnstedt and Gampe (2019).

Furthermore, as is well known, mortality data for the elderly are scarce (Yue, 2002). Therefore problems with the sample size are expected, making the likelihood ratio test impossible to detect heterogeneity, as developed by Böhnstedt and Gampe (2019). Therefore, the same simulation study was carried out considering the sample sizes $m = 20, 15, 10$, and the results obtained were similar to those presented in Table 1. Finally, the penalized likelihood is also indicated when there is a small sample size, especially if the goal is to detect heterogeneity.

3 Real data analysis

In this section, based on mortality datasets available on Human Mortality Database we estimate the gamma-Gompertz model parameter. In particular, we used the well-developed statistical software R (Team et al., 2021) to compute the ML and MAP estimates of $\theta = (a, b, \sigma^2)'$ by using the genetic algorithm (Michalewicz, 1995) through the library GA (Scrucca, 2013, 2016). This algorithm ensures the global maximum of the estimation process by the ML and MAP methods.

In the Human Mortality Database, only the mortality data for the country of Denmark in the years 1930, 1965, 1975, 1988, 2000 and 2004. The ML and MAP estimates of the model parameters are listed in Table 2. Furthermore, the average distance (\mathcal{D}) between the estimated force of mortality (log) and the mortality rate (log) was also calculated so that we can have a measure of the goodness of fit of the model.

Table 2: Parameter estimation.

Year	ML estimates				MAP estimates			
	\hat{a}	\hat{b}	$\hat{\sigma}^2$	$\mathcal{D}(\text{MLE})$	\hat{a}	\hat{b}	$\hat{\sigma}^2$	$\mathcal{D}(\text{MAP})$
1930	0.042757	0.143397	0.416511	0.246545	0.046899	0.094567	0	0.086222
1965	0.035867	0.110540	0.083146	0.112856	0.036499	0.105135	0.040756	0.110682
1975	0.031834	0.097972	0.023494	0.081010	0.032225	0.095130	0	0.082815
1988	0.029916	0.093917	0.000563	0.194960	0.030609	0.092380	0	0.198050
2000	0.027152	0.096958	0.017551	0.105074	0.027144	0.095940	0	0.103124
2004	0.024754	0.098578	0.000006	0.078447	0.024823	0.098147	0	0.077635

In Table 2, we can see that the MAP estimator has provided better results in 2/3 scenarios. That is, the model estimated by this method is closer to observed data than the MLE. On average, the MAP estimator provides results 11% better than MLE; both approaches generally provide similar results. However, the estimated values for the parameters b and σ^2 are pretty different. Finally, the MAP

estimates tell us that there is no heterogeneity in the years 1930, 1975, 1988, 2000, and 2004, and interestingly the data fit better when heterogeneity was not detected.

Changes in mortality dynamics are constantly observed and discussed, and many recent studies have provided strong evidence for the existence of mortality heterogeneity (see, for example, Alvarez et al. (2021); Barbi et al. (2018)). Thus, not finding heterogeneity does not mean that it does not exist, but rather that it could not be observed in the data used, mainly since the data is grouped at age 100+. As well presented by Barbi et al. (2018), after age 105, what we can expect to see is a force of mortality deceleration relative to earlier ages, or else a plateau. Such a change in the mortality regime can also be described as a change in the estimated b (b -hypothesis).

In addition to these six years, we also looked at the MAP and ML estimates between 1920 and 2020. The MAP estimate provided better results than the MLE. The MAP estimator proved to be superior to the MLE in 70% of the synthetic cohort, estimating models up to 65% better than the one estimated by the MLE. When MLE showed better results, this improvement was not more significant than 3%. In addition, the estimates from both approaches differ only for the parameter σ^2 . 78% of the times that heterogeneity is not detected, the model with $\sigma=0$ proved superior to the one that considers it. Furthermore, when heterogeneity is present in the population, the MAP estimator provides better results in 69% of cases.

4 Concluding remarks

It is well-known that Gompertz-related parametric models help describe mortality patterns in numerous species, including humans. An outstanding extension of the Gompertz model was introduced Vaupel et al. (1979), and it is known as the gamma-Gompertz model. The gamma-Gompertz model has been investigated in different areas of research and, in addition, has been applied successfully in all of them.

Still, several things have not been investigated about this critical mortality model, especially methods capable of detecting heterogeneity. This paper presents a brief study on estimating the gamma-Gompertz model, from computational results to real applications. First, we briefly present the heterogeneity detection problem and approach it as a model misspecification problem. Then, we move to statistical inference, presenting a solution that does not depend on asymptotic distributions and, consequently, is not strongly affected by the sample size. Next, the comparison of the proposed approach (MAP estimator) with the standard approach (MLE) was made from a simulation study. It was clear that under heterogeneity, both approaches produce similar results. However, in its absence, the proposed approach provided better results. In particular, the MAP estimator was able to detect the existence of heterogeneity or its absence. However, the empirical applications of the proposed approach to the mortality rate confirm that this approach is advantageous in detecting heterogeneity. Furthermore, the empirical application sheds light on other problems, especially those related to detecting the change in the parameter b over ages and time. Finally, it is vital to consider models that deal with data over-dispersion, once over-dispersion is extremely common in mortality data, and that has a severe consequence: the underestimation of standard errors in models and their parameters estimation (Macdonald et al., 2018).

Bibliography

- Alvarez, J.-A., Villavicencio, F., Strozza, C., and Camarda, C. G. (2021). Regularities in human mortality after age 105. *PloS one*, 16(7):e0253940.
- Barbi, E., Lagona, F., Marsili, M., Vaupel, J. W., and Wachter, K. W. (2018). The plateau of human mortality: Demography of longevity pioneers. *Science*, 360(6396):1459–1461.
- Belekos, S. P., Galatsanos, N. P., and Katsaggelos, A. K. (2010). Maximum a posteriori video super-resolution using a new multichannel image prior. *IEEE Transactions on Image Processing*, 19(6):1451–1464.
- Benaroya, L., Bimbot, F., and Gribonval, R. (2005). Audio source separation with a single sensor. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):191–199.
- Böhnstedt, M. and Gampe, J. (2019). Detecting mortality deceleration: Likelihood inference and model selection in the gamma-gompertz model. *Statistics & Probability Letters*, 150:68–73.
- Brillinger, D. R. (1986). A biometrics invited paper with discussion: the natural variability of vital rates and associated statistics. *Biometrics*, pages 693–734.
- Castellares, F., Patrício, S. C., and Lemonte, A. J. (2020a). On gamma-gompertz life expectancy. *Statistics & Probability Letters*, 165:108832.
- Castellares, F., Patrício, S. C., Lemonte, A. J., and Queiroz, B. L. (2020b). On closed-form expressions to gompertz–makeham life expectancy. *Theoretical population biology*, 134:53–60.
- Cutler, D., Deaton, A., and Lleras-Muney, A. (2006). The determinants of mortality. *Journal of economic perspectives*, 20(3):97–120.
- Dickson, D. C., Hardy, M. R., and Waters, H. R. (2019). *Actuarial mathematics for life contingent risks*. Cambridge University Press.
- Gauvain, J.-L. and Lee, C.-H. (1994). Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE transactions on speech and audio processing*, 2(2):291–298.
- Gavrilova, N. S. and Gavrilov, L. A. (2015). Biodemography of old-age mortality in humans and rodents. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 70(1):1–9.
- Gompertz, B. (1825). Xxiv. 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 &c. *Philosophical transactions of the Royal Society of London*, (115):513–583.
- Greig, D. M., Porteous, B. T., and Seheult, A. H. (1989). Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2):271–279.
- Haberman, S. (1996). Landmarks in the history of actuarial science (up to 1919). *Actuarial Research Paper 84, Dept. of Actuarial Science and Statistics, City University, London*.

- Levitani, E. and Herman, G. T. (1987). A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE transactions on medical imaging*, 6(3):185–192.
- Macdonald, A. S., Richards, S. J., and Currie, I. D. (2018). *Modelling mortality with actuarial applications*. Cambridge University Press.
- Makeham, W. M. (1860). On the law of mortality and the construction of annuity tables. *The Assurance Magazine, and Journal of the Institute of Actuaries*, 8(6):301–310.
- Makeham, W. M. (1867). On the law of mortality. *Journal of the Institute of Actuaries (1866-1867)*, 13(6):325–358.
- Makeham, W. M. (1890). On the further development of gompertz’s law (concluded). *Journal of the Institute of Actuaries (1886-1994)*, 28(4):316–332.
- Michalewicz, Z. (1995). Genetic algorithms, numerical optimization, and constraints. In *Proceedings of the sixth international conference on genetic algorithms*, volume 195, pages 151–158. Citeseer.
- Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313.
- Nester, M. R. (1996). An applied statistician’s creed. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 45(4):401–410.
- Newman, S. J. (2018). Errors as a primary cause of late-life mortality deceleration and plateaus. *PLoS biology*, 16(12):e2006776.
- Oeppen, J. and Vaupel, J. W. (2002). Broken limits to life expectancy.
- Scrucca, L. (2013). Ga: a package for genetic algorithms in r. *Journal of Statistical Software*, 53(1):1–37.
- Scrucca, L. (2016). On some extensions to ga package: hybrid optimisation, parallelisation and islands evolution. *arXiv preprint arXiv:1605.01931*.
- Team, R. C. et al. (2021). R: A language and environment for statistical computing.
- Teimouri, M. and Gupta, A. K. (2012). Estimation methods for the gompertz–makeham distribution under progressively type-i interval censoring scheme. *National Academy Science Letters*, 35(3):227–235.
- Thatcher, A. R., Kannisto, V., and Vaupel, J. W. (1998). The force of mortality at ages 80 to 120.
- Thompson, W. (2001). Problems with the hypothesis testing approach.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.
- Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., Yashin, A. I., Holm, N. V., Iachine, I. A., Kannisto, V., Khazaeli, A. A., Liedo, P., et al. (1998). Biodemographic trajectories of longevity. *Science*, 280(5365):855–860.

- Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.
- Vaupel, J. W., Villavicencio, F., and Bergeron-Boucher, M.-P. (2021). Demographic perspectives on the rise of longevity. *Proceedings of the National Academy of Sciences*, 118(9).
- Wienke, A. (2010). *Frailty models in survival analysis*. CRC press.
- Yue, J. C. (2002). Oldest-old mortality rates and the gompertz law: A theoretical and empirical study based on four countries. *Journal of Population Studies*, 24:33–57.