Statistical methods for frailty models: studies on old-age mortality and recurrent events

Böhnstedt, M.

Part I

Assessing mortality deceleration

2

Detecting mortality deceleration: Likelihood inference and model selection in the gamma-Gompertz model

Abstract

We study the asymptotic properties of the maximum likelihood estimator and the likelihood ratio test in the gamma-Gompertz model for local alternatives. We also show that the standard AIC is biased in this model due to the boundary parameter.

2.1 Introduction

Benjamin Gompertz (1825) pioneered human mortality research by demonstrating that death rates for adults increase exponentially with age. Since then, the Gompertz distribution has been widely used to model adult lifespans. It was not until more and better data at high ages became available that the overall validity of the Gompertz distribution was called into question. Downward deviations from an exponentially increasing hazard ('*mortality deceleration*') were observed when data for the oldest-old were analyzed

This chapter has been published as: M. Böhnstedt and J. Gampe (2019). Detecting mortality deceleration: Likelihood inference and model selection in the gamma-Gompertz model. *Statistics and Probability Letters* 150, 68–73, *Copyright Elsevier B.V. 2019.*

(Thatcher et al., 1998). This kind of mortality deceleration results if hazards differ between individuals, and a proportional hazards (PH) frailty model is the standard approach used to model such heterogeneity (Vaupel et al., 1979; Wienke, 2010).

If we denote by Y the continuous random variable that describes adult lifespans (typically above age 30) and its hazard by

$$h(y) = \lim_{\Delta y \searrow 0} \mathrm{P}(y \leq Y < y + \Delta y \,|\, Y \geq y)/\Delta y, \tag{2.1}$$

a PH frailty model is of the form $h(y|Z = z) = z \cdot h_0(y)$, where Z is a positive random effect ('*frailty*'), $h_0(y)$ is a baseline hazard, and $h(y|Z = z)$ denotes the conditional hazard of an individual at age y, given that his or her frailty is $Z = z$. If the baseline hazard is an exponential function, $h_0(y) = ae^{by}$ with parameters $a > 0, b > 0$, and the random effect Z has a gamma distribution with mean one and variance σ^2, we obtain the so-called gamma-Gompertz model. The variance parameter σ^2 describes the heterogeneity in the risks of death: i.e., individuals with higher frailty values tend to die earlier, while more robust individuals tend to survive. While all individual hazards $h(y|Z)$ are exponentially increasing, the resulting marginal hazard,

$$h(y) = \frac{ae^{by}}{1 + \sigma^2 \frac{a}{b}(e^{by} - 1)}, \tag{2.2}$$

shows a downward deviation from the exponential trajectory if $\sigma^2 > 0$. The deviation is discernible at high ages, when differential mortality has played out sufficiently. If $\sigma^2 = 0$ – i.e., if there is no heterogeneity – model (2.2) reduces to a plain Gompertz model with hazard $h(y) = ae^{by}$. The choice of the gamma distribution for the frailty Z is both mathematically convenient and theoretically justified (Abbring and van den Berg, 2007).

The question of whether mortality deceleration is present or an exponential hazard fits even at advanced ages has been repeatedly discussed (e.g., Gavrilova and Gavrilov, 2015). For the gamma-Gompertz model, the question is reduced to whether $\sigma^2 > 0$ or $\sigma^2 = 0$. In the latter case, the parameter σ^2 is on the boundary of the parameter space, which violates the standard assumptions that underlie the asymptotic properties of likelihood-based inference. Thus, asymptotic results need to be derived for this setting.

In this chapter, we consider the asymptotic distribution of the likelihood ratio test (LRT) statistic for $H_0 : \sigma^2 = 0$ in the gamma-Gompertz model. As we are interested in deriving large-sample approximations of the power of the LRT to detect $\sigma^2 > 0$, we will work within a framework of local alternatives (Lehmann, 1999). We will also consider model selection based on the Akaike information criterion (AIC, Akaike, 1974). In the gamma-Gompertz model, the standard AIC is not an asymptotically unbiased estimator of the Akaike information. We derive the bias using a local misspecification framework (Hjort and Claeskens, 2003).

The rest of the chapter is structured as follows. In Section 2.2, we introduce some further notations and assumptions. In Section 2.3, we derive the asymptotic distribution of the maximum likelihood estimator (MLE) in the gamma-Gompertz model. In Section 2.4, we establish the asymptotic distribution of the likelihood ratio test statistic, and provide a large-sample approximation to the power of the LRT to detect a positive σ^2. In Section 2.5, we show that the standard AIC is a biased estimator of the Akaike information in the gamma-Gompertz model. We conclude with a discussion in Section 2.6.

2.2 Preliminaries

We consider a sample of n iid lifespans y from a gamma-Gompertz density $f(y,\eta)$ where the parameter vector $\eta = (a, b, \sigma^2)^\top$ consists of the elements $\theta = (a, b)^\top$, resulting from the Gompertz baseline, and the gamma-variance σ^2. The density is

$$f(y,\eta) = \begin{cases} ae^{by}\ \left[1+\sigma^2 \frac{a}{b}\left(e^{by}-1\right)\right]^{-(1+\frac{1}{\sigma^2})} & \text{for } \sigma^2 > 0 \\ ae^{by}\ \exp\left\{-\frac{a}{b}(e^{by}-1)\right\} & \text{for } \sigma^2 = 0. \end{cases} \tag{2.3}$$

The framework of local alternatives assumes that the observations y are generated from a density

$$f_{\text{true}}(y) = f(y, \theta_0, \gamma_0 + \delta/\sqrt{n}), \tag{2.4}$$

where θ_0 is a p-dimensional parameter vector, and $\gamma = \gamma_0 + \delta/\sqrt{n}$ is a q-dimensional parameter vector perturbed around a (known) γ_0 in the direction of δ. In the gamma-Gompertz model (2.2), we have $\theta = (a, b)^\top$, so $p = 2$, and γ is the single ($q = 1$) boundary parameter $\gamma = \sigma^2$, with $\gamma_0 = 0$ and $\delta = \sqrt{n}\sigma^2 \geq 0$.

The asymptotic distribution of the MLEs in the general framework (2.4) was derived by Hjort and Claeskens (2003) under the usual regularity conditions, which require that the true parameter is an inner point of the parameter space. Hjort (1994) considered a boundary parameter, but exclusively dealt with the t-distribution as an extension of the normal model. Self and Liang (1987) presented asymptotic distributions of MLEs and likelihood ratio test statistics in the presence of boundary parameters, but while assuming a fixed true model, rather than allowing for a local specification as in (2.4).

To state our main results, we introduce some further notations, closely following Hjort and Claeskens (2003). We denote by $U(y) = \partial \ln f(y, \theta_0, \gamma_0)/\partial\theta$ and $V(y) = \partial \ln f(y, \theta_0, \gamma_0)/\partial\gamma$ the score functions with respect to θ and γ, respectively, of the log-likelihood of a single observation y from f_{true}, with both evaluated at the point $(\theta, \gamma) = (\theta_0, \gamma_0)$. Let J_{full} be the corresponding information matrix; i.e., J_{full} is the variance-covariance matrix of the score vector $(U(y)^\top, V(y))^\top$, with blocks J_{00}, J_{01}, J_{10}, and J_{11}.

In the following, $Y = (Y_1, \ldots, Y_n)$ denotes a random sample of size n from density (2.4), and we abbreviate $\eta = (\theta^\top, \gamma)^\top$ and $\eta_0 = (\theta_0^\top, \gamma_0)^\top$. We can express the log-likelihood as $\ell_n(\eta, Y) = \sum_{i=1}^n \ln f(Y_i, \theta, \gamma)$. Similarly, the averages of the score functions are denoted by $\bar{U}_n = n^{-1}\sum_{i=1}^n U(Y_i)$ and $\bar{V}_n = n^{-1}\sum_{i=1}^n V(Y_i)$, with the shorthand notation $\bar{W}_n = (\bar{U}_n^\top, \bar{V}_n)^\top$.

When γ could lie on the boundary of the parameter space, the derivatives of the log-likelihood have to be taken from the appropriate side. It is also important to note that these derivatives need to exist, and that they have to be bounded on intersections of neighborhoods of the true parameter value and the parameter space (see Self and Liang, 1987).

The derivation of the asymptotic distribution of the MLEs is based on the following result showing the weak convergence of the averaged score vector. It was given as Lemma 3.1 in Hjort and Claeskens (2003). Their proof carries over to the boundary setting considered here.

Lemma 2.1. *Under the sequence of local alternatives* (2.4)*, the score vector is asymptotically normally distributed,*

$$\begin{pmatrix}\sqrt{n}\bar{U}_n\\ \sqrt{n}\bar{V}_n\end{pmatrix} \xrightarrow{d} \begin{pmatrix}J_{01}\delta\\ J_{11}\delta\end{pmatrix} + \begin{pmatrix}M\\ N\end{pmatrix}, \quad \textit{with} \quad \begin{pmatrix}M\\ N\end{pmatrix} \sim \mathcal{N}_{p+1}(0, J_{\text{full}}).$$

2.3 Asymptotic distribution of maximum likelihood estimator

In the setting with one boundary parameter, the MLE is asymptotically distributed as the projection of a normal random vector onto the subspace of admissible parameter values.

Theorem 2.2. *Under the sequence of models* (2.4) *and with a normally distributed random vector* $(A^\top, B)^\top \sim \mathcal{N}(0, J_{\text{full}}^{-1})$*, it holds that*

$$\begin{pmatrix}\sqrt{n}(\hat{\theta}-\theta_0)\\ \sqrt{n}(\hat{\gamma}-\gamma_0)\end{pmatrix} \xrightarrow{d} \begin{pmatrix}A\\ B+\delta\end{pmatrix} \cdot \mathbb{1}_{\{B+\delta>0\}} + \begin{pmatrix}A + J_{00}^{-1}J_{01}(B+\delta)\\ 0\end{pmatrix} \cdot \mathbb{1}_{\{B+\delta\le 0\}},$$

where $\mathbb{1}_{\cdot}$ *is the indicator function.*

Remark 2.1. Theorem 2.2 applies not only to the gamma-Gompertz model, but more generally to parametric models with one boundary parameter. In particular, the theorem holds for other gamma-PH models if the parameters of the baseline hazard are not boundary parameters. This includes the gamma-exponential and the gamma-Weibull model.

Proof. Two cases need to be distinguished: namely, that the log-likelihood ℓ_n is maximized at $\hat{\gamma} > \gamma_0$ or at $\hat{\gamma} = \gamma_0$ (cf. Hjort, 1994). In the first case, we have $\partial\ell_n(\hat{\eta})/\partial\theta = 0$ and $\partial\ell_n(\hat{\eta})/\partial\gamma = 0$. If we apply, first, the usual Taylor arguments and, second, the result of Lemma 2.1 on the limiting distribution of the score vector, we find that for a maximum at $\hat{\gamma} > \gamma_0$,

$$\begin{pmatrix}\sqrt{n}(\hat{\theta}-\theta_0)\\ \sqrt{n}(\hat{\gamma}-\gamma_0)\end{pmatrix} \overset{d}{=} J_{\text{full}}^{-1}\begin{pmatrix}\sqrt{n}\bar{U}_n\\ \sqrt{n}\bar{V}_n\end{pmatrix} \xrightarrow{d} \begin{pmatrix}0\\ \delta\end{pmatrix} + J_{\text{full}}^{-1}\begin{pmatrix}M\\ N\end{pmatrix} = \begin{pmatrix}A\\ B+\delta\end{pmatrix}. \tag{2.5}$$

In the second case, when the maximum occurs at $\hat{\gamma} = \gamma_0$, we have $\partial\ell_n(\hat{\eta})/\partial\theta = 0$, but $\partial\ell_n(\hat{\eta})/\partial\gamma \le 0$ and $\hat{\gamma} = \gamma_0$, such that

$$\begin{pmatrix}\sqrt{n}(\hat{\theta}-\theta_0)\\ \sqrt{n}(\hat{\gamma}-\gamma_0)\end{pmatrix} \overset{d}{=} \begin{pmatrix}J_{00}^{-1}\sqrt{n}\bar{U}_n\\ 0\end{pmatrix} \xrightarrow{d} \begin{pmatrix}J_{00}^{-1}J_{01}\delta + J_{00}^{-1}M\\ 0\end{pmatrix}. \tag{2.6}$$

Direct calculations show that $J_{00}^{-1}J_{01}\delta + J_{00}^{-1}M$ can be expressed as $A + J_{00}^{-1}J_{01}(B+\delta)$. If the last component of $J_{\text{full}}^{-1}(\sqrt{n}\bar{U}_n^\top, \sqrt{n}\bar{V}_n)^\top$ in (2.5) is denoted by Δ_n, and if Ω_n denotes the set of samples for which the estimate $\hat{\gamma} > \gamma_0$, then Δ_n, and Ω_n are (asymptotically) equivalent in the following sense:

$$\mathbb{1}_{\Omega_n} - \mathbb{1}_{\{\Delta_n > 0\}} \xrightarrow{\text{P}} 0. \tag{2.7}$$

This convergence is verified in Section 2.7.

The following corollary explicitly gives the asymptotic distribution of the MLEs for the gamma-Gompertz model.

Corollary 2.3. *Under the sequence of true gamma-Gompertz models* (2.4) *with* $\theta_0 = (a,b)^\top$, $\gamma_0 = 0$, *and* $\delta = \sqrt{n}\sigma^2$, *the MLEs* $\hat{a}$, $\hat{b}$, *and* $\hat{\sigma}^2$ *asymptotically follow mixture distributions given by*

$$\begin{aligned}
\sqrt{n}(\hat{a}-a) &\xrightarrow{d} \Phi\left(\frac{\delta}{\kappa}\right)\mathcal{TN}_1(\mu_\delta, \Sigma_1, s, t) + \Phi\left(-\frac{\delta}{\kappa}\right)\mathcal{N}\left(-\frac{c_{13}}{c_{33}}\delta, c_{11} - \frac{c_{13}^2}{c_{33}}\right)\\
\sqrt{n}(\hat{b}-b) &\xrightarrow{d} \Phi\left(\frac{\delta}{\kappa}\right)\mathcal{TN}_1(\mu_\delta, \Sigma_2, s, t) + \Phi\left(-\frac{\delta}{\kappa}\right)\mathcal{N}\left(-\frac{c_{23}}{c_{33}}\delta, c_{22} - \frac{c_{23}^2}{c_{33}}\right)\\
\sqrt{n}\hat{\sigma}^2 &\xrightarrow{d} \Phi\left(\frac{\delta}{\kappa}\right)\mathcal{TN}(\delta, \kappa^2, 0, \infty) + \Phi\left(-\frac{\delta}{\kappa}\right)\chi_0^2.
\end{aligned}$$

Here, the c_{kl} *are the elements of* J_{full}^{-1}, *with* $\kappa^2 = c_{33}$, $\Phi(\cdot)$ *is the cdf of the standard normal distribution and* χ_0^2 *denotes a point mass at zero.* $\mathcal{TN}_1(\mu, \Sigma, s, t)$ *is the marginal distribution of the first component of a truncated bivariate normal distribution with mean vector* μ, *covariance matrix* Σ, *and the lower and upper truncation limits* $s = (-\infty, 0)^\top$ *and* $t = (\infty, \infty)^\top$, *respectively; where in particular,* $\mu_\delta = (0,\delta)^\top$, $\Sigma_1 = \begin{pmatrix} c_{11} & c_{13} \\ c_{13} & c_{33}\end{pmatrix}$, *and* $\Sigma_2 = \begin{pmatrix} c_{22} & c_{23} \\ c_{23} & c_{33}\end{pmatrix}$.

Remark 2.2. Two special cases of Corollary 2.3 are of particular interest: First, if $\delta = \sqrt{n}\sigma^2$ is sufficiently large, we obtain the standard result of asymptotic normality of the MLEs. Second, if $\delta = 0$ – i.e., if the true model is a pure Gompertz model – we obtain the typical mixture weights of 0.5.

Proof. Define $Z = (Z_1, Z_2, Z_3)^\top \sim \mathcal{N}((0,0,\delta)^\top, J_{\text{full}}^{-1})$, which is distributed as $(A^\top, B+\delta)^\top$, and the projection matrix

$$T = \begin{pmatrix} I & J_{00}^{-1}J_{01} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -c_{13}/c_{33} \\ 0 & 1 & -c_{23}/c_{33} \\ 0 & 0 & 0 \end{pmatrix},$$

to express the result of Theorem 2.2 as

$$\sqrt{n}(\hat{\eta} - \eta_0) \xrightarrow{d} \hat{\eta}_* = Z \cdot \mathbb{1}_{\{Z_3>0\}} + TZ \cdot \mathbb{1}_{\{Z_3\le 0\}}. \tag{2.8}$$

As the components of TZ are independent of $Z_3 \sim \mathcal{N}(\delta, \kappa^2)$, it follows that

$$\begin{aligned} \mathrm{P}[\hat{\eta}_* \le z] &= \mathrm{P}[\hat{\eta}_* \le z | Z_3 > 0]\mathrm{P}[Z_3 > 0] + \mathrm{P}[\hat{\eta}_* \le z | Z_3 \le 0]\mathrm{P}[Z_3 \le 0] \\ &= \mathrm{P}[Z \le z | Z_3 > 0]\Phi\left(\frac{\delta}{\kappa}\right) + \mathrm{P}[TZ \le z]\Phi\left(-\frac{\delta}{\kappa}\right). \end{aligned}$$

Now for $\hat{\eta}_n = (\hat{a}, \hat{b}, \hat{\sigma}^2)^\top$ and $\eta_0 = (a, b, 0)^\top$, a large-sample approximation to $\mathrm{P}[\sqrt{n}(\hat{\eta}_n - \eta_0) \le u]$ based on (2.8) is given by

$$\mathrm{P}[\hat{\eta}_* \le u] = \Phi\left(\frac{\delta}{\kappa}\right)\mathrm{P}[Z \le u | Z_3 > 0] + \Phi\left(-\frac{\delta}{\kappa}\right)\mathrm{P}[TZ \le u],$$

where TZ is normally distributed with mean $T(0,0,\delta)^\top = (-\delta c_{13}/c_{33}, -\delta c_{23}/c_{33}, 0)^\top$ and covariance matrix

$$TJ_{\text{full}}^{-1}T^\top = \begin{pmatrix} c_{11} - c_{13}^2/c_{33} & c_{12} - c_{13}c_{23}/c_{33} & 0 \\ c_{12} - c_{13}c_{23}/c_{33} & c_{22} - c_{23}^2/c_{33} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The result of Corollary 2.3 follows by determining the marginal distributions.

2.4 Asymptotics for the likelihood ratio test

This section deals with the asymptotics of the LRT for $H_0 : \sigma^2 = 0$ vs. $H_1 : \sigma^2 > 0$ in the gamma-Gompertz model in the framework of local alternatives (2.4). After presenting the limiting distribution of the likelihood ratio test statistic, we will provide a formula that approximates the power of the test to detect $\sigma^2 > 0$ in large samples.

The likelihood ratio test statistic based on a sample $Y = (Y_1, \ldots, Y_n)$ is defined as

$$-2\ln\lambda_n(Y) = 2[\ell_n(\hat{\eta}, Y) - \ell_n(\hat{\eta}_G, Y)],$$

where $\hat{\eta}_G = (\hat{a}_G, \hat{b}_G, 0)^\top$ with the MLE $(\hat{a}_G, \hat{b}_G)^\top$ for the Gompertz model. Under the non-standard condition of testing the boundary parameter σ^2, the test statistic does not converge to the standard chi-squared distribution, but rather to a mixture; see

Theorem 2.4. *Under the sequence of local alternatives* (2.4)*, the test statistic of a LRT for* $H_0 : \sigma^2 = 0$ *is asymptotically distributed as*

$$-2\ln\lambda_n(Y) \xrightarrow{d} \left\{\max\left(0, \frac{Z_3}{\kappa}\right)\right\}^2 \quad \textit{with } Z_3 \sim \mathcal{N}(\delta, \kappa^2).$$

Corollary 2.5. *In terms of the cumulative distribution function, we find that* $\mathrm{P}[-2\ln\lambda_n(Y) \le z]$ *asymptotically equals*

$$\frac{\Phi\left(\sqrt{z}-\frac{\delta}{\kappa}\right)-\Phi\left(-\frac{\delta}{\kappa}\right)}{\Phi\left(\sqrt{z}-\frac{\delta}{\kappa}\right)-\Phi\left(-\sqrt{z}-\frac{\delta}{\kappa}\right)} F_{\chi_1^2}\left(z, \mathrm{ncp}=\frac{\delta^2}{\kappa^2}\right)+\Phi\left(-\frac{\delta}{\kappa}\right) F_{\chi_0^2}(z),$$

which is a combination of a non-central chi-squared distribution with one degree of freedom and non-centrality parameter δ^2/κ^2*, and a point mass at zero.*

Remark 2.3. Setting $\delta = 0$ – i.e., $\sigma^2 = 0$ – we obtain the asymptotic distribution of the LRT statistic under H_0 as $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_0^2$, which is a 50:50 mixture of a chi-squared distribution with one degree of freedom and a point mass at zero.

Remark 2.4. Theorem 2.4 is also valid for other gamma-PH models, such as the gamma-Weibull model. The limiting distribution will depend on the respective model via the element κ^2 of the inverse information matrix.

Proof (of Theorem 2.4). We again distinguish two cases depending on whether the estimate $\hat{\sigma}^2$ does or does not lie on the boundary of the parameter space. If the likelihood is maximized at $\hat{\sigma}^2 = 0$, we have $\hat{\eta} = \hat{\eta}_G$, and, hence, $-2\ln\lambda_n(Y) = 0$. If the likelihood is maximized at some point with $\hat{\sigma}^2 > 0$, a second-order Taylor expansion yields (see also Hjort and Claeskens, 2003, Section 3.3)

$$2[\ell_n(\hat{\eta}, Y) - \ell_n(\eta_0, Y)] \overset{d}{=} \begin{pmatrix}\sqrt{n}\bar{U}_n \\ \sqrt{n}\bar{V}_n\end{pmatrix}^\top J_{\text{full}}^{-1} \begin{pmatrix}\sqrt{n}\bar{U}_n \\ \sqrt{n}\bar{V}_n\end{pmatrix} \text{ and}$$

$$2[\ell_n(\hat{\eta}_G, Y) - \ell_n(\eta_0, Y)] \overset{d}{=} \sqrt{n}\bar{U}_n^\top J_{00}^{-1}\sqrt{n}\bar{U}_n.$$

The quadratic form of the full score vector $\sqrt{n}(\bar{U}_n^\top, \bar{V}_n)^\top$ can be calculated as $n[\bar{U}_n^\top J_{00}^{-1}\bar{U}_n + \kappa^2(\bar{V}_n - J_{10}J_{00}^{-1}\bar{U}_n)^\top(\bar{V}_n - J_{10}J_{00}^{-1}\bar{U}_n)]$. Consequently, if $\hat{\sigma}^2 > 0$, the likelihood ratio test statistic, $-2\ln\lambda_n(Y)$, fulfills

$$-2\ln\lambda_n(Y) \overset{d}{=} n\kappa^2(\bar{V}_n - J_{10}J_{00}^{-1}\bar{U}_n)^\top(\bar{V}_n - J_{10}J_{00}^{-1}\bar{U}_n). \tag{2.9}$$

From the asymptotic normality of $\sqrt{n}(\bar{U}_n^\top, \bar{V}_n)^\top$ given in Lemma 2.1, we obtain that $\sqrt{n}(\bar{V}_n - J_{10}J_{00}^{-1}\bar{U}_n)$ is asymptotically normally distributed with mean δ/κ^2 and variance κ^{-2}. Hence, (2.9) is distributed as the square of a normal random variable with mean δ/κ and unit variance. The proof is completed by noting that $Z_3/\kappa \sim \mathcal{N}(\delta/\kappa, 1)$ and recalling from the proof of Corollary 2.3 that $\hat{\sigma}^2 > 0$ if $Z_3 > 0$ and $\hat{\sigma}^2 = 0$ if $Z_3 \le 0$.

One reason for working under the framework in (2.4) is that it allows us to derive large-sample approximations of the power of the LRT to detect $\sigma^2 > 0$.

Lemma 2.6. *Under the sequence of local alternatives* (2.4)*, the power* β_n *of the LRT for* $H_0: \sigma^2 = 0$ *at level* α *based on a gamma-Gompertz sample of size* n *is approximately equal to*

$$\beta_n(\delta) \approx 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\delta}{\kappa}\right) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\sqrt{n}\sigma^2}{\kappa}\right).$$

The proof is deferred to Section 2.7.

2.5 AIC for the gamma-Gompertz model

In this section, we study the AIC for the gamma-Gompertz model. The AIC is a popular tool for model selection. In the standard setting, the AIC is defined as

$$\text{AIC} = -2\,\ell_n(\hat{\eta}, Y) + 2(p+1). \tag{2.10}$$

The AIC aims to give an unbiased estimate of the Akaike information, the expected relative Kullback-Leibler distance between the true data-generating density and the best parametric approximation $f(y, \hat{\eta})$, which is determined by inserting the MLE $\hat{\eta}$ for η. To obtain such an unbiased estimate, the term $-2\,\ell_n(\hat{\eta}, Y)$ is penalized by twice the number of parameters in definition (2.10), which here is $2(p+1) = 2 \cdot 3$. This results from asymptotic considerations that, again, require regularity conditions that do not hold for the gamma-Gompertz model. We find that formula (2.10) needs to be modified, as stated in

Theorem 2.7. *Under the sequence of local alternatives* (2.4)*, an asymptotically unbiased estimator of the Akaike information of the gamma-Gompertz model is given by*

$$\text{AIC}^* = -2\,\ell_n(\hat{\eta}, Y) + 2 \cdot 3 - 2\,\Phi\left(-\frac{\delta}{\kappa}\right).$$

Proof. Here we give only an outline of the proof, and provide further details in Section 2.7.

The relative distance between the true underlying distribution g of X and an approximating parametric density $f(., \eta)$ is measured by $\mathbb{E}_X[\ln f(X, \eta)]$. In practice, the unknown parameter η is estimated from a sample Y generated from g, and the estimator is denoted by $\hat{\eta}(Y)$. The Akaike information, which is to be estimated by (2.10), takes the form (Akaike, 1974)

$$-\,2\,\mathbb{E}_Y[\mathbb{E}_X[\ln f(X, \hat{\eta}(Y))]]. \tag{2.11}$$

With the previous notation and $\hat{\eta} = \hat{\eta}(Y)$, an unbiased estimator of (2.11) is given by

$$-\,2\,\ell_n(\hat{\eta}, Y) + \underbrace{2\,\mathbb{E}_Y[\ell_n(\hat{\eta}, Y) - \ell_n(\eta_0, Y)]}_{=:S_1} + \underbrace{2\,\mathbb{E}_Y[\mathbb{E}_X[\ell_n(\eta_0, X) - \ell_n(\hat{\eta}, X)]]}_{=:S_2}. \tag{2.12}$$

Based on a Taylor expansion of the log-likelihood about η_0 and with the limiting distribution of the score given in Lemma 2.1, we can show that S_2 is asymptotically equal to

$$\underbrace{-2\mathbb{E}_Y[\sqrt{n}(\hat{\eta} - \eta_0)]^\top \begin{pmatrix} J_{01}\delta \\ J_{11}\delta \end{pmatrix}}_{=:S_{21}} + \underbrace{\mathbb{E}_Y[\sqrt{n}(\hat{\eta} - \eta_0)^\top J_{\text{full}} \sqrt{n}(\hat{\eta} - \eta_0)]}_{=:S_{22}}. \tag{2.13}$$

Exploiting the result on the limiting distribution of the MLE (2.8) and the arguments presented above yields

$$S_{21} \approx -2\left\{\left[\frac{\delta}{\kappa}\frac{\phi(-\delta/\kappa)}{[1-\Phi(-\delta/\kappa)]} + J_{11}\delta^2\right]\left[1 - \Phi\left(-\frac{\delta}{\kappa}\right)\right] + J_{10}J_{00}^{-1}J_{01}\delta^2\Phi\left(-\frac{\delta}{\kappa}\right)\right\}. \tag{2.14}$$

With regard to S_{22}, we have to combine (2.8) with the results on the mean of quadratic forms of random vectors, to arrive at

$$S_{22} \approx \left[1 - \Phi\left(-\frac{\delta}{\kappa}\right)\right] [3 + \delta^2 J_{11} + \frac{\delta}{\kappa} \frac{\phi(-\delta/\kappa)}{[1 - \Phi(-\delta/\kappa)]}] + [2 + \delta^2 J_{10} J_{00}^{-1} J_{01}] \Phi\left(-\frac{\delta}{\kappa}\right). \quad (2.15)$$

Finally, S_1 in (2.12) is shown to be asymptotically equivalent to S_{22} in (2.13); and by inserting (2.14) and (2.15) into (2.12), we obtain the postulated result. All of these steps are laid out in more detail in Section 2.7.

2.6 Discussion

In this chapter, we have studied likelihood inference in the gamma-Gompertz model. Standard likelihood theory does not apply in this case because of the boundary parameter σ^2. The limiting distributions of the MLEs and of the likelihood ratio test statistic were found to be mixtures. The results can be extended to other gamma-PH models, in which the baseline hazard differs from the exponential Gompertz hazard. Our findings further indicated that under the non-standard conditions of the gamma-Gompertz model, the common definition of the AIC was biased. We also found, however, that the bias term depended on the unknown parameter σ^2, which complicated immediate bias correction. Future work will investigate alternative model selection tools.

2.7 Supplementary material: Technical appendix and proofs

ad Theorem 2.2: *Proof of equation* (2.7)
Using a second-order Taylor expansion about η_0, the log-likelihood can be approximated by a parabola in γ, which has its maximum at $\hat{\gamma}$, satisfying

$$\sqrt{n}(\hat{\gamma} - \gamma_0) = \left[-\frac{1}{n} \frac{\partial^2 \ell_n}{\partial \gamma^2}(\hat{\theta}, \gamma_0)\right]^{-1} \sqrt{n} \bar{V}_n(\hat{\theta}, \gamma_0), \quad (2.16)$$

if the right-hand side is positive, and at $\hat{\gamma} = \gamma_0$ otherwise. While the first term, $-n^{-1} \cdot \partial^2 \ell_n(\hat{\theta}, \gamma_0)/\partial \gamma^2$, tends to J_{11}, the second term, $\sqrt{n}\bar{V}_n$, is expanded in a first-order Taylor series as

$$\sqrt{n}\bar{V}_n(\hat{\theta}, \gamma_0) \approx \sqrt{n}\bar{V}_n(\theta_0, \gamma_0) + \sqrt{n}(\hat{\theta} - \theta_0)^\top \frac{1}{n} \frac{\partial^2 \ell_n}{\partial \gamma \partial \theta}(\theta_0, \gamma_0).$$

Note that $-n^{-1}\partial^2 \ell_n(\theta_0, \gamma_0)/\partial \gamma \partial \theta$ tends to J_{10} and that $\sqrt{n}(\hat{\theta} - \theta_0)$ is distributed on Ω_n as the first components of $J_{\text{full}}^{-1}(\sqrt{n}\bar{U}_n^\top, \sqrt{n}V_n)^\top$ according to (2.5). From this it follows directly that on the set Ω_n the right-hand side of (2.16) can be approximated to first-order by Δ_n.

Proof of Lemma 2.6
The power β_n is the probability to correctly reject H_0, if H_0 does not hold. For a LRT at level α, we reject H_0 if the p-value is smaller than α. If t_{obs} is the value of the test statistic for the observed sample, the p-value is computed according to Corollary 2.5 as

$$p = \mathrm{P}[-2\ln\lambda_n(Y) \geq t_{\text{obs}}|H_0] \approx 1 - \left(\frac{1}{2}F_{\chi_1^2}(t_{\text{obs}}) + \frac{1}{2}F_{\chi_0^2}(t_{\text{obs}})\right).$$

Since $t_{\text{obs}} \geq 0$ by definition, $F_{\chi_0^2}(t_{\text{obs}}) = 1$. Furthermore,

$$F_{\chi_1^2}(t_{\text{obs}}) = \frac{\gamma\left(\frac{1}{2}, \frac{t_{\text{obs}}}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} = \frac{\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\frac{t_{\text{obs}}}{2}}\right)}{\sqrt{\pi}} = \mathrm{erf}\left(\sqrt{\frac{t_{\text{obs}}}{2}}\right),$$

using the gamma function $\Gamma(v) = \int_0^\infty u^{v-1}e^{-u}\,\mathrm{d}u$, the lower incomplete gamma function $\gamma(v, w) = \int_0^w u^{v-1}e^{-u}\,\mathrm{d}u$ and the error function $\mathrm{erf}(x) = 2/\sqrt{\pi}\int_0^x e^{-u^2}\,\mathrm{d}u$. Moreover, $1/2 + 1/2\,\mathrm{erf}(x/\sqrt{2}) = \Phi(x)$, and from this $p = 1 - \Phi(\sqrt{t_{\text{obs}}})$. Under the true model (2.4), we know from Theorem 2.4 that t_{obs} is a realization of a random variable that is asymptotically distributed as $\max\{0, Z_3/\kappa\}^2$. Thus,

$$\beta_n(\delta) = \mathrm{P}\left[p < \alpha|\text{true } \delta\right] \approx \mathrm{P}\left[1 - \Phi(\max\{0, Z_3/\kappa\}) < \alpha\right].$$

Rearranging the terms and exploiting the normality of Z_3/κ proves Lemma 2.6.

ad Theorem 2.7:
Proof of (2.13)*:*
A Taylor expansion of the log-likelihood about η_0 yields

$$\begin{aligned} S_2 &= 2\,\mathbb{E}_Y\left[\mathbb{E}_X\left[-(\hat{\eta}-\eta_0)^\top n\bar{W}_n(\eta_0, X) - \frac{1}{2}(\hat{\eta}-\eta_0)^\top\mathcal{H}(\eta_0, X)(\hat{\eta}-\eta_0) + R_1\right]\right] \\ &= \mathbb{E}_Y\left[2\,(\hat{\eta}-\eta_0)^\top\mathbb{E}_X\left[-n\bar{W}_n(\eta_0, X)\right] + (\hat{\eta}-\eta_0)^\top\mathbb{E}_X\left[-\mathcal{H}(\eta_0, X)\right](\hat{\eta}-\eta_0) + R_1\right] \end{aligned}$$

with $\mathcal{H}$ denoting the Hessian matrix of the log-likelihood and a remainder term $R_1 \overset{n\to\infty}{\longrightarrow} 0$. Inserting $\mathbb{E}_X[-\mathcal{H}(\eta_0, X)] = nJ_{\text{full}}$ and the asymptotic mean of $\bar{W}_n$ from Lemma 2.1, we obtain (2.13).

Proof of (2.14)*:*
Based on the limiting distribution of the MLE given in (2.8), the expectation $\mathbb{E}_Y[\sqrt{n}(\hat{\eta} - \eta_0)]$ in S_{21} can be approximated by

$$\mathbb{E}[Z \cdot \mathbb{1}_{\{Z_3>0\}} + TZ \cdot \mathbb{1}_{\{Z_3\leq 0\}}] = \mathbb{E}[Z \mid Z_3 > 0]\left[1 - \Phi\left(-\frac{\delta}{\kappa}\right)\right] + \mathbb{E}[TZ]\Phi\left(-\frac{\delta}{\kappa}\right).$$

Here, $\mathbb{E}[Z \mid Z_3 > 0]$ is the mean vector of a truncated trivariate normal distribution. Using the results of Gupta and Tracy (1978) on the moments of such distributions, it can be shown that

$$\mathbb{E}[Z \mid Z_3 > 0] = \kappa \begin{pmatrix} -J_{00}^{-1} J_{01} \\ 1 \end{pmatrix} \frac{\phi(-\delta/\kappa)}{[1 - \Phi(-\delta/\kappa)]} + \begin{pmatrix} 0 \\ \delta \end{pmatrix}.$$

Direct calculations then give (2.14).

Proof of (2.15)
According to (2.8) in combination with the continuous mapping theorem, the quadratic form $\sqrt{n}(\hat{\eta} - \eta_0)^\top J_{\text{full}} \sqrt{n}(\hat{\eta} - \eta_0)$ of the MLE in (2.13) converges in distribution to

$$Z^\top J_{\text{full}} Z \, \mathbb{1}_{\{Z_3 > 0\}} + (TZ)^\top J_{\text{full}} TZ \, \mathbb{1}_{\{Z_3 \le 0\}}.$$

Thus, the mean S_{22} of the quadratic form is asymptotically equal to

$$S_{22} \approx \mathbb{E}[(Z|Z_3 > 0)^\top J_{\text{full}} (Z|Z_3 > 0)] \left[1 - \Phi\left(-\frac{\delta}{\kappa}\right)\right] + \mathbb{E}[(TZ)^\top J_{\text{full}} (TZ)] \Phi\left(-\frac{\delta}{\kappa}\right),$$

which is a weighted sum of means of quadratic forms in the random vectors $(Z|Z_3 > 0)$ and TZ, respectively. For any k-dimensional random vector X with $\mathbb{E}[X] = \mu$, $\text{Cov}[X] = \Sigma$ and a constant, symmetric $(k \times k)$-matrix A, $\mathbb{E}[X^\top A X] = \text{tr}(A\Sigma) + \mu^\top A \mu$, where $\text{tr}(\cdot)$ denotes the trace of a matrix. To use this result here, the covariance matrix of the truncated $(Z|Z_3 > 0)$ is required. It can be obtained from the results of Gupta and Tracy (1978) as $J_{\text{full}}^{-1} + G$, where G equals

$$\left\{ \delta\kappa \frac{\phi(-\delta/\kappa)}{1 - \Phi(-\delta/\kappa)} - \kappa^2 \left[\frac{\phi(-\delta/\kappa)}{1 - \Phi(-\delta/\kappa)} \right]^2 \right\} \begin{pmatrix} J_{00}^{-1} J_{01} J_{10} J_{00}^{-1} & -J_{00}^{-1} J_{01} \\ -J_{10} J_{00}^{-1} & 1 \end{pmatrix}.$$

We omit the remaining straightforward computations, which result in (2.15).

Proof of the asymptotic equivalence of S_1 and S_{22}:
Another Taylor expansion of the log-likelihood about $\hat{\eta}$ yields

$$\ell_n(\eta_0, Y) = \ell_n(\hat{\eta}, Y) + (\eta_0 - \hat{\eta})^\top n \bar{W}_n(\hat{\eta}, Y) + \frac{1}{2} (\eta_0 - \hat{\eta})^\top \mathcal{H}(\hat{\eta}, Y)(\eta_0 - \hat{\eta}) + R_2,$$

with a remainder term $R_2 \overset{n\to\infty}{\longrightarrow} 0$. Thus, S_1 can be approximated by

$$\mathbb{E}_Y [2\, (\hat{\eta} - \eta_0)^\top n \bar{W}_n(\hat{\eta}, Y)] + \mathbb{E}_Y [(\hat{\eta} - \eta_0)^\top [-\mathcal{H}(\hat{\eta}, Y)](\hat{\eta} - \eta_0) + R_2]. \tag{2.17}$$

To evaluate the first summand, we study $\bar{W}_n(\hat{\eta}, Y)$ for the two cases $\hat{\sigma}^2 > 0$ and $\hat{\sigma}^2 = 0$. In the first case, the MLE is an inner point of the parameter space, and, given the definition of the MLE, we have $\bar{W}_n(\hat{\eta}, Y) = 0$, such that in this case $(\hat{\eta} - \eta_0)^\top n \bar{W}_n(\hat{\eta}, Y) = 0$. In the second case, we know that $\bar{U}_n(\hat{\eta}, Y) = 0$ and $\bar{V}_n(\hat{\eta}, Y) \le 0$, but as $(\hat{\gamma} - \gamma_0) = 0$, we still have $(\hat{\eta}-\eta_0)^\top n\bar{W}_n(\hat{\eta},Y)=0$. Consequently, the first term in (2.17) vanishes. With regard to the second summand, we argue that $-\mathcal{H}(\hat{\eta},Y)$ approximates reasonably well nJ_{full} (see also Hjort and Claeskens, 2003). Thus, S_1 is asymptotically equal to

$$\mathbb{E}_Y\,[\sqrt{n}(\hat{\eta}-\eta_0)^\top J_{\text{full}}\sqrt{n}(\hat{\eta}-\eta_0)],$$

which is equivalent to S_{22} in (2.13).

References

Abbring, J. H. and G. J. van den Berg (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika 94*(1), 87–99.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*(6), 716–723.

Gavrilova, N. S. and L. A. Gavrilov (2015). Biodemography of old-age mortality in humans and rodents. *The Journals of Gerontology: Series A 70*(1), 1–9.

Gompertz, B. (1825). 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. *Philosophical Transactions of the Royal Society of London 115*, 513–583.

Gupta, A. K. and D. S. Tracy (1978). Hermite polynomials and truncated trivariate normal distributions. *Journal of Statistical Computation and Simulation 7*, 269–286.

Hjort, N. (1994). The exact amount of t-ness that the normal model can tolerate. *Journal of the American Statistical Association 89*, 665–675.

Hjort, N. L. and G. Claeskens (2003). Frequentist model average estimators. *Journal of the American Statistical Association 98*, 879–899.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer texts in statistics. New York: Springer.

Self, S. G. and K.-Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association 82*(398), 605–610.

Thatcher, A. R., V. Kannisto, and J. W. Vaupel (1998). *The force of mortality at ages 80 to 120.* Odense Monographs on Population Aging 5. Odense University Press.

Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography 16*(3), 439–454.

Wienke, A. (2010). *Frailty Models in Survival Analysis*. Biostatistics Series. Chapman & Hall/CRC.