Statistical methods for frailty models: studies on old-age mortality and recurrent events
Böhnstedt, M.

4

Shifting attention to old age: Detecting mortality deceleration using focused model selection

Abstract

The decrease in the increase in death rates at old ages is a phenomenon that has repeatedly been discussed in demographic research. While mortality deceleration can be explained as an effect of selection in heterogeneous populations, this phenomenon can be difficult to assess statistically because it relates to the tail of the lifespan distribution. By using a focused information criterion (FIC) for model selection, we can directly target model performance at those advanced ages at death. We analyze this question in the framework of the gamma-Gompertz model that is reduced to the competing Gompertz model without mortality deceleration if the variance parameter lies on the boundary of the parameter space. We develop a new version of the FIC for this non-standard condition. In a simulation study, the new FIC is shown to outperform other methods in detecting mortality deceleration. We apply the approach to mortality data for extinct French-Canadian birth cohorts, and we extend the method to include additional covariate information.

4.1   Introduction

Almost two centuries ago, Benjamin Gompertz noted that death rates of humans increase exponentially with age from mid-life onwards (Gompertz, 1825). Makeham (1860) supplemented Gompertz's mortality model by a constant, and thus age-independent, component to provide a better fit at younger ages. By the 1930s, Perks (1932) had become aware that the Gompertz-Makeham hazard overestimated actual death rates at advanced ages, and suggested replacing the exponential part with a logistic function for graduation of mortality. The logistic hazard follows the Gompertz trajectory for the lower ages, but gradually deviates – that is, increases less quickly – at advanced ages. This slowdown in mortality rates late in life is now commonly known as mortality deceleration.

The study of death rates at high ages gained increased attention when, starting in the 1970s, progress against mortality became noticeable also among the elderly. This continuing improvement of death rates at old ages (Rau et al., 2008) is the primary reason for longevity increases in high income countries. Thus, accurately describing the trajectory of mortality at advanced ages is of great interest to actuaries, demographers, and aging researchers, as it has important implications for pension funds, life insurance, social support systems, and public health planning. Understanding and projecting the numbers of the oldest-old are essential for aging populations, and hinge on having proper estimates of mortality late in life.

As vital statistics improved and more detailed information became available for individuals who survive to very old ages, the early findings of mortality deceleration were replicated for more recent data (Horiuchi and Wilmoth, 1998; Thatcher et al., 1998; Richards, 2008; Feehan, 2018). However, the phenomenon of mortality deceleration has also been contested, with some scholars arguing that there is a continued exponential increase in mortality with age (Gavrilov and Gavrilova, 2011, 2019), and that the apparent slowdown in death rates with advancing age is primarily attributable to data errors (Newman, 2018). It is indeed the case that exaggeration in the reporting of age and the failure to remove deceased individuals from registers (due to unreported deaths) can result in an overestimation of the number of long-lived individuals, which will bias death rates downwards at the most advanced ages (Preston et al., 1999). Thus, for individuals who die at very old ages, a thorough scientific validation of the reported age at death is mandatory (Jeune and Vaupel, 1999).

Such age validation procedures involve linking individual birth (or baptism) records with death certificates, which is usually a tedious and time-consuming task. Thus, the amount of available data is often limited. Furthermore, mortality deceleration occurs in the tail of the survival distribution, where data are unavoidably scarce. For these reasons, the statistical assessment of this phenomenon is challenging, and standard methods may fail to identify deviations from the Gompertz hazard for the very old. The questions of whether and, if so, how statistical inference can be improved to tackle these challenges are addressed in this chapter.

To illustrate how rapidly the number of observations declines with age, Figure 4.1 shows for French-Canadians born between 1880 and 1896 the empirical death rates by age at ages 90 and above, as well as the number of deaths at each age. We can see that 75% of all deaths in this population had already occurred by age 96 for women and by age 95 for men. While we have sizable samples for this population (about 20,000 women and about 10,000 men), the information about (potential) mortality deceleration has to be extracted from relatively scant data in the tail.
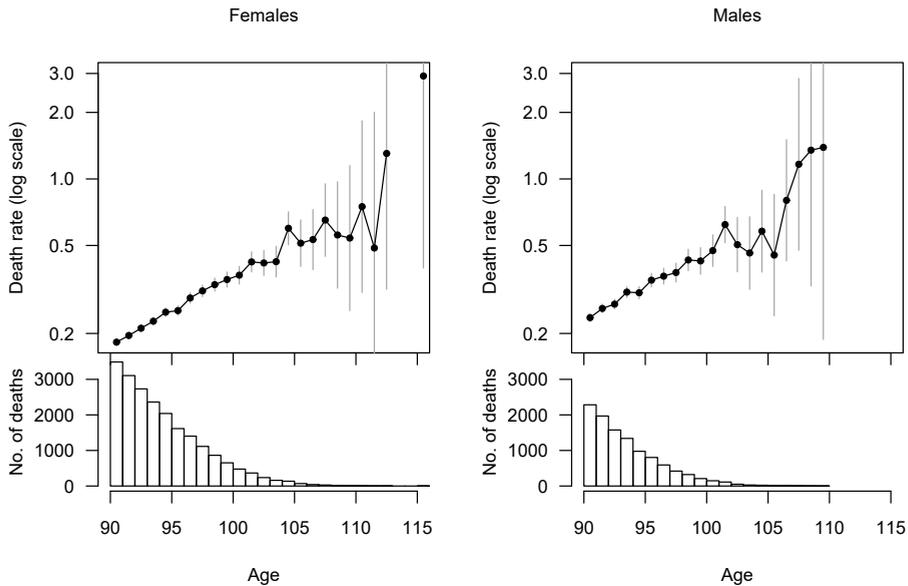
Figure 4.1: Top: Death rates (on log scale) with 95% confidence intervals for French-Canadian females (left) and males (right). Bottom: Frequency distribution of ages at death.

Much of present-day survival analysis is nonparametric, and thus avoids parametric distributional assumptions about the times to event. For the specific questions addressed here, a parametric framework proves both adequate and instrumental. The exponential increase of mortality over much of the adult lifespan has been established empirically for numerous populations, in different periods, and under varied social circumstances. Research results that dismiss mortality deceleration (Gavrilov and Gavrilova, 2011, 2019) support the use of a Gompertz hazard also at advanced ages. The possibility of treating this problem in a parsimonious parametric framework is attractive given the limited amount of data from which information on deceleration can be drawn.

In this chapter, we will discuss the statistical assessment of mortality deceleration in the framework of the gamma-Gompertz model, which is a particular parametrization of

the logistic hazard introduced by Perks (1932). This model belongs to the class of proportional hazards frailty models, which represent the standard approach for formalizing individually heterogeneous hazards of death (Vaupel et al., 1979; Wienke, 2011). As was already noted by Beard (1959), in such a model, frail individuals with higher mortality levels tend to die at younger ages, while more robust individuals with lower death risks tend to survive to higher ages, which leads to a deceleration of the average hazard with age.

In the gamma-Gompertz model, an exponentially increasing (Gompertz) baseline hazard is multiplied by a gamma distributed random effect (the frailty). If the variance σ^2 of the gamma frailty takes a positive value, then the population hazard shows a downward deviation from the Gompertz hazard at advanced ages. If the variance parameter takes the value of zero, the population hazard is exponentially increasing. Thus, answering the question of whether mortality does or does not decelerate at advanced ages corresponds to selecting the gamma-Gompertz model or the Gompertz model. However, the single additional parameter σ^2 of the gamma-Gompertz model lies on the boundary of the parameter space if the true model is the Gompertz model. As this boundary constraint on the parameter violates the usual regularity assumptions, the inference and the model selection have to be adapted to this non-standard condition.

Traditional approaches to inference in this context either employ a likelihood ratio test for $\sigma^2 = 0$, which has low power to detect actual deceleration, or model selection via Akaike's information criterion (AIC, Akaike, 1974). While information criteria, like the AIC, select a 'best' model regardless of the specific estimand that is of interest, we propose using a focused information criterion (FIC, Claeskens and Hjort, 2003) that selects the model that performs 'best' for a specific parameter of interest, called the focus parameter. Applying the FIC is particularly appealing in our context, as it will allow us to choose a focus parameter that is directly affected by the presence or the absence of mortality deceleration; for example, the hazard at some advanced age.

Technically, the FIC is constructed as an unbiased estimator of the limiting risk of an estimator of the focus parameter, and the candidate model with the smallest FIC value is selected. While the standard version of the FIC aims to minimize the mean squared error (MSE) of the estimator of the focus parameter, the criterion has been generalized to other risk measures, such as L_p-risks (Claeskens et al., 2006). Still, all of these model selection criteria have been developed based on general likelihood theory under the standard regularity assumptions, which are violated in our setting. Therefore, we will derive versions of the FIC that allow us to choose between two models in which the additional parameter may lie on the boundary of the parameter space.

The chapter is structured as follows. In Section 4.2, we summarize models for late-life mortality, and discuss the specific parametrization of the gamma-Gompertz model. We also present traditional methods for detecting mortality deceleration in this framework. Then, in Section 4.3, we propose the FIC as a new approach for assessing this phenomenon. We introduce the method for a single sample, and, in Section 4.4, we study the performance of the FIC in a simulation study, and compare it with the performance

of an AIC that is adjusted to the presence of the boundary constraint. In Section 4.5, we apply the new model selection criteria to the French-Canadian mortality data presented in Figure 4.1. We describe the data source and the age validation procedure, and apply the method to the samples of females and males separately. Additionally, we incorporate effects for different birth cohorts. We conclude with a discussion in Section 4.6.

4.2 Mortality at advanced ages

4.2.1 Hazard models

To model mortality at advanced ages, we consider the continuous random variable Y, which describes adult lifespans from mid-life onwards. Its distribution can be characterized by the hazard function, the instantaneous death rate at age y given survival up to y,

$$h(y) = \lim_{\Delta y \searrow 0} \frac{P(y < Y \leq y + \Delta y \,|\, Y > y)}{\Delta y}.$$

The value $y = 0$ corresponds to the age from which we start modeling, typically age 50 or 60. The Gompertz distribution is characterized by an exponentially increasing trajectory $h(y) = ae^{by}$, with two positive parameters a and b. Makeham (1860) extended the hazard by an age-independent constant $c \geq 0$ to $h(y) = a\,e^{by} + c$ to achieve a better fit at younger ages.

The replacement of the exponential component by a logistic term was suggested by Perks (1932) in order to allow for a slower-than-exponential increase in death rates at the highest ages, while keeping the proven Gompertz shape before that point. In its general form, the logistic hazard is

$$h(y) = \frac{Ae^{by}}{1 + Be^{by}} \qquad \text{with} \quad A > 0, B \geq 0. \tag{4.1}$$

If $B = 0$, then the hazard (4.1) reduces to the Gompertz model with $A = a$. The logistic hazard can be viewed as a simple device to capture potential mortality deceleration by one additional parameter. However, Beard (1959) already noted that a hazard of the form (4.1) arises when individuals are submitted to Gompertz hazards with individually varying parameters a. If the distribution of a follows a gamma distribution, then an average hazard of logistic shape results.

In contemporary statistical terminology, Beard's finding would be called a proportional hazards frailty model of the form $h(y|Z = z) = z \cdot h_0(y)$ (Duchateau and Janssen, 2008; Wienke, 2011). Here, a positive random effect Z (called frailty) acts multiplicatively on a common baseline hazard $h_0(y)$, such that $h(y|Z = z)$ denotes the conditional hazard of an individual at age y, given that his or her frailty is $Z = z$. The random effect Z assigns heterogeneous mortality risks to individuals in a cohort who, apart from those risks, share a common 'law' of mortality (the baseline). As a result of selection, frail individuals with higher mortality levels tend to die at younger ages, while the more

robust individuals with lower death risks tend to survive to higher ages. Consequently, the population hazard, averaged over the survivors at each age, deviates from the baseline shape; and the larger the initial heterogeneity, the stronger the deviation is.

The frailty Z is often assumed to follow a gamma distribution with mean one and variance σ^2 (Vaupel et al., 1979). The choice of the gamma distribution, which may be deemed ad hoc, is not only mathematically convenient, but is also theoretically justified. Abbring and van den Berg (2007) proved that the distribution of the heterogeneity among survivors, once selection took effect, converges to a gamma distribution for a large class of proportional hazards frailty models, even if the frailty is not gamma distributed from the outset.

The so-called gamma-Gompertz model is obtained if the gamma frailty is multiplied to an exponentially increasing Gompertz baseline hazard, $h_0(y) = ae^{by}$. The resulting marginal hazard is

$$h(y) = \frac{ae^{by}}{1 + \sigma^2 \frac{a}{b}(e^{by} - 1)}. \tag{4.2}$$

If $\sigma^2 > 0$, there is heterogeneity in the risk of death, and the selection of more robust individuals will take place. If $\sigma^2 = 0$, there is no heterogeneity, and the marginal hazard is exponentially increasing, such that $h(y) = ae^{by}$. Hence, in the framework of the gamma-Gompertz model, the statistical assessment of mortality deceleration is reduced to inference on the parameter σ^2.

The gamma-Gompertz hazard (4.2) is a straightforward reparametrization of the logistic hazard (4.1), with $A = a/(1 - \sigma^2 \frac{a}{b})$ and $B = \frac{\sigma^2 a}{b}/(1 - \sigma^2 \frac{a}{b})$; and in both versions the Gompertz hazard is retained if the third parameter equals zero, that is, $\sigma^2 = 0$ or $B = 0$, respectively. The gamma-Gompertz formulation explicitly signalizes the role of individual heterogeneity and its impact on and the progression of selection among survivors.

It is important to note that the parameter σ^2 measures population heterogeneity at the starting age of the model (corresponding to $y = 0$). Due to the continuing selection of robust individuals, the variance of frailty among the survivors decreases with age. Thus, the higher the age at which we start our observation, the lower the heterogeneity in mortality is among the individuals in the sample. The age at the beginning of the observation will, therefore, have an impact on the resulting inference.

The inference in the gamma-Gompertz model involves the frailty variance σ^2, which is a parameter that lies on the boundary of its parameter space in the absence of mortality deceleration ($\sigma^2 = 0$). This violates the standard assumptions that underlie the asymptotic properties of the likelihood-based inference, which, in turn, affects the traditional approaches for assessing mortality deceleration that are presented in the following section.

4.2.2   Traditional approaches to inference

Two methods are commonly used for assessing mortality deceleration in the framework of the gamma-Gompertz model: a likelihood ratio test for a zero frailty variance, and

model selection between the gamma-Gompertz model and the Gompertz model based on the AIC.

The likelihood ratio test for homogeneity in the gamma-Gompertz model, where $H_0 : \sigma^2 = 0$ and $H_1 : \sigma^2 > 0$, is non-standard in that, under the null hypothesis, the parameter σ^2 lies on the boundary of the parameter space. Consequently, the asymptotic distribution of the likelihood ratio test statistic under H_0 is no longer a chi-squared distribution with one degree of freedom. However, using the results of Self and Liang (1987), it can be shown that under the null hypothesis, the likelihood ratio test statistic asymptotically follows a 50:50 mixture of a point mass at zero and a chi-squared distribution with one degree of freedom, $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. Tests based on the wrong assumption of a χ_1^2-distribution of the test statistic occasionally appear in studies of mortality deceleration (Pletcher, 1999). Ignoring the issue of the boundary parameter and using the incorrect distribution of the test statistic lowers the power to (correctly) decide in favor of the gamma-Gompertz model. But even when the test statistic is correctly assumed to follow a $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$-distribution, the likelihood ratio test has low power to detect mortality deceleration in the gamma-Gompertz model. This is especially likely to be the case when the inference has to be based on age-restricted samples, such as a sample of individuals who survived beyond age 90 (see Section 4.7.2 for an illustration).

A popular alternative approach for assessing mortality deceleration is model selection based on the AIC (Richards, 2008; Gavrilova and Gavrilov, 2015). The AIC targets an unbiased estimate of the Akaike information; that is, of the expected relative Kullback-Leibler distance between the true data-generating mechanism and the best parametric approximation. Under standard conditions, the AIC is therefore defined as $-2\ell + 2k$, where the log-likelihood ℓ, evaluated at the maximum likelihood estimate, is penalized by the number k of parameters in the model. This common definition has, however, been found to be biased under the non-standard conditions of the gamma-Gompertz model (Böhnstedt and Gampe, 2019). Thus, the standard version of the AIC is not a valid tool for model selection in the setting of the gamma-Gompertz model. In Section 4.3.4, we will present a modified version of the AIC that is adjusted to the presence of a boundary parameter.

4.3 Focused information criterion for mortality deceleration

The preceding considerations indicate that neither a testing strategy, particularly if it is low-powered, nor an all-purpose model selection criterion will adequately assess the occurrence of mortality deceleration. Focused information criteria (FIC) have been introduced to address problems of this kind, and we propose selecting the model based on a new version of the FIC that takes the boundary constraint on the frailty variance into account.

4.3.1   Rationale for FIC

Statistical analyses are performed for particular purposes, and acknowledging the specific purpose when choosing the statistical model is the key concept of a FIC. In the following exposition, we use the terminology and notation of Claeskens and Hjort (2003).

Observations $y_i, i = 1, \ldots, n$ (here: ages at death) are assumed to be generated by a parametric density $f(y)$. The parameters of the model are split into a d-vector $\boldsymbol{\theta}$, which characterizes the narrow model, and an additional q-vector $\boldsymbol{\gamma}$ for the extended model. The narrow model is obtained for one particular value $\boldsymbol{\gamma}_0$, which is fixed and known. In the current application, the density of the gamma-Gompertz model (4.2) is

$$f(y) = ae^{by} \left\{ 1 + \sigma^2 \frac{a}{b}(e^{by} - 1) \right\}^{-\left(1 + \frac{1}{\sigma^2}\right)}.$$

The parameter $\boldsymbol{\theta} = (a, b)^\top$ is the Gompertz part of the model, so $d = 2$. The single additional parameter is $\gamma = \sigma^2$ with $\gamma_0 = 0$, so $q = 1$.

The original FIC is derived in a framework of local misspecification (Hjort and Claeskens, 2003), where a sample of size n is assumed to be generated from a density

$$f_{\text{true}}(y) = f(y, \boldsymbol{\theta}_0, \boldsymbol{\gamma}_0 + \boldsymbol{\delta}/\sqrt{n}), \tag{4.3}$$

with the parameter vector $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + \boldsymbol{\delta}/\sqrt{n}$ perturbed in the direction of $\boldsymbol{\delta}$. Selection is between the null model, where $\boldsymbol{\gamma}$ is fixed at the known value $\boldsymbol{\gamma}_0$; the full model, including both $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$; and, if $q > 1$, any model including $\boldsymbol{\theta}$, but only a subset of the components of $\boldsymbol{\gamma}$ and the remaining fixed at the respective values in $\boldsymbol{\gamma}_0$. For the current setting, selection is only between the null model with $\sigma^2 = 0$, that is, the Gompertz model; and the full model including σ^2, that is, the gamma-Gompertz model. Due to the boundary constraint on the frailty variance, $\delta = \sqrt{n}\sigma^2$ is subject to the a priori restriction $\delta \geq 0$. Therefore, we will restrict the framework in the following to the choice of including or not including a single parameter with a boundary constraint; that is, $q = 1$ and $\gamma \geq \gamma_0$.

The focus is the parameter of interest, which depends on the underlying density (4.3) via $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$. The focus is commonly denoted by μ, and we define $\mu_{\text{true}} = \mu(\boldsymbol{\theta}, \gamma_0 + \delta/\sqrt{n})$. Based on the maximum likelihood estimators $\hat{\boldsymbol{\theta}}_{\text{null}}$ in the null model and $(\hat{\boldsymbol{\theta}}_{\text{full}}, \hat{\gamma})$ in the full model, the focus parameter is estimated as $\hat{\mu}_{\text{null}} = \mu(\hat{\boldsymbol{\theta}}_{\text{null}}, \gamma_0)$ or $\hat{\mu}_{\text{full}} = \mu(\hat{\boldsymbol{\theta}}_{\text{full}}, \hat{\gamma})$. For each model M, $M \in \{\text{null, full}\}$, the estimator $\hat{\mu}_M$ converges in distribution, $\sqrt{n}(\hat{\mu}_M - \mu_{\text{true}}) \xrightarrow{d} \Lambda_M$.

The FIC selects the model that performs 'best' for the focus parameter μ. If it is based on the general L_p-loss, the FIC aims to estimate without bias the limiting L_p-risk of $\hat{\mu}_M$; that is, $r_p(M) = \mathbb{E}[|\Lambda_M|^p]$. The model for which this limiting risk is smaller is selected by the criterion. Of particular interest is a FIC based on the MSE ($p = 2$, as for the original version, Claeskens and Hjort, 2003), constructed as an estimator of $\mathbb{E}[\Lambda_M^2]$; and a FIC based on the mean absolute error (MAE, $p = 1$), constructed as an estimator of $\mathbb{E}[|\Lambda_M|]$.

4.3.2   FIC with a parameter on the boundary of the parameter space

Under standard regularity conditions, when general likelihood theory applies, the asymptotic normality of the maximum likelihood estimator implies that the Λ_M are normally distributed (Claeskens and Hjort, 2003). In the non-standard setting considered here, Λ_{full} is not normally distributed because the maximum likelihood estimator $(\hat{\boldsymbol{\theta}}_{\text{full}}, \hat{\gamma})$ converges in distribution to a mixture with two components (Böhnstedt and Gampe, 2019). The limiting distribution depends on the information matrix J_{full} of the full model evaluated at the null model $(\boldsymbol{\theta}_0, \gamma_0)$. We denote by J_{00}, J_{01}, J_{10}, and J_{11}, the four blocks of J_{full} corresponding to the components $\boldsymbol{\theta}$ and γ of the parameter vector; and by κ^2 the element of the inverse information matrix J_{full}^{-1}, which corresponds to γ. Then, the following convergence in distribution holds for the estimator of the frailty variance

$$\sqrt{n}(\hat{\gamma} - \gamma_0) \xrightarrow{d} \max(0, D) \quad \text{with } D \sim \mathcal{N}(\delta, \kappa^2).$$

For the limiting distribution of the estimator of the focus parameter, it can be shown that

$$\sqrt{n}(\hat{\mu}_{\text{null}} - \mu_{\text{true}}) \xrightarrow{d} \Lambda_{\text{null}} = \Lambda_0 + \omega\delta \quad \text{and}$$

$$\sqrt{n}(\hat{\mu}_{\text{full}} - \mu_{\text{true}}) \xrightarrow{d} \Lambda_{\text{full}} = \begin{cases} \Lambda_0 + \omega(\delta - D) & \text{if } D > 0 \\ \Lambda_0 + \omega\delta & \text{if } D \leq 0 \end{cases}, \tag{4.4}$$

where $\Lambda_0 \sim \mathcal{N}(0, \tau_0^2)$ is independent of D, $\tau_0^2 = \left(\frac{\partial\mu}{\partial\theta}\right)^{\top} J_{00}^{-1} \frac{\partial\mu}{\partial\theta}$ and $\omega = J_{10} J_{00}^{-1} \frac{\partial\mu}{\partial\theta} - \frac{\partial\mu}{\partial\gamma}$ (cf. Section 10.2 in Claeskens and Hjort, 2008).

To define a FIC, we need to derive $\mathbb{E}[|\Lambda|]$ or $\mathbb{E}[\Lambda^2]$ from (4.4), depending on whether we intend to base the criterion on the limiting L_1- or L_2-risk of the estimator $\hat{\mu}$.

As in the original version of the FIC, the limiting MSE of $\hat{\mu}$ is considered first. However, as we will demonstrate in the following, the FIC based on the L_2-risk has some drawbacks in the current setting, which makes the L_1-risk an attractive alternative.

From equation (4.4) we can determine $\mathbb{E}[\Lambda^2]$ for the null and the full model:

$$\mathbb{E}[\Lambda_{\text{null}}^2] = \tau_0^2 + \omega^2\delta^2 \quad \text{and}$$

$$\mathbb{E}[\Lambda_{\text{full}}^2] = \tau_0^2 + \omega^2 \left\{ \delta^2 \Phi\left(-\frac{\delta}{\kappa}\right) - \kappa\delta\phi\left(\frac{\delta}{\kappa}\right) + \kappa^2 \Phi\left(\frac{\delta}{\kappa}\right) \right\}, \tag{4.5}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cdf and the pdf of the standard normal distribution, respectively. The FIC_{MSE} would be constructed as an unbiased estimator of the MSEs in (4.5), and the model with the smaller FIC value would be selected.

As has already been pointed out by Claeskens and Hjort (2008, Section 5.3), in the case of a single additional parameter γ, the so-called tolerance radius does not depend on the focus μ. This radius signifies the deviation δ for which the MSE of the null model estimator is smaller than that of the full model estimator; that is, $\mathbb{E}[\Lambda_{\text{null}}^2] \leq \mathbb{E}[\Lambda_{\text{full}}^2]$. From (4.5), we see that the two risks are the same for $\omega = 0$, and that if $\omega \neq 0$ the

tolerance radius encompasses all δ with $\delta < 0.8399\kappa$. We can still define a pre-test strategy for assessing mortality deceleration, which is based on the quantity $\hat{\delta}/\hat{\kappa}$, where $\hat{\delta} = \sqrt{n}(\hat{\gamma} - \gamma_0) = \sqrt{n}\hat{\sigma}^2$ and $\hat{\kappa}$ is derived from the observed Fisher information. If $\hat{\delta}/\hat{\kappa} \leq 0.8399$, the estimator $\hat{\mu}_{\text{null}}$ based on the Gompertz model is used; whereas if $\hat{\delta}/\hat{\kappa} > 0.8399$, the estimator $\hat{\mu}_{\text{full}}$ based on the gamma-Gompertz model is used. We note here that $\hat{\delta}$ is not an unbiased estimator of δ, with the bias depending in a complex way on δ and κ. In appraising this pre-test-based model choice, we can see that for large samples, the local power of this strategy is approximately the same as the power of a likelihood ratio test for $H_0 \colon \sigma^2 = 0$ at the 20% level (cf. Section 4.7.3).

Although strategies based on the limiting L_2-risks of the estimator $\hat{\mu}$ are common, the derived pre-test strategy has drawbacks. On the one hand, the performance of this strategy does not depend on the chosen focus parameter; while on the other, the equal penalty for squared bias and variance of the estimators in the L_2-risk might not be suitable for choosing whether to include a heterogeneity parameter.

Consequently, using risk measures other than the L_2-risk can be more appropriate, as was already suggested in Claeskens et al. (2006). Formulas for the general limiting L_p-risk of $\hat{\mu}_M$ were derived there under regularity conditions where Λ_M follows a normal distribution for each of the models. In our non-standard setting, the limiting distribution of the full model estimator in (4.4) is not normal, but we can still derive the limiting L_1-risk of the estimators $\hat{\mu}_{\text{null}}$ and $\hat{\mu}_{\text{full}}$ as follows (see Section 4.7.4 for details):

$$\mathbb{E}[|\Lambda_{\text{null}}|] = 2\tau_0\phi\left(\frac{\omega\delta}{\tau_0}\right) + 2\omega\delta\left\{\Phi\left(\frac{\omega\delta}{\tau_0}\right) - \frac{1}{2}\right\} \quad \text{and}$$

$$\mathbb{E}[|\Lambda_{\text{full}}|] = \left[2\tau_0\phi\left(\frac{\omega\delta}{\tau_0}\right) + 2\omega\delta\left\{\Phi\left(\frac{\omega\delta}{\tau_0}\right) - \frac{1}{2}\right\}\right]\left\{1 - \Phi\left(\frac{\delta}{\kappa}\right)\right\} \qquad (4.6)$$

$$+ \sqrt{\tau_0^2 + \omega^2\kappa^2} \cdot \sqrt{\frac{2}{\pi}}\Phi\left(\frac{\delta}{\kappa} \cdot \frac{\sqrt{\tau_0^2 + \omega^2\kappa^2}}{\tau_0}\right) - \omega\kappa\,\phi\left(\frac{\delta}{\kappa}\right) \cdot 2\left\{\Phi\left(\frac{\omega\delta}{\tau_0}\right) - \frac{1}{2}\right\}.$$

Thus, we define the FIC_{MAE} of the null model and the full model as the estimators

$$\text{FIC}_{\text{MAE}}(\text{null}) = 2\hat{\tau}_0\phi\left(\frac{\hat{\omega}\hat{\delta}}{\hat{\tau}_0}\right) + 2\hat{\omega}\hat{\delta}\left\{\Phi\left(\frac{\hat{\omega}\hat{\delta}}{\hat{\tau}_0}\right) - \frac{1}{2}\right\} \quad \text{and}$$

$$\text{FIC}_{\text{MAE}}(\text{full}) = \left[2\hat{\tau}_0\phi\left(\frac{\hat{\omega}\hat{\delta}}{\hat{\tau}_0}\right) + 2\hat{\omega}\hat{\delta}\left\{\Phi\left(\frac{\hat{\omega}\hat{\delta}}{\hat{\tau}_0}\right) - \frac{1}{2}\right\}\right]\left\{1 - \Phi\left(\frac{\hat{\delta}}{\hat{\kappa}}\right)\right\}$$

$$+ \sqrt{\hat{\tau}_0^2 + \hat{\omega}^2\hat{\kappa}^2} \cdot \sqrt{\frac{2}{\pi}}\Phi\left(\frac{\hat{\delta}}{\hat{\kappa}} \cdot \frac{\sqrt{\hat{\tau}_0^2 + \hat{\omega}^2\hat{\kappa}^2}}{\hat{\tau}_0}\right) - \hat{\omega}\hat{\kappa}\,\phi\left(\frac{\hat{\delta}}{\hat{\kappa}}\right) \cdot 2\left\{\Phi\left(\frac{\hat{\omega}\hat{\delta}}{\hat{\tau}_0}\right) - \frac{1}{2}\right\},$$

respectively. Based on this new model selection criterion FIC_{MAE}, the full model is chosen if the estimated MAE of its estimator of the focus parameter μ is smaller than the MAE for the null model estimator. In contrast to the MSE, the tolerance radius determined by the MAE of $\hat{\mu}_M$ does depend on the focus parameter via ω and τ_0.

4.3.3 Choice of the focus parameter

The central concept and virtue of the FIC approach is that it allows us to consolidate a scientific question in a focus parameter, and to customize the model selection to the specific focus. In the context of mortality deceleration, two focus parameters suggest themselves. The first parameter is the frailty variance, since it determines whether mortality deceleration is present, so $\mu = \sigma^2$. The second focus parameter targets the deceleration of the hazard function, measured by the second derivative of the log-hazard at some (high) age y so that $\mu = [\ln h(y)]''$.

For $\mu = \sigma^2$ the expressions in (4.6) take the form

$$\mathbb{E}[|\Lambda_{\text{null}}|] = \delta \quad \text{and} \quad \mathbb{E}[|\Lambda_{\text{full}}|] = \kappa\sqrt{\frac{2}{\pi}} - \kappa\phi\left(-\frac{\delta}{\kappa}\right) + \delta\Phi\left(-\frac{\delta}{\kappa}\right).$$

Consequently, model choice based on the FIC_{MAE} results in the gamma-Gompertz model if $\hat{\delta}/\hat{\kappa} > 0.6399$. If we view this as a pre-test strategy, then it has asymptotically the same local power as the likelihood ratio test for $H_0 : \sigma^2 = 0$ at a level of 26%.

If we choose $\mu = [\ln h(y)]''$ the choice of the age y should be such that it marks an age in the tail of the distribution where deceleration occurs, but which still lies within the range of observed lifespans.

While the above choices of the focus parameter are natural and allow for immediate interpretations, we could also select as the focus any function that characterizes the distribution of lifespans, such as the survival function or the log-hazard. The effects of different focus parameters on the model selection will be briefly illustrated in the simulation study in Section 4.4, and recommendations will be given in Section 4.6.

4.3.4 A modified AIC for the gamma-Gompertz model

As we mentioned in Section 4.2.2, the standard AIC is biased as an estimator of the Akaike information in the presence of a boundary parameter, and should therefore not be used for assessing mortality deceleration. However, Böhnstedt and Gampe (2019) explicitly derived the bias of the standard AIC for the gamma-Gompertz model (4.2) under the local misspecification framework (4.3) as $2\,\Phi\left(-\delta/\kappa\right)$. This bias depends via $\delta = \sqrt{n}\sigma^2$ on the unknown value of the frailty variance, and it cannot be estimated without bias if the true variance is small. Thus, the bias cannot be removed completely, but it can be reduced if we correct the standard AIC using the estimator $2\,\Phi\left(-\hat{\delta}/\hat{\kappa}\right)$ of the bias term. Hence, we define a modified version of the AIC for the gamma-Gompertz model as

$$\text{AIC}^* = -2\,\ell + 2 \cdot 3 - 2\,\Phi\left(-\frac{\hat{\delta}}{\hat{\kappa}}\right). \tag{4.7}$$

The performance of this modified AIC^* for detecting mortality deceleration is studied in the next section.

4.4 Simulation study

To examine the performance of the proposed FIC_{MAE} in assessing mortality deceleration, we conducted a simulation study. In addition to considering different choices for the focus, the study compares the behavior of the FIC_{MAE} with that of the pre-test based on L_2-risks, and with that of the AIC^* defined in (4.7).

The following factors will affect the performance of the different strategies: the size of the true frailty variance σ^2; the sample size n; and the starting age used when observing lifespans, with a younger starting age being more favorable for detecting actual mortality deceleration.

For the frailty variance (at $y = 0$), three different scenarios were considered: $\sigma^2 = 0.0625$ (S_1) and $\sigma^2 = 0.03$ (S_2) with Gompertz parameters $a = 0.013$, $b = 0.092$. Scenario S_3 is a pure Gompertz model with $a = 0.0198$, $b = 0.0726$ (and $\sigma^2 = 0$). Scenario S_1 was chosen to resemble the mortality pattern in the data on French-Canadian females that are analyzed in the following section. Scenario S_2 mimics a population with the same Gompertz baseline hazard, but lower heterogeneity in the risk of death. By comparing Scenarios S_1 and S_2, we will therefore be able to single out the effect of the size of the frailty variance on the performance of the selection strategies. Note that the frailty variances were set to levels comparable to those estimated from the French-Canadian data. Lastly, Scenario S_3 was inspired by a fit of the Gompertz model to the female data.

To cover the latter two aspects, survival times were generated from the gamma-Gompertz model (4.2), with $y = 0$ corresponding to age 60. However, the model selection was based only on subsets of individuals reaching certain ages. Motivated by the French-Canadian data, we considered individuals who survived to ages 90 or higher (90+). Additional comparisons based on the larger subsets of individuals who survived to ages 85+ and 80+ are presented in Section 4.7.6.

For each scenario S_1 to S_3, three different initial sample sizes (at age 60) were chosen, such that the size of the 90+ subset approximately equals $n_{90+} = 10{,}000$ (small), $n_{90+} = 20{,}000$ (medium) or $n_{90+} = 105{,}000$ (large). The sample sizes may look unusually large, but they cover a realistic range of population-based data. Recall that the French-Canadian data presented in Figure 4.1 contain information on about 20,000 women and 10,000 men.

For each 90+ sample, the log-likelihoods for the Gompertz model and the gamma-Gompertz model were maximized using function `nlm()` in R (R Core Team, 2018); further computational details are given in Section 4.7.1. Then, the best model is selected based on the FIC_{MAE} for different focus parameters, the MSE pre-test of $\delta < 0.8399\kappa$, and the AIC^*. We ran 1,000 replications for each setting.

The left panel of Figure 4.2 compares the performance of the three selection approaches in Scenario S_1 ($\sigma^2 = 0.0625$) across the various sample sizes. The FIC_{MAE} with focus parameter $\mu = [\ln h(100)]''$ clearly outperforms the other two methods, as it detects mortality deceleration more often. The proportion of correct decisions in favor of the gamma-Gompertz model increases with the sample size for all three methods, and is close to one for the setting with a large sample size. However, for the setting with a small
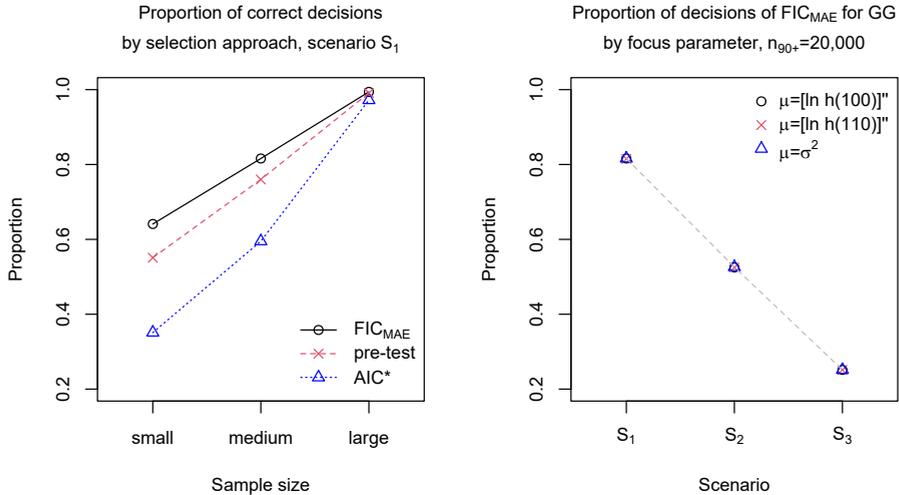
Figure 4.2: Proportion of decisions in favor of the gamma-Gompertz model. Left: Scenario S_1 for sample sizes $n_{90+} = 10{,}000$, $n_{90+} = 20{,}000$, and $n_{90+} = 105{,}000$ (left to right) based on FIC_{MAE} with $\mu = [\ln h(100)]''$ (black-solid-circle), pre-test (red-dashed-cross), and AIC^* (blue-dotted-triangle). Right: Scenarios S_1, S_2, and S_3 (left to right) all with $n_{90+} = 20{,}000$ based on FIC_{MAE} with $\mu = [\ln h(100)]''$ (black circle), $\mu = [\ln h(110)]''$ (red cross), and $\mu = \sigma^2$ (blue triangle).

(medium) sample size, the proportion of correct decisions based on the FIC_{MAE} is 82.6% (37.1%) higher than that based on the AIC^*.

The right panel of Figure 4.2 illustrates the performance of the FIC_{MAE} depending on the magnitude of the frailty variance, and on the choice of the focus parameter in the medium sample size setting. We display the results for the focus parameters $\mu = \sigma^2$, $\mu = [\ln h(100)]''$, and $\mu = [\ln h(110)]''$. The ability of the method to detect deviations from the Gompertz hazard naturally decreases when the frailty variance decreases. For Scenario S_2, in which the frailty variance is about half as large as it is in Scenario S_1, the proportion of correct decisions is about 35% smaller than it is for S_1. If the true model is the Gompertz model (S_3), then the proportion of decisions in favor of the gamma-Gompertz model is about 25% for the medium sample size. As the FIC_{MAE} performs equally well for all three focus parameters, the age y at which $\mu = [\ln h(y)]''$ is evaluated does not seem to matter. It also turns out that the focus parameters $\mu = \sigma^2$ and $\mu = [\ln h(y)]''$ perform better than, for instance, $\mu = \ln h(y)$ or $\mu = S(y)$; as is shown in Section 4.7.5. Although the focus age y did not affect the results in the simulation study, other aspects may render one choice more reasonable than another. In the medium-sized Scenario S_1, in which around 20,000 individuals reach age 90, more than a thousand will, on average, also reach age 100, but fewer than 10 will reach age 110. Consequently, a

focus age of $y = 100$ will probably produce more reliable results than a focus age of $y = 110$.
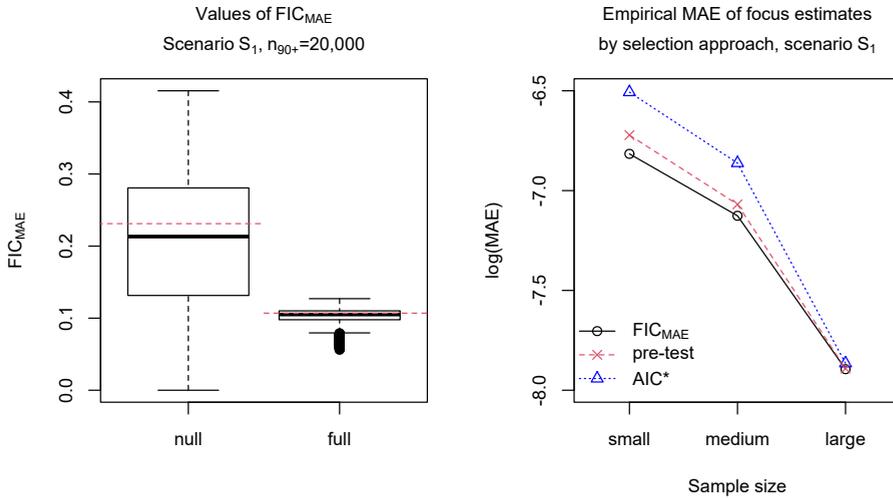
Figure 4.3: Left: Box plots of FIC_{MAE} values with $\mu = [\ln h(100)]''$ for the null and the full model in Scenario S_1 with $n_{90+} = 20{,}000$ and empirical MAE of focus estimates $\hat{\mu}$ (red-dashed). Right: Empirical MAE of selected $\hat{\mu}$ for $\mu = [\ln h(100)]''$ in Scenario S_1 for sample sizes $n_{90+} = 10{,}000$, $n_{90+} = 20{,}000$, and $n_{90+} = 105{,}000$ (left to right) based on FIC_{MAE} (black-solid-circle), pre-test (red-dashed-cross), and AIC* (blue-dotted-triangle).

The concept of the FIC_{MAE} as an estimator of the limiting MAE of $\hat{\mu}$ is illustrated in the left panel of Figure 4.3, which shows a box plot of the FIC_{MAE} values with $\mu = [\ln h(100)]''$ for 1,000 replications of the medium-sized Scenario S_1. We see that for both the null and the full model, the empirical MAEs of the estimators $\hat{\mu}_{\text{null}}$ and $\hat{\mu}_{\text{full}}$ are close to the average of the respective FIC scores. As a consequence, the empirical MAE of the selected estimators in the 1,000 replications – that is, $\hat{\mu}_{\text{full}}$ for those replications, where $\text{FIC}_{\text{MAE}}(\text{full}) < \text{FIC}_{\text{MAE}}(\text{null})$, and $\hat{\mu}_{\text{null}}$ otherwise – should be smaller than it is for other selection criteria. The right panel of Figure 4.3 verifies for Scenario S_1, that the estimator $\hat{\mu}$ of $\mu = [\ln h(100)]''$ has the smallest empirical MAE when the model selection is based on the FIC_{MAE} with $\mu = [\ln h(100)]''$, rather than on the pre-test or the AIC*.

Overall, the findings of the simulation study support the claim that the proposed FIC_{MAE} is a suitable tool for detecting mortality deceleration in the framework of the gamma-Gompertz model, which outperforms the competing approaches of the pre-test and AIC*.

4.5   Mortality of French-Canadians at high ages

4.5.1   Data and age validation

To demonstrate the performance of the proposed FIC, we analyze data on French-Canadian cohorts born between 1880 and 1896. The data were illustrated in Figure 4.1.

When French missionaries settled in Quebec in the 17th century, they followed the Catholic tradition of registering all baptisms, marriages, and burials in parish registers. Starting in 1679, two copies of the registrations were made, with one being kept at the parish, while the other was sent to the government body responsible for civil registration. This practice was continued until the end of the 20th century. Thus, this data collection is an invaluable resource for historical demography, among other disciplines (Desjardins, 1998). The data set that we analyze here covers virtually all Catholic French-Canadians (20,917 females and 10,878 males) who were born in the Province of Quebec during the 1880-1896 period, and who died at age 90 or older in Quebec between 1970 and 2009. These 1880-1896 birth cohorts were fully extinct by the end of 2009.

To confirm a reported age at death (date of death minus date of birth), the date of birth reported on the death certificate has to be verified using the baptism record in the corresponding parish register, and is often additionally verified using historic census returns. Although the historic French-Canadian parish registers contain information on all baptisms and are available on microfilm, the identification of individual entries in the registers continues to be time-consuming. The appropriate parish has to be identified using the information on the province and the names of the parents, and the individual register entry has to be looked up (for details, see Bourbeau and Desjardins, 2002).

The core of the data set was compiled by Beaudry-Godin (2010), who individually validated ages at death using the parish registers of all Catholic French-Canadians who died as centenarians – i.e., at age 100 or older – between 1970 and 2004. The data were further extended and validated to cover deaths until 2009 (Ouellette and Bourbeau, 2014; Ouellette, 2016), and the deaths of individuals aged 90-99 were added.

4.5.2   Comparison of information criteria

We fit the gamma-Gompertz model and the Gompertz model to the female and the male data separately via maximum likelihood. For that purpose, we set age 60 as the starting age of the models, and take into account left truncation at age 90. Then, we choose between the gamma-Gompertz model and the Gompertz model based on the AIC*, pretest, and FIC_{MAE} for the focus parameters $\mu = \sigma^2$ and $\mu = [\ln h(100)]''$.

Figure 4.4 shows the fit of the gamma-Gompertz model and the Gompertz model, respectively, to the empirical death rates (single years of age) for the French-Canadian cohorts. The estimated frailty variance in the gamma-Gompertz model is $\hat{\sigma}^2 = 0.043$ for the female population and $\hat{\sigma}^2 = 0.037$ for the male population. A likelihood ratio test for $H_0 : \sigma^2 = 0$ results in a p-value of 0.121 for females and 0.283 for males, such that the hypothesis of no mortality deceleration would not be rejected at the usual levels of

significance. Table 4.1 also shows that based on the modified AIC*, the Gompertz model is selected, if only just, for females and for males. By contrast, based on the pre-test and the FIC_{MAE}, the gamma-Gompertz model is selected for females, and the Gompertz model is selected for males. Hence, it appears that unlike other methods, the FIC_{MAE} detects mortality deceleration in the female sample. Figure 4.4 also supports this finding of a deceleration in the female mortality rates.
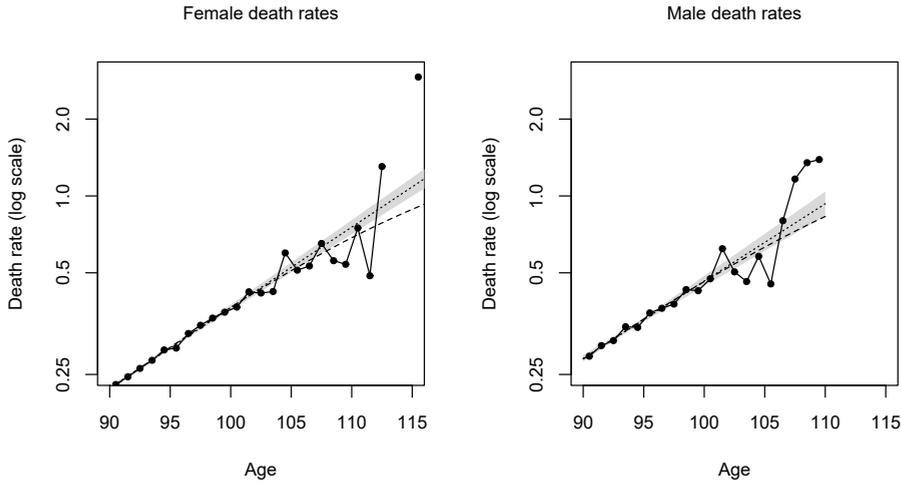
Figure 4.4: Death rates (on log scale) of French-Canadian females (left) and males (right): empirical death rates (solid-circle), gamma-Gompertz fit (dashed), Gompertz fit (dotted), and 95%-confidence band for the Gompertz log-hazard (gray).

Table 4.1: Values of different model selection criteria for the gamma-Gompertz model (GG) and the Gompertz model using data on French-Canadians.

	Females		Males	
	GG	Gompertz	GG	Gompertz
AIC*	101390.3	**101390.0**	48364.10	**48363.01**
$FIC_{MAE} : \mu = \sigma^2$	**4.065**	6.200	4.316	**3.890**
$FIC_{MAE} : \mu = [\ln h(100)]''$	**0.098**	0.149	0.120	**0.108**

4.5.3   Including effects of year of birth

The analysis so far has split the data by gender, but combined all birth cohorts in one model. This was done primarily in order to retain sufficiently large samples, which, as the simulations showed, is an important issue in the assessment of mortality deceleration, even if the samples can be considered sizable based on common statistical standards. But it is obvious that such a pooling of birth cohorts can increase the heterogeneity in mortality risks, and may thus yield a larger frailty variance, which, in turn, affects the assessment of mortality deceleration. This concern is particularly relevant if mortality patterns have improved over time at advanced ages (Vaupel et al., 2021). Including information on the year of birth has been shown to be important in different contexts (Richards et al., 2006).

We have, therefore, chosen to include the effect of year of birth in the analysis by allowing the parameters of the gamma-Gompertz model to vary by cohort, and have adapted our model selection approach accordingly. While the inclusion of cohort effects implies a more flexible approach to modeling, it also reduces the parsimony of the model, which is critical in this application. In the following, we will focus on the female sample in undertaking this considerably more complex analysis. The male sample is only half of the size of the female sample, and does not allow for a reliable statistical analysis in the extended setting introduced below.

The data cover 17 single-year birth cohorts made up of between about 700 and 2000 individuals. These cohort sizes, in conjunction with the restricted age range of 90+, are insufficient for an analysis by single years of birth. For this reason, we group the data into multi-year birth cohorts of roughly equal sizes by combining individuals with the following birth years: 1880-1884, 1885-1888, 1889-1891, 1892-1894, and 1895-1896. In the remainder of this section, we will refer to these five multi-year birth cohorts, which are, on average, made up of 4183.4 individuals.

As a gamma-Gompertz model in which all parameters are cohort-specific would still have five times as many parameters to be estimated as the plain gamma-Gompertz model considered in Section 4.5.2, we seek to identify a model that is more parsimonious, but is still appropriately flexible. To do so, we consult data for the respective cohorts between ages 60 and 109 from the Canadian Human Mortality Database (CHMD, 2020). The CHMD data are collected at the level of the Canadian provinces and territories, including the Province of Quebec. These data can be used to identify potential trends across cohorts in the Gompertz part of the gamma-Gompertz model, which is largely determined at the mid-adult ages. Although the data for Quebec in the CHMD cover a wider population, French-Canadians constitute the majority in the age groups considered here, so that these data can serve for this purpose. (A comparison of the two data sources has already been provided in Ouellette, 2016.) More precisely, we use the CHMD data (number of deaths and total exposure times at each age) for Quebec to estimate Gompertz models with a starting age of 60, separately for each of the five cohorts (maximizing a Poisson likelihood for death counts; see, for instance, Keiding, 1990). The parameter estimates suggest that there is a linear trend in the Gompertz parameter a across cohorts, while the Gompertz parameter b stays roughly constant for the cohorts under study (see Figure 4.9).

This gamma-Gompertz model, in which there is a linear trend in the parameter a and a single parameter b, but with each cohort having its own frailty variance, is estimated and analyzed in the following. This model is a restricted version of the gamma-Gompertz model in which all parameters vary by cohort. Nonetheless, it still provides an adequate description of the mortality patterns in the female French-Canadian cohorts studied here. More formally, we modify the gamma-Gompertz model with hazard (4.2) in the following way: The parameters for cohort c, $c = 1, ..., C$, are given by

$$a_c = a_0 + a_1 \cdot z_{1c}, \qquad b_c \equiv b_0, \qquad \text{and} \quad \sigma_c^2 = \sigma_0^2 \exp(z_2^\top \zeta), \qquad (4.8)$$

where a_0 and σ_0^2 are the parameters of a reference cohort c^* and $\zeta \in \mathbb{R}^{C-1}$ describes cohort-specific deviations from the frailty variance. The parameter $b_0 > 0$ applies to all cohorts, whereas the linear trend in the initial level of mortality a_c is governed by the slope $a_1 \in \mathbb{R}$ (but such that $a_c > 0$) with $z_{1c} = (c - c^*)$ and $z_1 = (z_{11}, ..., z_{1C})^\top$. The vector z_2 of length $(C - 1)$ consists of cohort dummies.

We fit model (4.8) to the individual survival times of the female cohorts in the French-Canadian data via maximum likelihood. The middle cohort is used as reference and, as before, age 60 is set as the starting age of the model, and left truncation at age 90 is taken into account. The parameter estimates and the corresponding standard errors are summarized in Table 4.2. The negative estimate of a_1 indicates a decrease in the initial level of mortality for the later cohorts, and, hence, that mortality has indeed improved. The estimated frailty variances are comparatively high for the first three cohorts, but are decreasing for the last two cohorts, reaching a value close to zero for the most recent cohort. These results are in line with the impression conveyed by the empirical death rates of the five cohorts displayed in Figure 4.5.

Table 4.2: Parameter estimates (with standard errors) for gamma-Gompertz model (4.8) with cohort effects for French-Canadian females.

$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	$\hat{\sigma}_4^2$	$\hat{\sigma}_5^2$
0.0122	−0.0006	0.0932	0.1089	0.0989	0.0757	0.0471	0.0089
(0.0035)	(0.0002)	(0.0113)	(0.0379)	(0.0399)	(0.0401)	(0.0406)	(0.0413)

Performing focused model selection is considerably more challenging in this extended setting with cohort effects, even in the still relatively parsimonious gamma-Gompertz model specified in (4.8). In the framework of local misspecification, as introduced in Section 4.3.1, the full model is now defined as the gamma-Gompertz model (4.8) with cohort effects. The null model – that is, the simplest candidate model – should still be the plain Gompertz model without cohort effects. Thus, the parameter vector of the full model is split into the Gompertz part $\theta = (a_0, b_0)^\top$, so still with $d = 2$, and the additional $\gamma = (\sigma_0^2, a_1, \zeta^\top)^\top$ with $\gamma_0 = (0, 0, \mathbf{0}^\top)^\top$, so $q = (C + 1)$.

Thus, selection is no longer only between the full model and the null model (as was the case in Section 4.5.2), as we can now choose from an extended list of candidate

Female death rates 1880-1884

Female death rates 1885-1888

Female death rates 1889-1891

Female death rates 1892-1894

Female death rates 1895-1896
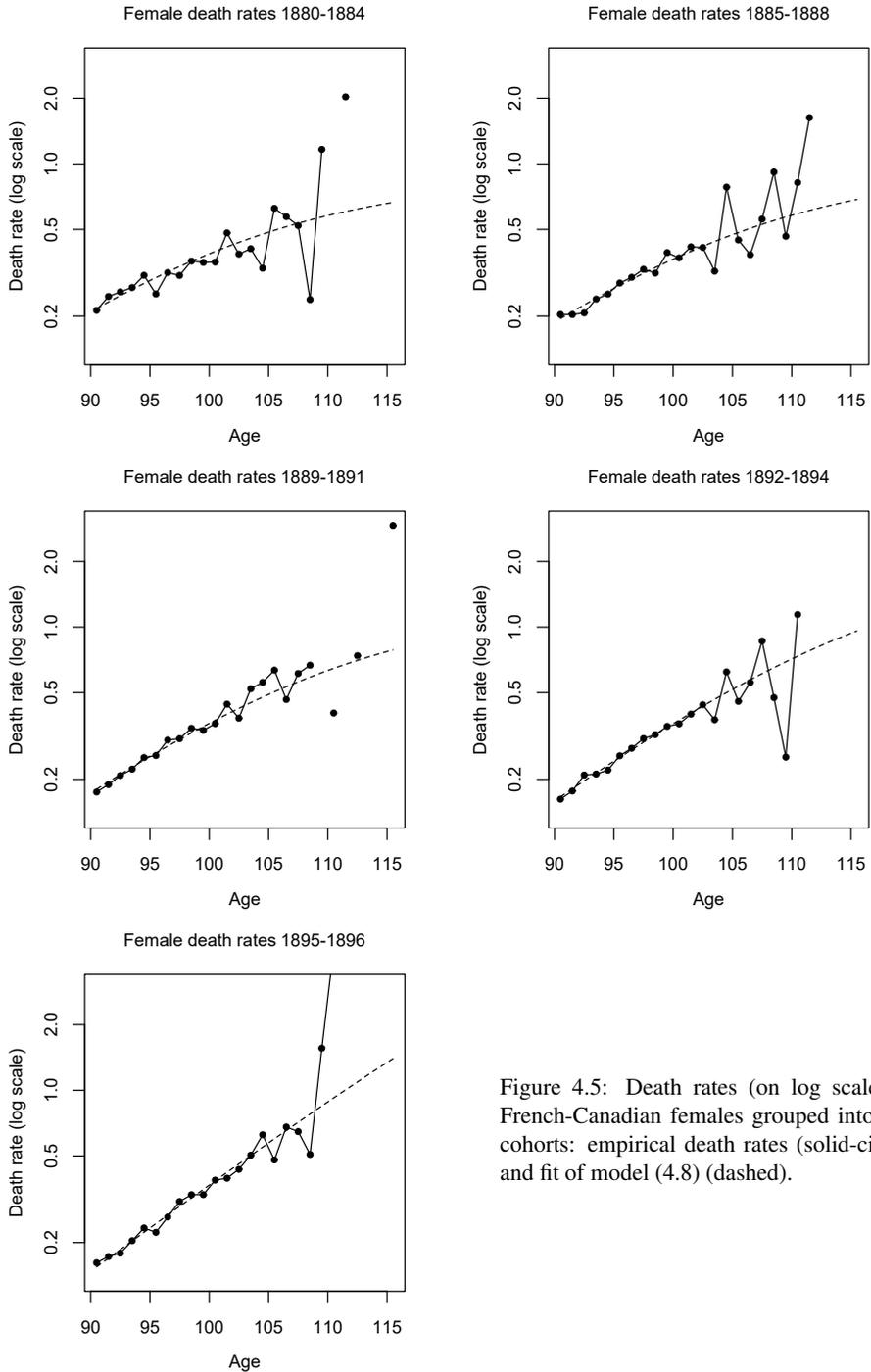
Figure 4.5: Death rates (on log scale) of French-Canadian females grouped into five cohorts: empirical death rates (solid-circle) and fit of model (4.8) (dashed).

models. On this list of candidate models are also the plain gamma-Gompertz model without cohort effects, as well as the Gompertz and gamma-Gompertz models in which only the parameter a varies linearly across cohorts, while the other parameters are constant. The relevant candidate models are shown in Table 4.3, where a model is denoted by M, $M \subseteq \{1, ..., q\}$, if it includes all parameters γ_j with $j \in M$. Mortality deceleration is generally present in those models M that include the parameter σ_0^2. But we have to be careful when drawing conclusions about this phenomenon from the models with cohort-specific frailty variances σ_c^2. If the frailty variances of some of the cohorts are relatively large, while those of other cohorts are close to zero, it may be assumed that mortality deceleration is present for the former, but not for the latter cohorts. The distinct advantage of focused model selection is that we can choose cohort-specific focus parameters, such as the frailty variance $\mu = \sigma_c^2$ of cohort c, which then allow us to assess mortality deceleration separately for each cohort.

Table 4.3: Overview of candidate models in the framework of model (4.8).

Model	M	Parameters			No. of
		a_c	b_c	σ_c^2	param.
Gompertz, plain (null)	$M = \emptyset$	a_0	b_0	0	2
Gompertz with linear trend in a	$M = \{2\}$	$a_0 + a_1 z_{1c}$	b_0	0	3
gamma-Gompertz (GG), plain	$M = \{1\}$	a_0	b_0	σ_0^2	3
GG with linear trend in a	$M = \{1, 2\}$	$a_0 + a_1 z_{1c}$	b_0	σ_0^2	4
GG with cohort-specific σ^2	$M = \{1, 3, ..., q\}$	a_0	b_0	σ_c^2	$2 + C$
GG as in (4.8) (full)	$M = \{1, ..., q\}$	$a_0 + a_1 z_{1c}$	b_0	σ_c^2	$3 + C$

The current selection framework also differs from the previous setting in Section 4.3 from a more technical point of view. Previously, we had to choose whether to include a single parameter with a boundary constraint ($q = 1$ and $\gamma \geq \gamma_0$). Here, we have to choose whether to include none, some, or all of the components of the parameter vector $\boldsymbol{\gamma}$ where only the first component is subject to a boundary constraint ($q > 1$ and $\gamma_1 \geq \gamma_{01}$). Consequently, the results from Section 4.3.2 do not apply in the current setting, and new formulas for the $\text{FIC}_{\text{MAE}}(M)$ for each of the candidate models M have to be derived. We refer the interested reader to Section 4.7.8 for these derivations of the limiting L_1-risks of the focus estimators.

To better evaluate the approach of focused model selection in the setting with cohort effects, we seek to compare model selection based on the new FIC_{MAE} with that based on the AIC. However, due to the presence of the boundary parameter σ_0^2 in some of the candidate models, the standard AIC is again found to be biased. As in the case of the modified AIC* for the plain gamma-Gompertz model presented in Section 4.3.4, we derived a modified AIC* with a bias correction term for the models listed in Table 4.3 (see Section 4.7.9 for a sketch of the derivations).

Finally, we apply the two model selection strategies to the five female cohorts in the French-Canadian data. We choose between the candidate models of Table 4.3 based on the AIC*, and based on the FIC$_{\text{MAE}}$ with cohort-specific focus parameters $\mu = \sigma_c^2$ and $\mu = [\ln h_c(100)]''$, where $h_c(y)$ is the gamma-Gompertz hazard (4.2) with parameters a_c, b_c, and σ_c^2 according to (4.8). The selected models are presented in Table 4.4.

Table 4.4: Selected models for data on French-Canadian females.

Multi-year	FIC$_{\text{MAE}}$		AIC*
birth cohort c	$\mu = \sigma_c^2$	$\mu = [\ln h_c(100)]''$	
1880-1884	gamma-Gompertz	full	
1885-1888	gamma-Gompertz	full	
1889-1891	gamma-Gompertz	gamma-Gompertz	full
1892-1894	Gompertz	Gompertz	
1895-1896	Gompertz	Gompertz, linear trend in a	

While the AIC* selects a single best model irrespective of whether the interest is in one particular cohort, the cohort-specific focus parameters for the FIC$_{\text{MAE}}$ make a clear differentiation, and may select different models for different cohorts of interest. More specifically, based on the AIC* the full gamma-Gompertz model (4.8) with a linear trend in a, constant b, and cohort-specific frailty variances σ_c^2 is selected. By contrast, based on the FIC$_{\text{MAE}}$, models with mortality deceleration are selected only for the first three cohorts, whereas models without mortality deceleration are selected for the two most recent cohorts. These selection results seem plausible given the parameter estimates reported in Table 4.2 and the mortality rates displayed in Figure 4.5. (The reader might notice that for the 1892-1894 cohort, the Gompertz model without mortality deceleration is selected based on the FIC$_{\text{MAE}}$, although the estimated frailty variance $\hat{\sigma}_4^2$ is larger than that for the pooled cohorts for which the gamma-Gompertz model was selected. Here, we have to keep in mind the smaller sample size of the cohort, which affects the performance of the FIC$_{\text{MAE}}$.)

Furthermore, the selection results nicely illustrate that based on the FIC$_{\text{MAE}}$, different models might be favored for the same cohort if different focus parameters are chosen. If the interest lies in the cohort's frailty variance, the plain gamma-Gompertz model is selected for the first two cohorts; but if the interest lies in the second derivative of the cohort's log-hazard, the full model is selected for these two cohorts. This is reasonable because the estimated frailty variances of the first two cohorts are close ($\hat{\sigma}_1^2 = 0.1089$ and $\hat{\sigma}_2^2 = 0.0989$), which suggests that cohort-specific deviations from the frailty variance would not markedly improve the estimator performance regarding the frailty variances σ_1^2 and σ_2^2. In contrast, all gamma-Gompertz parameters are involved in the estimation of the log-hazard, such that an estimator based on the full model with cohort-specific frailty variances and a trend in a, which consequently takes into account the higher initial mortality a_c in the first two cohorts, would, on average, perform better.

4.6   Discussion

Motivated by the issue of how mortality deceleration can be assessed at high ages, we have extended the FIC, as introduced by Claeskens and Hjort (2003), to a non-standard setting in which we are choosing between two models that differ by one parameter that takes a value on the boundary of the parameter space if the smaller model is the true model. We considered two versions of the FIC that aim to minimize the limiting MAE or MSE of the estimator of the focus, respectively. When targeting the MAE, we obtained the new model selection criterion FIC_{MAE}. When targeting the MSE, the model selection does not depend on the chosen focus, but a pre-test strategy was defined. In addition, we presented the new AIC*, which reduces the bias of the original AIC that occurs when the selection concerns a parameter that lies on the boundary of the parameter space in the narrow model.

The proposed model selection criteria provide new tools for the assessment of mortality deceleration in the framework of the gamma-Gompertz model. While traditional approaches either have low power to detect mortality deceleration or are not valid in the presence of boundary-constrained parameters, the methods developed here are adapted to the non-standard setting. An advantage of the FIC_{MAE} is that, by choosing an appropriate focus parameter, it can be targeted directly at the quantities that reveal mortality deceleration. We recommend using as the focus parameter the frailty variance or the second derivative of the log-hazard at some advanced age. Both potential choices readily translate into the presence or the absence of mortality deceleration, as the focus parameter takes a value of zero if there is no deceleration.

The results of our simulation studies indicate that the FIC_{MAE}, especially with the recommended choices of the focus parameter, outperforms the competing approaches of the pre-test and the AIC* in detecting mortality deceleration. This observation was made for different magnitudes of the frailty variance, and with different sample sizes. We found that the FIC_{MAE} performs substantially better than the AIC*, particularly for small samples.

Mortality deceleration addresses properties of the tail of the lifespan distribution, and inference about the tail behavior requires sufficiently large samples. Moreover, it is equally important that the models are sufficiently parsimonious so that the question of interest can be isolated in a few parameters that inference can center on. While the size of our data set on French-Canadian Catholics born at the end of the 19th century is relatively large, the male sample contains only half as many observations as the female sample. In contrast to the other methods, the FIC_{MAE} detected mortality deceleration in the sample of females. However, all of the methods point towards a Gompertz model, and hence towards no mortality deceleration for the French-Canadian males.

As we demonstrated, focused model selection allows us to include additional information, such as differences across birth cohorts. At the same time, however, the number of candidate models and the demands for data tend to increase. The analysis of the female sample with cohort effects pinpointed the differences between traditional (AIC*)

and focused criteria. The within-cohort heterogeneity was estimated to decline, and the gamma-Gompertz model was selected only for the three earlier cohorts.

Data on mortality at advanced ages are often limited due to the complexity of age validation procedures or the availability of the historic documents required for verifying alleged ages at death. As the ability to detect features that are contained in the distribution tail depends on having sufficient information in the data, it is worth investigating how our sampling and validation efforts can be optimized. This includes reconsidering both the required sample sizes and the age range over which data are collected. Here, we were limited to ages at death above 90, which is a rather high starting age compared to the starting ages in other studies. Thus, if further data collection efforts are planned, it may be more advantageous to extend the age range to lower ages. Böhnstedt et al. (2021) have offered some tools to address such questions.

While the set-up in this article was restricted to individual-level data, many studies on aging rely on aggregated data in which death counts and exposure times are available for given age-intervals. However, an extension of the approach to aggregated data is straightforward if we keep the assumption of the parametric model. Consequently, the new tools for the assessment of mortality deceleration presented here will be applicable to a variety of data sets collected for different human and non-human populations. For data on humans, the application of the Gompertz hazard is well-studied and well-established, both across time and across populations. For data on non-human species, we might want to consider relaxing the assumption of a parametric model for the hazard. More research is needed to understand how more flexible hazard shapes can be incorporated by, for example, using splines and penalized likelihood.

Although our development of the FIC_{MAE} was motivated by the specific problem of assessing mortality deceleration, the method could be used in a range of other contexts in which there is a need to choose between parametric models that differ only by one parameter with a boundary constraint, such as when assessing heterogeneity in other proportional hazards frailty models, or when choosing between a Poisson model and an over-dispersed negative binomial model. Linear mixed models are another model class where some parameters, in that case variance components, are restricted to be non-negative and where a focused search is useful (Cunen et al., 2020).

4.7   Supplementary material

4.7.1   Computational issues

The maximum likelihood estimates of the parameters of the gamma-Gompertz model are determined via numerical optimization in R (R Core Team, 2018). Non-negativity of the parameter estimates is achieved by maximizing the log-likelihood over the log-transformed parameters. Nevertheless, values of the frailty variance σ^2 that are close to zero cause numerical difficulties. Here, we briefly describe the steps that we took to increase the numerical stability of the estimation problem.

We maximize the log-likelihood via the R-function `nlm()`, where we can also supply the analytic gradient of the objective function. In addition, Taylor expansions of the log-likelihood and its gradient are used if the current value of σ^2 is smaller than 10^{-5}. The numerically identified maximum $\hat{\sigma}^2$ might still depend on the starting value that was provided to the optimization routine. We therefore recommend running the optimization with a number of different starting values for the frailty variance and choosing the fit with the largest value of the log-likelihood as the final estimate.

For calculating the $\mathrm{FIC_{MAE}}$ values, we need estimates not only of the model parameters, but also of the information matrix J_{full}. For that purpose, we analytically derive the matrix $H(a, b, \sigma^2)$ of second-order derivatives of the log-likelihood for the gamma-Gompertz model, and again use a Taylor expansion if $\sigma^2 < 10^{-5}$. J_{full} is then estimated as $-n^{-1}H(\hat{a}, \hat{b}, \hat{\sigma}^2)$; and $\hat{\kappa}^2$ is the bottom right element of its inverse.

4.7.2   Power of the likelihood ratio test

A likelihood ratio test (LRT) for homogeneity in the gamma-Gompertz model, where $H_0 : \sigma^2 = 0$ and $H_1 : \sigma^2 > 0$, may have low power to detect mortality deceleration. To illustrate this property, we summarize the results for two of the scenarios that were described in Section 4.4. In particular, we show the extent to which a smaller underlying frailty variance or a smaller sample size can decrease the power of the test, which is performed at a significance level of 5%. We also compare the power of the LRT in a situation in which only individuals who survived beyond age 90 can be studied to a situation in which observations for individuals who survived beyond age 80 or 85 are available.

Figure 4.6 illustrates how strongly the power of the test is affected by the three features. The left panel displays the results for Scenario S_1 (frailty variance $\sigma^2 = 0.0625$), while the right panel shows the results for Scenario S_2 in which the frailty variance was roughly halved ($\sigma^2 = 0.03$). Within each panel, we can see the loss in power that occurs if only individuals who survived beyond age 90 (90+) can be studied, instead of individuals who survived beyond age 80 (80+) or 85 (85+). For example, in the medium-sized Scenario S_1, the power of the LRT decreases by more than 45% if the test is based on the 90+ subset instead of on the 85+ subset.
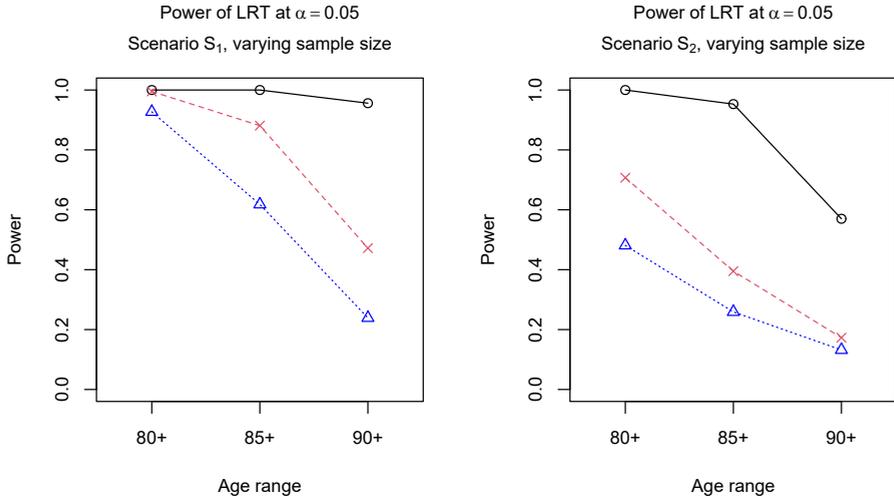
Figure 4.6: Power of the LRT at the 5% level to detect mortality deceleration in the gamma-Gompertz model depending on the age range of the data (left to right: 80+, 85+, or 90+). The depicted scenarios are S_1 (left) and S_2 (right) with the sample sizes $n_{90+} = 10,000$ (blue-dotted-triangle), $n_{90+} = 20,000$ (red-dashed-cross), and $n_{90+} = 105,000$ (black-solid-circle).

4.7.3   Local power of the LRT and the pre-test

The local power of the LRT for homogeneity in the gamma-Gompertz model is derived in Böhnstedt and Gampe (2019). Under the sequence of local alternatives (4.3), the power of the LRT for $H_0: \sigma^2 = 0$ at level α based on a gamma-Gompertz sample of size n can be approximated by

$$1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\delta}{\kappa}\right) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\sqrt{n}\sigma^2}{\kappa}\right). \tag{4.9}$$

The pre-test derived in Section 4.3.2 selects the gamma-Gompertz model if $\hat{\delta}/\hat{\kappa} > 0.8399$. Due to $\hat{\delta}/\hat{\kappa} \xrightarrow{d} \max(0, D/\kappa)$, we have $P\left[\hat{\delta}/\hat{\kappa} \leq z\right] \approx \Phi(z - \delta/\kappa) \mathbb{1}_{\{z \geq 0\}}$. As a consequence, the power of the pre-test with critical region $\hat{\delta}/\hat{\kappa} > 0.8399$ is determined as

$$P\left[\frac{\hat{\delta}}{\hat{\kappa}} > 0.8399 \Big| \text{fixed } \delta\right] \approx 1 - \Phi\left(0.8399 - \frac{\delta}{\kappa}\right). \tag{4.10}$$

Comparing (4.10) and (4.9), we find that for large samples, the pre-test has approximately the same power as the LRT for $H_0: \sigma^2 = 0$ at level $\tilde{\alpha}$ satisfying $\Phi^{-1}(1 - \tilde{\alpha}) = 0.8399$, which is $\tilde{\alpha} = 1 - \Phi(0.8399) \approx 0.2005$.

4.7.4 Derivation of FIC$_{MAE}$ with a single additional parameter on the boundary of the parameter space

The FIC$_{MAE}$ of a model with focus estimator $\hat{\mu}$, where $\sqrt{n}(\hat{\mu} - \mu_{\text{true}}) \xrightarrow{d} \Lambda$, is derived as an estimate of $\mathbb{E}[|\Lambda|]$. For the null model, $\hat{\mu}_{\text{null}}$ converges to a normal distribution, $\Lambda_{\text{null}} = (\Lambda_0 + \omega\delta) \sim \mathcal{N}(\omega\delta, \tau_0^2)$. Therefore, $\mathbb{E}[|\Lambda_{\text{null}}|]$ is calculated as the expected value of the folded normal random variable $|\Lambda_0 + \omega\delta|$; that is,

$$\mathbb{E}[|\Lambda_{\text{null}}|] = \mathbb{E}[|\Lambda_0 + \omega\delta|] = 2\tau_0\phi\left(\frac{\omega\delta}{\tau_0}\right) + 2\omega\delta\left\{\Phi\left(\frac{\omega\delta}{\tau_0}\right) - \frac{1}{2}\right\}. \qquad (4.11)$$

For the full model, we have $\Lambda_{\text{full}} = \Lambda_0 - \omega(D - \delta)\mathbb{1}_{\{D>0\}} + \omega\delta\mathbb{1}_{\{D\leq0\}}$, with $D \sim \mathcal{N}(\delta, \kappa^2)$ independent of Λ_0, such that

$$\mathbb{E}[|\Lambda_{\text{full}}|] = \mathbb{E}[|\Lambda_{\text{full}}| \mid D \leq 0]\mathrm{P}[D \leq 0] + \mathbb{E}[|\Lambda_{\text{full}}| \mid D > 0]\mathrm{P}[D > 0]$$

$$= \mathbb{E}[|\Lambda_0 + \omega\delta|]\Phi\left(-\frac{\delta}{\kappa}\right) + \mathbb{E}[|\Lambda_0 - \omega(D - \delta)| \mid D > 0]\Phi\left(\frac{\delta}{\kappa}\right). \qquad (4.12)$$

The first expectation is the same as (4.11). For the computation of the second expectation, we define the normally distributed random vector $X = (\Lambda_0, D)^\top$ and its affine transformation $Y = (\Lambda_0 - \omega(D - \delta), D)^\top$, which is also normally distributed, with mean vectors $\mu_X = \mu_Y = (0, \delta)^\top$ and covariance matrices

$$\mathrm{Cov}[X] = \begin{pmatrix} \tau_0^2 & 0 \\ 0 & \kappa^2 \end{pmatrix} \quad \text{and} \quad \mathrm{Cov}[Y] = \begin{pmatrix} \tau_0^2 + \omega^2\kappa^2 & -\omega\kappa^2 \\ -\omega\kappa^2 & \kappa^2 \end{pmatrix}.$$

Then, $\mathbb{E}[|\Lambda_0 - \omega(D - \delta)| \mid D > 0]$ can be rewritten as

$$\mathbb{E}[|Y_1| \mid Y_2 > 0] = \mathbb{E}[Y_1 \mid Y_1 > 0, Y_2 > 0]\frac{\mathrm{P}[Y_1 > 0, Y_2 > 0]}{\mathrm{P}[Y_2 > 0]}$$

$$+ \mathbb{E}[-Y_1 \mid -Y_1 \geq 0, Y_2 > 0]\frac{\mathrm{P}[Y_1 \leq 0, Y_2 > 0]}{\mathrm{P}[Y_2 > 0]}. \qquad (4.13)$$

The expected values of one component of a bivariate truncated normal distribution are more easily found for bivariate normal distributions with zero mean vectors, unit variances, and possible correlations. Transforming Y into such a normally distributed random vector $Z = ((\tau_0^2 + \omega^2\kappa^2)^{-1/2}Y_1, (Y_2 - \delta)/\kappa)^\top$ with covariances $-\omega\kappa/\sqrt{\tau_0^2 + \omega^2\kappa^2}$, and noting that

$$\mathbb{E}[Y_1 \mid Y_1 > 0, Y_2 > 0] = \sqrt{\tau_0^2 + \omega^2\kappa^2}\, \mathbb{E}\left[Z_1 \Big| Z_1 > 0, Z_2 > -\frac{\delta}{\kappa}\right],$$

we can apply the results of Tallis (1961) to obtain $\mathbb{E}[Y_1 \mid Y_1 > 0, Y_2 > 0]$ in (4.13) as

$$\frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\mathrm{P}\left[Y_1 > 0, Y_2 > 0\right]} \left\{ \frac{1}{\sqrt{2\pi}} \Phi\left(\frac{\delta}{\kappa} \cdot \frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\tau_0} \right) - \frac{\omega \kappa}{\sqrt{\tau_0^2 + \omega^2 \kappa^2}} \, \phi\left(\frac{\delta}{\kappa} \right) \Phi\left(\frac{\omega \delta}{\tau_0} \right) \right\}.$$

Analogously, $\mathbb{E}\left[-Y_1 \mid -Y_1 \geq 0, Y_2 > 0\right]$ in (4.13) is computed as

$$\frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\mathrm{P}\left[Y_1 \leq 0, Y_2 > 0\right]} \left\{ \frac{1}{\sqrt{2\pi}} \Phi\left(\frac{\delta}{\kappa} \cdot \frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\tau_0} \right) + \frac{\omega \kappa}{\sqrt{\tau_0^2 + \omega^2 \kappa^2}} \, \phi\left(\frac{\delta}{\kappa} \right) \Phi\left(-\frac{\omega \delta}{\tau_0} \right) \right\}.$$

Combining these two results, we find that $\mathbb{E}\left[\left|\Lambda_0 - \omega(D - \delta)\right| \mid D > 0\right]$ in (4.12) is equal to

$$\frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\Phi\left(\frac{\delta}{\kappa}\right)} \cdot \sqrt{\frac{2}{\pi}} \Phi\left(\frac{\delta}{\kappa} \cdot \frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\tau_0} \right) - \frac{\omega \kappa}{\Phi\left(\frac{\delta}{\kappa}\right)} \, \phi\left(\frac{\delta}{\kappa} \right) \cdot 2 \left\{ \Phi\left(\frac{\omega \delta}{\tau_0} \right) - \frac{1}{2} \right\}. \quad (4.14)$$

Inserting (4.11) and (4.14) into (4.12) yields the postulated result

$$\mathbb{E}\left[\left|\Lambda_{\mathrm{full}}\right|\right] = \left[2\tau_0 \phi\left(\frac{\omega \delta}{\tau_0} \right) + 2\omega \delta \left\{ \Phi\left(\frac{\omega \delta}{\tau_0} \right) - \frac{1}{2} \right\} \right] \left\{ 1 - \Phi\left(\frac{\delta}{\kappa} \right) \right\}$$

$$+ \sqrt{\tau_0^2 + \omega^2 \kappa^2} \cdot \sqrt{\frac{2}{\pi}} \Phi\left(\frac{\delta}{\kappa} \cdot \frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\tau_0} \right) - \omega \kappa \, \phi\left(\frac{\delta}{\kappa} \right) \cdot 2 \left\{ \Phi\left(\frac{\omega \delta}{\tau_0} \right) - \frac{1}{2} \right\}.$$

4.7.5   Other focus parameters

In our simulation studies, we also assessed the performance of the FIC_{MAE} for several other focus parameters, such as quantiles of the survival distribution or the log-hazard and the survival function at different advanced ages. Overall, the frailty variance, $\mu = \sigma^2$, and the second derivative of the log-hazard, $\mu = [\ln h(y)]''$, yielded the best results. Figure 4.7 illustrates the proportion of decisions in favor of the gamma-Gompertz model in several settings, when the focus is placed on the second derivative of the log-hazard at age 100, the log-hazard at age 100 or 110, or the survival function at age 100. While the choice of $\mu = [\ln h(100)]''$ again results in the highest proportion of decisions in favor of the gamma-Gompertz model, the choice of $\mu = S(100)$ performs almost as well. When the focus is put on the log-hazard, the age at which the function is evaluated apparently makes a difference, in that $\mu = \ln h(110)$ leads to a better performance of the FIC_{MAE} than $\mu = \ln h(100)$. However, survival beyond age 110 is relatively rare in some of our simulated settings, and we should be careful when putting the focus on ages for which there are too few data points.
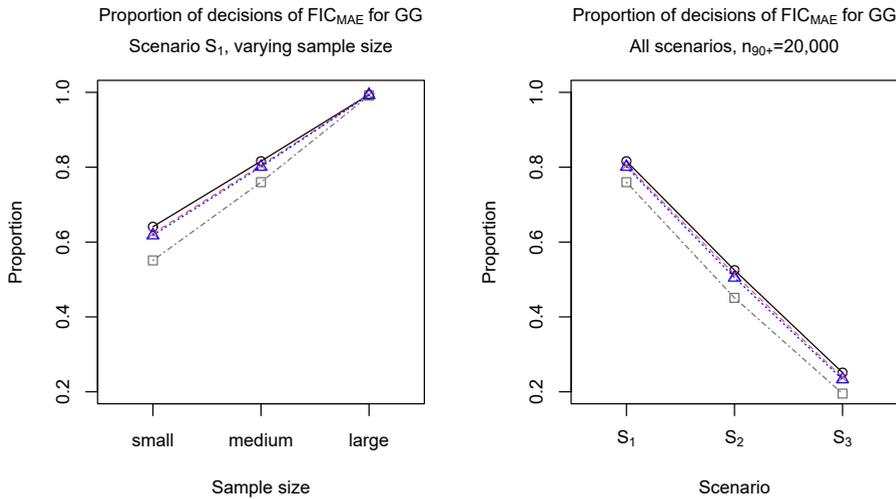
Figure 4.7: Proportion of decisions in favor of the gamma-Gompertz model based on FIC_{MAE} with $\mu = [\ln h(100)]''$ (black-solid-circle), $\mu = \ln h(100)$ (gray-dot-dashed-square), $\mu = \ln h(110)$ (red-dashed-cross), and $\mu = S(100)$ (blue-dotted-triangle). Left: Decisions in Scenario S_1 for sample sizes $n_{90+} = 10,000$, $n_{90+} = 20,000$, and $n_{90+} = 105,000$ (left to right). Right: Decisions in Scenarios S_1, S_2, and S_3 (left to right) all with $n_{90+} = 20,000$.

4.7.6   Impact of the age range on the performance of the FIC_{MAE}

So far, we have, motivated by the data application, considered only samples of individuals who survived beyond age 90. However, the amount of heterogeneity in mortality risk within the population decreases with age due to selection. Therefore, it is of interest to study the performance of the FIC_{MAE} according to the age range of the sample. Figure 4.8 depicts the proportion of correct decisions in favor of the gamma-Gompertz model based on the FIC_{MAE} with $\mu = [\ln h(100)]''$ in different settings when the sample consisted of all individuals who had reached at least age 80, 85, or 90. We see that, in general, the probability of detecting mortality deceleration increases if the sample covers a wider age range. For Scenario S_1 with the target sample size of $n_{90+} = 10{,}000$, the proportion of correct decisions increases by more than a third if we observe all individuals who had reached at least age 85 instead of only those individuals who had reached at least age 90. Both the larger sample size of the 85+ subset and the greater amount of heterogeneity in the mortality risk of this subset played a part in this result.
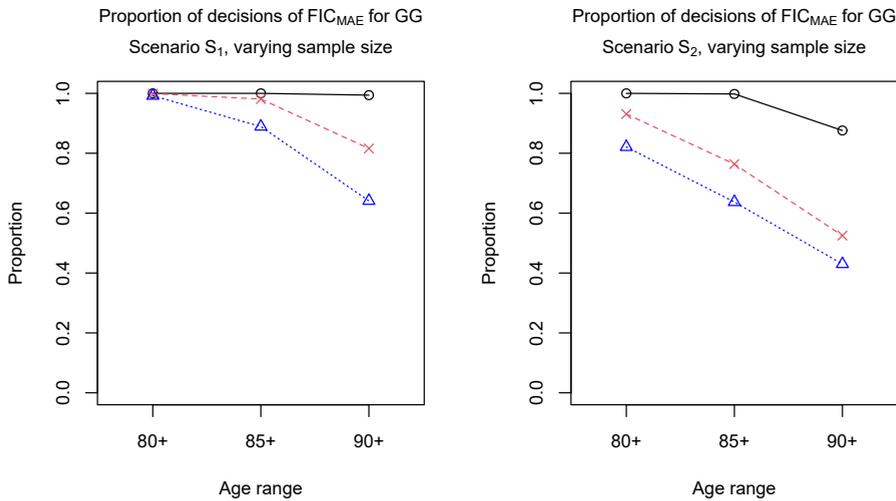
Figure 4.8: Proportion of correct decisions in favor of the gamma-Gompertz model based on the FIC_{MAE} with $\mu = [\ln h(100)]''$ depending on the age range of the data (left to right: 80+, 85+, or 90+). The depicted scenarios are S_1 (left) and S_2 (right) with sample sizes $n_{90+} = 10{,}000$ (blue-dotted-triangle), $n_{90+} = 20{,}000$ (red-dashed-cross), and $n_{90+} = 105{,}000$ (black-solid-circle).

4.7.7   Preliminary analysis of CHMD data
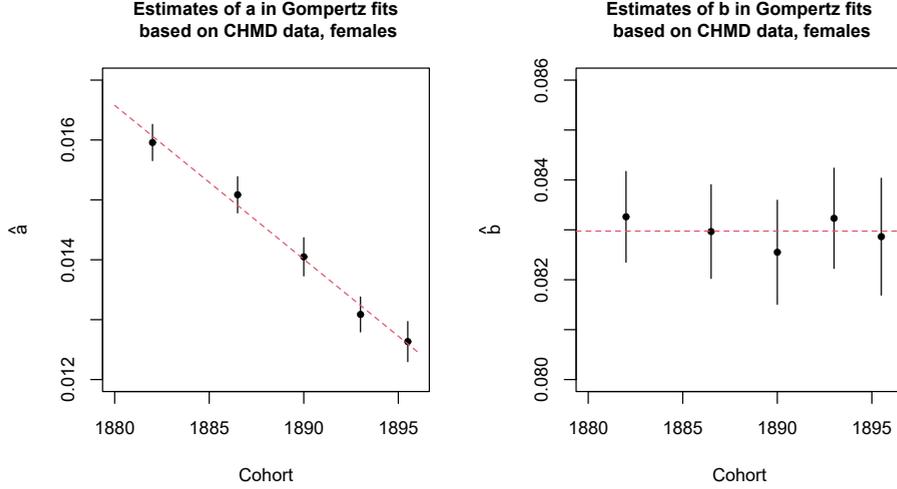
Figure 4.9: Estimates of Gompertz parameters a (left) and b (right) with 95% confidence intervals based on data for Quebec females between ages 60 and 109 from the CHMD. Fitted linear and constant trend, respectively, is marked by red, dashed line.

4.7.8   Derivation of FIC_{MAE} for the model with cohort effects

We derive formulas for the FIC_{MAE} under the framework of local misspecification (4.3) for the setting that the parameter vector γ has dimension $q > 1$, but only its first component is subject to a boundary constraint, $\gamma_1 \geq \gamma_{01}$.

A candidate model M includes the components γ_j for $j \in M$, while the remaining components, that is, γ_j for $j \in M^c = \{1, ..., q\} \backslash M$, are fixed at the respective values in γ_0. Thus, the focus estimator for model M is $\hat{\mu}_M = \mu(\hat{\theta}_M, \hat{\gamma}_M, \gamma_{0,M^c})$ with the maximum likelihood estimator $(\hat{\theta}_M^\top, \hat{\gamma}_M^\top)^\top$ for model M. The $\text{FIC}_{\text{MAE}}(M)$ of a model M is, as before, determined as an estimate of $\mathbb{E}[|\Lambda_M|]$, where $\sqrt{n}(\hat{\mu}_M - \mu_{\text{true}}) \xrightarrow{d} \Lambda_M$. The limiting results will again involve Λ_0 and ω, defined in Section 4.3.2, where ω is now a vector of length q. Similarly, we now consider normal random vectors

$$\boldsymbol{D} \sim \mathcal{N}_q(\boldsymbol{\delta}, Q) \text{ and } \boldsymbol{E} = Q^{-1}\boldsymbol{D} \sim \mathcal{N}_q(Q^{-1}\boldsymbol{\delta}, Q^{-1}),$$

independent of Λ_0, and the $q \times q$ lower-right submatrix Q of the inverse information matrix J_{full}^{-1}, which corresponds to γ.

The limiting distribution of focus estimators $\hat{\mu}_M$ under the framework (4.3) for the case that some or all of the components of $\boldsymbol{\gamma}$ are subject to a boundary constraint is given in Theorem 10.2 of Claeskens and Hjort (2008) as

$$\sqrt{n}(\hat{\mu}_M - \mu_{\text{true}}) \xrightarrow{d} \Lambda_M = \Lambda_0 + \boldsymbol{\omega}^\top(\boldsymbol{\delta} - \pi_M^\top \hat{\boldsymbol{t}}_M),$$

where $\hat{\boldsymbol{t}}_M$ is the random maximizer of $\boldsymbol{E}_M^\top \boldsymbol{t} - \frac{1}{2}\boldsymbol{t}^\top Q_M^{-1}\boldsymbol{t}$ over all $\boldsymbol{t} \in \Omega_M$. Here, the set $\Omega_M \subset \mathbb{R}^{|M|}$ describes the parameter space of $(\boldsymbol{\gamma}_M - \boldsymbol{\gamma}_{0,M})$ and π_M is a projection matrix of dimension $|M| \times q$ mapping a vector $\boldsymbol{v} = (v_1, v_2, ..., v_q)^\top$ onto the vector $\boldsymbol{v}_M$ that contains the $|M|$ components v_j for $j \in M$. In particular, we have $\boldsymbol{E}_M = \pi_M \boldsymbol{E} \sim \mathcal{N}_{|M|}(\pi_M Q^{-1}\boldsymbol{\delta}, Q_M^{-1})$ and $Q_M = (\pi_M Q^{-1}\pi_M^\top)^{-1}$.

In the following, we will consider different types of candidate models M and determine the limiting risk $\mathbb{E}[|\Lambda_M|]$ based on the specific Ω_M and the resulting $\hat{\boldsymbol{t}}_M$. The FIC$_{\text{MAE}}$ is obtained by replacing the unknowns $\boldsymbol{\delta}$, $\boldsymbol{\omega}$, τ_0^2, and Q in the formulas for $\mathbb{E}[|\Lambda_M|]$ by their estimates.

For some choices of the focus parameter μ, the derivations given below need to be adapted. If, for example, $\mu = \sigma_c^2$, it follows that $\frac{\partial \mu}{\partial \theta} = \boldsymbol{0}$, such that $\tau_0^2 = 0$ and the variable Λ_0 is deterministic.

$M = \emptyset$

If $M = \emptyset$, the focus estimator $\hat{\mu}_{\text{null}}$ converges to a normal distribution, $\Lambda_M = (\Lambda_0 + \boldsymbol{\omega}^\top\boldsymbol{\delta}) \sim \mathcal{N}(\boldsymbol{\omega}^\top\boldsymbol{\delta}, \tau_0^2)$. Hence, $\mathbb{E}[|\Lambda_M|]$ is calculated as the expected value of the folded normal random variable $|\Lambda_0 + \boldsymbol{\omega}^\top\boldsymbol{\delta}|$, so

$$\mathbb{E}[|\Lambda_\emptyset|] = 2\tau_0\phi\left(\frac{\boldsymbol{\omega}^\top\boldsymbol{\delta}}{\tau_0}\right) + 2\boldsymbol{\omega}^\top\boldsymbol{\delta}\left[\Phi\left(\frac{\boldsymbol{\omega}^\top\boldsymbol{\delta}}{\tau_0}\right) - \frac{1}{2}\right]. \tag{4.15}$$

$M \not\ni 1$

If $1 \notin M$, the parameter vector $\boldsymbol{\gamma}_M$ is not subject to boundary constraints, such that $\Omega_M = \mathbb{R}^{|M|}$ and $\hat{\boldsymbol{t}}_M = Q_M \boldsymbol{E}_M$. The limiting variable $\Lambda_M = \Lambda_0 + \boldsymbol{\omega}^\top(\boldsymbol{\delta} - \pi_M^\top Q_M \boldsymbol{E}_M)$ is a linear transformation of the normal random vector $(\Lambda_0, (Q_M \boldsymbol{E}_M)^\top)^\top$ and is therefore normally distributed with mean $m_M = \boldsymbol{\omega}^\top(I - V_M Q^{-1})\boldsymbol{\delta}$ and variance $\tau_M^2 = \tau_0^2 + \boldsymbol{\omega}^\top V_M \boldsymbol{\omega}$, where $V_M = \pi_M^\top Q_M \pi_M$. Thus, the limiting L_1-risk of Λ_M is again found as the mean of a folded normal distribution,

$$\mathbb{E}[|\Lambda_M|] = 2\tau_M\phi\left(\frac{m_M}{\tau_M}\right) + 2m_M\left[\Phi\left(\frac{m_M}{\tau_M}\right) - \frac{1}{2}\right].$$

$M = \{1\}$

If $M = \{1\}$, that is, the model M includes only the boundary parameter γ_1, the parameter space is restricted to the non-negative real numbers, $\Omega_{\{1\}} = \mathbb{R}_0^+$. Therefore, we have $\hat{t}_M = 0 \cdot \mathbb{1}_{\{Q_M E_M \leq 0\}} + Q_M E_M \cdot \mathbb{1}_{\{Q_M E_M > 0\}}$. The limiting variable is then given by

$$\Lambda_{\{1\}} = \Lambda_0 + \boldsymbol{\omega}^\top \boldsymbol{\delta} \cdot \mathbb{1}_{\{Q_M E_M \leq 0\}} + \boldsymbol{\omega}^\top (\boldsymbol{\delta} - \pi_M^\top Q_M E_M) \cdot \mathbb{1}_{\{Q_M E_M > 0\}},$$

and the risk $\mathbb{E}[|\Lambda_{\{1\}}|]$ can be calculated as

$$\mathbb{E}[|\Lambda_{\{1\}}| \mid Q_M E_M \leq 0]\mathrm{P}[Q_M E_M \leq 0] + \mathbb{E}[|\Lambda_{\{1\}}| \mid Q_M E_M > 0]\mathrm{P}[Q_M E_M > 0]$$
$$= \mathbb{E}[|\Lambda_0 + \boldsymbol{\omega}^\top \boldsymbol{\delta}|]\mathrm{P}[Q_M E_M \leq 0]$$
$$+ \mathbb{E}[|\Lambda_0 + \boldsymbol{\omega}^\top (\boldsymbol{\delta} - \pi_M^\top Q_M E_M)| \mid Q_M E_M > 0]\mathrm{P}[Q_M E_M > 0].$$

The first expectation was given in (4.15) and the computation of the second expectation follows along the lines of the computation of the second expectation in (4.12), see Section 4.7.4. We consider the normal random vector $\boldsymbol{Y} = (\Lambda_0 + \boldsymbol{\omega}^\top (\boldsymbol{\delta} - \pi_M^\top Q_M E_M),$ $Q_M E_M)^\top$, which is a linear transformation of $\boldsymbol{X} = (\Lambda_0, Q_M E_M)^\top$, and note that $\mathbb{E}[|\Lambda_0 + \boldsymbol{\omega}^\top (\boldsymbol{\delta} - \pi_M^\top Q_M E_M)| \mid Q_M E_M > 0]$ equals $\mathbb{E}[|Y_1| \mid Y_2 > 0]$ and thus,

$$\mathbb{E}[Y_1 \mid Y_1 > 0, Y_2 > 0]\mathrm{P}[Y_1 > 0 \mid Y_2 > 0] + \mathbb{E}[-Y_1 \mid -Y_1 > 0, Y_2 > 0]\mathrm{P}[Y_1 \leq 0 \mid Y_2 > 0].$$

In order to apply the results of Tallis (1961) on the expected values of the components of bivariate truncated normal distributions with zero mean vector and unit variances, we work with the transformed $\boldsymbol{Z}$ with components $Z_k = \tau_{Y_k}^{-1}(Y_k - m_{Y_k})$, where $m_{Y_k} = \mathbb{E}[Y_k]$ and $\tau_{Y_k}^2 = \mathrm{var}[Y_k]$, $k = 1, 2$. This yields

$$\mathbb{E}[Y_1 \mid Y_1 > 0, Y_2 > 0] = \tau_{Y_1}\mathbb{E}[Z_1 \mid Z_1 > u_1, Z_2 > u_2] + m_{Y_1}$$
$$= \tau_{Y_1} \frac{\phi(u_1)\Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) + \rho\phi(u_2)\Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1 - \rho^2}}\right)}{\mathrm{P}[Y_1 > 0, Y_2 > 0]} + m_{Y_1},$$

where $u_k = -m_{Y_k}/\tau_{Y_k}$, $\rho = \rho_Y \tau_{Y_1}^{-1}\tau_{Y_2}^{-1}$, and $\rho_Y = \mathrm{cov}(Y_1, Y_2)$, and, analogously,

$$\mathbb{E}[-Y_1 \mid -Y_1 > 0, Y_2 > 0] = \tau_{Y_1} \frac{\phi(-u_1)\Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) - \rho\phi(u_2)\Phi\left(\frac{u_1 - \rho u_2}{\sqrt{1 - \rho^2}}\right)}{\mathrm{P}[Y_1 \leq 0, Y_2 > 0]} - m_{Y_1}.$$

Combining these two results and using the symmetry of the normal distribution as well as that $\mathrm{P}[Y_1 \leq 0 \mid Y_2 > 0] = 1 - \mathrm{P}[Y_1 > 0 \mid Y_2 > 0]$, we find that $\mathbb{E}[|\Lambda_0 + \boldsymbol{\omega}^\top (\boldsymbol{\delta} - \pi_M^\top Q_M E_M)| \mid Q_M E_M > 0]$ is equal to

$$\frac{\tau_{Y_1}}{\mathrm{P}[Y_2 > 0]}\left\{2\phi(u_1)\Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) + \rho\phi(u_2)\left[2\Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1 - \rho^2}}\right) - 1\right]\right\}$$
$$+ m_{Y_1}\left(2\mathrm{P}[Y_1 > 0 \mid Y_2 > 0] - 1\right).$$

Consequently, the limiting risk $\mathbb{E}[|\Lambda_{\{1\}}|]$ is found to be

$$\mathbb{E}[|\Lambda_{\{1\}}|] = \left\{ 2\tau_0 \phi\left(\frac{\omega^\top \delta}{\tau_0}\right) + 2\omega^\top \delta \left[\Phi\left(\frac{\omega^\top \delta}{\tau_0}\right) - \frac{1}{2}\right] \right\} P[Y_2 \le 0]$$

$$+ 2\tau_{Y_1} \left\{ \phi(u_1)\Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1-\rho^2}}\right) + \rho\phi(u_2)\left[\Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1-\rho^2}}\right) - \frac{1}{2}\right] \right\}$$

$$+ 2m_{Y_1}\left(P[Y_1 > 0 \mid Y_2 > 0] - \frac{1}{2}\right) P[Y_2 > 0].$$

$M \supsetneq \{1\}$

In the last case, the model M includes the boundary parameter γ_1, but also some or all of the remaining components of $\boldsymbol{\gamma}$. The maximizer $\hat{\boldsymbol{t}}_M$ of $\boldsymbol{E}_M^\top \boldsymbol{t} - \frac{1}{2}\boldsymbol{t}^\top Q_M^{-1}\boldsymbol{t}$ has to be determined over the parameter space $\Omega_M = \mathbb{R}_0^+ \times \mathbb{R}^{|M|-1}$. If the first component of $Q_M \boldsymbol{E}_M$ is positive, then $\hat{\boldsymbol{t}}_M = Q_M \boldsymbol{E}_M$; but if the first component of $Q_M \boldsymbol{E}_M$ is negative or zero, then $\hat{t}_1$ is set to zero and the remaining components of $\hat{\boldsymbol{t}}_M$ maximize $\boldsymbol{E}_M^\top \boldsymbol{t} - \frac{1}{2}\boldsymbol{t}^\top Q_M^{-1}\boldsymbol{t}$ for $t_1 = 0$. More formally, we have

$$\hat{\boldsymbol{t}}_M = G_M^\top R_M G_M \boldsymbol{E}_M \cdot \mathbb{1}_{\{F_M Q_M \boldsymbol{E}_M \le 0\}} + Q_M \boldsymbol{E}_M \cdot \mathbb{1}_{\{F_M Q_M \boldsymbol{E}_M > 0\}},$$

where we defined a $1 \times |M|$ projection matrix F_M mapping a vector of length $|M|$ onto its first component, and a $(|M|-1) \times |M|$ projection matrix G_M mapping a vector of length $|M|$ onto all but its first component, as well as $R_M = (G_M Q_M^{-1} G_M^\top)^{-1}$. The limiting variable takes the form

$$\Lambda_M = \Lambda_0 + \omega^\top(\boldsymbol{\delta} - \pi_M^\top G_M^\top R_M G_M \boldsymbol{E}_M) \cdot \mathbb{1}_{\{F_M Q_M \boldsymbol{E}_M \le 0\}}$$
$$+ \omega^\top(\boldsymbol{\delta} - \pi_M^\top Q_M \boldsymbol{E}_M) \cdot \mathbb{1}_{\{F_M Q_M \boldsymbol{E}_M > 0\}},$$

such that the limiting risk of the focus estimator $\hat{\mu}_M$ is given by

$$\mathbb{E}[|\Lambda_M|] = \mathbb{E}[|\Lambda_0 + \omega^\top(\boldsymbol{\delta} - \pi_M^\top G_M^\top R_M G_M \boldsymbol{E}_M)| \mid F_M Q_M \boldsymbol{E}_M \le 0]P[F_M Q_M \boldsymbol{E}_M \le 0]$$
$$+ \mathbb{E}[|\Lambda_0 + \omega^\top(\boldsymbol{\delta} - \pi_M^\top Q_M \boldsymbol{E}_M)| \mid F_M Q_M \boldsymbol{E}_M > 0]P[F_M Q_M \boldsymbol{E}_M > 0].$$

For computing the first expectation, we consider the normal random vector

$$\boldsymbol{X} = \begin{pmatrix} \Lambda_0 \\ \boldsymbol{E}_M \end{pmatrix} \sim \mathcal{N}_{1+|M|}\left(\begin{pmatrix} 0 \\ \pi_M Q^{-1}\boldsymbol{\delta} \end{pmatrix}, \begin{pmatrix} \tau_0^2 & 0 \\ 0 & Q_M^{-1} \end{pmatrix}\right),$$

and its linear transformation

$$\boldsymbol{Y} = \begin{pmatrix} \Lambda_0 + \omega^\top(\boldsymbol{\delta} - \pi_M^\top G_M^\top R_M G_M \boldsymbol{E}_M) \\ F_M Q_M \boldsymbol{E}_M \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} m_{Y_1} \\ m_{Y_2} \end{pmatrix}, \begin{pmatrix} \tau_{Y_1}^2 & \rho_Y \\ \rho_Y & \tau_{Y_2}^2 \end{pmatrix}\right).$$

One can show that $\rho_Y = -\omega^\top \pi_M^\top G_M^\top R_M G_M F_M^\top = 0$. Consequently, the components of $\boldsymbol{Y}$ are jointly normally distributed and uncorrelated, and hence, independent. The first

expectation in the expression for $\mathbb{E}[|\Lambda_M|]$ therefore simplifies to $\mathbb{E}[|Y_1|]$, which is the mean of a folded normal random variable, that is,

$$\mathbb{E}[|\Lambda_0 + \boldsymbol{\omega}^\top(\boldsymbol{\delta} - \boldsymbol{\pi}_M^\top G_M^\top R_M G_M \boldsymbol{E}_M)|] = 2\tau_{Y_1}\phi\left(\frac{m_{Y_1}}{\tau_{Y_1}}\right) + 2m_{Y_1}\left[\Phi\left(\frac{m_{Y_1}}{\tau_{Y_1}}\right) - \frac{1}{2}\right].$$

For computing the second expectation, we define another linear transformation of $\boldsymbol{X}$, namely,

$$\boldsymbol{W} = \begin{pmatrix} \Lambda_0 + \boldsymbol{\omega}^\top(\boldsymbol{\delta} - \boldsymbol{\pi}_M^\top Q_M \boldsymbol{E}_M) \\ F_M Q_M \boldsymbol{E}_M \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} m_{W_1} \\ m_{W_2} \end{pmatrix}, \begin{pmatrix} \tau_{W_1}^2 & \rho_W \\ \rho_W & \tau_{W_2}^2 \end{pmatrix}\right),$$

such that

$$\mathbb{E}[|\Lambda_0 + \boldsymbol{\omega}^\top(\boldsymbol{\delta} - \boldsymbol{\pi}_M^\top Q_M \boldsymbol{E}_M)| \mid F_M Q_M \boldsymbol{E}_M > 0] = \mathbb{E}[|W_1| \mid W_2 > 0].$$

Using once again the law of total expectation, the latter can be calculated as

$$\mathbb{E}[W_1 \mid W_1 > 0, W_2 > 0]\mathrm{P}[W_1 > 0 \mid W_2 > 0]$$
$$+ \mathbb{E}[-W_1 \mid -W_1 > 0, W_2 > 0]\mathrm{P}[W_1 \le 0 \mid W_2 > 0].$$

We rewrite these expectations in terms of a centered random vector $\boldsymbol{Z}$ with unit variances, where $Z_k = \tau_{W_k}^{-1}(W_k - m_{W_k})$, $k = 1, 2$, to apply the results of Tallis (1961),

$$\mathbb{E}[W_1 \mid W_1 > 0, W_2 > 0] = \tau_{W_1}\mathbb{E}[Z_1 \mid Z_1 > u_1, Z_2 > u_2] + m_{W_1}$$
$$= \tau_{W_1}\frac{\phi(u_1)\Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1-\rho^2}}\right) + \rho\phi(u_2)\Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1-\rho^2}}\right)}{\mathrm{P}[W_1 > 0, W_2 > 0]} + m_{W_1},$$

with $u_k = -m_{W_k}/\tau_{W_k}$ and $\rho = \rho_W \tau_{W_1}^{-1}\tau_{W_2}^{-1}$, and

$$\mathbb{E}[-W_1 \mid -W_1 > 0, W_2 > 0] = \tau_{W_1}\frac{\phi(u_1)\Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1-\rho^2}}\right) - \rho\phi(u_2)\Phi\left(\frac{u_1 - \rho u_2}{\sqrt{1-\rho^2}}\right)}{\mathrm{P}[W_1 \le 0, W_2 > 0]} - m_{W_1}.$$

Combining these two expressions, we obtain that $\mathbb{E}[|W_1| \mid W_2 > 0]$ equals

$$\frac{\tau_{W_1}}{\mathrm{P}[W_2 > 0]}\left\{2\phi(u_1)\Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1-\rho^2}}\right) + \rho\phi(u_2)\left[2\Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1-\rho^2}}\right) - 1\right]\right\}$$
$$+ m_{W_1}\left(2\mathrm{P}[W_1 > 0 \mid W_2 > 0] - 1\right).$$

Finally, the above results lead to

$$\mathbb{E}[|\Lambda_M|] = \left\{2\tau_{Y_1}\phi\left(\frac{m_{Y_1}}{\tau_{Y_1}}\right) + 2m_{Y_1}\left[\Phi\left(\frac{m_{Y_1}}{\tau_{Y_1}}\right) - \frac{1}{2}\right]\right\}\mathrm{P}[W_2 \le 0]$$

$$+ 2\tau_{W_1}\left\{\phi(u_1)\Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1-\rho^2}}\right) + \rho\phi(u_2)\left[\Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1-\rho^2}}\right) - \frac{1}{2}\right]\right\}$$

$$+ 2m_{W_1}\left(\mathrm{P}[W_1 > 0 \mid W_2 > 0] - \frac{1}{2}\right)\mathrm{P}[W_2 > 0].$$

4.7.9   A modified AIC for the model with cohort effects

In this section, we study the AIC for the (gamma-)Gompertz models with cohort effects which are listed in Table 4.3. We show that the standard AIC is biased as an estimator of the Akaike information for the models that include the boundary parameter σ_0^2 and define a modified AIC* with a bias correction term.

The modified AIC

We work within the framework of local misspecification (4.3) for the setting that the parameter vector γ has dimension $q > 1$, and only its first component is subject to a boundary constraint, $\gamma_1 \ge \gamma_{01}$. The candidate models M always include the full d-dimensional parameter vector θ, but only those components γ_j of γ for which $j \in M$. The standard AIC for such a model M is defined as

$$\mathrm{AIC}_M = -2\ell_M(\hat{\eta}_M; Y) + 2(d + |M|),$$

where the log-likelihood ℓ_M for a sample Y under model M is evaluated at the maximum likelihood estimate under model M, that is, $\hat{\eta}_M = (\hat{\theta}_M^\top, \hat{\gamma}_M^\top)^\top$. Let us denote by π_M a $|M| \times q$ projection matrix mapping a vector $v = (v_1, v_2, ..., v_q)^\top$ onto the vector v_M that contains the $|M|$ components v_j for $j \in M$, and by Q the $q \times q$ lower-right submatrix of the information matrix J_{full}^{-1}, which corresponds to γ, and finally $Q_M = (\pi_M Q^{-1}\pi_M^\top)^{-1}$. We will find that for models M that include the boundary parameter $\gamma_1 = \sigma_0^2$, the standard AIC has bias $2\Phi\left(-[Q_M]_{11}^{-1/2}[Q_M\pi_M Q^{-1}\delta]_1\right)$, where $[Q_M]_{11}$ is the top-left element of Q_M and $[Q_M\pi_M Q^{-1}\delta]_1$ is the first component of $Q_M\pi_M Q^{-1}\delta$. We therefore define a modified version of the AIC for the models M, listed in Table 4.3, as

$$\mathrm{AIC}_M^* = \begin{cases} -2\ell_M(\hat{\eta}_M; Y) + 2(d + |M|), & \text{if } M \not\ni 1, \\ -2\ell_M(\hat{\eta}_M; Y) + 2(d + |M|) - 2\Phi\left(-[\hat{Q}_M]_{11}^{-1/2}[\hat{Q}_M\pi_M \hat{Q}^{-1}\hat{\delta}]_1\right), & \text{if } M \ni 1. \end{cases}$$
$$(4.16)$$

Outline of proof

We give a sketch of the derivations leading to the modified AIC in (4.16) here and provide further details on pp. 101 ff.

The AIC of a model M is derived as an asymptotically unbiased estimator of the expected relative Kullback-Leibler distance

$$-2\,\mathbb{E}_Y\left[\mathbb{E}_X\left[\ln f(X;\hat{\boldsymbol{\eta}}_M(Y))\right]\right], \tag{4.17}$$

which measures the distance between the true underlying distribution from which X and Y are generated and the best parametric approximation $f(\cdot,\hat{\boldsymbol{\eta}}_M)$ (Akaike, 1974). An unbiased estimator of the Akaike information (4.17) is given by

$$-2\,\ell_M(\hat{\boldsymbol{\eta}}_M;Y) + \underbrace{2\,\mathbb{E}_Y\left[\ell_M(\hat{\boldsymbol{\eta}}_M;Y) - \ell_M(\boldsymbol{\eta}_{0,M};Y)\right]}_{=:A_1}$$
$$+ \underbrace{2\,\mathbb{E}_Y\left[\mathbb{E}_X\left[\ell_M(\boldsymbol{\eta}_{0,M};X) - \ell_M(\hat{\boldsymbol{\eta}}_M;X)\right]\right]}_{=:A_2}, \tag{4.18}$$

where the MLE $\hat{\boldsymbol{\eta}}_M = \hat{\boldsymbol{\eta}}_M(Y)$ is based on the sample Y of size n. Using a Taylor expansion of the log-likelihood of model M evaluated at the MLE $\hat{\boldsymbol{\eta}}_M$ about the null point $\boldsymbol{\eta}_{0,M} = (\boldsymbol{\theta}_0^\top, \boldsymbol{\gamma}_{0,M}^\top)^\top$, the term A_2 is found to be asymptotically equivalent to

$$\underbrace{-2\,\mathbb{E}_Y\left[\sqrt{n}(\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M})^\top\right]\begin{pmatrix} J_{01}\boldsymbol{\delta} \\ \pi_M J_{11}\boldsymbol{\delta} \end{pmatrix}}_{=:A_{21}} + \underbrace{\mathbb{E}_Y\left[\sqrt{n}(\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M})^\top J_M \sqrt{n}(\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M})\right]}_{=:A_{22}}.$$
$$\tag{4.19}$$

Here, J_M is the information matrix for model M, with blocks J_{00}, $J_{01,M} = J_{01}\pi_M^\top$, $J_{10,M} = \pi_M J_{10}$, and $J_{11,M} = \pi_M J_{11}\pi_M^\top$. The derivation of result (4.19) involves the limiting distribution of the score vector under model M. Let us denote by $\boldsymbol{U}(y)$ and $\boldsymbol{V}(y)$ the score functions with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ of the log-likelihood of a single observation y from f_{true} in (4.3), where $\boldsymbol{U}$ and $\boldsymbol{V}$ are evaluated at the null model $(\boldsymbol{\theta}_0^\top, \boldsymbol{\gamma}_0^\top)^\top$. For a sample Y of size n, the averaged score vectors are $\bar{\boldsymbol{U}}_n = n^{-1}\sum_{i=1}^n \boldsymbol{U}(Y_i)$ and $\bar{\boldsymbol{V}}_n = n^{-1}\sum_{i=1}^n \boldsymbol{V}(Y_i)$. According to the multivariate central limit theorem, and under the framework (4.3), the score vector of model M converges in distribution to a normal random vector (see Hjort and Claeskens, 2003, for a proof in the regular setting that can be seen to carry over to the current setting),

$$\begin{pmatrix} \sqrt{n}\bar{\boldsymbol{U}}_n \\ \sqrt{n}\bar{\boldsymbol{V}}_{n,M} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} J_{01}\boldsymbol{\delta} \\ \pi_M J_{11}\boldsymbol{\delta} \end{pmatrix} + \begin{pmatrix} \boldsymbol{K} \\ \boldsymbol{N}_M \end{pmatrix}, \qquad \text{with } \begin{pmatrix} \boldsymbol{K} \\ \boldsymbol{N}_M \end{pmatrix} \sim \mathcal{N}_{d+|M|}(\boldsymbol{0}, J_M). \tag{4.20}$$

Further approximations of the expressions A_{21} and A_{22} in (4.19) rely on the limiting distribution of the MLE $\hat{\boldsymbol{\eta}}_M$ under model M. As stated in Theorem 10.2 in Claeskens and Hjort (2008), under the framework (4.3) in case some or all components of $\boldsymbol{\gamma}$ are subject to boundary constraints, the MLE tends to the following limiting variable,

$$\begin{pmatrix} \sqrt{n}(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}_0) \\ \sqrt{n}(\hat{\boldsymbol{\gamma}}_M - \boldsymbol{\gamma}_{0,M}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} J_{00}^{-1}(J_{01}\boldsymbol{\delta} + \boldsymbol{K} - J_{01,M}\hat{\boldsymbol{t}}_M) \\ \hat{\boldsymbol{t}}_M \end{pmatrix}, \tag{4.21}$$

where $\hat{\boldsymbol{t}}_M$ is the random maximizer of $\boldsymbol{E}_M^\top \boldsymbol{t} - \frac{1}{2}\boldsymbol{t}^\top Q_M^{-1}\boldsymbol{t}$ over all $\boldsymbol{t} \in \Omega_M$, and Ω_M is the parameter space of $(\boldsymbol{\gamma}_M - \boldsymbol{\gamma}_{0,M})$. Based on this, we obtain

$$
\begin{aligned}
A_{21} &\approx \boldsymbol{\delta}^\top J_{10} J_{00}^{-1} J_{01} \boldsymbol{\delta} + \mathbb{E}[\hat{\boldsymbol{t}}_M]^\top \pi_M Q^{-1} \boldsymbol{\delta}, \\
A_{22} &\approx \boldsymbol{\delta}^\top J_{10} J_{00}^{-1} J_{01} \boldsymbol{\delta} + \mathbb{E}[\boldsymbol{K}^\top J_{00}^{-1} \boldsymbol{K}] + \mathbb{E}[\hat{\boldsymbol{t}}_M^\top Q_M^{-1} \hat{\boldsymbol{t}}_M].
\end{aligned}
\tag{4.22}
$$

As for a k-dimensional random vector X with $\mathbb{E}[X] = \boldsymbol{m}$, $\text{Cov}[X] = \Sigma$, and a constant $k \times k$ symmetric matrix B, it holds that

$$
\mathbb{E}[X^\top B X] = \text{tr}(B\Sigma) + \boldsymbol{m}^\top B \boldsymbol{m},
\tag{4.23}
$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix, we can rewrite A_{22} in (4.22) as

$$
A_{22} \approx \boldsymbol{\delta}^\top J_{10} J_{00}^{-1} J_{01} \boldsymbol{\delta} + d + \text{tr}(Q_M^{-1} \mathbb{E}[\hat{\boldsymbol{t}}_M \hat{\boldsymbol{t}}_M^\top]).
\tag{4.24}
$$

Regarding A_1 in (4.18), it can be shown to be asymptotically equivalent to A_{22} in (4.19). Combining the above results yields

$$
-2\ell_M(\hat{\boldsymbol{\eta}}_M; Y) + 2d - 2\,\mathbb{E}[\hat{\boldsymbol{t}}_M]^\top \pi_M Q^{-1} \boldsymbol{\delta} + 2\text{tr}(Q_M^{-1} \mathbb{E}[\hat{\boldsymbol{t}}_M \hat{\boldsymbol{t}}_M^\top])
\tag{4.25}
$$

as an asymptotically unbiased estimator of the Akaike information (4.17) for model M. Similar to the derivations of the FIC_{MAE} in Section 4.7.8, we need to consider the different types of candidate models M with the specific Ω_M and resulting $\hat{\boldsymbol{t}}_M$ to determine more specific forms of (4.25).

$M = \emptyset$

If $M = \emptyset$, then (4.25) simplifies to the standard AIC,

$$
\text{AIC}_\emptyset^* = -2\ell_{\text{null}}(\hat{\boldsymbol{\eta}}_{\text{null}}; Y) + 2d.
$$

$M \not\ni 1$

If $1 \notin M$, the parameter vector $\boldsymbol{\gamma}_M$ is not subject to boundary constraints, such that $\Omega_M = \mathbb{R}^{|M|}$ and $\hat{\boldsymbol{t}}_M = Q_M \boldsymbol{E}_M \sim \mathcal{N}_{|M|}(Q_M \pi_M Q^{-1} \boldsymbol{\delta}, Q_M)$. The term $\text{tr}(Q_M^{-1} \mathbb{E}[\hat{\boldsymbol{t}}_M \hat{\boldsymbol{t}}_M^\top])$ is then calculated as

$$
\text{tr}(Q_M^{-1} \text{Cov}[\hat{\boldsymbol{t}}_M]) + \text{tr}(Q_M^{-1} \mathbb{E}[\hat{\boldsymbol{t}}_M] \mathbb{E}[\hat{\boldsymbol{t}}_M]^\top) = |M| + \boldsymbol{\delta}^\top Q^{-1} \pi_M^\top Q_M \pi_M Q^{-1} \boldsymbol{\delta},
$$

because the trace is a linear mapping and invariant under cyclic permutations. Hence, the asymptotically unbiased estimator (4.25) in this case again takes the form of the standard AIC,

$$
\text{AIC}_M^* = -2\ell_M(\hat{\boldsymbol{\eta}}_M; Y) + 2(d + |M|).
$$

$M = \{1\}$

If the model M includes only the boundary parameter γ_1, then

$$\hat{t}_M = 0 \cdot \mathbb{1}_{\{Q_M E_M \le 0\}} + Q_M E_M \cdot \mathbb{1}_{\{Q_M E_M > 0\}},$$

with the indicator function $\mathbb{1}_{\{\cdot\}}$, a scalar Q_M, and $Q_M E_M \sim \mathcal{N}_1(Q_M \pi_M Q^{-1} \delta, Q_M)$. Consequently, the moments of $\hat{t}_M$ can be computed as

$$\mathbb{E}[\hat{t}_M] = 0 \cdot \mathrm{P}[Q_M E_M \le 0] + \mathbb{E}[Q_M E_M \mid Q_M E_M > 0] \cdot \mathrm{P}[Q_M E_M > 0],$$
$$\mathbb{E}[\hat{t}_M^2] = \mathbb{E}[(Q_M E_M)^2 \mid Q_M E_M > 0] \cdot \mathrm{P}[Q_M E_M > 0].$$

Applying the formulas for the moments of the truncated normal distribution, we find that

$$\mathbb{E}[\hat{t}_M] = Q_M^{1/2} u \left[1 - \Phi(-u)\right] + Q_M^{1/2} \phi(-u),$$
$$\mathbb{E}[\hat{t}_M^2] = \left\{ Q_M \left[1 - \frac{u\phi(-u)}{1 - \Phi(-u)} - \left(\frac{\phi(-u)}{1 - \Phi(-u)} \right)^2 \right] + \left[Q_M^{1/2} u + Q_M^{1/2} \frac{\phi(-u)}{1 - \Phi(-u)} \right]^2 \right\}$$
$$\cdot \left[1 - \Phi(-u)\right],$$

where $u = \mathbb{E}[Q_M E_M]/\sqrt{\mathrm{var}[Q_M E_M]} = Q_M^{1/2} \pi_M Q^{-1} \delta$.

After inserting these expressions into formula (4.25), some straightforward calculations and replacing the unknowns by their estimates lead to

$$\mathrm{AIC}^*_{\{1\}} = -2\ell_{\{1\}}(\hat{\boldsymbol{\eta}}_{\{1\}}; Y) + 2(d + 1) - 2\Phi(-\hat{Q}_{\{1\}}^{1/2} \pi_{\{1\}} \hat{Q}^{-1} \hat{\boldsymbol{\delta}}).$$

$M \supsetneq \{1\}$

If the model M includes the boundary parameter γ_1 and some or all of the remaining components of $\boldsymbol{\gamma}$, then

$$\hat{t}_M = G_M^\top R_M G_M E_M \cdot \mathbb{1}_{\{F_M Q_M E_M \le 0\}} + Q_M E_M \cdot \mathbb{1}_{\{F_M Q_M E_M > 0\}},$$

where F_M is a $1 \times |M|$ projection matrix mapping a vector of length $|M|$ onto its first component, G_M is a $(|M| - 1) \times |M|$ projection matrix mapping such a vector onto all but its first component, and $R_M = (G_M Q_M^{-1} G_M^\top)^{-1}$. The moments of $\hat{t}_M$ can again be determined based on the law of total expectation, that is,

$$\mathbb{E}[\hat{t}_M] = \mathbb{E}[G_M^\top R_M G_M E_M \mid F_M Q_M E_M \le 0] \cdot \mathrm{P}[F_M Q_M E_M \le 0]$$
$$+ \mathbb{E}[Q_M E_M \mid F_M Q_M E_M > 0] \cdot \mathrm{P}[F_M Q_M E_M > 0].$$

We first note that $G_M^\top R_M G_M E_M$ and $F_M Q_M E_M$ are independent (cf. p. 95), such that the condition $F_M Q_M E_M \le 0$ in the first expectation can be dropped. Moreover, from $F_M Q_M E_M \sim \mathcal{N}_1(F_M Q_M \pi_M Q^{-1} \delta, F_M Q_M F_M^\top)$, we have $\mathrm{P}[F_M Q_M E_M \le 0] = \Phi(-u_M)$

with $u_M = (F_M Q_M F_M^\top)^{-1/2} F_M Q_M \pi_M Q^{-1} \delta = [Q_M]_{11}^{-1/2} [Q_M \pi_M Q^{-1} \delta]_1$. Finally, the moments of $Q_M E_M$ conditional on $F_M Q_M E_M > 0$ are found by applying the results of Gupta and Tracy (1978) on truncated trivariate normal distributions. In the end, we arrive at

$$
\begin{aligned}
\mathbb{E}[\hat{t}_M] \quad &= G_M^\top R_M G_M \pi_M Q^{-1} \delta \cdot \Phi(-u_M) \\
&\quad + Q_M \pi_M Q^{-1} \delta \cdot [1 - \Phi(-u_M)] + (F_M Q_M F_M^\top)^{-1/2} Q_M F_M^\top \phi(-u_M), \\
\mathbb{E}[\hat{t}_M \hat{t}_M^\top] &= G_M^\top R_M G_M \Phi(-u_M) + G_M^\top R_M G_M \pi_M Q^{-1} \delta \delta^\top Q^{-1} \pi_M^\top G_M^\top R_M G_M \Phi(-u_M) \\
&\quad + Q_M [1 - \Phi(-u_M)] - \frac{Q_M F_M^\top F_M Q_M}{F_M Q_M F_M^\top} u_M \phi(-u_M) \\
&\quad + 2 Q_M \pi_M Q^{-1} \delta \frac{F_M Q_M}{(F_M Q_M F_M^\top)^{1/2}} \phi(-u_M) \\
&\quad + Q_M \pi_M Q^{-1} \delta \delta^\top Q^{-1} \pi_M^\top Q_M [1 - \Phi(-u_M)]. \quad (4.26)
\end{aligned}
$$

Inserting these expressions in (4.25), using some algebra, and replacing the unknowns by their estimates results in

$$
\mathrm{AIC}_M^* = -2\ell_M(\hat{\eta}_M; Y) + 2(d + |M|) - 2\Phi\left(-(F_M \hat{Q}_M F_M^\top)^{-1/2} F_M \hat{Q}_M \pi_M \hat{Q}^{-1} \hat{\delta}\right).
$$

Additional proofs and derivations

Proof of (4.19)

A Taylor expansion of the log-likelihood of model M at the MLE $\hat{\eta}_M$ about $\eta_{0,M}$ yields

$$
\begin{aligned}
A_2 &= 2 \mathbb{E}_Y \left[\mathbb{E}_X \left[-(\hat{\eta}_M - \eta_{0,M})^\top n \bar{W}_{n,M}(\eta_{0,M}; X) \right. \right. \\
&\quad \left. \left. - \frac{1}{2} (\hat{\eta}_M - \eta_{0,M})^\top \mathcal{H}_M(\eta_{0,M}; X)(\hat{\eta}_M - \eta_{0,M}) + R_2 \right] \right] \\
&\approx \mathbb{E}_Y \left[2\sqrt{n}(\hat{\eta}_M - \eta_{0,M})^\top \mathbb{E}_X \left[-\sqrt{n} \bar{W}_{n,M}(\eta_{0,M}; X) \right] \right. \\
&\quad \left. + (\hat{\eta}_M - \eta_{0,M})^\top \mathbb{E}_X \left[-\mathcal{H}_M(\eta_{0,M}; X) \right] (\hat{\eta}_M - \eta_{0,M}) \right],
\end{aligned}
$$

where $\bar{W}_{n,M} = (\bar{U}_n^\top, \bar{V}_{n,M}^\top)^\top$, $\mathcal{H}_M$ is the Hessian matrix of the log-likelihood of model M, and R_2 is a remainder term with $R_2 \xrightarrow{n\to\infty} 0$. Using the limiting distribution of the score (4.20) and that $\mathbb{E}_X[-\mathcal{H}_M(\eta_{0,M}; X)] = n J_M$, we obtain (4.19).

Proof of (4.22)

Using the limiting distribution of the MLE given in (4.21), A_{21} can be rewritten as,

$$
\begin{aligned}
A_{21} &\approx \mathbb{E}\left[(J_{01}\delta + K - J_{01,M}\hat{t}_M)^\top J_{00}^{-1} \right] J_{01}\delta + \mathbb{E}[\hat{t}_M]^\top \pi_M J_{11}\delta \\
&= \delta^\top J_{10} J_{00}^{-1} J_{01}\delta + \mathbb{E}[K]^\top J_{00}^{-1} J_{01}\delta + \mathbb{E}[\hat{t}_M]^\top \pi_M (J_{11} - J_{10} J_{00}^{-1} J_{01})\delta.
\end{aligned}
$$

Because $\mathbb{E}[K] = 0$ and noting that $Q = (J_{11} - J_{10}J_{00}^{-1}J_{01})^{-1}$, this is the first part of (4.22). Similarly, we can approximate A_{22},

$$
\begin{aligned}
A_{22} &\approx \mathbb{E}\left[(J_{01}\delta + K - J_{01,M}\hat{t}_M)^\top J_{00}^{-1}J_{00}J_{00}^{-1}(J_{01}\delta + K - J_{01,M}\hat{t}_M)\right. \\
&\qquad\qquad \left. + 2(J_{01}\delta + K - J_{01,M}\hat{t}_M)^\top J_{00}^{-1}J_{01,M}\hat{t}_M + \hat{t}_M^\top J_{11,M}\hat{t}_M\right] \\
&= \mathbb{E}\left[(J_{01}\delta + K)^\top J_{00}^{-1}(J_{01}\delta + K)\right] - 2\,\mathbb{E}\left[(J_{01}\delta + K)^\top J_{00}^{-1}J_{01,M}\hat{t}_M\right] \\
&\qquad + \mathbb{E}[\hat{t}_M^\top J_{10,M}J_{00}^{-1}J_{01,M}\hat{t}_M] + 2\,\mathbb{E}\left[(J_{01}\delta + K)^\top J_{00}^{-1}J_{01,M}\hat{t}_M\right] \\
&\qquad - 2\,\mathbb{E}[\hat{t}_M^\top J_{10,M}J_{00}^{-1}J_{01,M}\hat{t}_M] + \mathbb{E}[\hat{t}_M^\top J_{11,M}\hat{t}_M] \\
&= \delta^\top J_{10}J_{00}^{-1}J_{01}\delta + 2\delta^\top J_{10}J_{00}^{-1}\mathbb{E}[K] + \mathbb{E}[K^\top J_{00}^{-1}K] \\
&\qquad + \mathbb{E}[\hat{t}_M^\top(J_{11,M} - J_{10,M}J_{00}^{-1}J_{01,M})\hat{t}_M],
\end{aligned}
$$

which with $\mathbb{E}[K] = 0$ and $Q_M = (J_{11,M} - J_{10,M}J_{00}^{-1}J_{01,M})^{-1}$ is the second part of (4.22).

Proof of (4.24)

According to (4.23), we have

$$
\begin{aligned}
\mathbb{E}[K^\top J_{00}^{-1}K] &= \operatorname{tr}(J_{00}^{-1}J_{00}) + \mathbb{E}[K]^\top J_{00}^{-1}\mathbb{E}[K] = d, \\
\mathbb{E}[\hat{t}_M^\top Q_M^{-1}\hat{t}_M] &= \operatorname{tr}(Q_M^{-1}\operatorname{Cov}[\hat{t}_M]) + \mathbb{E}[\hat{t}_M]^\top Q_M^{-1}\mathbb{E}[\hat{t}_M] \\
&= \operatorname{tr}(Q_M^{-1}\mathbb{E}[\hat{t}_M\hat{t}_M^\top]) - \operatorname{tr}(Q_M^{-1}\mathbb{E}[\hat{t}_M]\mathbb{E}[\hat{t}_M]^\top) + \mathbb{E}[\hat{t}_M]^\top Q_M^{-1}\mathbb{E}[\hat{t}_M],
\end{aligned}
$$

where the last two expressions cancel because the trace is invariant under cyclic permutations and the last term is scalar.

Proof of the asymptotic equivalence of A_1 and A_{22} in (4.19)

To show that A_1 in (4.18) is asymptotically equivalent to A_{22} in (4.19), we consider the following Taylor expansion of the log-likelihood of model M at the null point about the MLE,

$$
\begin{aligned}
\ell_M(\eta_{0,M};Y) = \ell_M(\hat{\eta}_M;Y) &+ (\eta_{0,M} - \hat{\eta}_M)^\top n\bar{W}_{n,M}(\hat{\eta}_M;Y) \\
&+ \frac{1}{2}(\eta_{0,M} - \hat{\eta}_M)^\top \mathcal{H}_M(\hat{\eta}_M;Y)(\eta_{0,M} - \hat{\eta}_M) + R_1,
\end{aligned}
$$

with a remainder term $R_1 \overset{n\to\infty}{\longrightarrow} 0$. Then, we see that

$$
\begin{aligned}
A_1 &= 2\,\mathbb{E}_Y\left[\ell_M(\hat{\eta}_M;Y) - \ell_M(\eta_{0,M};Y)\right] \\
&= 2\,\mathbb{E}_Y\left[(\hat{\eta}_M - \eta_{0,M})^\top n\bar{W}_{n,M}(\hat{\eta}_M;Y)\right] \\
&\qquad + \mathbb{E}_Y\left[(\hat{\eta}_M - \eta_{0,M})^\top[-\mathcal{H}_M(\hat{\eta}_M;Y)](\hat{\eta}_M - \eta_{0,M})\right] + R_1.
\end{aligned}
$$

The term involving the score vector is equal to zero. If the MLE $\hat{\boldsymbol{\eta}}_M$ is an inner point of the parameter space, then by definition $\bar{W}_{n,M}(\hat{\boldsymbol{\eta}}_M; Y) = \mathbf{0}$. If some components of the MLE, say $\hat{\gamma}_j$ for $j \in B \subset M$, lie on the boundary of the parameter space, then $\bar{U}_n(\hat{\boldsymbol{\eta}}_M; Y) = \mathbf{0}$ and $\bar{V}_{n,B^c}(\hat{\boldsymbol{\eta}}_M; Y) = \mathbf{0}$, while $\bar{V}_{n,B}(\hat{\boldsymbol{\eta}}_M; Y) \neq \mathbf{0}$, but because $(\hat{\boldsymbol{\gamma}}_B - \boldsymbol{\gamma}_{0,B}) = \mathbf{0}$, we still have $(\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M})^\top n \bar{W}_{n,M}(\hat{\boldsymbol{\eta}}_M; Y) = 0$.

Regarding the term involving the Hessian matrix, one can argue that $-\mathcal{H}_M(\hat{\boldsymbol{\eta}}_M; Y)$ reasonably well approximates nJ_M (see also Claeskens and Hjort, 2008, Section 6.5), which proves the assertion.

Proof of (4.26)

The first lines in the expressions for $\mathbb{E}[\hat{\boldsymbol{t}}_M]$ and $\mathbb{E}[\hat{\boldsymbol{t}}_M \hat{\boldsymbol{t}}_M^\top]$, respectively, in (4.26) follow from $G_M^\top R_M G_M \boldsymbol{E}_M \sim \mathcal{N}_{|M|}(G_M^\top R_M G_M \pi_M Q^{-1} \boldsymbol{\delta}, G_M^\top R_M G_M)$.

The remaining lines correspond to the product of $\mathrm{P}[F_M Q_M \boldsymbol{E}_M > 0]$ and $\mathbb{E}[Q_M \boldsymbol{E}_M \mid F_M Q_M \boldsymbol{E}_M > 0]$ or $\mathbb{E}[(Q_M \boldsymbol{E}_M)(Q_M \boldsymbol{E}_M)^\top \mid F_M Q_M \boldsymbol{E}_M > 0]$, respectively. The conditional expectations can be computed componentwise using the results of Gupta and Tracy (1978). For that purpose, we define random vectors $X = ([Q_M \boldsymbol{E}_M]_1, [Q_M \boldsymbol{E}_M]_k, [Q_M \boldsymbol{E}_M]_l)^\top$ made up of the first, k^{th}, and l^{th} component of $Q_M \boldsymbol{E}_M$, for $k, l \in M$. Then, the X have a trivariate normal distribution with mean vector given by the corresponding components of $Q_M \pi_M Q^{-1} \boldsymbol{\delta}$ and covariance matrix equal to the corresponding submatrix of Q_M. The transformed vectors $Y = \mathrm{diag}([Q_M]_{11}^{-1/2}, [Q_M]_{kk}^{-1/2}, [Q_M]_{ll}^{-1/2})(X - \mathbb{E}[X])$ have a trivariate normal distribution with zero mean vector, unit variances, and correlations $\rho_{st} = [Q_M]_{st}[Q_M]_{ss}^{-1/2}[Q_M]_{tt}^{-1/2}$, for $s, t \in \{1, k, l\}$. The kth component of $\mathbb{E}[Q_M \boldsymbol{E}_M \mid F_M Q_M \boldsymbol{E}_M > 0]$ can then be expressed as

$$\mathbb{E}[X_2 \mid X_1 > 0] = [Q_M]_{kk}^{1/2} \mathbb{E}[Y_2 \mid Y_1 > -u_M] + [Q_M \pi_M Q^{-1} \boldsymbol{\delta}]_k, \quad k \in M,$$

where still $u_M = [Q_M]_{11}^{-1/2}[Q_M \pi_M Q^{-1} \boldsymbol{\delta}]_1$. According to the formulas in Section 4 of Gupta and Tracy (1978),

$$\mathbb{E}[Y_2 \mid Y_1 > -u_M] = \rho_{1k} \frac{\phi(-u_M)}{1 - \Phi(-u_M)},$$

such that

$$\mathbb{E}[Q_M \boldsymbol{E}_M \mid F_M Q_M \boldsymbol{E}_M > 0] = Q_M \pi_M Q^{-1} \boldsymbol{\delta} + (F_M Q_M F_M^\top)^{-1/2} Q_M F_M^\top \frac{\phi(-u_M)}{1 - \Phi(-u_M)}.$$

Similarly, the components $\mathbb{E}[[Q_M \boldsymbol{E}_M]_k [Q_M \boldsymbol{E}_M]_l \mid [Q_M \boldsymbol{E}_M]_1 > 0]$ of the conditional expectation of $(Q_M \boldsymbol{E}_M)(Q_M \boldsymbol{E}_M)^\top$ correspond to $\mathbb{E}[X_2 X_3 \mid X_1 > 0]$, and hence, equal

$$[Q_M]_{kk}^{1/2}[Q_M]_{ll}^{1/2} \mathbb{E}[Y_2 Y_3 \mid Y_1 > -u_M] + [Q_M]_{kk}^{1/2} \mathbb{E}[Y_2 \mid Y_1 > -u_M] \cdot [Q_M \pi_M Q^{-1} \boldsymbol{\delta}]_l$$

$$+ [Q_M]_{ll}^{1/2} \mathbb{E}[Y_3 \mid Y_1 > -u_M] \cdot [Q_M \pi_M Q^{-1} \boldsymbol{\delta}]_k + [Q_M \pi_M Q^{-1} \boldsymbol{\delta}]_k \cdot [Q_M \pi_M Q^{-1} \boldsymbol{\delta}]_l.$$

Consulting again Section 4 in Gupta and Tracy (1978), to find

$$\mathbb{E}[Y_2 Y_3 \mid Y_1 > -u_M] = \rho_{kl} - u_M \rho_{1k}\rho_{1l} \frac{\phi(-u_M)}{1 - \Phi(-u_M)},$$

this leads to $\mathbb{E}[(Q_M \boldsymbol{E}_M)(Q_M \boldsymbol{E}_M)^\top \mid F_M Q_M \boldsymbol{E} M > 0]$ being equal to

$$Q_M - \frac{Q_M F_M^\top F_M Q_M}{F_M Q_M F_M^\top} \cdot u_M \frac{\phi(-u_M)}{1 - \Phi(-u_M)}$$

$$+ 2Q_M \pi_M Q^{-1}\boldsymbol{\delta} \frac{F_M Q_M}{(F_M Q_M F_M^\top)^{1/2}} \cdot \frac{\phi(-u_M)}{1 - \Phi(-u_M)} + Q_M \pi_M Q^{-1}\boldsymbol{\delta}\boldsymbol{\delta}^\top Q^{-1}\pi_M^\top Q_M.$$

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.

Beard, R. E. (1959). Note on some mathematical mortality models. In G. E. W. Wolstenholme and M. O'Connor (Eds.), *The Lifespan of Animals*, Ciba Foundation Colloquia on Ageing, pp. 302–311. Boston: Little, Brown.

Beaudry-Godin, M. (2010). *La démographie des centenaires québécois: validation des âges au décès, mesure de la mortalité et composante familiale de la longévité.* Ph. D. thesis, Université de Montréal.

Böhnstedt, M. and J. Gampe (2019). Detecting mortality deceleration: Likelihood inference and model selection in the gamma-Gompertz model. *Statistics and Probability Letters 150*, 68–73.

Böhnstedt, M., J. Gampe, and H. Putter (2021). Information measures and design issues in the study of mortality deceleration: findings for the gamma-Gompertz model. *Lifetime Data Analysis 27*, 333–356.

Bourbeau, R. and B. Desjardins (2002). Dealing with problems in data quality for the measurement of mortality at advanced ages in Canada. *North American Actuarial Journal 6*(3), 1–13.

Canadian Human Mortality Database (2020). Department of Demography, Université de Montréal (Canada). Available at www.bdlc.umontreal.ca/CHMD/ (data downloaded on March 30, 2020).

Claeskens, G., C. Croux, and J. Van Kerckhoven (2006). Variable selection for logistic regression using a prediction-focused information criterion. *Biometrics 62*(4), 972–979.

Claeskens, G. and N. L. Hjort (2003). The focused information criterion. *Journal of the American Statistical Association 98*, 900–916.

Claeskens, G. and N. L. Hjort (2008). *Model Selection and Model Averaging*. New York: Cambridge University Press.

Cunen, C., L. Walløe, and N. L. Hjort (2020). Focused model selection for linear mixed models with an application to whale ecology. *Annals of Applied Statistics 14*, 872–904.

Desjardins, B. (1998). Le registre de la population du Québec ancien. *Annales de démographie historique 2*, 215–226.

Duchateau, L. and P. Janssen (2008). *The Frailty Model*. New York: Springer.

Feehan, D. M. (2018). Separating the signal from the noise: Evidence for deceleration in old-age death rates. *Demography 55*, 2025–2044.

Gavrilov, L. A. and N. S. Gavrilova (2011). Mortality measurement at advanced ages: A study of the Social Security Administration Death Master File. *North American Actuarial Journal 15*(3), 432–447.

Gavrilov, L. A. and N. S. Gavrilova (2019). New trend in old-age mortality: Gompertzialization of mortality trajectory. *Gerontology 65*, 451–457.

Gavrilova, N. S. and L. A. Gavrilov (2015). Biodemography of old-age mortality in humans and rodents. *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. L. and G. Claeskens (2003). Frequentist model average estimators. *Journal of the American Statistical Association 98*, 879–899.

Horiuchi, S. and J. R. Wilmoth (1998). Deceleration in the age pattern of mortality at older ages. *Demography 35*(4), 391–412.

Jeune, B. and J. W. Vaupel (Eds.) (1999). *Validation of exceptional longevity*. Odense Monographs on Population Aging 6. Odense University Press.

Keiding, N. (1990). Statistical inference in the Lexis diagram. *Philosophical Transactions: Physical Sciences and Engineering 332*(1627), 487–509.

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*, 301–310.

Newman, S. J. (2018). Errors as a primary cause of late-life mortality deceleration and plateaus. *PLOS Biology 16*(12), 1–12.

Ouellette, N. (2016). La forme de la courbe de mortalité des centenaires canadiens-français. *Gérontologie et société 38*(151), 41–53.

Ouellette, N. and R. Bourbeau (2014). Measurement of mortality among centenarians in Canada. In *Living to 100 Monograph*, pp. 17 pages. Society of Actuaries.

Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries 61*(1), 12–57.

Pletcher, S. D. (1999). Model fitting and hypothesis testing for age-specific mortality data. *Journal of Evolutionary Biology 12*, 430–439.

Preston, S. H., I. T. Elo, and Q. Stewart (1999). Effects of age misreporting on mortality estimates at older ages. *Population Studies 53*, 165–177.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rau, R., E. Soroko, D. Jasilionis, and J. W. Vaupel (2008). Continued reductions in mortality at advanced ages. *Population and Development Review 34*(4), 747–768.

Richards, S. J. (2008). Applying survival models to pensioner mortality data. *British Actuarial Journal 14*(2), 257–303.

Richards, S. J., J. G. Kirkby, and I. D. Currie (2006). The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal 12*(1), 5–61.

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.

Tallis, G. M. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society. Series B 23*(1), 223–229.

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.

Vaupel, J. W., F. Villavicencio, and M.-P. Bergeron-Boucher (2021). Demographic perspectives on the rise of longevity. *PNAS 118*(9). doi: 10.1073/pnas.2019536118.

Wienke, A. (2011). *Frailty Models in Survival Analysis*. Biostatistics Series. Chapman & Hall/CRC.