# THE SAINT MODEL: A DECADE LATER

BY

SØREN F. JARNER AND SNORRE JALLBJØRN

## ABSTRACT

While many of the prevalent stochastic mortality models provide adequate short- to medium-term forecasts, only few provide biologically plausible descriptions of mortality on longer horizons and are sufficiently stable to be of practical use in smaller populations. Among the very first to address the issue of modelling adult mortality in small populations was the SAINT model, which has been used for pricing, reserving and longevity risk management by the Danish Labour Market Supplementary Pension Fund (ATP) for more than a decade. The lessons learned have broadened our understanding of desirable model properties from the practitioner's point of view and have led to a revision of model components to address accuracy, stability, flexibility, explainability and credibility concerns. This paper serves as an update to the original version published 10 years ago and presents the SAINT model with its modifications and the rationale behind them. The main improvement is the generalization of frailty models from deterministic structures to a flexible class of stochastic models. We show by example how the SAINT framework is used for modelling mortality at ATP and make comparisons to the Lee-Carter model.

## KEYWORDS

SAINT model, frailty, EM algorithm, Lee-Carter model, mortality modelling, multi-population modelling.

## 1. INTRODUCTION

Since the beginning of the 21st century, life annuity providers have faced an upsurge of pensioners to provide for and the need for reliable, long-term mortality projections is perhaps greater than ever. Indeed, the world-wide increases in life expectancy show no signs of slowing down, and populations where

*Astin Bulletin*, 52(2), 483–517. doi:10.1017/asb.2021.37 

mortality rates are already low still experience rates of improvements of the same, or even higher, magnitude than historically. The situation accentuates the importance of powerful predictive models to handle the consequences of an ever older population.

From a practical point of view, there are two partly conflicting aims: (1) producing accurate forecasts and (2) producing forecasts stable under (annual) updates. Accurate forecasting has been the long-standing objective in actuarial and demographic literature and is, broadly speaking, the goal of the academic, while stability is a more recent requirement pertaining to the needs of the practitioner. When applied in practice, the prevailing market value accounting regime dictates that a mortality model should be updated annually to reflect the latest trends in the data. However, many mortality modelling paradigms are very sensitive to the data period used for calibration, and forecasts can therefore vary substantially from year to year. For an annuity provider, large fluctuations or systematic underperformance of a mortality model can lead to significant shifts in liabilities and capital requirements, resulting in huge costs for either the company or the risk collective. Moreover, throughout Europe, mortality models have become an integral part of policy-making as statutory retirement ages are directly linked to gains in life expectancy. For these decisions, stable short-, medium- and long-term forecasts are not just a requirement, but a necessity.

Stability requirements are particularly difficult to meet when forecasting concerns small populations, including, in fact, many countries. Because improvement patterns in these populations exhibit a great deal of variability, simple extrapolations of past trends tend to have poor predictive power over long horizons and projections are prone to dramatic changes following data updates. This holds true for many of the popular projection methodologies such as the model of Lee and Carter (1992). In view of these accuracy and stability concerns, Jarner and Kryger (2011) developed the SAINT model that has been used by the Danish, nationwide pension fund ATP, since 2008.

The SAINT model was designed with the purpose of producing stable, biologically plausible long-term projections for adult mortality. More precisely, projections with smooth, increasing age-profiles and gradually changing rates of improvement over time. With the main application of pricing and reserving for long-term pension liabilities in mind, capturing long-term trends reliably were deemed more important than, for example, short-term fit. However, more than a decade's worth of experience with the SAINT model in use has broadened our understanding of model requirements from the practitioner's point of view. Even though the overarching structure of the SAINT model has not changed, its components have been revised over the years to address not only accuracy and stability concerns but also the model's flexibility, explainability and credibility. In this paper, we describe how the SAINT model has evolved since Jarner and Kryger (2011) in response to changing demands arising from practical use and user feedback.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

In the years following the introduction of the first version of the SAINT model, quantification of longevity risk became a regulatory requirement. As the deterministic trend component used in Jarner and Kryger (2011) was not able to adequately assess this risk, a version of the SAINT model with a stochastic trend was developed. Eventually, this work led to a generally applicable class of stochastic frailty models, which we present in this paper. This methodology constitutes the main theoretical contribution of the paper.

The rest of the paper is organized as follows. Section 2 contains a survey of the evolution of the SAINT model over the last decade. In Section 3 we discuss how changing rates of improvements can be modelled using frailty. Section 4 formalizes the notion of a stochastic frailty model and develops estimation and forecasting procedures. This is followed by Section 5 presenting a comprehensive application of the SAINT model to international and Danish data. The findings are discussed in light of comparable results from the Lee-Carter model. Finally, Section 6 offers some concluding remarks.

## 2. The evolution of the SAINT model

ATP is a funded supplement to the Danish state pension, guaranteeing most of the population a whole-life annuity. In 2008, ATP introduced a new market value (whole-life) annuity. The main characteristic of this annuity is that contributions are converted to pension entitlements on a tariff based on prevailing market rates and an annually updated, best estimate, cohort-specific mortality forecast. Once acquired, pension entitlements are guaranteed for life. The structure gives a high degree of certainty for the members, but it leaves ATP with a substantial longevity risk.

The SAINT model was developed as part of the market value annuity, with the specific aim of producing accurate and stable forecasts in order to manage the longevity risk of ATP. Clearly, accuracy is desirable to avoid long-run deficits due to life expectancy increasing faster than expected (or, in the case of life expectancy increasing slower than expected, pension entitlements being too small). However, stability of forecasts is of equal importance. Each update of the model causes a change in the size of technical provisions, which in turn affects the risk capital allocated to cover longevity risk. Effectively, a stable mortality model is "cheaper" than a volatile model, because the former frees up risk capital to be used more efficiently elsewhere, for example, to cover a higher market exposure.

Over the course of the last decade, the SAINT model has undergone a number of changes due to changing demands and feedback from its users. Below, we give a brief presentation of the original SAINT model, followed by a survey of the subsequent major changes and their rationale.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

### 2.1. The original SAINT model

The original SAINT model, as described in Jarner and Kryger (2011), has two core model components:

- A reference population consisting of a large, pooled, international data set, and;
- A frailty model for modelling increasing rates of improvements over time.

The rationale for using a reference population, in addition to the target population, is that it is easier to extract a long-term trend from a large dataset, than a small dataset, since idiosyncratic features are typically more pronounced in the latter. The mortality of the target population is subsequently linked to the long-term trend. A similar idea, although differently implemented, was introduced by Li and Lee (2005) in their multi-population extension of the Lee-Carter model. Since Jarner and Kryger (2011), the concept of a reference population has appeared in a number of models and applications, for example Cairns *et al*. (2011), Dowd *et al*. (2011), Börger *et al*. (2014), Villegas and Haberman (2014), Wan and Bertschi (2015), Hunt and Blake (2017b), Villegas *et al*. (2017), Menzietti *et al*., 2019), Li and Liu (2019), Li *et al*. (2019).

In the notation of the present paper, the SAINT model is of the form:

$$\mu_{\text{target}}(t,x) = \mu_{\text{ref}}(t,x)\exp\left(y_t^\top r_x\right), \tag{2.1}$$

$$\mu_{\text{ref}}(t,x) = \mathbb{E}[Z|t,x]\mu_0(t,x) + \mu_b(t), \tag{2.2}$$

where $\mu_{\text{target}}(t,x)$ and $\mu_{\text{ref}}(t,x)$ are the force of mortality at age $x$ and time $t$ of the target and reference population, respectively. The target mortality is linked to the reference mortality by a set of age-dependent regressors, $r_x$, with time-dependent coefficients, $y_t$, termed *spread* parameters. Further, the reference mortality is modelled by the sum of a multiplicative frailty model with baseline mortality $\mu_0$, and age-independent background mortality, $\mu_b$. The term $\mathbb{E}[Z|t,x]$ denotes the (conditional) mean frailty of the given cohort, and we will discuss this in detail later. All mortality rates are gender-specific, although this is not shown explicitly in the notation.

Formally, the SAINT model in use today is still of form (2.1)–(2.2), but with a different interpretation of the frailty term, and different specifications of time-series dynamics, regressors, and baseline and background mortality. From a methodological point of view, the new frailty model represents by far the greatest of these changes.

### 2.2. From deterministic to stochastic frailty

The SAINT model uses frailty to forecast increasing rates of improvement in age-specific mortality, thereby reducing the risk of underestimating future life

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

expectancy gains. Loosely speaking, changes in baseline mortality affect selection and thereby mean frailty, $\mathbb{E}[Z|t, x]$, which in turn modifies the way baseline mortality affect population-level mortality.

In the original SAINT model, $\mu_0$ and $\mu_b$ were assumed to be of a form equivalent to

$$\mu_0(t, x) = \exp\left(\theta_1 + \theta_2 t + \theta_3 x + \theta_4 tx + \theta_5 x^2\right), \quad \mu_b(t) = \exp\left(\theta_6 + \theta_7 t\right). \quad (2.3)$$

This allowed an explicit calculation of $\mathbb{E}[Z|t, x]$ and thereby of $\mu_{\text{ref}}$, assuming gamma-distributed frailties, see (14)–(18) in Jarner and Kryger (2011) for details. The resulting model for $\mu_{\text{ref}}$ had 8 parameters ($\theta_1 - \theta_7$ and a frailty parameter) for each sex, which were estimated by maximizing a standard Poisson likelihood. When forecasting, the parametric form was used to extrapolate $\mu_{\text{ref}}$ to form a deterministic trend around which $\mu_{\text{target}}$ would vary. We refer to this as a deterministic frailty model.

Despite its parsimonious structure, the estimated $\mu_{\text{ref}}$-surface provided a remarkable fit to the international dataset. However, it soon became clear that the model was not able to fit other datasets equally well. More importantly, assessment of longevity risk was becoming a regulatory requirement, and for this purpose, a deterministic trend model was insufficient.

The current version of the SAINT model features a stochastic frailty model, in which $\mu_0$ and $\mu_b$ are stochastic processes. This implies that the frailty term, $\mathbb{E}[Z|t, x]$, also becomes stochastic, and explicit expressions are no longer available. In order to disentangle the dependence between $\mu_0$ and $\mathbb{E}[Z|t, x]$, a pseudo-likelihood approach, reinterpreting the frailty term in terms of observable quantities, had to be devised. This, in turn, led to a generally applicable "fragilization" method by which essentially any frailty distribution can be combined with any baseline and background mortality to form a model, cf. Section 4.

### 2.3. Cointegrating gender dynamics

In the original SAINT model, male and female mortality were modelled separately, which resulted in a sex differential of just over 10 years in forecasted cohort life expectancies at birth. When the first version of the stochastic frailty model was implemented, it was therefore decided to model the new stochastic processes via cointegration to ensure better aligned forecasts. At the time, the baseline and background mortality were modelled as

$$\mu_0(t, x) = \exp\left(\alpha_t + \beta_t x + 2x^2/10^4\right), \quad \mu_b(t) = \exp\left(\zeta_t\right), \quad (2.4)$$

with the processes $\{\alpha_t, \beta_t\}$ governing $\mu_0$ being modelled as in Equation (5.5), but with an unrestricted $A$-matrix. Cointegration reduced the life expectancy difference at birth to about 5 years, and subsequent model development brought it further down to 3.6 years.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

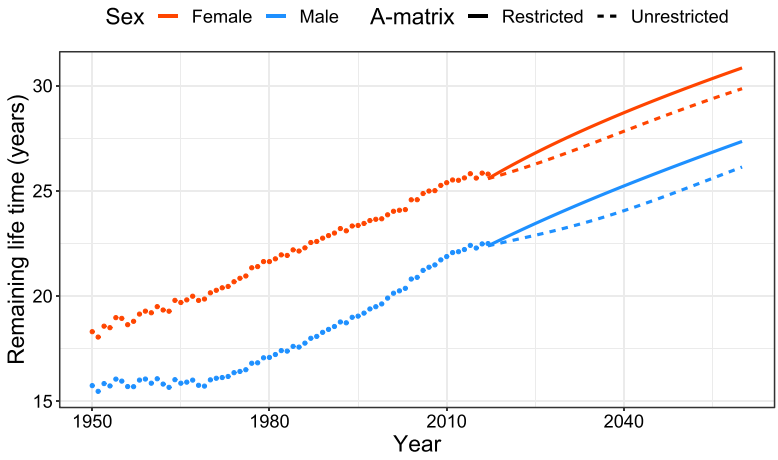


FIGURE 1. Age 60 actual (dots) and forecasted (lines) period life expectancy using the SAINT model with and without restrictions on the A-matrix from Equation (5.5).

In Equation (5.5), the B-matrix controls the cointegrating relations, while the A-matrix controls the adjustments to these over time. It later became apparent that an unrestricted $A$-matrix could lead to complex transitory effects when reestablishing equilibrium relations, as seen on the dashed lines in Figure 1. The resulting projections were hard to justify and communicate, and eventually structural zeros were introduced in $A$ to generate more linear projections. Allowing only pairwise dependence between parameters also offered a greater degree of explainability as the forecasting distribution simplified.

The cointegrating relations and the restrictions placed on the $A$-matrix were imposed, rather than formally tested. Even though formal testing could be done using the comprehensive statistical framework developed by Johansen (1995), it was deemed problematic that the underlying time dynamics could change annually following data updates as this would destabilize projections and potentially damage the model's credibility. Cointegration is therefore used merely as a modelling tool to achieve reasonable projections, rather than to gain insights into the joint behaviour of the time-varying mortality indices, see also Jarner and Jallbjørn (2020) for further discussion on this point.

## 2.4. Improving the fit

The specification of $\mu_0$ in (2.4) was the result of an extensive model search among "simple" models. At the time, it provided a reasonable fit, but eventually failed to adequately capture old-age mortality. Generalizations of linear mortality models typically involve adding a quadratic term to the age effect or introducing cohort components, see for example Cairns *et al*. (2009). The natural candidate to replace (2.4) was therefore the log-quadratic model $\mu_0(t, x) =$

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

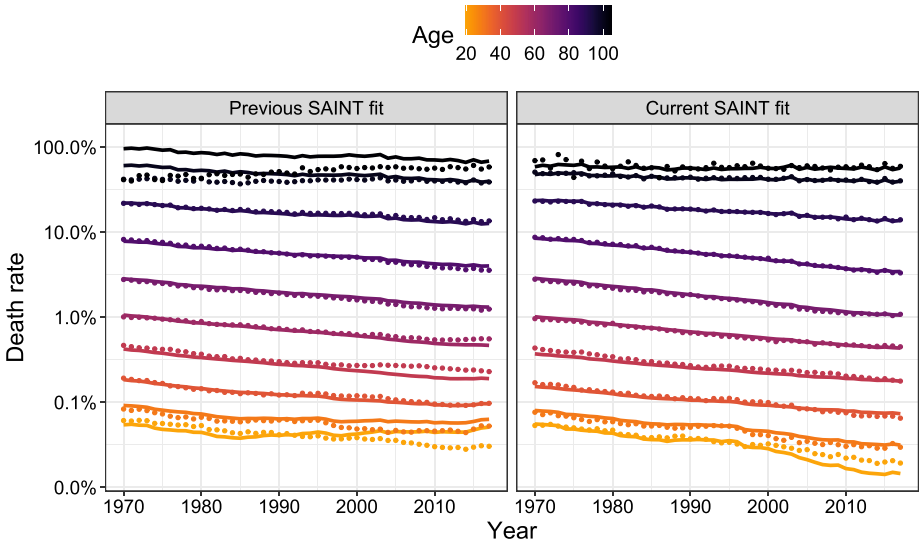


FIGURE 2. Observed (dots) female death rates for select ages with SAINT fits (solid lines) superimposed. The left panel shows the previous version of SAINT with $\mu_0$ as in (2.4), estimated on a dataset with the US included and the window of calibration starting in 1950. The right panel shows the current version of SAINT with $\mu_0$ as in (2.5), estimated on a dataset with the US excluded and the window of calibration starting in 1970.

$\exp(\alpha_t + \beta_t x + \kappa_t x^2)$. But despite a clearly superior fit, its parameter estimates turned out exceedingly difficult to forecast.

After further research into the shape of the mortality age profile, the lack of fit was found to be caused by an inflexibility of (2.4) at the younger ages, a typical problem when fitting parsimonious models over a large age span. By replacing the fixed quadratic term in (2.4) with an excess slope parameter, the low mortality rates of the young were prevented from influencing the trend of the old. The baseline model therefore became

$$\mu_0(t,x) = \exp\left(\widetilde{\alpha}_t + \widetilde{\beta}_t x + \widetilde{\kappa}_t (x-75)\mathbb{1}_{\{x\geq 75\}}\right). \tag{2.5}$$

The model's parameters are linearized for forecasting purposes through reparametrization, achieved by setting $\alpha_t = \widetilde{\alpha}_t + 75\widetilde{\beta}_t$, $\beta_t = \widetilde{\beta}_t + \widetilde{\kappa}_t$, and $\kappa_t = -\widetilde{\kappa}_t$ whereby

$$\mu_0(t,x) = \exp\left(\alpha_t + \beta_t (x-75) + \kappa_t (x-75)\mathbb{1}_{\{x<75\}}\right). \tag{2.6}$$

Figure 2 shows the clear improvement in the model's fit. The fit is particularly impressive at the higher ages and also matches the logistic type behaviour seen in the data for the oldest-old as the frailty component comes into play.

### *2.4.1. Revisiting the reference population and data window*

In the original paper, Jarner and Kryger (2011) used an aggregate of international data over the years 1933–2005, including, notably, data from the US.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

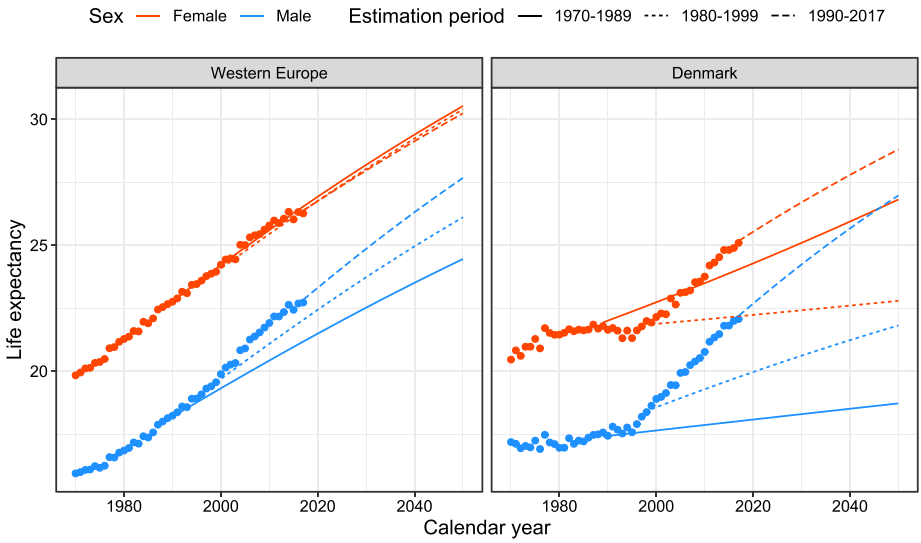


FIGURE 3. Age 60 actual (dots) and forecasted (lines) period life expectancy using a Lee-Carter model based on a rolling estimation window.

Following the first version of the stochastic frailty model, the left cut point of the data window was updated to 1950, while the right cut point was updated annually to match the most recent available data.

Over the years, it became apparent that the slowdown of the improvement rates in the US had begun to manifest itself in the long-term trend. At the time, the US constituted more than 40% of the reference population. In response, an extensive review of the demographic transitions in the Western world was conducted country for country. This work led to a new paradigm for putting together a more homogeneous and balanced data pool, necessitating an exclusion of the US. In the new dataset, two mortality regimes emerged, and, accordingly, the left cut point of the data window was updated to 1970. We note that other countries have also shown recent stagnating rates of improvement, for example the UK, but overall there has been a continued improvement throughout the period, see Figure 3.

## 3. MODELLING CHANGING RATES OF IMPROVEMENTS

The mortality experience of several countries has shown increasing rates of improvements for older age groups while rates for younger groups have been decelerating, see for example Kannisto *et al*. (1994), Lee and Miller (2001), Booth *et al*. (2002), Bongaarts (2005), Li *et al*. (2013), Vékás (2019). For long-term projections, in particular, it is important to model these changes to reduce the risk of underestimating future life expectancy gains.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

### 3.1. Motivating example

Although a plethora of models for modelling and forecasting mortality have been proposed in recent years, see for example Booth (2008) and Janssen (2018) for an overview, the model of Lee and Carter (1992) is still by far the most widely used. Lee and Carter (1992) model the (observed) death rate, $m$, at time $t$ for age $x$ in a log-bilinear fashion

$$\log m(t, x) = a_x + b_x k_t + \varepsilon_{t,x}, \tag{3.1}$$

where $a$ and $b$ are age-specific parameters and $k$ is a time-varying index in which all temporal trends leading to improvements in mortality are encapsulated. The $k$-index is typically modelled as a random walk with drift, extrapolating the index linearly from the first to the last data point, resulting in constant rates of improvement in forecasted age-specific mortality.

Figure 3 illustrates the problem with assuming constant rates of improvements in forecasts. The figure shows actual and Lee-Carter forecasted period remaining life expectancies at age 60 for Western Europe and Denmark. Except for Western European females, the forecasts vary substantially with the estimation period due to changing rates of improvement. Moreover, the forecasts generally underestimate the future gains in life expectancy because rates of improvement for older age groups tend to increase over time.

This phenomenon is not specific to the Lee-Carter model, but pertains to all models with constant rates of improvements. Fitting these models to shorter, more recent periods of data alleviates the downward bias to some extent, but it does not address the fundamental issue of changing rates. Coherent, multi-population mortality models, for example the model of Li and Lee (2005), can in principle produce changing rates of improvement while the individual populations "lock on" to common rates of improvement. However, the common rates are typically constant over time. Thus, coherence in itself does not guarantee the type of ongoing change in improvement rates that we advocate. For a more detailed discussion of coherent models and their pros and cons, see Jarner and Jallbjørn (2020) and the references therein.

### 3.2. Frailty theory

Frailty theory rests on the assumption that cohorts are heterogeneous and that some people are more susceptible to death (frail) than others. The difference in frailty causes selection effects in the population and leads to old cohorts being dominated by low mortality individuals.

Frailty theory is well-established in biostatistics and survival analysis, and several monographs are devoted to the topic, for example Duchateau and Janssen (2008), Wienke (2010), Hougaard (2012). In demographic and actuarial science, frailty models are also known as heterogeneity models. They have been used in mortality modelling to fit the logistic form of old-age mortality, see for example Wang and Brown (1998), Thatcher (1999), Butt and Haberman

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

(2004), Olivieri (2006), Cairns *et al*. (2006), Spreeuw *et al*. (2013), Li and Liu (2019), and to allow for overdispersion in mortality data, cf. Li *et al*. (2009). The SAINT model, however, employs frailty theory with the dual purpose of fitting old-age mortality and generating changing rates of improvement.

Below, we present a flexible class of continuous time models spanning multiple birth cohorts, with additive frailty and non-frailty components. With additive models, we can distinguish between "selective" mortality influenced by frailty and "background" mortality not affected by frailty, for example accidents. Following Vaupel *et al*. (1979), individual frailty is a non-negative stochastic quantity $Z$ that acts multiplicatively on an underlying baseline mortality rate. We assume that frailty is assigned at birth (according to some distribution) and remains constant throughout an individual's life span. In this context, we can interpret frailty as individual (congenital) genetic differences. Conditionally on frailty being $Z$, mortality at age $x$ at time $t$ takes the form

$$\mu(t, x|Z) = Z\mu_0(t, x) + \mu_b(t, x), \tag{3.2}$$

where $\mu_0$ is the baseline rate describing age-period effects influenced by individual frailty and $\mu_b$ is background mortality common to all individuals regardless of their respective frailties.

Equation (3.2) describes the mortality rate of an individual, but this quantity is not observable in population-level data. In fact, we only observe an aggregate of the death rates. We can derive an explicit expression for this aggregate, namely the population-level rate, by writing up the survival function

$$S(t, x) = e^{-\int_0^x \mu(t-x+u,u)\,\mathrm{d}u} = \mathbb{E}\left[e^{-\int_0^x \mu(t-x+u,u|Z)\,\mathrm{d}u}\right] \tag{3.3}$$

and differentiating $-\log S(t, x)$ to get

$$\mu(t, x) = \mathbb{E}[Z|t, x]\mu_0(t, x) + \mu_b(t, x). \tag{3.4}$$

Here, $\mathbb{E}[Z|t, x]$ is the mean frailty among the survivors of the cohort born at time $t - x$. As a matter of convention, we assume without loss of generality that mean frailty is one at birth. It is useful to introduce the Laplace transform $\mathcal{L}(s) = \mathbb{E}[\exp(-sZ)]$ of the common frailty distribution in which case

$$\mathbb{E}[Z|t, x] = \frac{-\mathcal{L}'(\mathcal{M}_0(t, x))}{\mathcal{L}(\mathcal{M}_0(t, x))}, \tag{3.5}$$

where $\mathcal{M}_0(t, x) = \int_0^x \mu_0(t - x + u, u)\,\mathrm{d}u$ is the cumulated baseline rate.

So far, the expressions above relating mean frailty to the baseline rate are standard in survival analysis. For later use, we establish an additional relationship between mean frailty and the cumulated cohort rate adjusted for background mortality, namely

$$\mathcal{M}(t, x) = \int_0^x (\mu(t - x + u, u) - \mu_b(t - x + u, u))\,\mathrm{d}u, \tag{3.6}$$

via its survival function

$$\exp(-\mathcal{M}(t, x)) = S(t, x)e^{\int_0^x \mu_b(t-x+u,u)\,\mathrm{d}u} = \mathcal{L}(\mathcal{M}_0(t, x)). \tag{3.7}$$

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

Introducing the function $\nu(\cdot) = -\log \mathcal{L}(\cdot)$, we have $\mathcal{M}(t,x) = \nu(\mathcal{M}_0(t,x))$ and $\mathcal{M}_0(t,x) = \nu^{-1}(\mathcal{M}(t,x))$ which gives us

$$\mathbb{E}[Z|t,x] = \nu'(\mathcal{M}_0(t,x)) = \nu'(\nu^{-1}(\mathcal{M}(t,x))), \tag{3.8}$$

upon insertion into (3.5).

The relation between mean frailty and mortality described by Equation (3.8) will be central to our estimation approach. Whereas $\mathcal{M}_0$ is given solely in terms of the baseline rate, $\mathcal{M}$ can be estimated using empirical death rates. Substituting $\mathcal{M}$ by such an estimate disentangles the frailty distribution from the baseline rate which greatly simplifies the estimation procedure whenever parametric structures have been imposed on $\mu_0$ and $\mu_b$. We return to the specifics in Section 4.

To apply frailty theory in practice, we must identify suitable choices of frailty distributions. A brief account of appropriate distributions is given in Appendix A. Essentially, useful distributions are the ones with an explicit Laplace transform. The most commonly used distribution is the Gamma distribution which has a tractable Laplace transform, see Example 4.1, along with other desirable properties.

### 3.3. Frailty leads to changing rates of improvements

To clarify how the inclusion of frailty leads to changing rates of improvements, we define the rate of improvement in selective mortality as

$$\rho_s(t,x) = -\frac{\partial}{\partial t}\log \mathbb{E}[Z|t,x] - \frac{\partial}{\partial t}\log \mu_0(t,x), \tag{3.9}$$

where $\rho_0(t,x) = -\frac{\partial}{\partial t}\log \mu_0(t,x)$ is the rate of improvement in baseline mortality.

Suppose that we model the period effect so that baseline mortality is decreasing over time, that is, $\mu_0(t,x) \to 0$ as $t \to \infty$ for fixed age $x$. Then, the cumulated baseline will also be decreasing over time, that is $\mathcal{M}_0(t,x) \to 0$ as $t \to \infty$, while the mean frailty in successive cohorts will be *increasing* over time due to less and less selection of frail individuals, so $\mathbb{E}[Z|t,x] \to 1$ as $t \to \infty$. From (3.9), we see that improvements in cohort mortality are smaller than, but gradually increasing to, the rate of improvement at the individual level.

This line of thinking offers an explanation to the small improvement rates observed in old-age mortality, and it suggests that we might expect to see higher rates of improvements in the future. At old ages where death rates – and thereby selection – are high, the change in mean frailty over time can substantially offset improvements in baseline mortality. This makes improvements in observed mortality close to zero, cf. Equation (3.9), a behaviour illustrated in Figure 4. As improvements in baseline mortality continue to occur at the individual level, the selection mechanism gradually weakens and improvements in observed mortality get closer to the underlying improvement rates. The pattern

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

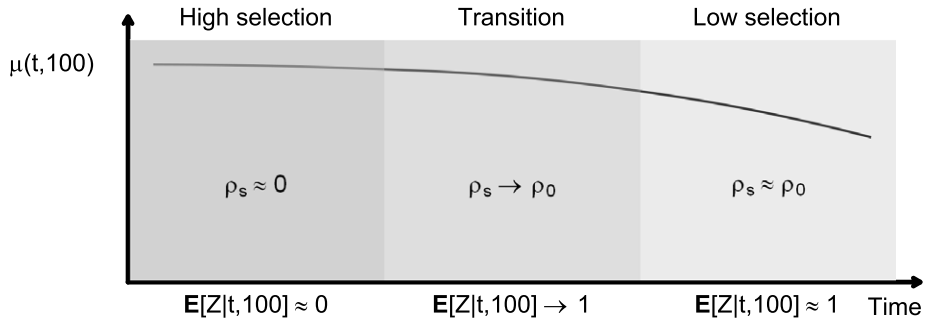


FIGURE 4. Illustration of population-level mortality at age 100 over time. When selection is high, observed mortality rates do not improve much even though rates are assumed to be decreasing at the individual level. This is due to improvements in baseline mortality being (partially) offset by increases in mean frailty. As mean frailty eventually approaches one, observed improvements and improvements in the underlying mortality are approximately equal.

of gradually changing improvement rates of old-age mortality resembles what is seen in the data. We will return to this point in Section 5.

*3.3.1. The difference between frailty and the traditional cohort perspective*

While conditional mean frailty $\mathbb{E}[Z|t, x]$ may be regarded as a cohort component in the sense that it focuses on the evolution of a cohort through time, the notion is qualitatively different from the traditional cohort perspective in mortality modelling. As an illustrative example, consider the cohort extended Lee-Carter model of Renshaw and Haberman (2006), that is

$$\log \mu(t, x) = a_x + b_x k_t + c_{t-x}. \tag{3.10}$$

The cohort component $\exp(c_{t-x})$ assumes the role of a dummy variable and offsets mortality by the same factor throughout the life span of individuals born at time $t - x$. Although the multiple $\mathbb{E}[Z|t, x]\mu_0(t, x)$ does in principle contain the structure (3.10) with $\mu_0(t, x) = \exp(a_x + b_x k_t)$ being the Lee-Carter model, $\mathbb{E}[Z|t, x]$ is not constant over time except in degenerate cases. Indeed, $\mathbb{E}[Z|t, x]$ progressively decreases for a given cohort as selection intensifies and should therefore not be regarded as a cohort component in the traditional sense.

### 3.4. Alternative ways of addressing changing improvement rates

With a growing body of empirical evidence against the assumption of constant age-specific rates of improvements, several other approaches have been developed to address the issue. One suggestion involves finding an "optimal" calibration period for which model assumptions are not violated. This usually involves fitting (Lee-Carter) models to shorter and more recent periods of data, see for example Lee and Miller (2001), Tuljapurkar *et al*. (2000), Booth *et al*.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

(2002), Li and Li (2017). For projections over longer horizons, the problem persists; future increases in especially old-age mortality beyond what has been seen historically cannot be captured by making an informed decision about the period of calibration.

Within the Lee-Carter framework, another solution is to extend the original model by replacing the time-invariant age-response with a time-varying version, so instead of Equation (3.1) we would have

$$\log m(t, x) = a_x + B_{t,x}k_t + \varepsilon_{t,x}. \tag{3.11}$$

A freely varying $B_{t,x}$ introduces far too many parameters, and some restrictions have to be imposed. Li *et al*. (2013) suggest that $B_{t,x}$ describe the "rotating" pattern of mortality improvements, namely that improvement rates are declining for the young but increasing for the old. Consequently, Equation (3.11) is sometimes referred to as the rotated Lee-Carter model. Li *et al*. (2013) achieve rotation by letting the $b_x$'s from the original model converge (smoothly) to some assumed long-term target $B_x$. The approach has since been adopted and adapted by a number of authors, for example Vékás (2019) and Gao and Shi (2021).

Another idea is to model and project improvement rates directly as opposed to projecting death rates, see for example Haberman and Renshaw (2012). Various forms of projections built from age-specific improvement rates applied to reference mortality tables are also becoming popular among actuarial practitioners, see for example Jarner and Møller (2015) for a detailed account of the longevity benchmark employed by the Danish financial supervisory authority or the model used by the Continuous Mortality Investigation (2016).

The alternative approaches mentioned above each have their own merits as ways of addressing changing rates of improvements. Frailty, however, has the unique advantage that it can be introduced into any existing mortality model to forecast improvement rates higher than those observed historically, while preserving both the original model structure and the underlying driver of the system.

## 4. Stochastic frailty models

In the following, we detail how frailty can be used with any stochastic mortality model and we show how to estimate and forecast these models. The extension from deterministic to stochastic frailty is a fundamental point in practical applications where the ability to describe uncertainty of forecasts is essential for managing longevity risk.

### 4.1. Data and terminology

Data are assumed to be on the form of death counts, $D(t,x)$, and corresponding exposures, $E(t,x)$, over time-age cells of the form $[t, t+1) \times [x, x+1)$

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

for a range of calendar years $t \in \{t_{\min}, \ldots, t_{\max}\} = \mathcal{T} \subseteq \mathbb{N}$ and ages $x \in \{x_{\min}, \ldots, x_{\max}\} = \mathcal{X} \subseteq \mathbb{N}_0$. From the death counts and risk exposures, we define the observed (empirical) death rate as the ratio

$$m(t, x) = D(t, x)/E(t, x). \tag{4.1}$$

The empirical death rate is a nonparametric estimate of the underlying cohort rate, $\mu(t, x)$, which for modelling purposes is also assumed constant over $[t, t+1) \times [x, x+1)$.

### 4.2. Modelling framework

For ease of presentation, we will consider stochastic models for baseline and background mortality of the form

$$\mu_0(t, x) = F(\theta_t, \eta_x), \tag{4.2}$$

$$\mu_b(t, x) = G(\zeta_t, \xi_x), \tag{4.3}$$

where $F$ and $G$ are functional forms taking parameters in the age-period dimension as input. All quantities may be multidimensional, and further generalizations to include cohort effects and general dependence structures are possible if so desired. We assume that parameters are to be estimated from data, but they can also be fixed or empty.

We define a (generalized) stochastic frailty model as a model of the form

$$D(t, x) \sim \text{Poisson}(\mu(t, x)E(t, x)), \tag{4.4}$$

$$\mu(t, x) = \mathbb{E}[Z|t, x]\mu_0(t, x) + \mu_b(t, x), \tag{4.5}$$

with $\mu_0$ and $\mu_b$ given by (4.2)–(4.3) while $\mathbb{E}[Z|t, x]$ denotes conditional mean frailty as in Section 3.2. Note that $\mathbb{E}[Z|t, x]$ is now a stochastic quantity since it depends on $\mu_0$. The frailty distribution at birth is the same for all cohorts, and it is assumed to belong to a family indexed by $\psi$ with Laplace transform $\mathcal{L}_\psi$ available in explicit form. The parameters of the model are thus $(\psi, \theta, \eta, \zeta, \xi)$ where all components can be vectors.

Based on (3.8), we can write

$$\mu(t, x) = \nu'_\psi\left(\mathcal{M}_0(t, x)\right) F\left(\theta_t, \eta_x\right) + G\left(\zeta_t, \xi_x\right) \tag{4.6}$$

with $\nu_\psi(\cdot) = -\log \mathcal{L}_\psi(\cdot)$ and $\mathcal{M}_0(t, x) = \sum_{u=0}^{x-1} F(\theta_{u+t-x}, \eta_x)$. Inserting the above into (4.4), all parameters can be estimated jointly from the resulting likelihood. This likelihood is, however, rather intractable with frailty and remaining parameters occurring in a complex mix. Consequently, estimation has to be handled on a case-by-case basis depending on the choice of frailty distribution and mortality model. Below we propose an alternative, generally applicable pseudo-likelihood approach which greatly simplifies the estimation task.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

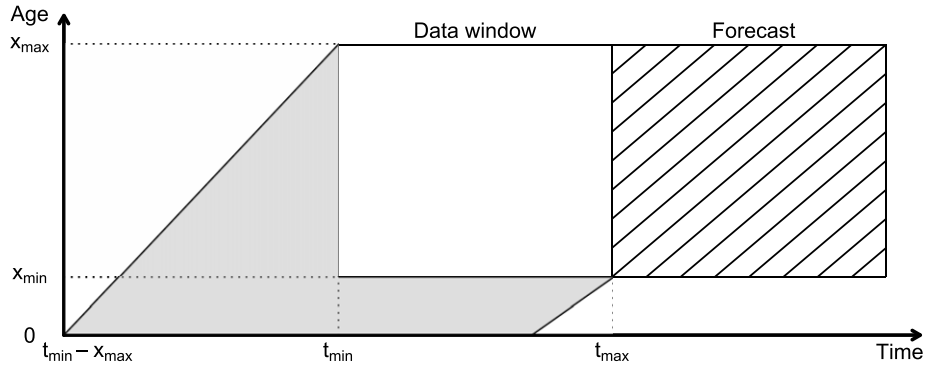


FIGURE 5. Data are available for years between $t_{\min}$ and $t_{\max}$ and for ages between $x_{\min}$ and $x_{\max}$. The grey area below and to the left of the data window illustrates the part of the trajectories needed for calculation of $\widetilde{\mathcal{M}}$ that falls outside the data window. The cross-hatched area to the right illustrates the years and ages for which we wish to forecast mortality.

### 4.3. Pseudo-likelihood function

We seek to replace the problematic term $\mathbb{E}[Z|t,x]$ with an estimate that does not depend on baseline parameters. From (3.8), we have that $\mathbb{E}[Z|t,x] = \nu_\psi'(\nu_\psi^{-1}(\mathcal{M}(t,x)))$, where $\mathcal{M}$ is the cumulated frailty-dependent part of mortality. At first sight, this does not seem to help much since $\mathcal{M}$ is even more complicated than $\mathcal{M}_0$. However, in contrast to $\mathcal{M}_0$ we can obtain an estimate of $\mathcal{M}$ which does not involve the baseline parameters. With a slight abuse of notation, suppressing the dependency on the background parameters, we estimate $\mathcal{M}$ by

$$\widetilde{\mathcal{M}}(t,x) = \sum_{u=0}^{x-1} \widetilde{m}(t-x+u,u), \tag{4.7}$$

where

$$\widetilde{m}(t,x) = \begin{cases} m(t_{\min},x) - G(\zeta_{t_{\min}},\xi_x) & \text{for } x_{\min} \le x \le x_{\max} \text{ and } t < t_{\min}, \\ m(t,x) - G(\zeta_t,\xi_x) & \text{for } x_{\min} \le x \le x_{\max} \text{ and } t_{\min} \le t \le t_{\max}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.8}$$

is an extension of the empirical death rates (with background mortality subtracted). The extension is required because the summation in (4.7) falls partly outside the data window; we need to know the death rates from birth to the present or maximum age for all cohorts entering the estimation. The gray area of Figure 5 illustrates the "missing" death rates. The extension implies that selection prior to $t_{\min}$ has happened according to initial rates (rather than actual rates) and that all cohorts have mean frailty one at age $x_{\min}$ (rather than at birth).

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

We propose to base estimation of (4.4)–(4.5) on a likelihood function in which the term $\mathbb{E}[Z|t,x]$ is replaced by $\nu_\psi'(\nu_\psi^{-1}(\widetilde{\mathcal{M}}(t,x)))$. The resulting approximate likelihood function is referred to as a pseudo-likelihood function, cf. Besag (1975), and corresponds to estimating the modified model

$$D(t,x) \sim \text{Poisson}(\mu(t,x)E(t,x)), \tag{4.9}$$

$$\mu(t,x) = \nu_\psi'(\nu_\psi^{-1}(\widetilde{\mathcal{M}}(t,x)))F(\theta_t,\eta_x) + G(\zeta_t,\xi_x). \tag{4.10}$$

Importantly, the cohort rate $\mu$ is separable in frailty and baseline parameters. The model is therefore considerably easier to handle than (4.4)–(4.5) in the sense that joint estimation can be based on marginal estimation procedures for the baseline and background intensities, $\mu_0$ and $\mu_b$. Equation (4.10) might still look daunting, but it simplifies in specific cases.

**Example 4.1** When Z follows a Gamma distribution with mean one and variance $\sigma^2$, the Laplace transform and conditional mean frailty are given by

$$\mathcal{L}(s) = \left(1+\sigma^2 s\right)^{-1/\sigma^2}, \tag{4.11}$$

$$\mathbb{E}[Z|t,x] = \left(1+\sigma^2\mathcal{M}_0(t,x)\right)^{-1} = \exp\left(-\sigma^2\mathcal{M}(t,x)\right), \tag{4.12}$$

whereby Equation (4.10) reads

$$\mu(t,x) = \exp\left(-\sigma^2\widetilde{\mathcal{M}}(t,x)\right)F(\theta_t,\eta_x) + G(\zeta_t,\xi_x). \tag{4.13}$$

### 4.4. Maximum likelihood estimation

Maximum likelihood estimates of the model (4.9)–(4.10) can be obtained by optimizing the profile (pseudo) log-likelihood function,

$$\ell(\psi) = \log L\left(\psi, \hat\theta(\psi), \hat\eta(\psi), \hat\zeta(\psi), \hat\xi(\psi)\right), \tag{4.14}$$

where $L$ is the likelihood resulting from (4.9)–(4.10) and $\hat\theta(\psi)$, $\hat\eta(\psi)$, $\hat\zeta(\psi)$ and $\hat\xi(\psi)$ denote the maximum likelihood estimates for fixed value of $\psi$. Since the frailty family is typically of low dimension, the profile log-likelihood function can usually be optimized reliably by general purpose optimization routines. This is particularly simple when $\mu_b(t,x) = 0$ for all $t \in \mathcal{T}$ and $x \in \mathcal{X}$.

In the general setting with non-zero background mortality, the model describes a situation of competing risks. The interpretation is that individuals of age $x$ at time $t$ are susceptible to death from two different, mutually exclusive sources with intensities $\mu_s$ and $\mu_b$. This structure is natural to consider in many contexts, but the likelihood function is complicated and direct estimation of the parameters may prove difficult even for fixed $\psi$.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

Assuming that we have separate routines available for estimating the baseline and background mortality models, we can exploit the additive structure of (4.9)–(4.10) using the EM algorithm of Dempster *et al*. (1977). It can be shown that the likelihood is increased in each step of the EM algorithm and thus converges to a local maximum, although it may do so rather slowly. It is, however, a great advantage that we can use the same top-level procedure to estimate virtually any combination of mortality models, especially when estimation routines for the underlying models are readily available. It is also straightforward to extend the EM algorithm to more than two competing risks.

Alternatively, if computational efficiency is essential, we can carry out estimation by Newton-Raphson sweeps over frailty and mortality model parameters, but this requires a substantial amount of tailor-made code. We find that optimizing the log-likelihood using the EM algorithm is both flexible, easy to implement and in our experience sufficiently fast and robust to be of practical use. The algorithm is detailed in Appendix B.

### 4.5. Forecasting

Suppose that we have estimates $(\hat{\psi}, \hat{\theta}, \hat{\eta}, \hat{\zeta}, \hat{\xi})$ available. Following the usual approach in stochastic mortality modelling, death rates are to be projected using a time-series model for the time-varying parameters $\{\theta_t, \zeta_t\}_{t\in\mathcal{T}}$. Typically, a (multidimensional) random walk with drift is used, see for example Lee and Carter (1992) or Cairns *et al*. (2006), but models with more complex structure can also be applied. Let an overbar denote projected parameters and assume that these are available for $t \in \{t_{\max}+1, \ldots, t_{\max}+h\}$ given a forecasting horizon $h \in \mathbb{N}_+$. The forecast region is illustrated as the cross-hatched box in Figure 5.

Baseline and background mortality rates are readily projected by inserting $\bar{\theta}_t$ and $\hat{\eta}_x$ into (4.2) and $\bar{\zeta}_t$ and $\hat{\xi}_x$ into (4.3). Forecasting mean frailty is slightly more involved. We notice that while it was convenient to specify mean frailty in terms of $\mathcal{M}$ for estimation purposes, it is practical to express it in terms of $\mathcal{M}_0$ when forecasting, because $\mathcal{M}_0$ can be computed directly from $F$ throughout the forecast region. Mortality is thus projected via

$$\mu(t,x) = \nu'_{\hat{\psi}}\left(\widetilde{\mathcal{M}}_0(t,x)\right) F\left(\bar{\theta}_t, \hat{\eta}_x\right) + G\left(\bar{\zeta}_t, \hat{\xi}_x\right), \tag{4.15}$$

where $\widetilde{\mathcal{M}}_0(t,x)$ in the forecast region is given by the recursion

$$\widetilde{\mathcal{M}}_0(t,x) = \begin{cases} \widetilde{\mathcal{M}}_0\left(t_{\max}, x-1\right) + F\left(\hat{\theta}_{t_{\max}}, \hat{\eta}_{x-1}\right) & \text{for } x_{\min} < x \text{ and } t_{\max}+1 = t, \\ \widetilde{\mathcal{M}}_0(t-1, x-1) + F\left(\bar{\theta}_{t_{\max}}, \hat{\eta}_{x-1}\right) & \text{for } x_{\min} < x \text{ and } t_{\max}+1 < t, \\ 0 & \text{for } x_{\min} = x. \end{cases} \tag{4.16}$$

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

We underline that $G$ does not enter $\widetilde{\mathcal{M}}_0$ in the forecast, whereas in the data window $\widetilde{\mathcal{M}}_0$ is defined by the transformation $\widetilde{\mathcal{M}}_0(t,x) = \nu_{\hat{\psi}}^{-1}(\widetilde{\mathcal{M}}(t,x))$ to ensure consistency with the estimated model.

**Example 4.2** Continuing Example 4.1 where frailty is Gamma distributed with mean one and estimated variance $\hat{\sigma}^2$, we get using (4.12) in conjunction with (4.15) that mortality should be forecasted via

$$\mu(t,x) = \left(1 + \hat{\sigma}^2 \widetilde{\mathcal{M}}_0(t,x)\right)^{-1} F(\bar{\theta}, \hat{\eta}_x) + G(\bar{\zeta}_t, \hat{\xi}_x), \tag{4.17}$$

with $\widetilde{\mathcal{M}}_0$ given by (4.16) in the forecast region. In the data window, $\widetilde{\mathcal{M}}_0(t,x)$ can be expressed as $\widetilde{\mathcal{M}}_0(t,x) = [\exp(\hat{\sigma}^2 \widetilde{\mathcal{M}}(t,x)) - 1]/\hat{\sigma}^2$ using (4.12).

## 5. An application to international mortality

We consider the implementation of the stochastic frailty model currently used at ATP and make comparisons to the usual Lee-Carter methodology. The application is based on mortality data retrieved from the Human Mortality Database (2021). To allow the reader to reproduce the results, we use Danish data to model the spread, rather than proprietary ATP data. Sections 5.1–5.3 cover reference population mortality, while Section 5.4 discusses spread modelling for target population mortality.

### 5.1. An international reference trend

The reference mortality trend, denoted $\mu_{\text{ref}}$, belongs to the class of stochastic frailty models (4.2)–(4.3) and is gender-specific although this is not shown explicitly in the notation. The model assumes Gamma distributed frailty with mean one and variance $\sigma^2$ and takes the functional form

$$\mu_{\text{ref}}(t,x) = \exp\left(-\sigma^2 \widetilde{\mathcal{M}}(t,x)\right) \mu_0(t,x) + \mu_b(t), \tag{5.1}$$

$$\mu_0(t,x) = \exp\left(\alpha_t + \beta_t(x-75) + \kappa_t(x-75)\mathbb{1}_{\{x<75\}}\right), \tag{5.2}$$

$$\mu_b(t) = \exp(\zeta_t). \tag{5.3}$$

In (5.1), the variance of the frailty distribution expresses the amount of heterogeneity in the population, but since any estimate depends on the choice of $\mu_0$, the quantity can only be interpreted in a model-specific context. On the other hand, its influence on the mortality curve is clear. If death rates increase with age, the function $x \mapsto \exp(-\sigma^2 \widetilde{\mathcal{M}}(t,x))$ decreases from one towards zero and describes how $\mu_0$ is "dragged" down by the frailty component. If $\sigma^2$ is close to zero, then mean frailty is close to one for all ages. As the variance grows, the decline in mean frailty steepens. This drags down the old-age part of the mortality curve and eventually so much that the rates fall into a decline.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

*5.1.1. A Lee-Carter baseline?*

Instead of the Gompertzian model applied in (5.2), one could use a Lee-Carter model for $\mu_0$, see Jarner (2014) for such an application. However, we find that the parametric structure in (5.2) is favourable in terms of stability and preserving the overall structure of the data and, in particular, smooth and increasing age-profiles, which is not guaranteed in long-term Lee-Carter forecasts.

Moreover, the parameters of the Lee-Carter model are not identifiable without additional constraints which precludes the use of more flexible time dynamics such as error-correction models, a limitation the parametric (identifiable) structure does not have, see for example Hunt and Blake (2017a) and Jarner and Jallbjørn (2020) for a detailed discussion.

Lastly, unlike the parametric form (5.2) that readily expands in the age dimension, that is the model extends to ages not part of the estimation, the Lee-Carter model applies only to the age span used in the calibration. This is a pronounced problem at the highest ages where data are often sparse. To obtain reliable and stable rates in both model and forecasts, one typically employs a Kannisto extension (Kannisto *et al*., 1994), or a similar procedure, for example the methods discussed in Pitacco *et al*. (2009), for the oldest-old. Irrespective of the extrapolation procedure, the coupling of two separate models adds an additional layer of complexity and defeats part of the purpose for introducing frailty, namely to capture the logistic type old-age mortality behaviour seen in the data.

*5.1.2. Selecting a suitable reference population*

To establish an international reference population, we have to decide on a suitable list of countries to use. While all countries appear to follow the same long-term trend, mortality improvements occur at different times for individual countries and variation in improvement rates differ as well. Ideally, the reference trend should consist of countries that reflect the prevalent mortality regime, but their historical development ought to be comparable as well. That is, the countries chosen ought to have undergone the same stages of demographic transition at roughly the same time.

It proves quite difficult to find a set of rules for selecting countries satisfying these broad criteria. Kjærgaard *et al*. (2016) propose an out-of-sample selection criterion as a way of constructing an "optimal" set of countries. Others have established hard inclusion criteria based on various socio-demographic-indicators, for example the Dutch actuarial society who base their official projections on a peer group of all European countries with a per capita GDP above the European average (Antonio *et al*. 2017).

We remain agnostic with regard to specific rules. Having populations entering or leaving the data pool following (annual) data updates is almost certainly bound to cause problems in terms of model stability. Our advice is to choose a wide range of countries with the intention of sticking with them in the long term. With an outset in the countries available from the Human Mortality

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

Database and excluding countries only if they clearly violate the criteria above, we are left with primarily the western part of Europe. In particular, the SAINT model is currently based on pooled data from the following 18 countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany (West), Iceland, Italy, Luxembourg, Netherlands, Norway, Scotland, Spain, Sweden, Switzerland and UK (England and Wales).

## 5.2. Time dynamics

The SAINT model (5.2)–(5.3) has four time-varying parameters that need to be forecasted (for each sex). The Makeham component $\zeta$ and the excess slope $\kappa$ are nuisance parameters included in the model to add sufficient flexibility to fit the historical data and to enhance interpretability of the level $\alpha$ and the slope $\beta$. Even though $\{\zeta_t\}_{t\in\mathcal{T}}$ and $\{\kappa_t\}_{t\in\mathcal{T}}$ appear non-stationary, see Figure 6, modelling their trending behaviour has little effect on mortality projections. Striking a compromise between model complexity and performance, we forecast these parameters using a random walk without drift, so for any horizon $h \in \mathbb{N}_0$ we have

$$\kappa_{t_{\max}+h}|\kappa_{t_{\max}} \sim \mathcal{N}\left(\kappa_{t_{\max}}, h\sigma_\kappa^2\right) \quad \text{and} \quad \zeta_{t_{\max}+h}|\zeta_{t_{\max}} \sim \mathcal{N}\left(\zeta_{t_{\max}}, h\sigma_\zeta^2\right), \quad (5.4)$$

where $\sigma_\kappa, \sigma_\zeta \in \mathbb{R}_+$.

Projecting $\{\alpha_t, \beta_t\}$ is a more delicate task. In particular, some thought should go into the joint behaviour of the gender-specific forecasts. It is a well-established fact that women live longer than men and while this gender gap varies over time it is believed to persist. Since forecasting even closely related populations independently will lead to diverging forecasts, cf. Tuljapurkar *et al*. (2000), we must deal with the problem through joint modelling in order not to have undesirable scenarios such as projecting men to live longer than women.

### *5.2.1. An error-correction model*

To ensure that female and male parameters "stay together," not just in a median forecast but also for every stochastic realization, we need parameters to cointegrate, that is a given linear combination of them should be stationary. This is achieved by forecasting from the error-correction model

$$\Delta Y_t = AB^\top Y_{t-1} + C + \omega_t, \quad (5.5)$$

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

where $\omega_t \overset{\text{iid}}{\sim} \mathcal{N}_4(0, \Omega)$, $\Delta Y_t = Y_t - Y_{t-1}$ and

$$Y_t = \begin{pmatrix} \alpha_t^f \\ \alpha_t^m \\ \beta_t^f \\ \beta_t^m \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & 0 \\ a_{21} & 0 \\ 0 & a_{32} \\ 0 & a_{42} \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, \quad C = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}, \tag{5.6}$$

with superscript $f$ and $m$ denoting female and male parameter values, respectively. Equation (5.6) contains two critical assumptions:

1. Structural zero's have been placed in the $A$-matrix making parameter pairs weakly exogenous, so that the relation between the $\alpha$'s will not affect the relation between the $\beta$'s and vice versa.
2. The $B$-matrix imposes two cointegrating relations, one between each parameter pair, with coefficients of unity so that the model corrects any disequilibrium that may arise in the difference between the parameters.

The error-correction behaviour of (5.5) can be made more precise by decomposing the drift term. For a $p \times q$ matrix $X$ of full rank, we say that another matrix $X_\perp$ of full rank and dimension $p \times (p-q)$ such that $X^\top X_\perp = 0$ is its orthogonal complement. With $M = A(B^\top A)^{-1} B^\top$, it can be shown that $I - M = B_\perp (A_\perp^\top B_\perp)^{-1} A_\perp^\top$, cf. Chapter 3 of Johansen (1995). Using this identity, we can rewrite (5.5) so that

$$\Delta Y_t = A\left(B^\top Y_{t-1} - C_D\right) + C_\Delta + \omega_t, \tag{5.7}$$

where $C_D = -(B^\top A)^{-1} B^\top C = \left(C_{D_\alpha}, C_{D_\beta}\right)^\top$ is the parameter difference in stationarity with $C_{D_\alpha} = \frac{c_1 - c_2}{a_{21} - a_{11}}$ and $C_{D_\beta} = \frac{c_3 - c_4}{a_{42} - a_{32}}$, while $C_\Delta = (I - M)C = (C_{\Delta_\alpha}, C_{\Delta_\alpha}, C_{\Delta_\beta}, C_{\Delta_\beta})^\top$ is the common drift with $C_{\Delta_\alpha} = \frac{a_{21}c_1 - a_{11}c_2}{a_{21} - a_{11}}$ and $C_{\Delta_\beta} = \frac{a_{42}c_3 - a_{32}c_4}{a_{42} - a_{32}}$. From Equation (5.7), it is clear $Y_t$ is updated in response to the disequilibrium error $B^\top Y_{t-1} - C_D$ with a force depending on the magnitude of the $a$'s.

An explicit representation of the (median) forecast for a horizon $h \in \mathbb{N}_0$ can be discerned via the Granger representation theorem, see for example Jarner and Jallbjørn (2020). We have

$$\begin{aligned} \mathbb{E}\left[Y_{t_{\max}+h} \middle| Y_{t_{\max}}\right] &= Y_{t_{\max}} + C_\Delta h \\ &\quad - A\left(B^\top A\right)^{-1} \begin{pmatrix} 1 - \lambda_\alpha^h & 0 \\ 0 & 1 - \lambda_\beta^h \end{pmatrix} \left(B^\top Y_{t_{\max}} - C_D\right), \end{aligned} \tag{5.8}$$

where $\lambda_\alpha = 1 + a_{11} - a_{21}$ and $\lambda_\beta = 1 + a_{32} - a_{42}$ are eigenvalues of $I + AB^\top$. Equation (5.8) highlights the asymptotic random walk behaviour, with the

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

initial disequilibrium error decaying exponentially to zero provided that $\lambda_\alpha, \lambda_\beta < 1$.

Writing out the error-correction behaviour helps us identify means of ensuring stable and robust forecasts. With this aim in mind, adjusting $C_D$ and $C_\Delta$ to equal desired long-term values might be preferable compared to unconstrained estimation. Equation (5.8) details how the target equilibrium may severely affect projections. It shows that the error-correction model can bring about a number of undesirable features in the median forecast, like short- to medium-term increases in mortality for one of the genders during the restoration of the equilibrium. It is difficult to justify such behaviours in best estimate projections. To avoid them, we assume that $B^\top Y_t$ is distributed according to its stationary distribution by equating $C_D$ to the difference between jump-off values, namely $\widehat{C}_{D_\theta} = \widehat{\theta}^f_{t_{\max}} - \widehat{\theta}^m_{t_{\max}}$ where $\theta$ is either $\alpha$ or $\beta$. This makes the median forecast of the error-correction model coincide with the median forecast of a random walk model, since the ultimate term in (5.8) vanishes. The error-correction interpretation still applies, however, when considering a stochastic realization of the process. Moreover, looking at the evolution of mortality through the years, for example Figure 3, it is clear that female mortality has developed more steadily than mortality for the males. We therefore use the empirical average of the female parameters to model the common slope, that is $\widehat{C}_{\Delta_\theta} = \frac{1}{t_{\max} - t_{\min} + 1} \sum_t \Delta\widehat{\theta}^f_t$ where $\theta$ is either $\alpha$ or $\beta$.

### 5.3. Model fit and forecast

Since international mortality develops more steadily than country-specific mortality, we are able to use a relatively wide window of data for model calibration. The endpoints should be chosen such that we do not introduce structural breaks if data for some countries are missing. Balancing these concerns, we apply the model to international mortality data, ages 20–100 and calendar years 1970–2017 with 2017 being the last year where data exist for all the countries considered at the time of writing. The model is estimated separately for females and males, and follows the EM algorithm described in Appendix B.

The estimated parameters of the SAINT model (5.2)–(5.3) are shown in Figure 6. To justify the use of the error-correction model (5.5), the series $\{\alpha_t\}_{t\in\mathcal{T}}$ and $\{\beta_t\}_{t\in\mathcal{T}}$ must be integrated of order 1 for both genders, that is the stochastic part of these processes must be non-stationary. We conclude from unit root tests (test results not shown) that this is indeed the case. Further, we take the observed stable difference between parameters as (weak) evidence that they engage in an equilibrium correcting relationship.

In mortality forecasting applications, the choice of jump-off point is a key consideration to match the start of the projection with recently observed data (Lee and Miller 2001). Because the SAINT model fits the empirical death rates very well, cf. Figure 2, no jump-off correction is needed and we use the

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

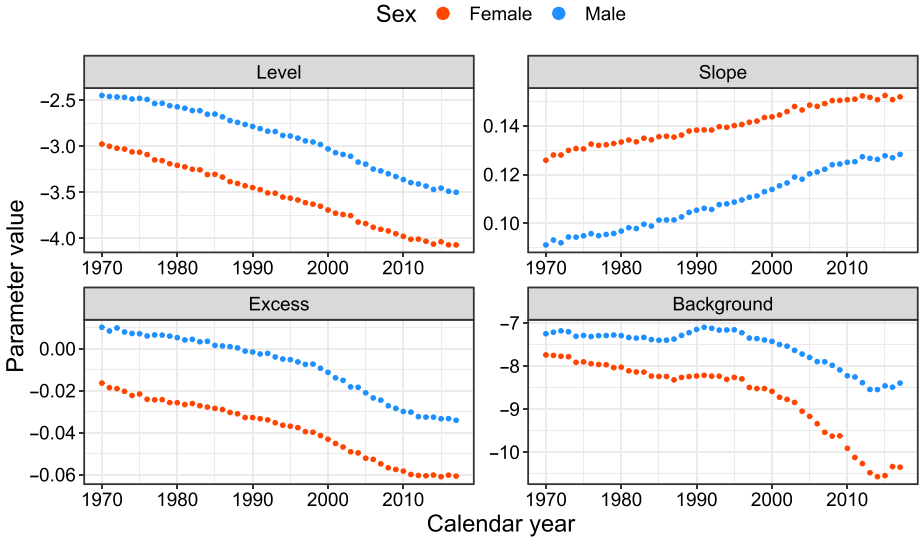


FIGURE 6. Estimated parameters for baseline (5.2) and background (5.3) mortality in the SAINT model.

model values in the jump-off year to determine the jump-off rates. Moreover, the Poisson assumption (4.4) ensures that the model approximates the total number of deaths in the data, see also the discussion in Brouhns *et al*. (2002).

*5.3.1. Comparing SAINT and Lee-Carter forecasts*

To put forecasting into perspective, the SAINT projections are compared to projections generated by a Lee-Carter model. The Lee-Carter model is, in the spirit of this paper, estimated under the Poisson assumption as in Brouhns *et al*. (2002), and we use a random walk with drift for forecasting as is customary. Extrapolating the time-varying index linearly implies that age-specific death rates decay exponentially at a *constant* rate. We shall see that this assumption causes forecasts based on the Lee-Carter methodology to have a tendency of underestimating the actual gains in old-age mortality. The purpose of the comparison is not to show superiority of SAINT over other models, but to illustrate the beneficial effects of cointegration and frailty.

Forecasts are compared by considering period life expectancies. While cohort life expectancies taking future improvements of mortality into account are generally of more interest, the period life expectancy summarizes the level of mortality at a given time and is better suited for illustrative purposes as it can be compared with the actual experience. Recently, Arnold *et al*. (2019) added perspective on the stability of period versus cohort tables, arguing that the former might be preferable for practitioners looking to minimize capital adjustments following life tables updates.

Historical and forecasted life expectancies of a 60-year-old are shown in Figure 7. The figure shows that both models produce highly similar median

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

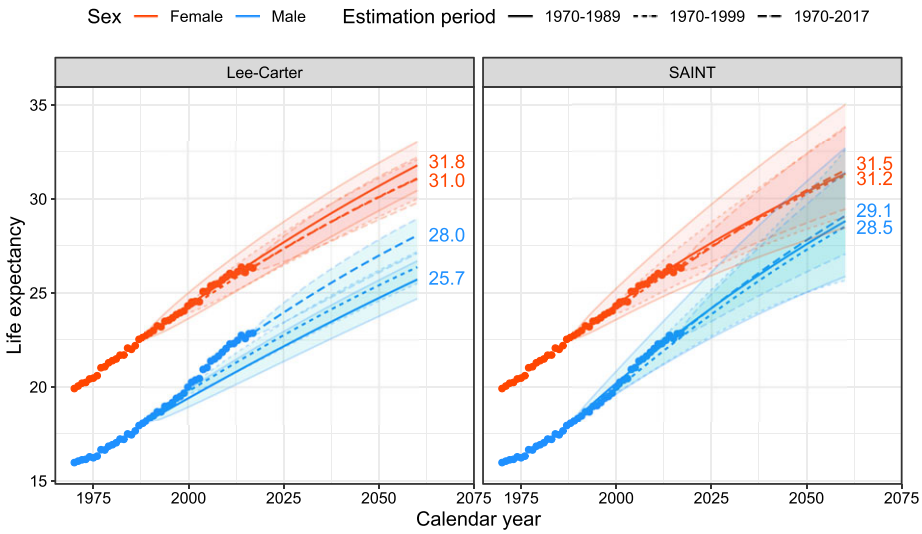


FIGURE 7. The panels show actual (dots) and forecasted (lines) remaining life expectancies for age 60 females and males with pointwise 95% confidence bands based on expanding estimation windows. Even though a rolling fixed-length data window is a more common back test approach in the literature, an expanding data window corresponds to how a mortality model is typically updated in practice.

forecasts for female life expectancy and that predictions are quite robust to the choice of the data window used to calibrate the model. This similarity can be attributed to the stable rates of improvements observed since the 1970s. For the males, however, the improvement rates considered during the period of estimation are considerably lower than the actual rates in the periods that follow. The Lee-Carter model fails to capture this development, and the median forecasts are a good way from agreeing with the actual experience. Since improvement rates are increasing over time, predictions from the Lee-Carter model grow increasingly optimistic as we include more recent data.

The SAINT model lends its strength in the stable female improvement rates by the coupling of the genders described in Section 5.2. This leads to the forecasts being decidedly more robust to the choice of estimation period, and the resulting almost linear median life expectancy projections resemble the actual experience better than the scattered Lee-Carter forecasts. In fact, it is quite remarkable how well even the first SAINT forecast based on 1970-1989 predicts present day male life expectancy.

To contrast the forecasting uncertainty of the two methods, Table 1 reports summary statistics for the projected life time distributions based on 10,000 simulations. Both the table and the figure reveal that there is a major difference in the forecasting uncertainty between the two methods, even when point estimates are similar. It is evident that the uncertainty is greatest in the SAINT model, while it is, at least with hindsight, worryingly low for the Lee-Carter model. Specifically for the males using the two earliest calibrations, we get no

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

TABLE 1.

EMPIRICAL MEAN (AND STANDARD DEVIATION IN PARENTHESES) OF PROJECTED PERIOD LIFE EXPECTANCIES FOR A 60-YEAR-OLD BASED ON 10,000 SIMULATIONS FOR SELECT YEARS.

| | Female | | | Male | | |
|---|---|---|---|---|---|---|
| Model | 2020 | 2040 | 2060 | 2020 | 2040 | 2060 |
| Calibration period: 1970–1989 | | | | | | |
| Lee-Carter | 27.01 (0.54) | 29.50 (0.62) | 31.74 (0.65) | 21.63 (0.38) | 23.73 (0.46) | 25.71 (0.51) |
| SAINT | 27.06 (0.81) | 29.38 (1.15) | 31.40 (1.47) | 23.46 (0.76) | 26.35 (1.14) | 28.88 (1.51) |
| Calibration period: 1970–1999 | | | | | | |
| Lee-Carter | 26.73 (0.41) | 29.04 (0.49) | 31.04 (0.52) | 22.18 (0.31) | 24.39 (0.40) | 26.36 (0.43) |
| SAINT | 26.86 (0.53) | 29.22 (0.78) | 31.26 (1.07) | 23.12 (0.53) | 26.04 (0.90) | 28.65 (1.49) |
| Calibration period: 1970–2017 | | | | | | |
| Lee-Carter | 26.72 (0.22) | 29.02 (0.52) | 31.01 (0.61) | 23.28 (0.16) | 25.83 (0.40) | 28.04 (0.47) |
| SAINT | 26.75 (0.25) | 29.31 (0.71) | 31.51 (1.00) | 23.39 (0.25) | 26.44 (0.71) | 29.09 (1.04) |

warning that the projections might be well off target; the models do not capture the final years of observed data, let alone the 1970–2017 model's median projection, within their 95%-confidence bands.

Figure 8 compares the average improvement rates over a long horizon for select high ages. For the females, improvement rates are similar in both models over the short term; the height of the light-red bars align with the dashed lines. On longer horizons, however, the SAINT model gives rise to increasing rates of improvements endogenously within the model as the frailty composition changes over time. The Lee-Carter model and many of its descendants cannot predict rates of improvements higher than those observed historically and are therefore likely to be understating future increases in old-age mortality, even for large and stable datasets. In fact, the likeness between forecasts seen in Figure 7 is deceptive; the cohort life expectancy for a newborn female in 2017 is 94.81 years in the SAINT model, but nearly 2 years less in the Lee-Carter prediction at 92.96 years. For the males, the overall level of the improvement rates is higher in the SAINT forecast as a result of the genders being tied together compared to the Lee-Carter predictions where this is not so.

### 5.4. Spread modelling

Given an underlying model for reference population mortality, we model mortality in the target population as deviations from the trend. As the name suggests, the trend should capture the main features of the mortality evolution and the spread should therefore be a model flexible enough to adequately fit observed mortality in the target population but introduce as little complexity as possible.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

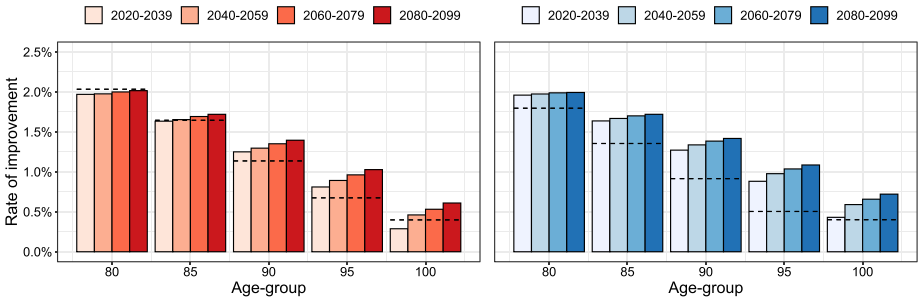


FIGURE 8. The panels show predicted average rates of improvement for select ages using the SAINT model estimated on the full data period 1970–2017. Female rates are shown in the left panel and male rates in the right. The corresponding improvement rates for the Lee-Carter model are superimposed as the dashed lines.

Since Plat (2009), common practice is to take a regression approach using a linear structure

$$D_{\text{target}}(t,x) \sim \text{Poisson}\left(\mu_{\text{target}}(t,x)E_{\text{target}}(t,x)\right), \tag{5.9}$$

$$\mu_{\text{target}}(t,x) = \mu_{\text{ref}}(t,x)\exp\left(y_t^\top r_x\right), \tag{5.10}$$

with parameters $y_t = (y_{1,t},\ldots,y_{p,t})^\top \in \mathbb{R}^p$ describing the evolution of the spread over time and age-dependent regressors $r_x = (r_{1,x},\ldots,r_{p,x})^\top \in \mathbb{R}^p$. Any standard GLM-routine can be used to fit the model by specifying a Poisson family with a log-link.

The original version of SAINT used a parsimonious model for the spread describing just its level, slope and curvature. While this model performed respectably, it became clear from a practical point of view that the fit of target population mortality was simply unconvincing when plotted on log-scale against empirical data, a plot frequently reported to the FSA and the Board of Representatives. To improve the fit, the three regressor model was replaced with a five regressor model

$$r_{i,x} = \min\left(1, \max\left(0, x_i - x\right)/20\right), \quad i \in \{1,\ldots,5\}, \tag{5.11}$$

where $(x_1, x_2, x_3, x_4, x_5) = (40, 60, 80, 100, 120)$, so that regressors are one until a given breakpoint after which they decrease linearly to zero over the course of 20 years. Even though a model with evenly spaced knots has obvious practical advantages, there are now methods available to explore the optimal choice of the number and location of the knots in spline models, see Kaishev *et al.* (2016). The chief reason for choosing the regressors above is that $r_{2,x}$, $r_{3,x}$, and $r_{4,x}$ are equivalent to the three regressor model specified by the Danish FSA for their longevity benchmark, cf. Jarner and Møller (2015), giving credence to the credibility of (5.11) in the eyes of the regulator.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

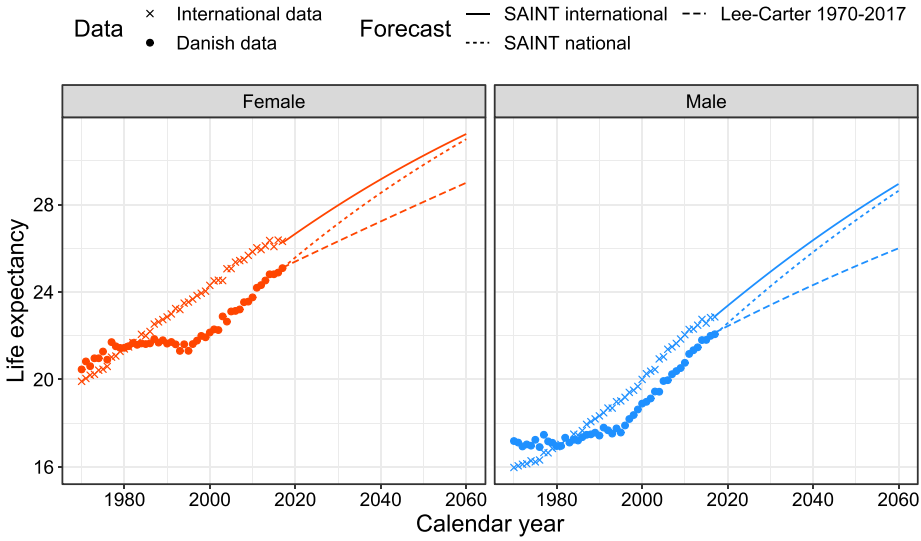


FIGURE 9. Actual and forecasted remaining period life expectancy for a 60-year-old.

*5.4.1. Spread forecast*

Since it is the trend that governs long-term mortality behaviour, the time dynamics used for forecasting should ensure that the spread remains bounded in probability. We use a (stable) vector autoregressive model,

$$y_t = Ay_{t-1} + \upsilon_t, \tag{5.12}$$

with $A$ being a $p \times p$ matrix and $\upsilon_t$ mean-zero Gaussian errors. We note that even if data appear stationary, for example Figure 9, maximum likelihood estimation of the VAR-model (5.12) does not necessarily yield stationarity and more sophisticated models, for example including additional lags, may have to be used. Alternatively, stationarity can be imposed by putting a curb on the eigenvalues of $A$ such that they lie within the unit circle.

In practice, restricting $A$ to be a diagonal matrix with diagonal elements $0 \leq A_{ii} < 1$, works well in terms of stability and is easy to communicate to non-specialists as deviation half-life. For instance, the Danish mortality curve moves more or less in unison with the international trend and there are no signs of an impending catch-up. Simply imposing $A = 0.99I$, with $I$ being the identity matrix of size $p$, corresponds to a deviation half-life of about 69 years and results in the forecast depicted in Figure 9.

The figure epitomizes the advantage of the SAINT projection methodology. In a typical country-specific projection, exemplified by the Lee-Carter forecast, the recent stagnation in Danish life expectancy makes the data window a critical component of the analysis. The Lee-Carter model extrapolates past trends, and even if based on very recent data periods, it is likely to understate future improvements in especially old-age mortality. Using the SAINT methodology, the long-term trend is determined by a large pooled set of countries which

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

serves as a stable reference for small populations that exhibit substantial variability. Rather than basing long-term projections on irregular national rates of improvements, we can frame the development as short- to medium-term deviations from the trend. Having a point of reference also makes it possible – even for non-specialists – to visually gauge if the projection is reasonable.

Although only the Danish data are modelled here, the SAINT framework is easily adapted to other populations of interest. All that is required is a separate VAR model for the spread between the (new) target population and the trend. The assumed stationarity of the spread will ensure that deviations from the trend are bounded. Multi-population analyses will therefore not exaggerate short-term differences or lead to diverging projections. All in all the SAINT methodology opens for a unified, flexible and robust forecasting approach, which is applicable to a wide range of populations despite their possibly unsteady development.

## 6. Concluding remarks

A large and growing body of literature focuses on theoretical properties of mortality models and achieving accurate mortality forecasts. However, there are a number of desirable model properties besides theoretical ones that determine the success and applicability of a mortality model in practice. In this paper, we reviewed the lessons learned from more than a decade of experience with the SAINT model in use at ATP and the model modifications that have followed.

The main improvement over the original SAINT model was the generalization of frailty models from deterministic structures to a flexible class of stochastic mortality models, offering a general way of combining essentially any mortality model with frailty. This sets the work apart from the typical use of frailty which rely on matching parametric forms with closed-form expressions. We demonstrated how frailty extended mortality models dramatically improve the historical fit of even simple age-period structures and provide more realistic projections of improvement patterns on longer horizons. Obviously, frailty on its own is not enough to explain the complex dynamics of mortality, but helps capture essential features observed in the data otherwise addressed by ad hoc methods.

Although the original SAINT model was explicitly designed with stability in mind, we underestimated the effect that annual data updates could have on best estimate projections. This lack of robustness was primarily rooted in the model's time dynamics whose parameters were estimated freely. In the trend, the long-run equilibrium relation between male and female mortality was overly sensitive to small changes in observed patterns, while the autoregressive model used to forecast the spread between reference and target population mortality did not guarantee stationarity. The issues were resolved by imposing sensible structural restrictions on the time dynamics of the model.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

Even with well-behaved time dynamics, model stability is still challenged by outliers. In light of the COVID-19 crisis, it may not come as a surprise that a reference population – even when built from a large and geographically dispersed group of countries – will not guarantee stability under annual data updates. Nevertheless, we were taken by surprise by how deeply events like severe flu seasons impact reference mortality levels. Because of our globally connected world, infectious disease outbreaks cannot be diversified away by simply adding more countries to the data pool. We expect these types of fluctuations to be temporary rather than part of a lasting trend and measures to dampen their effect are indispensable in practical applications. This is an important issue for future research.

While accuracy, stability and flexibility requirements all pertain to model performance, there are certain properties in practical applications that do not. Governance rules entail an obligation on the part of the modeller to report, explain and justify outputs, some of which might not even have any direct practical implication, for example, in a pension fund context, how fitted death rates among the very young compare to actual rates. A poor fit in this age group will not affect the calculation of technical provisions in any way but may seriously detract from the model's credibility in the eyes of non-specialists. As modellers, we should be aware of this, partly external, desire for model explainability.

## Acknowledgements

The authors are indebted to Esben Masotti Kryger for numerous stimulating discussions. The authors wish to thank two anonymous referees for their valuable input which helped improve the manuscript. The work was partly funded by Innovation Fund Denmark (IFD) under File No. 9065-00135B.

## Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/asb.2021.37.

## References

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

Antonio, K., Devriendt, S., Boer, W., Vries, R., Waegenaere, A., Kan, H.-K., Kromme, E., Ouburg, W., Schulteis, T., Slagter, E., Winden, M., Iersel, C. and Vellekoop, M. (2017) Producing the Dutch and Belgian mortality projections: A stochastic multi-population standard. *European Actuarial Journal*, **7**(2), 297–336.

Arnold, S., Jijiie, A., Jondeau, E. and Rockinger, M. (2019) Periodic or generational actuarial tables: Which one to choose? *European Actuarial Journal*, **9**(2), 519–554.

BESAG, J. (1975) Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **24**(3), 179–195.
BONGAARTS, J. (2005) Long-range trends in adult mortality: Models and projection methods. *Demography*, **42**(1), 23–49.
BOOTH, H. (2008) Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, **3**(1–2), 3–43.
BOOTH, H., MAINDONALD, J. and SMITH, L. (2002) Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, **56**(3), 325–336.
BÖRGER, M., FLEISCHER, D. and KUKSIN, N. (2014) Modeling the mortality trend under modern solvency regimes. *ASTIN Bulletin*, **44**(1), 1–38.
BROUHNS, N., DENUIT, M. and VERMUNT, J.K. (2002) A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance Mathematics and Economics*, **31**(3), 373–393.
BUTT, Z. and HABERMAN, S. (2004) Application of frailty-based mortality models using generalized linear models. *ASTIN Bulletin*, **34**(1), 175–197.
CAIRNS, A., BLAKE, D., DOWD, K., COUGHLAN, G., EPSTEIN, D., ONG, A. and BALEVICH, I. (2009) A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *North American Actuarial Journal*, **13**, 1–35.
CAIRNS, A.J., BLAKE, D., DOWD, K., COUGHLAN, G.D. and KHALAF-ALLAH, M. (2011) Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, **41**(1), 29–59.
CAIRNS, A.J.G., BLAKE, D. and DOWD, K. (2006) A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, **73**(4), 687–718.
Continuous Mortality Investigation (2016) CMI Mortality Projections Model-Working Paper 90. Technical report, Institute and Faculty of Actuaries. https://www.actuaries.org.uk/system/files/field/document/CMI%20WP090%20v03%202016-08-31%20-%20CMI%20Model%20consultation.pdf.
DEMPSTER, A.P., LAIRD, N.M. and RUBIN, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**(1), 1–22.
DOWD, K., CAIRNS, A.J.G., BLAKE, D., COUGHLAN, G.D. and KHALAF-ALLAH, M. (2011) A gravity model of mortality rates for two related populations. *North American Actuarial Journal*, **15**(2), 334–356.
DUCHATEAU, L. and JANSSEN, P. (2008) *The Frailty Model*, 1st ed. Statistics for Biology and Health. New York, NY: Springer.
GAO, G. and SHI, Y. (2021) Age-Coherent extensions of the Lee–Carter model. *Scandinavian Actuarial Journal*, **2021**(10), 1–19.
HABERMAN, S. and RENSHAW, A. (2012) Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and Economics*, **50**(3), 309–333.
HOUGAARD, P. (1984) Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika*, **71**(1), 75–83.
HOUGAARD, P. (1986) Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**(2), 387–396.
HOUGAARD, P. (2012) Analysis of Multivariate Survival Data. Berlin: Springer Science & Business Media.
Human Mortality Database (2021) http://www.mortality.org. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Data downloaded May 2021.
HUNT, A. and BLAKE, D. (2017a) Identifiability, cointegration and the gravity model. *Insurance: Mathematics and Economics*, **78**, 360–368.
HUNT, A. and BLAKE, D. (2017b) Modelling mortality for pension schemes. *ASTIN Bulletin*, **47**(2), 601–629.
JANSSEN, F. (2018) Advances in mortality forecasting: Introduction. *Genus*, **74**(1), 1–12.
JARNER, S.F. (2014) Stochastic frailty models for modeling and forecasting mortality. arXiv preprint arXiv:2109.02584, pp. 1–37.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

JARNER, S.F. and JALLBJØRN, S. (2020) Pitfalls and merits of cointegration-based mortality models. *Insurance: Mathematics and Economics*, **90**, 80–93.
JARNER, S.F. and KRYGER, E. (2011) Modelling adult mortality in small populations: The SAINT model. *ASTIN Bulletin*, **41**, 377–418.
JARNER, S.F. and MØLLER, T. (2015) A partial internal model for longevity risk. *Scandinavian Actuarial Journal*, **2015**(4), 352–382.
JOHANSEN, S. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.
KAISHEV, V., DIMITROVA, D., HABERMAN, S. and VERRALL, R. (2016) Geometrically designed, variable knot regression splines. *Computational Statistics*, **31**(3), 1079–1105.
KANNISTO, V., LAURITSEN, J., THATCHER, A.R. and VAUPEL, J.W. (1994) Reductions in mortality at advanced ages: Several decades of evidence from 27 countries. *Population and Development Review*, **20**(4), 793–810.
KJÆRGAARD, S., CANUDAS-ROMO, V. and VAUPEL, J. (2016) The importance of the reference populations for coherent mortality forecasting models. European Population Conference; Conference date: 31-08-2016 Through 03-09-2016.
LEE, R. and MILLER, T. (2001) Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, **38**(4), 537–549.
LEE, R.D. and CARTER, L.R. (1992) Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, **87**(419), 659–671.
LI, H. and LI, J. (2017) Optimizing the Lee-Carter approach in the presence of structural changes in time and age patterns of mortality improvements. *Demography*, **54**(3), 1073–1095.
LI, J., LI, J.S.-H., TAN, C.I. and TICKLE, L. (2019) Assessing basis risk in index-based longevity swap transactions. *Annals of Actuarial Science*, **13**(1), 166–197.
LI, J. and LIU, J. (2019) A logistic two-population mortality projection model for modelling mortality at advanced ages for both sexes. *Scandinavian Actuarial Journal*, **2019**(2), 97–112.
LI, J.S.-H., HARDY, M.R. and TAN, K.S. (2009) Uncertainty in mortality forecasting: An extension to the classical Lee-Carter approach. *ASTIN Bulletin*, **39**(1), 137–164.
LI, N. and LEE, R. (2005) Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, **42**(3), 575–594.
LI, N., LEE, R. and GERLAND, P. (2013) Extending the Lee-Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography*, **50**(6), 2037–2051.
MENZIETTI, M., MORABITO, M.F. and STRANGES, M. (2019) Mortality projections for small populations: An application to the Maltese elderly. *Risks*, **7**(2), 35.
Olivieri, A. (2006) Heterogeneity in survival models - Applications to pensions and life annuities. *Belgian Actuarial Bulletin*, **6**, 23–39.
PITACCO, E., DENUIT, M., HABERMAN, S. and OLIVIERI, A. (2009) Modelling Longevity Dynamics for Pensions and Annuity Business. London: Oxford University Press.
PLAT, R. (2009) Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance Mathematics and Economics*, **45**(1), 123–132.
RENSHAW, A. and HABERMAN, S. (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, **38**, 556–570.
SPREEUW, J., NIELSEN, J. and JARNER, S.F. (2013) A nonparametric visual test of mixed hazard models. *Statistics and Operations Research Transactions*, **37**, 149–170.
THATCHER, A.R. (1999) The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society: Series A, (Statistics in Society)*, **162**(1), 5–43.
TULJAPURKAR, S., LI, N. and BOE, C. (2000) A universal pattern of mortality decline in the G7 countries. *Nature*, **405**(6788), 789–792.
VAUPEL, J.W., MANTON, K.G. and STALLARD, E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**(3), 439–454.
VILLEGAS, A.M. and HABERMAN, S. (2014) On the modeling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in England. *North American Actuarial Journal*, **18**(1), 168–193.
VILLEGAS, A.M., HABERMAN, S., KAISHEV, V.K. and MILLOSSOVICH, P. (2017) A comparative study of two-population models for the assessment of basis risk in longevity hedges. *ASTIN Bulletin*, **47**(3), 631–679.

VÉKÁS, P. (2019) Rotation of the age pattern of mortality improvements in the European Union. *Central European Journal of Operations Research*, **28**(3), 1–18.
WAN, C. and BERTSCHI, L. (2015) Swiss coherent mortality model as a basis for developing longevity de-risking solutions for Swiss pension funds: A practical approach. *Insurance: Mathematics and Economics*, **63**, 66–75.
WANG, S. and BROWN, R. (1998) A frailty model for projection of human mortality improvements. *Journal of Actuarial Practice*, **6**, 221–241.
WIENKE, A. (2010) *Frailty models in Survival Analysis*. Chapman & Hall/CRC Biostatistics Series. Boca, Raton: Taylor & Francis.

SØREN F. JARNER
*Department of Mathematical Science, University of Copenhagen, Copenhagen, Denmark*
*E-Mail: soren@jarner.dk*

SNORRE JALLBJØRN (Corresponding author)
*Danish Labour Market Supplementary Pension Fund (ATP), Kongens Vænge 8, 3400 Hillerød, Denmark*
*E-Mail: sjb@atp.dk*

## A. POSITIVE STABLE FRAILTY

Hougaard (1986) introduced a family of generalized stable laws which includes the two most often used frailty distributions for mortality modelling, namely the Gamma and inverse Gaussian distributions, see for example Vaupel *et al*. (1979), Hougaard (1984), Butt and Haberman (2004), Jarner and Kryger (2011), Spreeuw *et al*. (2013). The family is obtained by exponential tilting of stable densities with index $\alpha \in [0, 1)$. The stable laws themselves only have moments of order strictly less than $\alpha$, while moments of all orders exist for the exponentially tilted densities. From the original three-parameter family, we obtain a two-parameter family by imposing the condition that mean frailty is one.

When $Z$ follows a generalized stable law with index $\alpha \in [0, 1)$, mean one and variance $\sigma^2$ the Laplace transform and mean frailty are given by

$$\mathcal{L}(s) = \exp\left(\frac{1-\alpha}{\alpha}\left[\frac{1-\left(1+\frac{\sigma^2 s}{1-\alpha}\right)^{\alpha}}{\sigma^2}\right]\right), \tag{A1}$$

$$\mathbb{E}[Z|t,x] = \left(1+\frac{\sigma^2}{1-\alpha}\mathcal{M}_0(t,x)\right)^{\alpha-1} = \left(1+\frac{\alpha}{1-\alpha}\sigma^2\mathcal{M}(t,x)\right)^{\frac{\alpha-1}{\alpha}}. \tag{A2}$$

The stated formulas are obtained from Hougaard (1986) using the parametrization $\theta = (1-\alpha)/\sigma^2$ and $\delta = [(1-\alpha)/\sigma^2]^{1-a}$.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

The generalized stable law specializes to the inverse Gaussian distribution for $\alpha = 1/2$ and to the Gamma distribution for the limiting case $\alpha = 0$ defined by continuity. While both Gamma and inverse Gaussian densities exist in closed form, generally (tilted) stable densities exist only as a series representation, cf. Hougaard (1984). Since the closed form expressions for the Laplace transform and mean frailty (A1)–(A2) are all we need for estimation and forecasting purposes, this is not problematic.

Arguably, the Gamma distribution is the most widely used frailty distribution. It is well-known that Gamma frailty in combination with Gompertz or Makeham baseline intensities leads to a cohort rate of the logistic type and has been found to describe old-age mortality very well, see for example Thatcher (1999), Cairns *et al*. (2006). Gamma distributed frailty is also mathematically tractable and allows explicit calculations of many quantities of interest, for example frailty among survivors at a given age is Gamma distributed with known scale and shape parameters, cf. Vaupel *et al*. (1979).

It is generally found that Gamma frailty and the associated logistic form provides a better description of old-age mortality than inverse Gaussian frailty, see for example Butt and Haberman (2004), Spreeuw *et al*. (2013). Furthermore, Abbring and van den Berg (2007) show that for a large class of initial frailty distributions the frailty distribution among survivors converges to a Gamma distribution as the cumulated rate tends to infinity. Thus, overall the Gamma distribution is a good default choice.

## B. Estimation of a competing risks model

We consider the competing risk model of (4.9)–(4.10), that is

$$D(t,x) \sim \text{Poisson}\left(\left[\mu_s\left(\psi,\theta_t,\eta_x,\zeta,\xi\right)+\mu_b\left(\zeta_t,\xi_x\right)\right]E(t,x)\right), \tag{B1}$$

where we have made $\mu_s$'s dependence on the vector of background parameters explicit. Selective and background mortality are given by

$$\mu_s\left(\psi,\theta_t,\eta_x,\zeta,\xi\right) = \nu_\psi{}'\left(\nu_\psi^{-1}\left\{\widetilde{\mathcal{M}}\left(t,x,\zeta,\xi\right)\right\}\right)F\left(\theta_t,\eta_x\right), \tag{B2}$$

$$\mu_b(\zeta_t,\xi_x) = G(\zeta_t,\xi_x), \tag{B3}$$

for a fixed value of $\psi$. Imagine that deaths were recorded according to the (hidden) sources

$$D_s(t,x) \sim \text{Poisson}\left(E(t,x)\mu_s(\psi,\theta_t,\eta_x,\zeta,\xi)\right), \tag{B4}$$

$$D_b(t,x) \sim \text{Poisson}\left(E(t,x)\mu_b(\zeta_t,\xi_x)\right), \tag{B5}$$

with $D_s$ and $D_b$ mutually exclusive so that $D(t,x) = D_s(t,x) + D_b(t,x)$. Even though $D_s$ and $D_b$ do not necessarily exist and hence are not "missing" in the usual sense of the word, we can still use the EM algorithm based on the missing data interpretation of the model. Note that $\zeta$ and $\xi$ are estimated from (B5),

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

while $\theta$ and $\eta$ are estimated from (B4) with $\zeta$ and $\xi$ kept fixed at their current value.

Omitting time and age indices for ease of notation the expectation step is then to compute the complete data log-likelihood given death counts $D$ and current parameter estimates from iteration $i-1$,

$$Q(\theta,\eta,\zeta,\xi)=\mathbb{E}[\ell(\theta,\eta,\zeta,\xi)\mid D,\theta^{i-1},\eta^{i-1},\zeta^{i-1},\xi^{i-1}], \tag{B6}$$

where a death is distributed according to a Bernoulli trial with a success parameter depending on the weight of the cause-specific death rate, that is

$$D_s|D,\theta^{i-1},\eta^{i-1},\zeta^{i-1},\xi^{i-1} \sim \text{Bin}\left(D,\frac{\mu_s\left(\theta^{i-1},\eta^{i-1},\zeta^{i-1},\xi^{i-1}\right)}{\mu_s\left(\theta^{i-1},\eta^{i-1},\zeta^{i-1},\xi^{i-1}\right)+\mu_b\left(\zeta^{i-1},\xi^{i-1}\right)}\right), \tag{B7}$$

$$D_b|D,\theta^{i-1},\eta^{i-1},\zeta^{i-1},\xi^{i-1} \sim \text{Bin}\left(D,\frac{\mu_b\left(\zeta^{i-1},\xi^{i-1}\right)}{\mu_s\left(\theta^{i-1},\eta^{i-1},\zeta^{i-1},\xi^{i-1}\right)+\mu_b\left(\zeta^{i-1},\xi^{i-1}\right)}\right), \tag{B8}$$

under the assumption that the likelihood factorizes so that

$$\ell(\theta,\eta,\zeta,\xi)=\ell_s(\theta,\eta\mid D_s,\zeta^{i-1},\xi^{i-1})+\ell_b(\zeta,\xi\mid D_b). \tag{B9}$$

The maximization step consists of maximizing (B6) to obtain the $i$'th parameter estimate, that is estimating the two marginal mortality models with death counts

$$D_s=\mathbb{E}\left[D_s\mid D,\theta^{i-1},\eta^{i-1},\zeta^{i-1},\xi^{i-1}\right], \tag{B10}$$

$$D_b=\mathbb{E}\left[D_b\mid D,\theta^{i-1},\eta^{i-1},\zeta^{i-1},\xi^{i-1}\right]. \tag{B11}$$

The E-step and M-step are iterated until converge.

In summary, the model (4.9)–(4.10) is estimated by optimizing the profile log-likelihood function (4.14), which for fixed frailty parameter $\psi$ is computed by the following algorithm.

1. Initialize $\theta^0,\eta^0,\zeta^0,\xi^0$ and set $i=1$.
2. For all $t$ and $x$ in the data window compute $\widetilde{\mathcal{M}}(t,x,\zeta^{i-1},\xi^{i-1})$ and set
$$c(t,x)=\nu'_\psi\left(\nu_\psi^{-1}\left\{\widetilde{\mathcal{M}}\left(t,x,\zeta^{i-1},\xi^{i-1}\right)\right\}\right). \tag{B12}$$
3. Compute $\theta^i$ and $\eta^i$ as the maximum likelihood estimates of the model
$$D_s(t,x)\sim\text{Poisson}(c(t,x)\,E(t,x)\,F(\theta_t,\eta_x)) \tag{B13}$$

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press

with death counts

$$D_s(t,x) = D(t,x)\,\frac{c(t,x)\,F\left(\theta_t^{i-1},\eta_x^{i-1}\right)}{c(t,x)\,F\left(\theta_t^{i-1},\eta_x^{i-1}\right) - G\left(\zeta_t^{i-1},\xi_x^{i-1}\right)}. \tag{B14}$$

4. Compute $\zeta^i$ and $\xi^i$ as the maximum likelihood estimates of the model

$$D_b(t,x) \sim \text{Poisson}(E(t,x)G(\zeta_t,\xi_x)) \tag{B15}$$

with death counts

$$D_b(t,x) = D(t,x)\,\frac{G\left(\zeta_t^{i-1},\xi_x^{i-1}\right)}{c(t,x)\,F\left(\theta_t^{i-1},\eta_x^{i-1}\right) - G\left(\zeta_t^{i-1},\xi_x^{i-1}\right)}. \tag{B16}$$

5. Increase $i$ by one.
6. Repeat steps 2–5 until convergence.

https://doi.org/10.1017/asb.2021.37 Published online by Cambridge University Press