Stochastic frailty models for modeling and forecasting mortality

Søren Fiig Jarner[*]

October 3, 2014

Working paper

Abstract. In many countries life expectancy gains have been substantially higher than predicted by even recent forecasts. This is primarily due to increasing rates of improvement in old-age mortality not captured by existing models. In this paper we show how the concept of frailty can be used to model both changing rates of improvement and the deceleration of mortality at old ages, also seen in data.

We present a "fragilization" method by which frailty can be added to standard mortality models. The aim is to improve the modeling and forecasting of old-age mortality while preserving the structure of the original model and the underlying stochastic processes. Estimation is based on a general pseudo-likelihood approach which allows the use of essentially any frailty distribution and mortality model. We also consider a class of generalized stochastic frailty models with both frailty and non-frailty terms, and we describe how these models can be estimated by the EM-algorithm.

The method is applied to the Lee-Carter model and a parametric time-series model. For both applications the effect of adding frailty is illustrated with mortality data for US males.

Key words and phrases. Frailty theory, old-age mortality, increasing rates of improvement, stochastic mortality models, pseudo-likelihood, EM-algorithm.

---

[*]The Danish Labour Market Supplementary Pension Fund, Denmark; Department of Mathematical Sciences, University of Copenhagen, Denmark; Email: sj@atp.dk

1   Introduction

Mortality rates have been steadily declining in most of the industrialized world throughout the 20th century and there are no signs of improvements decelerating. On the contrary, populations where mortality rates are already very low still experience rates of improvement of the same, or even higher, magnitude than historically.

A great deal of models for modeling and projecting mortality rates have been proposed in recent years. However, the model formulated by Lee and Carter (1992) is still the most widely used. The Lee-Carter model describes the age-specific death rates (ASDRs) by the log-bilinear relation

$$\log \mu(t, x) = a_x + b_x k_t, \tag{1}$$

where a and b are age-specific parameters and k is a time-varying index. Projections are made by estimating the parameters of the model from observed data and extrapolate the time index k by standard time series methods. Typically k is modeled as a random walk with drift. The combination of a linear forecast of k and structure (1) implies that ASDRs are forecasted to improve by the same age-specific factor each year.

The Lee-Carter model has gained widespread popularity due to its simplicity and ease of interpretation. Essentially, ASDRs are forecasted to improve by the average rate of improvement over the estimation period. However, the assumption that rates of improvement are constant over time is not generally satisfied. Indeed, the mortality experience of several countries has shown increasing mortality improvement rates over time, in particular for older age groups, see e.g. Lee and Miller (2001); Booth et al. (2002); Renshaw and Haberman (2003); Bongaarts (2005). Consequently, Lee-Carter forecasts have in many situations underestimated the gains in old-age mortality.

Left panel of Figure 1 shows ASDRs for US males from 1950 to 2010 based on data from the Human Mortality Database (www.mortality.org). The death rates are plotted on a logarithmic scale and increasing rates of improvement can be observed by the concavity, rather than linearity, of the age-specific curves. The consequence for Lee-Carter forecasts is exemplified in the right panel of Figure 1 which shows actual and Lee-Carter forecasted period remaining life expectancy of 60-year old US males.[1] Due to increasing rates of improvement the forecasts increase when a later estimation period is used. Still, all historic forecasts are below the actual life expectancy experience.

In the light of changing rates of mortality improvement it has been suggested to fit the Lee-Carter model to shorter periods of data for which death rates comply better with the log-linearity assumption, see e.g. Lee and Miller (2001); Tuljapurkar et al. (2000); Booth et al. (2002). In an environment of increasing rates of improvement the use of shorter, more recent data periods can indeed result in better forecasts. However, future increases in old-age mortality improvement rates — beyond what has been seen historically — cannot be captured by this method.

---

[1]The period (remaining) life expectancy is a summary measure for the mortality profile of a population in a given year. It is calculated on the basis of the age-specific mortality rates for that year. Assuming death rates continue to decline the life expectancy experienced by a given cohort will be higher. This quantity is termed the cohort life expectancy.
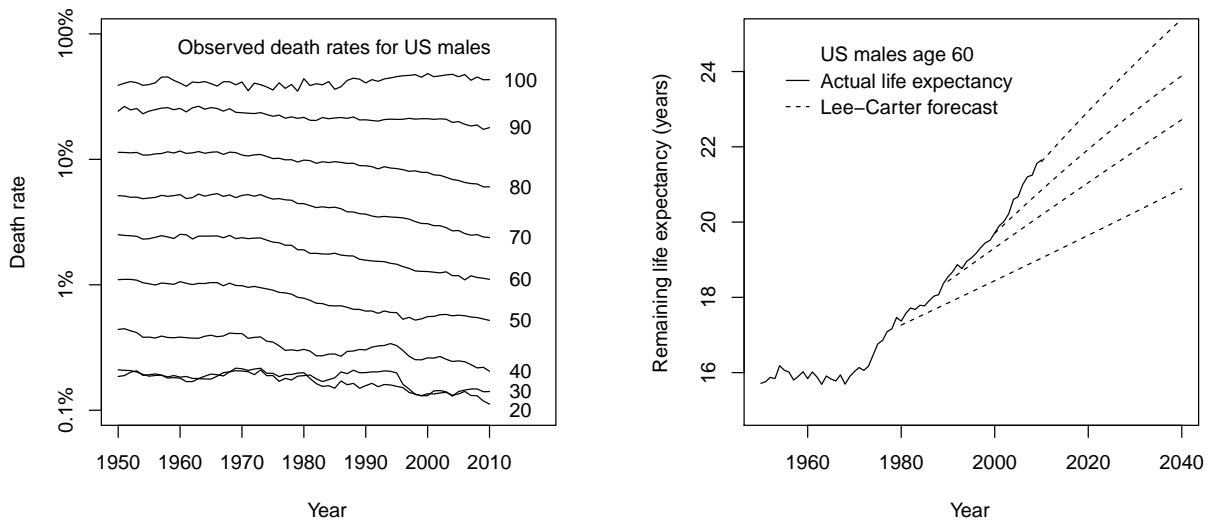
Figure 1: Left panel shows observed death rates for US males from 1950 to 2010 for ages 20, 30, ..., 100 years. Right panel shows remaining period life expectancy of US males age 60. The solid line is the actual life expectancy and the dashed lines are forecasts based on the Lee-Carter model with estimation periods 1950–1980, 1960–1990, 1970–2000 and 1980–2010.

1.1   Stochastic frailty models

In this paper we investigate the use of frailty theory to address the pattern of changing rates of improvement. Frailty theory rests on the assumption that cohorts are heterogeneous and that some people are more susceptible to death (frail) than others. Frail individuals tend to die sooner than stronger individuals leading to old cohorts being dominated by low mortality individuals. In effect, the selection mechanism causes the cohort intensity to "slow down", i.e. to increase less rapidly than the individual intensity. This has been used to explain the logistic form of old-age mortality observed in data, see e.g. Thatcher (1999).

In the model introduced by Vaupel et al. (1979) frailty enters as an unobservable, positive random quantify Z acting multiplicatively on an underlying baseline mortality intensity. The observable (aggregate) mortality intensity takes the form

$$\mu(t, x) = \mathsf{E}[Z|t, x]\, \mu_{\text{base}}(t, x), \tag{2}$$

where $\mathsf{E}[Z|t, x]$ denotes mean frailty at time t for age x, and $\mu_{\text{base}}(t, x)$ is the underlying baseline intensity for an individual with frailty one. We assume without loss of generality that mean frailty at birth is one. For young ages with low mortality mean frailty is close to one, while for old ages mean frailty gradually decreases as selection takes effect.

Frailty theory offers an explantation to the observed change in improvement rates of old-age mortality and it suggests that we might expect to see higher rates of improvements in the future. As the probability of attaining a given (old) age increases over time the selection effect weakens and mean frailty increases towards one. This increase partly offsets the underlying improvements in individual baseline intensity, cf. (2). Initially when selection is high old-age mortality rates will improve slowly (or even increase), but

over time as selection weakens improvements in old-age mortality will gradually increase towards baseline improvement rates.

In this paper we will use structure (2) as the basis for mortality modeling. We will show how a given model for aggregate mortality, e.g. the Lee-Carter model, can be "fragilized" by using it as a model for the underlying individual mortality instead. This procedure will enable us to model logistic-type old-age mortality and changing rates of improvement while preserving the structure of the original model and the underlying time-series dynamics. As an example, a "fragilized" Lee-Carter model takes the form

$$\mu(t, x) = \mathsf{E}[Z|t, x] \exp[a_x + b_x k_t]. \tag{3}$$

In Section 4.2 we will compare this model to the ordinary Lee-Carter model.

We show how to perform maximum likelihood estimation and how to forecast in models of form (2). Estimation will be based on a Poisson pseudo-likelihood which allows for general specifications of the frailty distribution, e.g. the generalized stable laws of Hougaard (1986). We also generalize (2) to include an additive (non-frailty) term and show how models of this form can be estimated by the EM-algorithm. Methods for joint maximum likelihood estimation of the frailty distribution and the baseline intensity are derived based on the (conditional) likelihood for fixed frailty distribution.

In contrast to the typical use of frailty theory in mortality modeling our approach does not rely on a specific parametric model. Indeed, we show how essentially any combination of frailty distribution and parametric or semi-parametric model for baseline intensity can be estimated. The only requirement is that the underlying baseline mortality model can be estimated (by maximum likelihood) and that the Laplace transform of the frailty distribution is known.

Frailty theory is well-established in biostatics and survival analysis and several monographs are devoted to the topic, e.g. Wienke (2010); Hanagal (2011). In demographic and actuarial science frailty models are also known as heterogeneity models. They have been used in mortality modeling to fit the logistic form of old-age mortality, see e.g. Wang and Brown (1998); Butt and Haberman (2004); Olivieri (2006); Cairns et al. (2006); Spreeuw et al. (2013), and to allow for overdispersion in mortality data, cf. Li et al. (2009).

The present use of frailty theory is closest in spirit to Jarner and Kryger (2011) in which a deterministic frailty model is used to describe the trend in international mortality. However, the methods developed in this paper greatly extend that work by incorporating frailty into any standard stochastic mortality model.

1.2   Outline

The rest of the paper is organized as follows. In Section 2 we give a brief introduction to frailty theory to establish the fundamental relation between cohort and individual mortality on which our approach relies. Section 3 is the main theoretical chapter in which stochastic frailty models are formalized and estimation and forecasting discussed. This is followed in Section4 by applications to a parametric time-series model and the Lee-Carter model. Section 5 contains a generalization of stochastic frailty models and explains how they can be estimated by the EM-algorithm. Finally, Section 6 offers some concluding remarks.

4

2   Frailty theory

In this section we go through the basic theory and establish relations to be used in the following. For ease of exposition we consider only a single birth cohort in a continuous-time model. Later on we will add a time parameter. As our starting point we take the frailty model of Vaupel et al. (1979) in which frailty is modeled as an unobservable, non-negative quantity Z acting multiplicatively on an underlying, baseline (mortality) intensity. The intensity, or death rate, of an individual of age x with *frailty z* is given by

$$\mu(x; z) = z \, \mu_{\text{base}}(x), \qquad (4)$$

where μ_{base} is the baseline intensity describing the age effect. The corresponding survival function, i.e. the probability of the person surviving to age x, is

$$S(x; z) = \exp\left(-\int_0^x \mu(u; z) du\right) = \exp\left(-z I(x)\right), \qquad (5)$$

where I denotes the integrated baseline intensity,

$$I(x) = \int_0^x \mu_{\text{base}}(u) du. \qquad (6)$$

The survival function of the cohort, i.e. the expected fraction of the cohort surviving to age x, is then

$$S(x) = \mathsf{E}\left[S(x; Z)\right] = \mathsf{E}\left[\exp\left(-Z I(x)\right)\right] = L\left(I(x)\right), \qquad (7)$$

where L denotes the Laplace transform of the frailty distribution (at birth),

$$L(s) = \mathsf{E}\left[\exp(-sZ)\right]. \qquad (8)$$

Assuming sufficient regularity the (conditional) mean frailty of the cohort at age x is given by

$$\mathsf{E}[Z|x] = \frac{\mathsf{E}\left[Z S(x; Z)\right]}{S(x)} = \frac{\mathsf{E}\left[Z \exp\left(-Z I(x)\right)\right]}{L(I(x))} = -\frac{L'(I(x))}{L(I(x))} = \nu'(I(x)), \qquad (9)$$

where $'$ denotes differentiation and $\nu(s) = -\log L(s)$. It follows that the cohort intensity can be written as

$$\mu(x) = -\frac{d}{dx} \log S(x) = \nu'(I(x)) I'(x) = \mathsf{E}[Z|x] \, \mu_{\text{base}}(x). \qquad (10)$$

The derivations above relating mean frailty and cohort intensity to the frailty distribution and baseline intensity provide the usual tools for frailty modeling. For later use we make the additional observation that mean frailty can also be expressed in terms of the integrated cohort intensity,

$$H(x) = \int_0^x \mu(u) du. \qquad (11)$$

The cohort survival function (7) provides the following link between H and I,

$$S(x) = \exp\left(-H(x)\right) = L\left(I(x)\right), \qquad (12)$$

from which we obtain $H(x) = \nu(I(x))$, and thereby $I(x) = \nu^{-1}(H(x))$. By (9) and (10) we then get the expression

$$\mu(x) = \nu'(\nu^{-1}\{H(x)\})\,\mu_{\text{base}}(x). \tag{13}$$

The idea later on is to replace H by a non-parametric estimate obtained from data, i.e. the integrated observed (empirical) death rates. The substitution disentangles the frailty distribution and the baseline intensity which greatly simplifies the estimation procedure.

In order to apply the theory we need to identity suitable frailty distributions and their Laplace transforms. Typically, the baseline intensity contains a scale parameter in which case it is customary to assume that the frailty distribution has mean one to ensure identifiability. We will also make this assumption here.

2.1   Gamma and inverse Gaussian frailty

The frailty distributions most often employed in mortality modeling are the Gamma and inverse Gaussian distributions, see e.g. Vaupel et al. (1979); Hougaard (1984); Manton et al. (1986); Butt and Haberman (2004); Jarner and Kryger (2011); Spreeuw et al. (2013). Below we state their Laplace transforms and comment on their use.

When frailty is Gamma-distributed with mean one and variance σ^2 the Laplace transform and mean frailty are given by

$$L(s) = \left(1 + \sigma^2 s\right)^{-1/\sigma^2}, \tag{14}$$

$$\mathsf{E}[Z|x] = \left(1 + \sigma^2 I(x)\right)^{-1} = \exp\left(-\sigma^2 H(x)\right). \tag{15}$$

It is well-known that Gamma frailty in combination with Gompertz or Makeham baseline intensity leads to a cohort intensity of logistic type, see e.g. Example 2.1 of Jarner and Kryger (2011) for details.[2] This is also known as the Perks model, Perks (1932), and it has been found to describe old-age mortality very well, see e.g. Thatcher et al. (1998); Thatcher (1999); Cairns et al. (2006). Gamma-distributed frailty is also mathematically tractable and allows explicit calculations of many quantities of interest, e.g. frailty among survivors at a given age is Gamma-distributed with known scale and shape parameters, cf. Vaupel et al. (1979).

When frailty follows an inverse Gaussian distribution with mean one and variance σ^2 we have for the Laplace transform and mean frailty

$$L(s) = \exp\left[\frac{1 - \sqrt{1 + 2\sigma^2 s}}{\sigma^2}\right], \tag{16}$$

$$\mathsf{E}[Z|x] = \left(1 + 2\sigma^2 I(x)\right)^{-1/2} = \left(1 + \sigma^2 H(x)\right)^{-1}. \tag{17}$$

Similar to Gamma frailty inverse Gaussian frailty enjoys a number of nice theoretical properties, e.g. the frailty distribution among survivors of a given age is again inverse

---

[2]A Gompertz intensity is of the form $\mu_{\text{base}}(x) = \exp(a + bx)$, while a Makeham intensity is of the form $\mu_{\text{base}}(x) = \exp(a + bx) + c$. A Makeham intensity is also sometimes referred to as a Gompertz-Makeham intensity.

Gaussian and an explicit formula for the density exists, see Hougaard (1984). Hougaard (1984) shows further that for inverse Gaussian frailty the coefficient of variation among survivors is decreasing with age making the cohort more homogeneous as it gets older, while for Gamma frailty the coefficient of variation is constant. Thus the impact of frailty is smaller for inverse Gaussian frailty than for Gamma frailty. From (15) and (17) we also see directly how Gamma frailty "slows" down baseline intensity more than inverse Gaussian frailty.

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 e.g. Butt and Haberman (2004); Spreeuw et al. (2013). Furthermore, Abbring and van der Berg (2007) show that for a large class of initial frailty distributions the frailty distribution among survivors converges to a Gamma distribution as the integrated intensity tends to infinity. Thus overall the Gamma distribution is a good default choice.

That being said, it is not immediately clear that this conclusion carries over to the case where frailty is introduced to model increasing rates of improvement over time.

2.2   Positive stable frailty

Hougaard (1986) introduced a family of generalized stable laws which include the Gamma and inverse Gaussian distributions as special cases. 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 α, 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 σ^2 the Laplace transform and mean frailty are given by[3]

$$L(s) = \exp\left[\frac{1-\alpha}{\alpha}\left\{\frac{1 - [1 + \sigma^2 s/(1-\alpha)]^{\alpha}}{\sigma^2}\right\}\right], \tag{18}$$

$$\mathsf{E}[Z|x] = \left[1 + \frac{\sigma^2}{1-\alpha}I(x)\right]^{\alpha-1} = \left[1 + \frac{\alpha}{1-\alpha}\sigma^2 H(x)\right]^{(\alpha-1)/\alpha}. \tag{19}$$

We note that for $\alpha = 0$ (defined by continuity) the generalized stable law specializes to the Gamma distribution, while for $\alpha = 1/2$ it specializes to the inverse Gaussian distribution. While Gamma and inverse Gaussian densities are available in closed form, generally only series representations of (tilted) stable densities exist, cf. Hougaard (1986) and Lemma XVII.6.1 of Feller (1971).

In either case, we have simple closed form expressions for mean frailty in terms of integrated baseline and cohort intensity which is all we need for estimation and forecasting purposes. Moreover, the fact that the parametric family spans the two most commonly used frailty distributions, Gamma and inverse Gaussian, opens for a unified approach for joint estimation of frailty and baseline parameters.

---

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

The family has been further generalized by Aalen (1988, 1992) to include also negative α. Negative values of α correspond to compound Poisson distributions with positive probability of zero frailty. This is important for some applications when modeling the time to non-certain events, e.g. divorce or unemployment. The generalization is of no use in our context where it implies immortality for some individuals. However, it could be useful for course-specific mortality modeling as demonstrated by Aalen (1988).

3 Stochastic frailty models

We first include frailty in a continuous-time model spanning multiple birth cohorts and show the effect on the dynamics of age-specific death rates over time. Then we add frailty to stochastic mortality models and discuss estimation and forecasting of the resulting models.

Again, we start with the multiplicative frailty model. With a slight abuse of notation we will reuse the notation of Section 2 with an added time parameter. Assume that the intensity for an individual of age x at time t with frailty z has the form

$$\mu(t, x; z) = z \, \mu_{\text{base}}(t, x), \tag{20}$$

where μ_{base} is a baseline intensity describing the period (time) and age effect. The cohort intensity is then given by

$$\mu(t, x) = \mathsf{E}[Z|t, x] \, \mu_{\text{base}}(t, x), \tag{21}$$

where $\mathsf{E}[Z|t, x]$ denotes the (conditional) mean frailty of the cohort of age x at time t, i.e. the cohort born at time $t - x$.

Assume that all cohorts have the same frailty distribution at birth, and denote the Laplace transform of this common distribution by L. Let $\nu(s) = -\log L(s)$, and define the integrated baseline and cohort intensities as

$$I(t, x) = \int_0^x \mu_{\text{base}}(u + t - x, u) du, \tag{22}$$

$$H(t, x) = \int_0^x \mu(u + t - x, u) du. \tag{23}$$

That is, we follow a specific birth cohort over time and age. We have as before the relations $H(t, x) = \nu \left(I(t, x) \right)$ and

$$\mathsf{E}[Z|t, x] = \nu'(I(t, x)) = \nu' \left(\nu^{-1} \left\{ H(t, x) \right\} \right). \tag{24}$$

Following the notation of Bongaarts (2005) we define the rate of improvement in (senescent) mortality as $\rho(t, x) = -\partial \log \mu(t, x)/\partial t$. It follows from (21) that

$$\rho(t, x) = -\frac{\partial \log \mathsf{E}[Z|t, x]}{\partial t} + \rho_{\text{base}}(t, x), \tag{25}$$

where $\rho_{\text{base}}(t, x) = -\partial \log \mu_{\text{base}}(t, x)/\partial t$ denotes the rate of improvement of baseline intensity.

Suppose we model the period effect of mortality improvements by decreasing age-specific baseline mortality, i.e. $\rho_{\text{base}} > 0$. For fixed x, the integrated baseline intensity will then also be decreasing over time, while the mean frailty will be *increasing* over time due to less and less selection. From (25) we see that this implies that the rate of improvement of cohort mortality will be smaller than the rate of improvement at the individual level.

At old ages where death rates, and thereby selection, is high the change in mean frailty over time can substantially offset improvements in baseline mortality causing rates of improvement of cohort mortality to be close to zero. As improvements continue to occur at the individual level the selection effect gradually disappears and rates of improvement of cohort mortality get closer to baseline rates of improvement. The resulting pattern of gradually changing rates of improvement of old-age mortality resembles what is seen in data. This will be illustrated in the applications in Section 4.

3.1 Data and terminology

Data are assumed to be of the form of death counts, $D(t, x)$, and corresponding exposures, $E(t, x)$, for a range of years, $t_{\min} \leq t \leq t_{\max}$, and ages, $x_{\min} \leq x \leq x_{\max}$. Data may or may not be gender-specific, but that will not be part of the notation. $D(t, x)$ denotes the number of deaths occurring in calender year t among people aged $[x, x+1)$. Correspondingly, $E(t, x)$ denotes the total number of years lived during calender year t by people of age $[x, x+1)$. For readers familiar with the Lexis diagram, $D(t, x)$ counts the number of deaths in the square $[t, t+1) \times [x, x+1)$ of the Lexis diagram and $E(t, x)$ gives the corresponding exposure, i.e. we work with so-called *A-groups*.

From the death counts and exposures we form the observed (empirical) death rates

$$m(t, x) = \frac{D(t, x)}{E(t, x)}. \tag{26}$$

The death rate is an estimate of the cohort intensity, $\mu(t, x)$, which for modeling purposes is assumed constant over the square $[t, t+1) \times [x, x+1)$ in the following.

3.2 Fragilization of mortality models

Consider as given a stochastic model for baseline mortality, $\mu_{\text{base}}(t, x)$. For ease of presentation we will assume that it is of the form

$$\mu_{\text{base}}(t, x) = F(\theta_t, \eta_x), \tag{27}$$

where θ_t and η_x are, possibly multi-dimensional, time- and age-dependent quantities and F describes the functional dependence. We could have included also cohort effects and more general dependence structures if so desired.

The form (27) is chosen with two applications mind: with $\theta_t = k_t$, $\eta_x = (a_x, b_x)$ and $F(k, (a, b)) = \exp(a + bk)$ we get the Lee-Carter model; with $\eta_x = x$ we get parametric time-series models where the functional form of the age-profile is determined by F. In the parametric case we could use for instance a Gompertz law, $F(\theta_t, x) = \exp(\theta_t^1 + \theta_t^2 x)$, or a logistic function, $F(\theta_t, x) = \exp(\theta_t^1 + \theta_t^2 x)/(1 + \exp(\theta_t^1 + \theta_t^2 x))$. The latter is the model

considered by Cairns et al. (2006) for aggregate mortality. As mentioned in Section 2.1 the logistic model already has a frailty interpretation, and it is therefore not an obvious candidate for further fragilization; but it can nevertheless be done.

For presentational convenience we assume that both θ_t and η_x need to be estimated from data, although as just seen η_x may in fact be fixed. Once estimated we can forecast θ_t by ARIMA time-series methods and thereby obtain forecasts of μ_{base} by inserting the forecasted values of θ_t together with the estimated values of η_x into (27).

In order to proceed we assume that we have available a procedure for maximum likelihood estimation of (θ_t, η_x) in the model where death counts are independent with

$$D(t,x) \sim \text{Poisson} \left(\mu_{\text{base}}(t,x) E(t,x) \right). \tag{28}$$

In the Lee-Carter case such a procedure is described in Brouhns et al. (2002).

With these building blocks in place we now consider the following "fragilized" version of the baseline model in which death counts are independent with

$$D(t,x) \sim \text{Poisson} \left(\mu(t,x) E(t,x) \right), \tag{29}$$

$$\mu(t,x) = \mathsf{E}[Z|t,x]\, \mu_{\text{base}}(t,x), \tag{30}$$

where μ_{base} is given by (27) and $\mathsf{E}[Z|t,x]$ denotes the conditional mean frailty of the cohort of age x at time t. It is assumed that the frailty distribution at birth is the same for all cohorts and that the distribution belongs to a family indexed by ϕ. The parameters of the model are thus (ϕ, θ_t, η_x). The Laplace transform of the frailty distribution with index ϕ is denoted L_ϕ, and this is assumed available in explicit form. Further, as a matter of convention we assume that mean frailty is one at birth and we define $\nu_\phi = -\log L_\phi$.

Based on (22) and (24) we can write

$$\mu(t,x) = \nu'_\phi(I(t,x)) F(\theta_t, \eta_x), \tag{31}$$

$$I(t,x) = \sum_{u=0}^{x-1} F(\theta_{u+t-x}, \eta_u), \tag{32}$$

and insert this into (29). In principle, we can estimate all parameters jointly from the resulting likelihood function. However, the likelihood function is intractable with frailty and baseline parameters occurring in a complex mix. Consequently, estimation has to be handled on a case-by-case basis depending on the choice of frailty and baseline model. Below we propose an alternative, generally applicable pseudo-likelihood approach which greatly simplifies the estimation task.

3.3   Pseudo-likelihood function

From (24) we know that $\mathsf{E}[Z|t,x] = \nu'\left(\nu^{-1}\left\{H(t,x)\right\}\right)$, where H is the integrated cohort intensity μ. At first sight this does not seem to help much since H is at least as complicated as I. However, in contrast to I we can obtain a (model free) estimate of H directly from data as

$$\tilde{H}(t,x) = \sum_{u=0}^{x-1} m(u+t-x, u). \tag{33}$$

We thus propose to base estimation of (29)–(30) on a likelihood function in which the term $\mathsf{E}[Z|t,x]$ is replaced by $\nu'(\nu^{-1}\{\tilde{H}(t,x)\})$. The resulting approximate likelihood function is referred to as the pseudo-likelihood function,[4]

$$L(\phi, \theta, \eta) = \prod_{t,x} \frac{\lambda(t,x)^{D(t,x)}}{D(t,x)!} \exp(-\lambda(t,x)), \tag{34}$$

$$\lambda(t,x) = \nu'_\phi(\nu_\phi^{-1}\{\tilde{H}(t,x)\})F(\theta_t, \eta_x)E(t,x). \tag{35}$$

Formally, it corresponds to estimating the modified model

$$D(t,x) \sim \text{Poisson}\left(\mu(t,x)E(t,x)\right), \tag{36}$$

$$\mu(t,x) = \nu'_\phi(\nu_\phi^{-1}\{\tilde{H}(t,x)\})F(\theta_t, \eta_x). \tag{37}$$

In model (36)–(37) μ is separable in frailty and baseline parameters and the model is therefore considerably easier to handle than (29)–(32).[5] The separability allows for simple joint estimation based on the estimation procedure for the baseline model.

One issue remains before turning to estimation in Section 3.4. To form the (pseudo) likelihood function we need to compute $\tilde{H}(t,x)$ for all t and x in the data window. However, these quantities depend in part on data outside the data window. In principle, we need to know the death rates from birth to the present/maximum age for all cohorts entering the estimation. The gray area of Figure 2 illustrates the "missing" death rates.
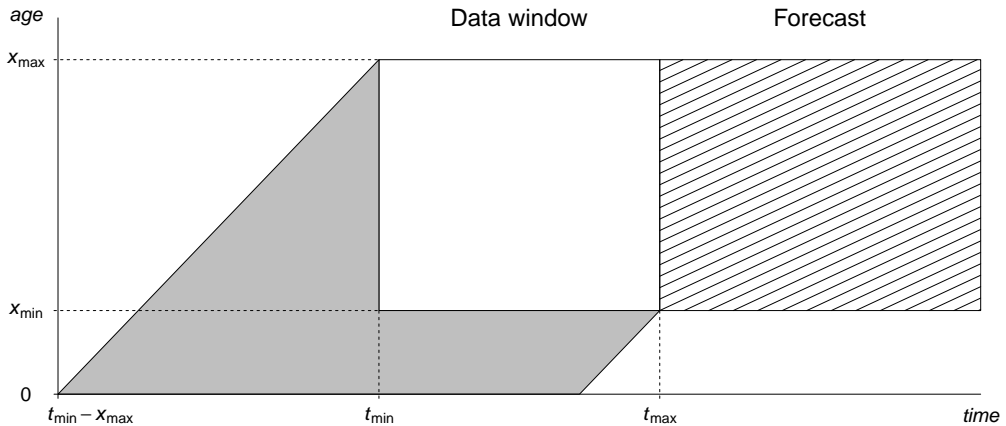

Figure 2: Data are available for years between $t_{\min}$ and $t_{\max}$ and for ages between $x_{\min}$ and $x_{\max}$. The gray area below and to the left of the data window illustrates the part of the trajectories needed for calculation of $\tilde{H}$ that falls outside the data window. The shaded area to the right illustrates the years and ages for which we wish to forecast mortality intensities.

---

[4]The idea of basing inference on a pseudo-likelihood function was first introduced in spatial statistics, see Besag (1975).

[5]Equation (37) might still look daunting, but is simplifies in specific cases. If frailty is assumed Gamma distributed with variance σ^2 (and mean one) we get $\mu(t,x) = \exp(-\sigma^2 \tilde{H}(t,x))F(\theta_t, \eta_x)$, while inverse Gaussian frailty with variance σ^2 (and mean one) yields $\mu(t,x) = F(\theta_t, \eta_x)/(1 + \sigma^2 \tilde{H}(t,x))$.

For the purpose of calculating $\tilde{H}$ we suggest to define the death rates before and below the data window by

$$m(t,x) = \begin{cases} m(t_{\min}, x) & \text{for } t < t_{\min} \text{ and } x_{\min} \leq x \leq x_{\max}, \\ 0 & \text{for } 0 \leq x < x_{\min}. \end{cases} \tag{38}$$

This corresponds to saying that selection prior to $t_{\min}$ have 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). The latter assumption was also used by Vaupel in the contributed discussion part of Thatcher (1999). With the extension of death rates $\tilde{H}$ and hence the likelihood function can be computed.

We note that in some situations, e.g. for many national data sets, we may have more data than we choose to use for estimation. In these situations some or all of the "missing" death rates may in fact be available and hence could be used when calculating $\tilde{H}$. In the present paper, however, we will calculate $\tilde{H}$ using the extension (38) even in this case. Assuming no selection prior to $x_{\min}$ also makes forecasting more straightforward.

3.4   Maximum likelihood estimation

We here discuss various ways to obtain maximum likelihood estimates of model (36)–(37) of Section 3.3. For computational efficiency and numeric stability one often considers the log-likelihood function (rather than the likelihood function),

$$l(\phi, \theta, \eta) = \log L(\phi, \theta, \eta) = \sum_{t,x} \{ D(t,x) \log \mu(t,x) - \mu(t,x) E(t,x) \} + \text{constant}, \tag{39}$$

where μ is given by (37) and the constant depends on data only. The simplest and most general method to optimize (39) is to form the profile log-likelihood function

$$l(\phi) = l(\phi, \hat{\theta}(\phi), \hat{\eta}(\phi)), \tag{40}$$

where $\hat{\theta}(\phi)$ and $\hat{\eta}(\phi)$ denote the maximum likelihood estimates of θ and η for fixed frailty parameter ϕ. Assuming that we know how to perform maximum likelihood estimation of the baseline model, $\hat{\theta}(\phi)$ and $\hat{\eta}(\phi)$ can be obtained by estimating the baseline model with death counts $D(t,x)$ and exposure $\nu'_\phi(\nu_\phi^{-1}\{\tilde{H}(t,x)\})E(t,x)$. Since the frailty family is typically of low dimension, e.g. one or two dimensions, the profile log-likelihood function can normally be optimized reliably by general purpose optimization routines, e.g. the `optimize` or `optim` routines in the free statistical software package R.

For a one-dimensional frailty family, e.g. Gamma with mean one and variance σ^2, we could use the following skeleton R-code

```
# model is a list with functions for calculating
# Htilde, mean frailty and MLE of the baseline model
frail.optim <- function(D,E,model) {

   Htilde <- model$Htilde(D,E)
```

```
    profile.loglike <- function(phi) {
        EZ   <- model$meanfrail(phi,Htilde)
        ret <- model$MLEbase(D,EZ*E)
        ret$value   # value of maximized log-likelihood
    }
    optret  <- optimize(profile.loglike,interval=c(0,2),maximum=TRUE)
    MLEphi  <- optret$maximum
    EZ       <- model$meanfrail(MLEphi,Htilde)
    ret      <- model$MLEbase(D,EZ*E)
    MLEbase <- ret$par  # MLE of baseline parameters
    list(phi=MLEphi,base=MLEbase)
}
```

Alternatively, maximum likelihood estimation can also be implemented by a switching algorithm. Starting from an initial value, ϕ_0, of the frailty parameter and $i = 1$, the algorithm consists in repeating the following two steps until convergence. On each iteration i is increased by one.

1. Calculate (θ_i, η_i) as the maximum likelihood estimate with $\phi = \phi_{i-1}$ fixed.

2. Calculate ϕ_i as the maximum likelihood estimate with $(\theta, \eta) = (\theta_i, \eta_i)$ fixed.

The estimation in the first step is, as before, performed by estimating the baseline model with death counts $D(t, x)$ and exposure $\nu'_{\phi_{i-1}}(\nu^{-1}_{\phi_{i-1}}\{\tilde{H}(t, x)\})E(t, x)$. The estimation in the second step is performed by maximizing $l(\phi, \theta_i, \eta_i)$ as a function of ϕ. With the dimension of ϕ typically low, this is easy to do by either Newton-Raphson, steepest ascent, or, in the one-dimensional case, even a simple bisection algorithm.

For the case of Gamma frailty with mean one and variance σ^2 we have for l and its first two partial derivatives with respect to σ^2

$$l(\sigma^2, \theta, \eta) = \sum_{t,x} \left\{ -\sigma^2 D\tilde{H} - \exp(-\sigma^2 \tilde{H})F(\theta_t, \eta_x)E \right\} + \text{const.},$$

$$\frac{\partial l}{\partial \sigma^2}(\sigma^2, \theta, \eta) = \sum_{t,x} \left\{ -D\tilde{H} + \tilde{H}\exp(-\sigma^2 \tilde{H})F(\theta_t, \eta_x)E \right\},$$

$$\frac{\partial^2 l}{\partial(\sigma^2)^2}(\sigma^2, \theta, \eta) = -\sum_{t,x} \tilde{H}^2 \exp(-\sigma^2 \tilde{H})F(\theta_t, \eta_x)E < 0,$$

where we have suppressed the dependence on t and x of D, E and $\tilde{H}$ to simplify notation. The calculations show that l is strictly concave as a function of σ^2, and thereby that a local maximum is also a global maximum. Hence any gradient method is guaranteed to converge to the global maximum. Note however that the maximum may not be interior and therefore we do not necessarily have $\partial l/\partial \sigma^2 = 0$ at the maximum.

The switching algorithm will always converge to a (local) maximum. However, many iterations may be required in particular if the likelihood has a "diagonal" ridge along frailty and baseline parameters, i.e. if frailty and baseline parameters to some extent describe the same feature of data. The advantage of the switching algorithm is that it

allows a more detailed analysis which can guide the choice of optimization routine, like in the Gamma example above.

We finally note that estimation can also be carried out by Newton-Raphson sweeps over frailty and baseline parameters, similar to the algorithm described in Brouhns et al. (2002) for the Lee-Carter model. This may be more efficient in terms of computing time, but it requires a substantial amount of tailor-made code for each combination of frailty and baseline model.

Overall, we find that optimization of the profile log-likelihood function is the method of choice. It is particularly well-suited in the exploratory phase due to its flexibility and ease of implementation. In our experience it is also sufficiently fast and robust to be of practical use.

3.5 Forecasting

We now assume that we have (maximum likelihood) estimates of (ϕ, θ, η) and we denote these by $(\hat{\phi}, \hat{\theta}, \hat{\eta})$. We will follow the usual approach in stochastic mortality modeling and forecast mortality based on a time-series model for (θ_t). The time-series model is estimated on the basis of $(\hat{\theta}_t)_{t_{\min} \leq t \leq t_{\max}}$ treating these as observed, rather than estimated, quantities. Typically a simple (multi-dimensional) random walk with drift is used, see e.g. Lee and Carter (1992) and Cairns et al. (2006), but models with more structure can also be used.

Assume that we have a forecast $(\bar{\theta}_t)_{t_{\max} < t \leq t_{\max}+h}$ for a given forecast horizon h. The forecast can be either a deterministic (mean) forecast, or a stochastic realization from the time-series model. The aim is to forecast baseline and cohort mortality for ages $x_{\min} \leq x \leq x_{\max}$; the forecast region is illustrated as the shaded box to the right in Figure 2.

We first note that the estimated μ can be written

$$\hat{\mu}(t,x) = \nu'_{\hat{\phi}}(\nu^{-1}_{\hat{\phi}}\{\tilde{H}(t,x)\})F(\hat{\theta}_t, \hat{\eta}_x) = \nu'_{\hat{\phi}}(\tilde{I}(t,x))F(\hat{\theta}_t, \hat{\eta}_x), \tag{41}$$

where $\tilde{I}(t,x) = \nu^{-1}_{\hat{\phi}}\{\tilde{H}(t,x)\}$. Expressing mean frailty in terms of $\tilde{I}$ provides the necessary link for forecasting mean frailty, and thereby cohort mortality.

Baseline mortality is readily forecasted by inserting $\bar{\theta}$ and $\hat{\eta}$ into (27),

$$\mu_{\text{base}}(t,x) = F(\bar{\theta}_t, \hat{\eta}_x), \tag{42}$$

while cohort mortality is forecasted by

$$\mu(t,x) = \nu'_{\hat{\phi}}(\tilde{I}(t,x))F(\bar{\theta}_t, \hat{\eta}_x), \tag{43}$$

where $\tilde{I}$ in the forecast region is given by the recursion

$$\tilde{I}(t,x) = \begin{cases} 0 & \text{for } x = x_{\min}, \\ \tilde{I}(t_{\max}, x-1) + F(\hat{\theta}_{t_{\max}}, \hat{\eta}_{x-1}) & \text{for } x > x_{\min}, t = t_{\max} + 1, \\ \tilde{I}(t-1, x-1) + F(\bar{\theta}_{t-1}, \hat{\eta}_{x-1}) & \text{for } x > x_{\min}, t > t_{\max} + 1. \end{cases} \tag{44}$$

14

Note that in the data window $\tilde{I}$ is defined by transformation of $\tilde{H}$ to ensure consistency with the estimated model, while in the forecast region it is defined recursively in terms of the forecasted baseline mortality.

For Gamma frailty with mean one and estimated variance $\hat{\sigma}^2$ we will forecast cohort mortality by

$$\mu(t, x) = \frac{F(\bar{\theta}_t, \hat{\eta}_x)}{1 + \hat{\sigma}^2 \tilde{I}(t, x)}, \tag{45}$$

where $\tilde{I}$ in the data window is given by $\tilde{I} = [\exp(\hat{\sigma}^2 H) - 1]/\hat{\sigma}^2$. Inverse Gaussian frailty with mean one and estimated variance $\hat{\sigma}^2$ yields the cohort mortality forecast

$$\mu(t, x) = \frac{F(\bar{\theta}_t, \hat{\eta}_x)}{\sqrt{1 + 2\hat{\sigma}^2 \tilde{I}(t, x)}}, \tag{46}$$

where $\tilde{I}$ in the data window is given by $\tilde{I} = [(1 + \hat{\sigma}^2 H)^2 - 1]/2\hat{\sigma}^2$.

4   Applications to US male mortality

Broadly speaking stochastic mortality models fall in two main categories: parametric time-series models and semi-parametric models of Lee-Carter type. We consider two applications, one for each of these model classes. The applications are illustrative examples highlighting different aspects of the theory rather than full-blown statistical analyses.

In the first application we take a standard Makeham mortality intensity and show how the fit can be improved for different age groups by adding frailty. To give a flavor of the possibilities we consider models of the generalized, additive form presented in Section 5 in combination with the two-dimensional family of stable laws of Section 2.2. In the second application we add Gamma frailty to the Lee-Carter model and illustrate the effect on the forecasts.

Both applications use US male mortality data from 1950–2010 available at the Human Mortality Database (www.mortality.org). The first application compares a forecast based on the initial half of data with the actual mortality experience. The second application uses the latter half of data and makes a forecast into the future.

4.1   Parametric time-series application

The classical Makeham, or Gompertz-Makeham, mortality law states that age-specific mortality can be described by the functional form

$$\mu(x) = \exp(\theta^1 + \theta^2 x) + \exp(\zeta), \tag{47}$$

where x denotes age. The additive structure suggests an interpretation of mortality as composed of an age-dependent (senescent) part describing mortality as people grow older, and a part describing background mortality, e.g. accidents, which happen independent of age. Despite its simplicity the Makeham intensity provides a reasonable description for a wide range of adult ages. However, for the very old and also for younger adults the

description is not adequate. The right panel of Figure 3 shows observed death rates in 1980 for US males aged 20 to 100 with a fitted Makeham curve imposed (and two other curves which will be explained below). It is seen that the Makeham curve overstates old-age mortality and that the curvature is wrong for the younger ages.
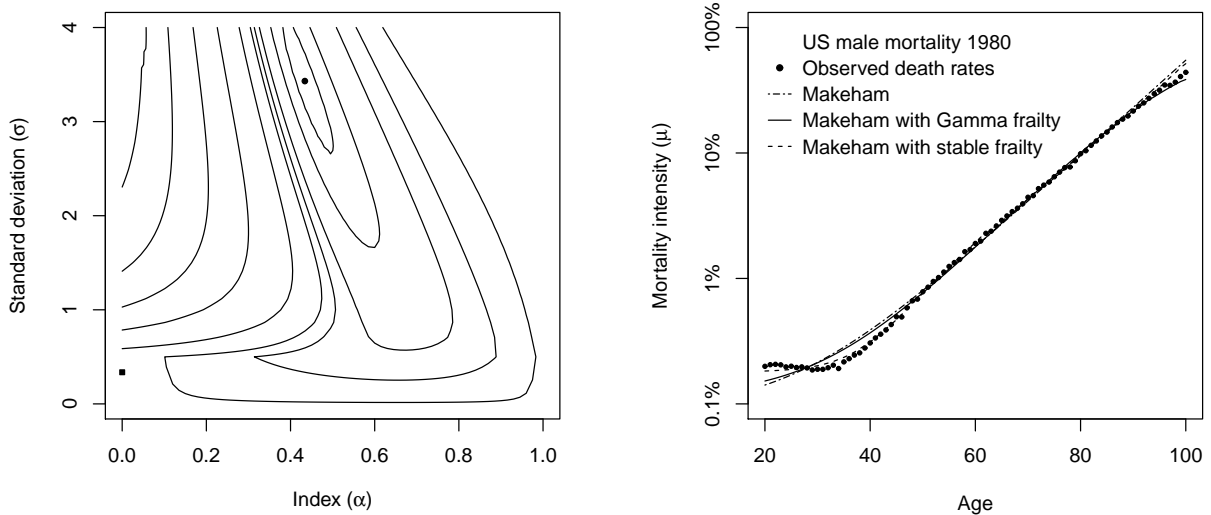

Figure 3: Left panel shows a contour plot of the profile log-likelihood function, $l(\alpha, \sigma)$, of model (48)–(49). The dot marks the maximal value, $(\hat{\alpha}, \hat{\sigma})$, and the box marks the maximal value along the axis $\alpha = 0$ (Gamma frailty), $\hat{\sigma}_\Gamma$. Right panel shows observed death rates in 1980 for US males aged 20 to 100 with a Makeham fit superimposed. Also shown are the fits of model (48)–(49) with optimal Gamma and stable frailty, respectively.

For old-age mortality, the lack of fit of the Makeham law is well-known and logistic-type models have been proposed instead, see e.g. Thatcher et al. (1998). In Cairns et al. (2006) a logit-model for death probabilities (rather than rates) is suggested and applied to England & Wales males above age 60 for which it fits well. However, in Cairns et al. (2009) potential problems of fitting US male data with the pure logit-model is reported and a variant is considered.

4.1.1 Stochastic frailty model

As an alternative to the logit-model and its variants we here consider the stochastic frailty model

$$D(t, x) \sim \text{Poisson}\left(\mu(t, x)E(t, x)\right), \tag{48}$$

$$\mu(t, x) = \mathsf{E}[Z|t, x]\exp(\theta_t^1 + \theta_t^2 x) + \exp(\zeta_t), \tag{49}$$

where the distribution of Z belongs to the positive stable family with index $\alpha \in [0, 1)$ and variance σ^2, described in Section 2.2. The model is of the additive form introduced in Section 5 with frailty acting only on the senescent part of mortality. With this structure we keep the interpretation of the second term as background mortality to which everyone is equally susceptible.

16

We recall from Section 2.2 that the stable family includes both Gamma ($\alpha = 0$) and inverse Gaussian ($\alpha = 1/2$) as special cases. Gamma frailty leads to a logistic-type model, akin to the one used in Thatcher (1999), while values of α greater than zero yield old-age mortality "in between" logistic and exponential. The Makeham model with time-varying parameters is also included and corresponds to $\sigma^2 = 0$ (for all values of α).

We apply the model to US male mortality data for the period 1950–1980 and ages 20–100. The left panel of Figure 3 shows a contour plot of the profile log-likelihood function,

$$l(\alpha, \sigma) = \log L(\alpha, \sigma, \hat{\theta}(\alpha, \sigma), \hat{\eta}(\alpha, \sigma)), \tag{50}$$

where L is the pseudo-likelihood function of Section 5.2 and $\hat{\theta}(\alpha, \sigma)$ and $\hat{\eta}(\alpha, \sigma)$ denote the maximum likelihood estimates for fixed value of frailty parameters. The profile log-likelihood function is calculated by the EM-algorithm of Section 5.3.

Interestingly, the profile log-likelihood is maximized for $(\hat{\alpha}, \hat{\sigma}^2) = (0.434, 11.770)$. This implies that the best fit is obtained for a model which is quite far from a logistic form. A closer look at data reveals why this is the case. The right panel of Figure 3 shows the observed death rates in 1980 with three fitted curves imposed. The curve labeled 'Makeham with stable frailty' is the one corresponding to the MLE, while the curve labeled 'Makeham' is the fitted Makeham curve mentioned earlier. If we restrict attention to Gamma frailty, i.e. maximize the profile log-likelihood along the axis $\alpha = 0$, we find the maximum $\hat{\sigma}_\Gamma^2 = 0.113$. The corresponding fitted curve is the one labeled 'Makeham with Gamma frailty'.

Compared to the Makeham curve, the MLE curve gives an improved fit at young ages and a similar fit at old ages, while the Gamma curve gives an improved fit at old ages and a similar fit at young ages. Since the exposure is much bigger at young ages than at old ages the global MLE is the one fitting young ages the best. Ideally we would like a model which combines the MLE fit for young ages with the Gamma fit at old ages. Note that the observed death rates are actually decreasing from age 20 to around age 30. This however cannot be captured by any of the models considered.

4.1.2 Forecast

Figure 4 shows the estimated and forecasted level, θ^1, and slope, θ^2, parameters of individual senescent mortality. The forecast is produced as the mean forecast of a fitted two-dimensional random walk with drift,

$$\theta_t = \theta_{t-1} + \xi + U_t, \tag{51}$$

where ξ is the drift and U_t are independent, identically distributed two-dimensional normal variables with mean zero. Qualitatively similar plots hold for the Gamma frailty model (not shown).

It appears that there is a structural break around 1970 where the annual drift in level and slope changes. With hindsight we now know that this change was genuine and that using only the period 1970–1980 would have resulted in a better forecast. Imagine standing in 1980, however, it is less obvious that we would have used only the last 10 years of data when forecasting. To illustrate the forecast the method would likely have

produced in 1980 we therefore use the mean forecast based on the average drift for the full period 1950–1980, as shown in Figure 4.
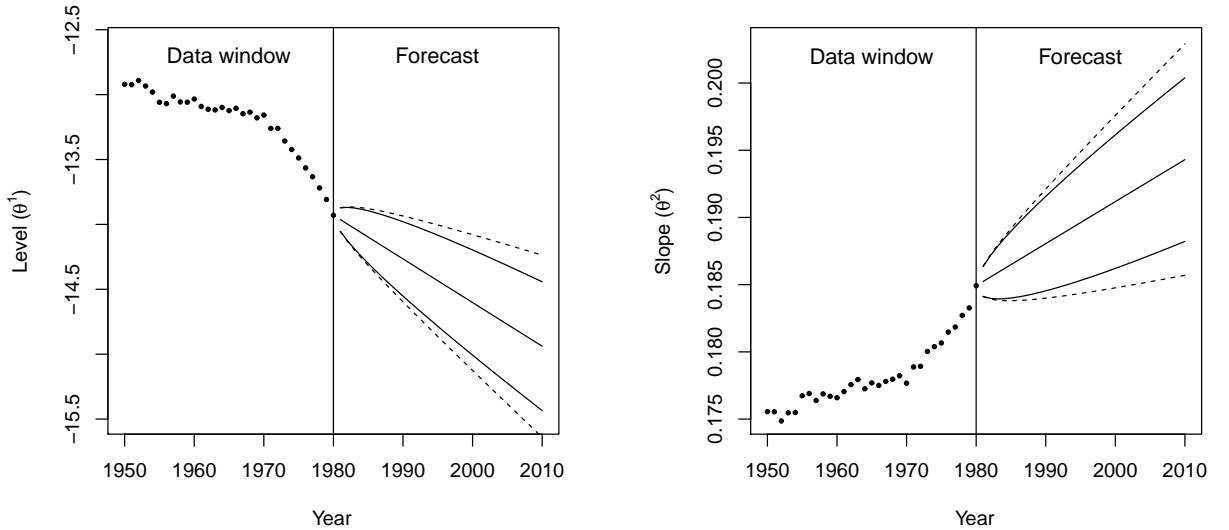

Figure 4: The dots show maximum likelihood estimates of level, θ^1, and slope, θ^2, parameters of model (48)–(49) with stable frailty. Mean forecasts based on model (51) are shown as solid lines. 95%-confidence intervals with and without parameter uncertainty on ξ (drift) are shown as dashed and solid lines respectively.

Cohort mortality with stable frailty is forecasted by

$$\mu(t,x) = \exp(\bar{\theta}_t^1 + \bar{\theta}_t^2 x) \left[1 + \frac{11.770}{1 - 0.434} \tilde{I}(t,x) \right]^{0.434-1} + \exp(\bar{\zeta}_t), \qquad (52)$$

where $\bar{\theta}$ denotes the mean forecast of θ, cf. Section 5.4. Background mortality shows very little variation over the data window and is therefore kept fixed at the last value in the forecast, $\bar{\zeta}_t = \hat{\zeta}_{1980}$. Similarly, cohort mortality with Gamma frailty is forecasted by

$$\mu(t,x) = \frac{\exp(\bar{\theta}_t^1 + \bar{\theta}_t^2 x)}{1 + 0.113 \tilde{I}(t,x)} + \exp(\bar{\zeta}_t). \qquad (53)$$

The fit and forecast with the two frailty specifications are shown in Figure 5 together with observed death rates from 1950 to 2010. In the data window, the model with stable frailty is seen to provide a very good fit for all ages except the oldest (age 100), while Gamma frailty leads to a good fit for all ages above 50, but a poor fit below. None of the models are able to fully predict the improvements that occurred from 1980 onwards in particular for ages 60–80. Of the two models the Gamma frailty model does in fact come closest due to its more log-concave forecasts, cf. the discussion of Gamma versus inverse Gaussian frailty in Section 2.1.

Overall, the example shows that adding frailty can substantially improve the fit of a simple baseline model. However, as with all parsimonious models the fit is not perfect, in particular when we consider an age span as wide as 20 to 100 years. The example also shows that regarding both fit and forecast of old-age mortality Gamma frailty performs better than (other) stable frailties.
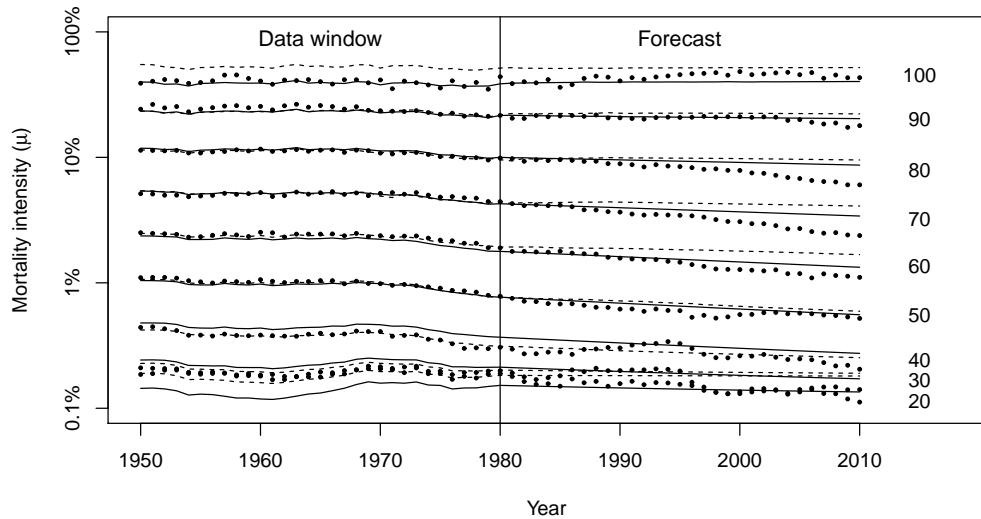
Figure 5: The dots show observed death rates for US males from 1950 to 2010 for ages 20, 30, ..., 100 years. The fit and forecast of model (48)–(49) with optimal Gamma and stable frailty are shown as solid and dashed lines respectively

4.2 Lee-Carter application

The model of Lee and Carter (1992) assumes a log-bilinear structure of mortality,

$$\log \mu(t, x) = a_x + b_x k_t, \tag{54}$$

where a and b are age-specific parameters and k is a time-varying index. The index is typically modeled as a random walk with drift,

$$k_t = k_{t-1} + \xi + U_t, \tag{55}$$

where ξ is the drift and U_t are independent, identically distributed normal variables with mean zero. The combination of (54) and (55) implies that forecasted age-specific death rates decay exponentially at a constant rate. However, the experience of old-age mortality has shown increasing rates of improvement in many countries and, consequently, forecasts based on the Lee-Carter methodology have a tendency to underestimate the actual gains. In the example to follow we illustrate how the addition of frailty can be used to improve the forecast of old-age mortality.

We consider the Poisson version of the Lee-Carter model,

$$D(t, x) \sim \text{Poisson}(\exp(a_x + b_x k_t) E(t, x)), \tag{56}$$

and we use the algorithm of Brouhns et al. (2002) to obtain the maximum likelihood estimates of the parameters under the usual identifiability constraints,

$$\sum_t k_t = 0 \quad \text{and} \quad \sum_x b_x = 1. \tag{57}$$

19

We will use the Lee-Carter model with and without frailty to model mortality for ages 0 to 90, and assume a logistic form for higher ages,[6]

$$\mu(t, x) = \frac{\exp(c_t + d_t x)}{1 + \exp(c_t + d_t x)} \quad \text{for } x = 91, \ldots, 110. \tag{58}$$

The extrapolation at the oldest ages is necessary to obtain reliable and stable rates, both historically where data are sparse and in forecasts. The left panel of Figure 6 shows the fit in 2010 for the full age-span from 0 to 110 years. The plot illustrates the flexibility and better fit to data of semi-parametric models, like the Lee-Carter model, over parametric models like the ones considered in Section 4.1. Parametric models, on the other hand, generally produce forecasts which better preserve the overall structure of data.
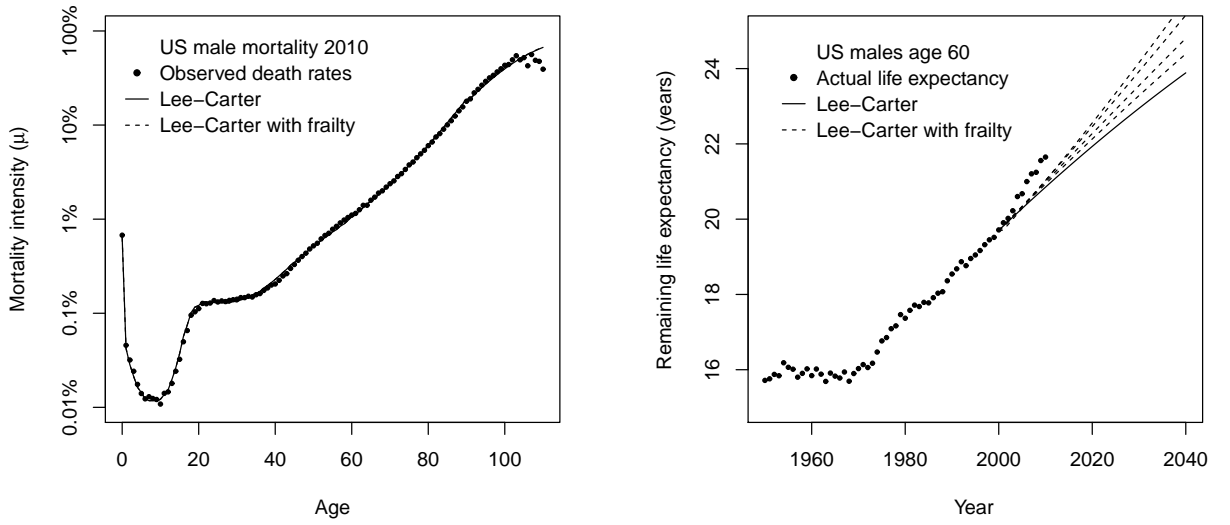

Figure 6: Left panel shows observed death rates in 2010 for US males ages 0 to 110. The fitted value for model (60)–(61) without frailty (solid) and with $\hat{\sigma}^2 = 0.73$ (dashed) are superimposed. Logistic extension above age 90. Right panel shows actual and forecasted life expectancy for US males age 60. Forecasts are based on model (60)–(61) without (solid) and with (dashed) frailty and estimation period 1970–2000. The frailty forecasts are for $\sigma^2 = 0.25$, 0.5, 1.0 and 1.5 with higher values of σ^2 corresponding to higher forecasts.

The four Lee-Carter forecasts used in the introductory example in Section 1 are produced as described above. The mean of the index, $\bar{k}_t$, is used for forecasting assuming an underlying random walk with drift. The period life expectancy used for illustration is calculated by the following formula with $x_0 = 60$

$$e_{x_0}(t) = \int_{x_0}^{110} \exp\left(-\int_{x_0}^{x} \mu(t, \lfloor y \rfloor) dy\right) dx, \tag{59}$$

where $\lfloor y \rfloor$ denotes the integer part of y. Arguably, the cohort life expectancy taking future improvements into account is of more interested in practise. However, for our purposes the period life expectancy is more useful since it can be compared with the actual experience.

---

[6] For each t, c_t and d_t are estimated by least squares based on the relation logit $\mu(t, x) = c_t + d_t x + \text{error}$, for ages $x = 70, 71, \ldots, 90$.

4.2.1 Stochastic frailty model

Consider the following multiplicative stochastic frailty model

$$D(t,x) \sim \text{Poisson}\left(\mu(t,x)E(t,x)\right), \tag{60}$$

$$\mu(t,x) = \mathsf{E}[Z|t,x]\exp(a_x + b_x k_t), \tag{61}$$

where Z is assumed to be Gamma distributed with mean one and variance σ^2. Due to the assumption of Gamma frailty we have the relations

$$\mathsf{E}[Z|t,x] = \left(1 + \sigma^2 I(t,x)\right)^{-1} = \exp\left(-\sigma^2 H(t,x)\right), \tag{62}$$

where H and I denote integrated cohort and baseline mortality respectively. For $\sigma^2 = 0$ we have the ordinary Lee-Carter model.

We will deviate slightly from the theory presented so far and use the period version of H and I rather than the cohort version previously used. Hence we will replace H by the quantity

$$\check{H}(t,x) = \sum_{u=0}^{x-1} m(t,u), \tag{63}$$

when estimating the model. Similarly, we calculate and forecast I by its period version $\check{I}$. Note that with the period version we do not need to extend the data window.

In the context of the Lee-Carter model the period version appears to stabilize the estimates and forecasts, and this is why we use it. We also note that the use of Gamma frailty on a period basis is consistent with the logistic extension above age 90.

For fixed value of σ^2 model (60)–(61) is estimated as an ordinary Lee-Carter model with exposure $\exp(-\sigma^2 \check{H}(t,x))E(t,x)$. With the large number of parameters of the Lee-Carter model the addition of σ^2 makes little difference for the fit. This is illustrated in the left panel of Figure 6 where the fit of the Lee-Carter model with and without frailty is indistinguishable.

Since frailty is introduced with the specific aim of improving the forecast we propose to estimate σ^2 with a back-test approach, rather than by maximum likelihood estimation. Specifically, we define the forecast fit by

$$f(\sigma^2) = \sum_{t=2001}^{2010} \sum_{x=0}^{90} \left\{ D(t,x)\log\mu(t,x;\sigma^2) - \mu(t,x;\sigma^2)E(t,x) \right\}, \tag{64}$$

where $\mu(t,x;\sigma^2)$ denotes the forecast of model (60)–(61) estimated with data from 1970–2000 and ages 0–90 for fixed value of σ^2. The model is forecasted as described in Section 4.2.2 below. Plotting f shows a unimodal function which is maximized at $\hat{\sigma}^2 = 0.73$ (plot not shown). The chosen approach is simple and serves as illustration, but many other choices for measure of fit and data period could have been made.

The right panel of Figure 6 shows the life expectancy of US males age 60 resulting from the forecast $\mu(t,x;\sigma^2)$ for different values of σ^2. The impact of frailty is initially modest but it gradually leads to substantial differences. Yet, even the presence of frailty cannot fully explain the rapid increase in life expectancy of the last decade. Note that the optimal frailty variance is not the one leading to the highest forecast, but the one giving the best fit overall for ages 0–90 in the ten years, 2001–2010, following the estimation period.

4.2.2 Forecast

Forecasts of model (60)–(61) are obtained by

$$\mu(t,x) = \frac{\exp(\hat{a}_x + \hat{b}_x \bar{k}_t)}{1 + \hat{\sigma}^2 \check{I}(t,x)}, \tag{65}$$

where $\hat{a}_x$, $\hat{b}_x$ and $\hat{\sigma}^2$ denote estimated parameters and $\bar{k}_t$ is the forecasted index, either deterministic or stochastic. The integrated baseline intensity is given by $\check{I}(t,x) = [\exp(\hat{\sigma}^2 \check{H}(t,x)) - 1]/\hat{\sigma}^2$ in the data window and by

$$\check{I}(t,x) = \sum_{u=0}^{x-1} \exp(\hat{a}_u + \hat{b}_u \bar{k}_t) \tag{66}$$

in the forecast.

The left panel of Figure 7 shows the estimated index for the estimation period 1980–2010 and frailty variance $\hat{\sigma}^2 = 0.73$. The estimated index looks close to linear and a linear forecast seems reasonable. The right panel of Figure 7 shows the corresponding life expectancy forecast for US males age 60 as the dashed line starting in 2010. The forecast is almost linear while the ordinary Lee-Carter forecast (solid line) curves downwards over time. For comparison the plot also shows Lee-Carter forecasts with and without frailty based on the estimation periods 1950–1980, 1960–1990 and 1970–2000.[7] The pattern is the same with almost linear frailty forecasts and lower, downward-curving Lee-Carter forecasts. Note that all historic forecasts, both with and without frailty, are below the actual life expectancy evolution.

The effect of frailty in terms of the forecasted age-specific death rates is illustrated in Figure 8. For ages below 70 the Lee-Carter forecasts with and without frailty are almost identical. For higher ages the frailty effect gradually increases and gives rise to the previously announced curving forecasts. In particular, the frailty forecast predicts substantial improvements in mortality for the very old over the next 50 years while the ordinary Lee-Carter forecasts predicts essentially no improvements above age 90. The plot also illustrates that the forecasted age-profile is somewhat distorted. This is a general feature of semi-parametric methods, like the Lee-Carter model, due to the lack of structure.

History shows that mortality rates generally decline over time, but it also shows changing rates of improvement. Clearly, there are substantial period effects due to medical breakthroughs, changes in nutrition and working conditions etc., and frailty can at best be a contributing factor to the observed increasing rates of improvement. Indeed, the example shows that adding frailty to historic forecasts is not enough to explain the actual evolution. However, the addition of frailty leads to projected old-age mortality rates which at least partly accommodate future higher rates of improvement. The resulting almost linear life expectancy projections resemble the actual experience better than the downward-curving Lee-Carter forecasts.

---

[7] To show the effect of the estimation period rather than the level of frailty the same value of $\hat{\sigma}^2 = 0.73$ is used for all forecasts.
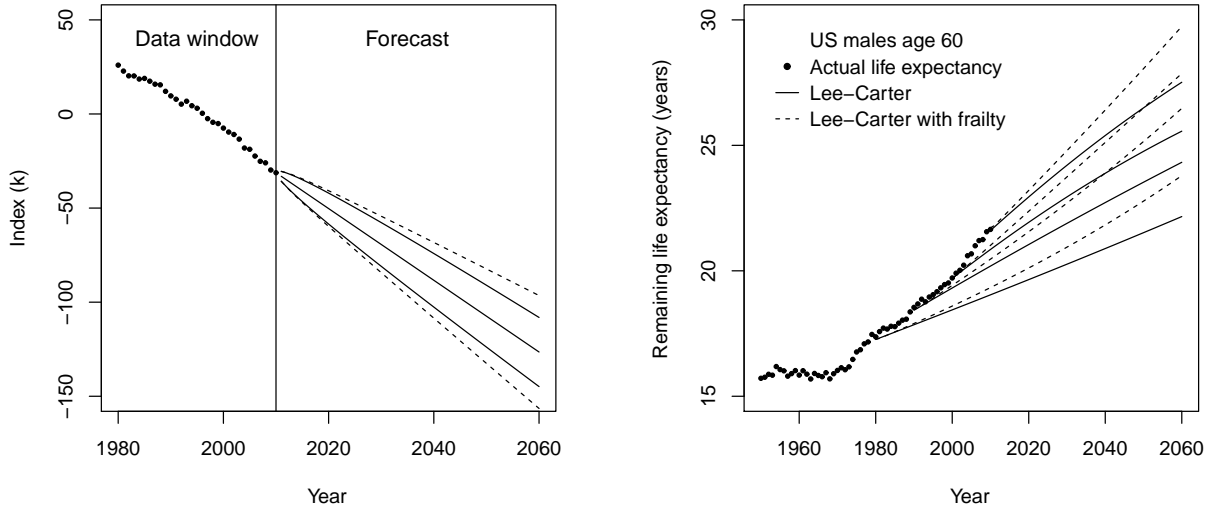
Figure 7: Left panel shows estimated (dots) and forecasted index, k, of model (60)–(61) with $\hat{\sigma}^2 = 0.73$ and estimation period 1980–2010. 95%-confidence intervals with and without parameter uncertainty on ξ (drift) are shown as dashed and solid lines respectively. Right panel shows actual and forecasted life expectancy for US males age 60. Forecasts are based on model (60)–(61) without frailty (solid) and with $\hat{\sigma}^2 = 0.73$ (dashed) for four different estimation periods, 1950–1980, 1960–1990, 1970–2000 and 1980–2010.

5   Generalized stochastic frailty models

The theory presented in Section 3 allows for estimation and forecasting of multiplicative frailty models. This is a useful class of models, but it can also be relevant to consider a larger and more flexible class of additive models with both frailty and non-frailty components. With additive models we can distinguish between senescent mortality influenced by frailty and selection and "background" mortality due to e.g. accidents with no selection effects. The class introduced below contains e.g. logistic-type models with constant terms which is useful when modeling mortality for a wide age span.

In the following we define the class of generalized stochastic frailty models and we show how to estimate and forecast these models. Additive models complicates the theory somewhat and we will have to resort to an EM-algorithm for estimation. The exposition is rather brief and focuses on the areas where the theory differs from the one presented in Section 3. Many of the general remarks still apply, but they will not be repeated here.

The models we consider are based on the underlying assumption that the intensity for an individual of age x at time t with frailty z is of the form

$$\mu(t, x; z) = z\, \mu_{\text{base}}(t, x) + \mu_{\text{back}}(t, x), \tag{67}$$

where μ_{base} is the baseline intensity influenced by individual frailty and μ_{back} is the background mortality common to all individuals independent of their frailty. The cohort intensity is then given by

$$\mu(t, x) = \mathsf{E}[Z|t, x]\, \mu_{\text{base}}(t, x) + \mu_{\text{back}}(t, x), \tag{68}$$
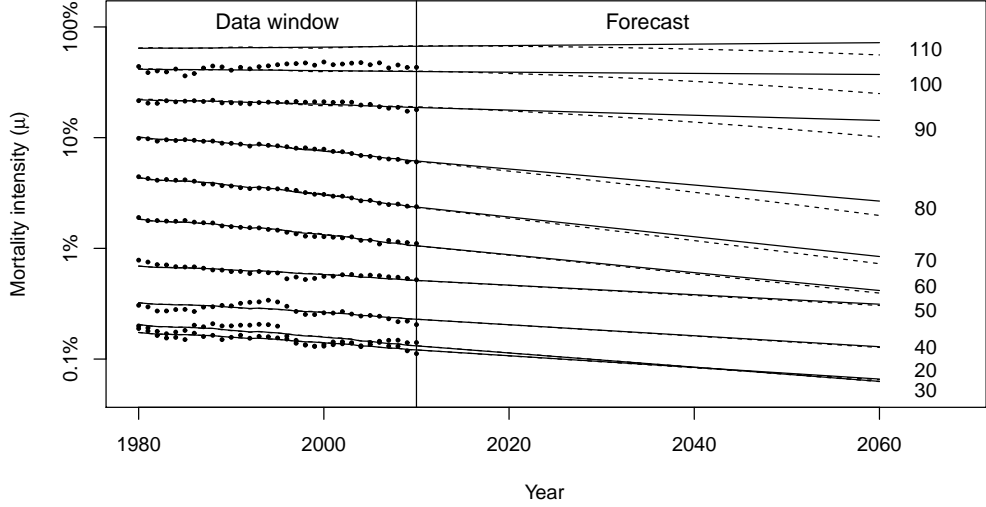
23

Figure 8: The dots show observed death rates for US males from 1980 to 2010 for ages 20, 30, ..., 100 years. The fit and forecast of model (60)–(61) without frailty (Lee-Carter model) and $\hat{\sigma}^2 = 0.73$ are shown as solid and dashed lines respectively. The model is fitted to data from 1980–2010 and ages 0–90. Mortality for higher ages is obtained by logistic extrapolation.

where $\mathsf{E}[Z|t,x]$ denotes the (conditional) mean frailty of the cohort of age x at time t, i.e. the cohort born at time $t - x$.

Assume that all cohorts have the same frailty distribution at birth and denote its Laplace transform by L. By a straightforward generalization of the results in Section 2 we have $H(t,x) = \nu(I(t,x))$ and

$$\mathsf{E}[Z|t,x] = \nu'(I(t,x)) = \nu'\left(\nu^{-1}\left\{H(t,x)\right\}\right), \qquad (69)$$

where $\nu = -\log L$ and

$$I(t,x) = \int_0^x \mu_{\text{base}}(u + t - x, u)du, \qquad (70)$$

$$H(t,x) = \int_0^x \mu(u + t - x, u) - \mu_{\text{back}}(u + t - x, u)du. \qquad (71)$$

Note that in contrast to earlier H is not just the integrated cohort intensity. The frailty independent part of mortality, μ_{back}, needs to be subtracted to get the usual connection between I and H.

5.1 Model structure

Assume we have models for baseline and background mortality of the form

$$\mu_{\text{base}}(t,x) = F(\theta_t, \eta_x), \qquad (72)$$

$$\mu_{\text{back}}(t,x) = G(\zeta_t, \omega_x), \qquad (73)$$

24

where θ_t and η_x together with F describe the period and age effects of frailty dependent mortality, and ζ_t and ω_x together with G describe the period and age effects of frailty independent mortality. All quantities can be multi-dimensional. Further generalizations are possible, but the chosen structure is sufficient to illustrate the idea.

For ease of presentation we assume that η_x and ω_x are to be estimated from data, but they may in fact be fixed. As a simple example we could use a Gompertz law for baseline mortality, $F(\theta_t, x) = \exp(\theta_t^1 + \theta_t^2 x)$, and a constant for background mortality, $G(\zeta_t) = \exp(\zeta_t)$. With no frailty this is the Gompertz-Makeham model, and when frailty is added to baseline mortality we get the logistic-type model used in Section 4.1.

A generalized stochastic frailty model is a model where death counts are independent with

$$D(t,x) \sim \text{Poisson}\left(\mu(t,x)E(t,x)\right), \tag{74}$$

$$\mu(t,x) = \mathsf{E}[Z|t,x]\,\mu_{\text{base}}(t,x) + \mu_{\text{back}}(t,x), \tag{75}$$

where μ_{base} and μ_{back} are given by (72)–(73) and $\mathsf{E}[Z|t,x]$ denotes the conditional mean frailty of the cohort of age x at time t. The frailty distribution at birth is the same for all cohorts and it is assumed to belong to a family indexed by ϕ. The Laplace transform of the frailty distribution with index ϕ is denoted L_ϕ, and this is assumed available in explicit form. Further, as a matter of convention we assume that mean frailty is one at birth and we define $\nu_\phi = -\log L_\phi$.

5.2 Pseudo-likelihood function

Estimation will be based on a pseudo-likelihood function in which the problematic term $\mathsf{E}[Z|t,x]$ is replaced by an estimate. This opens for a general, (relatively) easy estimation procedure which can be implemented whenever the baseline and background models can be estimated separately.

The idea is to estimate H in the data window by

$$\tilde{H}(t,x,\zeta,\omega) = \sum_{u=0}^{x-1} \tilde{m}(u+t-x, u), \tag{76}$$

where

$$\tilde{m}(t,x) = \begin{cases} \tilde{m}(t_{\min}, x) & \text{for } t < t_{\min} \text{ and } x_{\min} \leq x \leq x_{\max}, \\ m(t,x) - G(\zeta_t, \omega_x) & \text{for } t_{\min} \leq t \leq t_{\max} \text{ and } x_{\min} \leq x \leq x_{\max}, \\ 0 & \text{for } 0 \leq x < x_{\min}. \end{cases} \tag{77}$$

The reason we need to extend $\tilde{m}$ by (77) is that the summation in (76) falls partly outside the data window. Apart from the subtraction of background mortality the construction is the same as in Section 3.3.

Appealing to (69) we propose to base estimation of model (74)–(75) on a likelihood function in which $\mathsf{E}[Z|t,x]$ is replaced by $\nu'(\nu^{-1}\{\tilde{H}(t,x,\zeta,\omega)\})$. The resulting approximate

likelihood function is referred to as the pseudo-likelihood function,

$$L(\phi, \theta, \eta, \zeta, \omega) = \prod_{t,x} \frac{\lambda(t,x)^{D(t,x)}}{D(t,x)!} \exp(-\lambda(t,x)), \tag{78}$$

$$\lambda(t,x) = \left[\nu'_\phi(\nu_\phi^{-1}\{\tilde{H}(t,x,\zeta,\omega)\}) F(\theta_t, \eta_x) + G(\zeta_t, \omega_x) \right] E(t,x). \tag{79}$$

Formally, it corresponds to estimating the modified model

$$D(t,x) \sim \text{Poisson}\left(\mu(t,x)E(t,x)\right), \tag{80}$$

$$\mu(t,x) = \nu'_\phi(\nu_\phi^{-1}\{\tilde{H}(t,x,\zeta,\omega)\}) F(\theta_t, \eta_x) + G(\zeta_t, \omega_x). \tag{81}$$

5.3 Maximum likelihood estimation

Maximum likelihood estimates of model (80)–(81) can be obtained by optimization of the profile log-likelihood function,

$$l(\phi) = \log L(\phi, \hat{\theta}(\phi), \hat{\eta}(\phi), \hat{\zeta}(\phi), \hat{\omega}(\phi)), \tag{82}$$

where L is given by (78) and $\hat{\theta}(\phi)$, $\hat{\eta}(\phi)$, $\hat{\zeta}(\phi)$ and $\hat{\omega}(\phi)$ denote the maximum likelihood estimates for fixed value of the frailty parameter ϕ. As argued in Section 3.4 the frailty parameter is typically of low dimension, and the profile log-likelihood function can therefore be optimized by standard optimization routines. The issue remains of how to compute the profile log-likelihood function.

Assume that we have available routines for maximum likelihood estimation of the baseline and background mortality models separately. Utilizing that model (80)–(81) is a special case of a competing risks model, we can compute the estimates $\hat{\theta}(\phi)$, $\hat{\eta}(\phi)$, $\hat{\zeta}(\phi)$ and $\hat{\omega}(\phi)$ by the following algorithm. The algorithm is an adapted version of the EM-algorithm described in Appendix A. Let $i = 1$ and proceed as follows

1. Choose initial values for baseline and background parameters, θ^0, η^0, ζ^0 and ω^0.

2. For all t and x in the data window compute $\tilde{H}(t,x,\zeta^{i-1},\omega^{i-1})$ by (76), and let $c(t,x) = \nu'_\phi(\nu_\phi^{-1}\{\tilde{H}(t,x,\zeta^{i-1},\omega^{i-1})\})$.

3. Calculate θ^i and η^i as the maximum likelihood estimates for the model

$$D_{\text{base}}(t,x) \sim \text{Poisson}\left(c(t,x)E(t,x)F(\theta_t,\eta_x)\right) \tag{83}$$

   with "data"

$$D_{\text{base}}(t,x) = D(t,x)\frac{c(t,x)F(\theta_t^{i-1},\eta_x^{i-1})}{c(t,x)F(\theta_t^{i-1},\eta_x^{i-1}) + G(\zeta_t^{i-1},\omega_x^{i-1})}. \tag{84}$$

4. Calculate ζ^i and ω^i as the maximum likelihood estimates for the model

$$D_{\text{back}}(t,x) \sim \text{Poisson}\left(E(t,x)G(\zeta_t,\omega_x)\right) \tag{85}$$

   with "data"

$$D_{\text{back}}(t,x) = D(t,x)\frac{G(\zeta_t^{i-1},\omega_x^{i-1})}{c(t,x)F(\theta_t^{i-1},\eta_x^{i-1}) + G(\zeta_t^{i-1},\omega_x^{i-1})}. \tag{86}$$

5. Increase i by one.

6. Repeat steps 2–5 until convergence.

Note that the two Poisson models are only formal in the sense that D_{base} and D_{back} are not integer-valued, see Appendix A for details.

5.4 Forecasting

Forecasting is performed essentially as for the multiplicative case. Assume that we have maximum likelihood estimates $(\hat{\phi}, \hat{\theta}, \hat{\eta}, \hat{\zeta}, \hat{\omega})$ of model (80)–(81). Also, assume that we have a forecast, either deterministic or stochastic, $(\bar{\theta}_t, \bar{\zeta}_t)_{t_{\max} < t \leq t_{\max}+h}$ for a given forecast horizon h. The aim is to forecast baseline, background and cohort mortality for ages $x_{\min} \leq x \leq x_{\max}$ and years $t_{\max} < t \leq t_{\max} + h$.

The estimated intensity can be written

$$\hat{\mu}(t,x) = \nu'_{\hat{\phi}}(\tilde{I}(t,x))F(\hat{\theta}_t, \hat{\eta}_x) + G(\hat{\zeta}_t, \hat{\omega}_x), \tag{87}$$

where $\tilde{I}(t,x) = \nu_{\hat{\phi}}^{-1}(\tilde{H}(t,x,\hat{\zeta},\hat{\omega}))$. Expressing mean frailty in terms of $\tilde{I}$ facilitates forecasting of cohort mortality as in Section 3.5.

Baseline and background mortality are forecasted by inserting $\bar{\theta}$ and $\hat{\eta}$ into (72) and $\bar{\zeta}$ and $\hat{\omega}$ into (73),

$$\mu_{\text{base}}(t,x) = F(\bar{\theta}_t, \hat{\eta}_x), \tag{88}$$
$$\mu_{\text{back}}(t,x) = G(\bar{\zeta}_t, \hat{\omega}_x), \tag{89}$$

while cohort mortality is forecasted by

$$\mu(t,x) = \nu'_{\hat{\phi}}(\tilde{I}(t,x))F(\bar{\theta}_t, \hat{\eta}_x) + G(\bar{\zeta}_t, \hat{\omega}_x), \tag{90}$$

where $\tilde{I}$ in the forecast region is given by recursion (44).

Note that in the data window $\tilde{I}$ is given by a transformation of observed death rates with estimated background mortality subtracted, while in the forecast region it is defined recursively in terms of the forecasted baseline mortality. In particular, G does not enter $\tilde{I}$ in the forecast.

6 Final remarks

Populations are without doubt heterogeneous. In this paper we have investigated the use of multiplicative and additive frailty models taking this fact into account. We have shown how frailty can improve the fit of simple parametric models, and how it can be combined with the Lee-Carter model to generate more plausible old-age mortality projections. Obviously, frailty on its own is not enough to explain the complex dynamics of mortality, but the models can help capture certain essential features observed in data, e.g. changing rates of improvements, otherwise addressed by ad hoc methods.

Stochastic frailty models as here presented offer a general way to combine essentially any frailty distribution with a parametric or semi-parametric baseline model. This sets the work apart from the typical use of frailty which rely on matching parametric forms and closed-form expressions.

The proposed pseudo-likelihood approach is easy to implement for multiplicative models. Additive models are somewhat harder, but they can be handled by use of the EM-algorithm. In principle, additive models with more than two components, and multiple sources of frailty, can also be analyzed. However, unless the components target different segments of data identification might be problematic.

For ease of exposition we have presented the models as single population models. However, it is often desirable to consider joint models to obtain coherent forecasts for related populations, e.g. males and females, subpopulations of a given populations, or similar national populations believed to share a common trend. A number of models of this type has been proposed, see e.g. Li et al. (2004); Li and Lee (2005); Plat (2009); Biatat and Currie (2010); Li and Hardy (2011). Apart from notational changes the "fragilization" methodology can be used in this context also.

The models considered are related to the cohort models of e.g. Renshaw and Haberman (2006) and Cairns et al. (2009) in the sense that they all focus on the evolution of cohorts through time. The cohort models are designed to capture the so-called cohort effect seen in some data sets, i.e. the phenomenon that the mortality experience of cohorts born in certain periods differ markedly from the experience of neighboring cohorts, see e.g. Willets (2004); Richards et al. (2006). Cohort effects could be incorporated in the present framework by allowing the frailty distribution to change over time.

We generally rely on maximum likelihood estimation for determining parameter values. However, if the primary aim is mortality projection it can be argued that some parameters, e.g. frailty parameters, should not be determined solely on the basis of historic fit but rather on their forecast ability. We gave an example of how this could be done in the Lee-Carter application, but a more systematic analysis of how best to determine the frailty distribution remains to be done.

Acknowledgements

A   Estimation of competing risks model

Consider the competing risks model

$$D(x) \sim \text{Poisson}\left(E(x)[\mu_1(\alpha, x) + \mu_2(\beta, x)]\right), \tag{91}$$

where D and E denote, respectively, death counts and exposures indexed by age x, and μ_1 and μ_2 are mortality intensities with parameters α and β respectively. For notational

simplicity we do not include time and we not specify exactly how the parameters enter the two mortality intensities. Apart from notational changes these omissions are of no consequences for the following discussion.

The interpretation of model (91) is that persons of age x can die from two different, independent sources with intensities μ_1 and μ_2 respectively. The structure is natural to consider in many contexts, but the likelihood function is complicated and direct estimation of α and β can be difficult. This appendix describes how maximum likelihood estimation can be implemented by the EM-algorithm of Dempster et al. (1977) assuming each individual model can be estimated. The following is a standard application of the EM-algorithm, but since we rely on it in Section 5 we here give a brief self-contained exposition for ease of reference.

The EM-algorithm is an algorithm for maximum likelihood estimation for incomplete, or missing, data. It consists of an E-step in which the expectation of the full log-likelihood is computed with respect to the missing data, followed by an M-step in which the expectation is maximized with respect to the parameters. The steps are iterated till convergence.

Imagine in the present setup that deaths had been recorded according to source such that

$$D_1(x) \sim \text{Poisson} \left(E(x)\mu_1(\alpha, x) \right), \tag{92}$$

$$D_2(x) \sim \text{Poisson} \left(E(x)\mu_2(\beta, x) \right), \tag{93}$$

with $D_1(x)$ and $D_2(x)$ independent and $D(x) = D_1(x) + D_2(x)$. Then $D(x)$ would be distributed according to (91) and estimation of α and β would be easy. The full log-likelihood function is given by

$$l(\alpha, \beta; D_1, D_2) = l_1(\alpha; D_1) + l_2(\beta; D_2), \tag{94}$$

where (omitting x for ease of notation)

$$l_1(\alpha; D_1) = \sum_x \{ D_1 \log \mu_1(\alpha) - E\mu_1(\alpha) \} + \text{constant}, \tag{95}$$

$$l_2(\beta; D_2) = \sum_x \{ D_2 \log \mu_2(\beta) - E\mu_2(\beta) \} + \text{constant}. \tag{96}$$

Even though D_1 and D_2 do not necessarily exist and hence are not "missing" in the normal sense of the word, we can still use the EM-algorithm based on the missing data interpretation of the model. The algorithm is as follows:

Choose initial values α_0 and β_0 for the parameters.

E-step Treat D_1 and D_2 as missing data and compute the expected value of the full log-likelihood given data D and current parameter estimates

$$Q(\alpha, \beta) = \mathsf{E}[l(\alpha, \beta; D_1, D_2)|D, \alpha_{i-1}, \beta_{i-1}], \tag{97}$$

where

$$D_1|D, \alpha_{i-1}, \beta_{i-1} \sim \text{Binom} \left(D, \frac{\mu_1(\alpha_{i-1})}{\mu_1(\alpha_{i-1}) + \mu_2(\beta_{i-1})} \right), \tag{98}$$

$$D_2|D, \alpha_{i-1}, \beta_{i-1} \sim \text{Binom} \left(D, \frac{\mu_2(\beta_{i-1})}{\mu_1(\alpha_{i-1}) + \mu_2(\beta_{i-1})} \right). \tag{99}$$

M-step Maximize Q to obtain new estimates α_i and β_i.

Iterate E-step and M-step until parameter estimates converge.

Both the E-step and the M-step are easy to perform. The E-step merely consists of replacing D_1 and D_2 in (95)–(96) by their conditional expectations, and the M-step consists in estimating the two marginal models. Formally, the EM-algorithm corresponds to iteratively estimating the models (92)–(93) with "data"

$$D_1(x) = D(x)\frac{\mu_1(\alpha_{i-1}, x)}{\mu_1(\alpha_{i-1}, x) + \mu_2(\beta_{i-1}, x)}, \tag{100}$$

$$D_2(x) = D(x)\frac{\mu_2(\beta_{i-1}, x)}{\mu_1(\alpha_{i-1}, x) + \mu_2(\beta_{i-1}, x)}. \tag{101}$$

On each iteration "data" are updated according to the latest estimate and the parameters reestimated. Note, this holds only formally since D_1 and D_2 computed above are not integer-valued.

It can be shown that the likelihood is increased in each step of the EM-algorithm. The algorithm hence converges to a local maximum, but it can be rather slow. The advantage is that we can use the same algorithm to estimate all combinations of models as long as we know how to estimate the models separately. It is also straightforward to generalize the algorithm to more than two competing risks.

References

O. O. Aalen. Heterogeneity in survival analysis. *Statistics in Medicine*, 7:1121–1137, 1988.

O. O. Aalen. Modelling heterogeneity in survival analysis by the compound Poisson distribution. *The Annals of Applied Probability*, 2:951–972, 1992.

J. H. Abbring and G. J. van der Berg. The unobserved heterogeneity distribution in duration analysis. *Biometrika*, 94:87–99, 2007.

J. Besag. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 24:179–195, 1975.

V. D. Biatat and I. D. Currie. Joint models for classification and comparison of mortality in different countries. Proceedings of 25rd International Workshop on Statistical Modelling, Glasgow, pages 89–94, 2010.

J. Bongaarts. Long-range trends in adult mortality: Models and projection methods. *Demography*, 42:23–49, 2005.

H. Booth, J. Maindonald, and L. Smith. Applying Lee-Carter under conditions of varying mortality decline. *Population Studies*, 56:325–336, 2002.

N. Brouhns, M. Denuit, and J. K. Vermunt. A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31: 373–393, 2002.

Z. Butt and S. Haberman. Application of frailty-based mortality models using generalized linear models. *ASTIN Bulletin*, 34:175–197, 2004.

A. J. Cairns, D. Blake, and K. Dowd. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73: 687–718, 2006.

A. J. Cairns, D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, A. Ong, and I. Balevich. A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *North American Actuarial Journal*, 13:1–35, 2009.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 39:1–38, 1977.

W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. II, second edition*. Wiley, New York, 1971.

D. D. Hanagal. *Modeling survival data using frailty models*. Chapman & Hall, 2011.

P. Hougaard. Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika*, 71:75–83, 1984.

P. Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73:387–396, 1986.

S. F. Jarner and E. M. Kryger. Modelling adult mortality in small populations: the SAINT model. *ASTIN Bulletin*, 41:377–418, 2011.

R. D. Lee and L. R. Carter. Modeling and Forecasting of U.S. Mortality. *Journal of the American Statistical Association*, 87:659–675, 1992.

R. D. Lee and T. Miller. Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38:537–549, 2001.

J. S.-H. Li and M. R. Hardy. Measuring basis risk in longevity hedges. *North American Actuarial Journal*, pages 177–200, 2011.

J. S.-H. Li, M. R. Hardy, and K. S. Tan. Uncertainty in mortality forecasting: An extension to the classic Lee-Carter approach. *ASTIN Bulletin*, 39:137–164, 2009.

N. Li and R. Lee. Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42:575–594, 2005.

N. Li, R. Lee, and S. Tuljapurkar. Using the Lee-Carter method to forecast mortality for populations with limited data. *International Statistical Review*, 72:19–36, 2004.

K. G. Manton, E. Stallard, and J. W. Vaupel. Alternative models for the heterogeneity of mortality risks among the aged. *Journal of the American Statistical Association*, 81: 635–644, 1986.

A. Olivieri. Heterogeneity in survival models. applications to pensions and life annuities. *Belgian Actuarial Journal*, 6:23–39, 2006.

W. Perks. On some experiments on the graduation of mortality statistics. *Journal of the Institute of Actuaries*, 63:12–40, 1932.

R. Plat. Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance: Mathematics and Economics*, 45:123–132, 2009.

A. E. Renshaw and S. Haberman. Lee-Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33:255–272, 2003.

A. E. Renshaw and S. Haberman. A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38:556–570, 2006.

S. J. Richards, J. G. Kirkby, and I. D. Currie. The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal*, 12:5–38, 2006.

J. Spreeuw, J. P. Nielsen, and S. F. Jarner. A nonparametric visual test of mixed hazard models. *Statistics and Operations Research Transactions*, 37:149–170, 2013.

A. R. Thatcher. The Long-Term Pattern of Adult Mortality and the Highest Attained Age. *Journal of the Royal Statistical Society. Series A*, 162:5–43, 1999.

A. R. Thatcher, V. Kannisto, and J. W. Vaupel. *The Force of Mortality at Ages 80-120. Monographs on Population Aging 5*. Odense University Press, Odense, 1998.

S. Tuljapurkar, N. Li, and C. Boe. A universal pattern of mortality decline in the G7 countries. *Nature*, 405:789–792, 2000.

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

S. S. Wang and R. L. Brown. A frailty model for projection of human mortality improvements. *Journal of Actuarial Practise*, 6:221–241, 1998.

A. Wienke. *Frailty models in survival analysis*. Chapman & Hall, 2010.

R. C. Willets. The cohort effect: Insights and explanations. *British Actuarial Journal*, 10:833–877, 2004.