DISCUSSION PAPER PI-0902

Modelling Adult Mortality in Small Populations: The Saint Model

Søren Fiig Jarner & Esben Masotti Kryger

January 2009

ISSN 1367-580X

The Pensions Institute
Cass Business School
City University
106 Bunhill Row London
EC1Y 8TZ
UNITED KINGDOM

http://www.pensions-institute.org/

MODELLING ADULT MORTALITY IN SMALL POPULATIONS: THE SAINT MODEL

SØREN FIIG JARNER AND ESBEN MASOTTI KRYGER

Abstract. The mortality evolution of small populations often exhibits substantial variability and irregular improvement patterns making it hard to identify underlying trends and produce plausible projections. We propose a methodology for robust forecasting based on the existence of a larger reference population sharing the same long-term trend as the population of interest. The reference population is used to estimate the parameters in a frailty model for the underlying intensity surface. A multivariate time series model describing the deviations of the small population mortality from the underlying mortality is then fitted and forecasted. Coherent long-term forecasts are ensured by the underlying frailty model while the size and variability of short- to medium-term deviations are quantified by the time series model. The frailty model is particularly well suited to describe the changing improvement patterns in old age mortality. We apply the method to Danish mortality data with a pooled international data set as reference population.

1. Introduction

Mortality projections are of great importance for public financing decisions, health care planning and the pension industry. A large number of forecasts are being produced on a regular basis by government agencies and pension funds for various populations of interest. In many situations the population of interest is quite small, e.g. the population of a small region or the members of a specific pension fund, and historic data shows substantial variability and irregular patterns. Also, historic data may be available only for a relatively short period of time.

The prevailing methodology for making mortality projections is the method proposed by Lee and Carter (1992). The model describes the evolution in age-specific death rates (ASDRs) by a single time-varying index together with age-specific responses to the index. The structure implies that all ASDRs move up and down together, although not by the same amounts. The method has gained widespread popularity due to its simplicity and ease of interpretation and there has been a wealth of applications, see e.g. Tuljapurkar *et al.* (2000); Booth *et al.* (2006) and references therein. A number of extensions and improvements have been proposed, e.g. Brouhns *et al.* (2002); Lee

Date: December 2, 2008.

Key words and phrases. SAINT model, stochastic modelling, small populations, adult mortality, frailty, changing rates of improvement, Danish and international mortality.

and Miller (2001); Renshaw and Haberman (2006); Renshaw and Haberman (2003); de Jong and Tickle (2006); Currie *et al.* (2004); Cairns *et al.* (2006), but the original Lee-Carter method still serves as the point of reference.

1.1. **Small population mortality.** The structure of the Lee-Carter model makes it well suited to extrapolate regular patterns with constant improvement rates over time. However, while the mortality experience of large populations often conforms with this pattern the mortality evolution of small populations is generally much more irregular. Lack of fit of the Lee-Carter model for small populations, including the Nordic countries, was reported by Booth *et al.* (2006) in a comparative study. In Denmark, for instance, improvement rates have varied considerably over time within age groups and there has been periods with improvements in some age groups and stagnation or even slight increases in other age groups violating the Lee-Carter assumptions; see Jarner *et al.* (2008); Andreev (2002) for detailed accounts of the evolution of Danish mortality.

The characteristics of small population mortality makes forecasting based on past trends problematic and very sensitive to the fitting period. Naive extrapolation of historic trends in ASDRs is likely to lead to implausible projections and unrealistic age-profiles, e.g. old age mortality dropping below that of younger ages. Often, however, the population under study can be regarded as a subpopulation of a (larger) reference population obeying a more regular pattern of improvement, e.g. a region within a country, or a small country in a larger geographical area etc. Furthermore, it will often be reasonable to assume that the study and reference populations will share the same long-term trend.

In this paper we propose a methodology for robust small population mortality projection based on the identification of a reference population. The method consists of two steps: First, the reference population is used to estimate a parametric, underlying intensity surface which determines the long-term trend. Second, a multivariate, stationary time series model describing the deviations of the small population mortality from the trend is fitted. Projections are obtained by combining extrapolations of the parametric surface with forecasts of the time series model for deviations. Coherent long-term mortality profiles are guaranteed by the parametric surface while the purpose of the time series is to quantify the short- to medium-term variability in improvement rates of the small population.

It is common to base mortality forecasts on time series models. Typically, parameters describing the evolution of period life tables are estimated assuming either a parametric or non-parametric age-profile. A time series model, often a random-walk with drift, is then fitted and forecasted for each of the parameters. The role of the time series in these methods is to capture both the trend in parameters and their uncertain evolvement around the trend. The structure implies that large short-term variability invariably lead to even larger long-term variability. In contrast, the proposed method

by treating deviations from the trend as stationary allows for substantial short-term variability without inflating the long-term uncertainty.

1.2. **Old age mortality.** The modelling of old age mortality presents perhaps the most challenging part of mortality modelling. The historic development in Danish old age mortality, say age 70 and above, shows very modest improvement rates, far below those observed for the younger ages. However, over the last decade or so improvement rates have picked up and currently 70-year-olds experience improvement rates equalling those of younger ages. The picture is the same in most developed countries: old age mortality has historically improved at a slower pace than young and middle age mortality, but recently improvements rates have gradually risen.

The fact that old age improvement rates increase over time cause forecasts based on the Lee-Carter methodology to systematically under-predict the gains in old age mortality, cf. Lee and Miller (2001). As a consequence it has been recommended to use a shorter fitting period over which data conforms better with the assumptions of time-invariant improvement rates, see e.g. Lee and Miller (2001); Tuljapurkar *et al.* (2000); Booth *et al.* (2002). Although this approach clearly forecasts higher improvement rates in old age mortality it seems somewhat ad-hoc and not entirely satisfying. In this paper we take a different route.

Our ambition is to derive a simple model for the population dynamics in which changing mortality patterns naturally arise. This will allow us to characterize how mortality improvements change over time and to make predictions for future improvements in old and oldest-old mortality. Inspired by frailty theory we assume that the population consists of a heterogeneous group of individuals with varying degrees of frailty. Frail individuals have a tendency to die first causing a concentration of robust individuals at high ages. Taken this selection mechanism into account and assuming continued improvements over time a model for the entire intensity surface of the population over time can be derived. We will use the resulting parametric surface to describe the development of the reference population.

The frailty model offers an explanation to the observed lack of improvement in old age mortality. The mortality for a given age group at a given time is influenced by two factors: the current level of health and the average frailty. Over time the level of health improves but so does the average frailty. In effect, as mortality improves the selection mechanism to reach a given age weakens causing healthier but more frail individuals to become old. In the transition from high to low selection the two effects partly offset each other such that the aggregate mortality for the age group is almost constant. Eventually the health effect will dominate the selection effect and improvements will be seen. We will explore these effects in some detail and derive the asymptotic improvement pattern implied by the model.

For two reasons we focus on adult mortality only in our modelling, i.e. mortality for age 20 and above. First, the nature of infant and child mortality

is rather different from adult mortality and more complexity will have to be added to the model to fit the historic evolution adequately. Second, current levels of infant and child mortality are so low that their future course has very little impact on life expectancy and other aggregate measures. Hence, from a forecasting perspective not much is gained by the added complexity. In fact, with current mortality levels already very low up to age 60, say, future life expectancy gains will be driven almost exclusively by the development in old age mortality. However, by modelling the full adult mortality surface we are able to extract information about the general nature of improvement patterns and to predict when improvements will start to occur for age groups where none have been seen historically.

1.3. **Outline.** The rest of the paper is organized as follows: in Section 2.1 we present the proposed methodology for small population mortality modelling consisting of a separation of trends and deviations; in Section 2.2 we derive a parametric, frailty model for the underlying intensity surface and study some implications and asymptotic properties; and Section 2.3 contains a description of the time series model for deviations. In Section 3 we give an application to Danish data taken a large international data set as reference population, Section 4 contains a study of the fit and forecasting performance of the model, and in Section 5 we offer some concluding remarks and indicate further lines of research. All proofs are in the Appendix.

2. The model

2.1. **Methodology.** In the following we suggest a methodology for robust forecasting of small population mortality. The evolution of small population mortality is characterized by being more volatile and having less regular improvement patterns compared to what is observed in larger populations. These features make simple projection methodologies very sensitive to the choice of fitting period and lead to very uncertain long-term forecasts. The fundamental idea in the proposed method is to use a large population to estimate the underlying long-term trend and use the small population to estimate the deviations from the trend.

We distinguish between unsystematic and systematic variability. Unsystematic variability refers to the variability associated with the randomness of the time of deaths in a population with a known mortality intensity, while systematic variability refers to the variability of the mortality intensity itself. Since the populations we are interested in are small by assumption we expect noticeable unsystematic variability. For instance, we do not expect crude (unsmoothed) death rates, constructed from the mortality experience of a single year, to be strictly increasing with age although we believe it to hold for the underlying intensity (at least from some age).

However, even taken the larger unsystematic variability into account it appears that small populations also have a greater *systematic* variability

than larger populations. Presumably small populations are more homogeneous and thereby more influenced by specific effects. There are a number of reasons why this might be. Consider for instance the members of an occupational pension scheme. The members have the same or similar education and job, and probably also to some extent similar economic status and life style compared to the population at large. Similarly, the population in a specific country is influenced by common factors, e.g. the health care system and social habits such as smoking. The homogeneity implies that specific changes in e.g. socioeconomic conditions will have a greater impact on the mortality in a small population compared to a large population with greater diversity in background factors.

We will assume that the population under study can be regarded as a subpopulation of a larger population, e.g. the population of a province is a subpopulation of the national population, and a national population can be regarded a subpopulation of the total population of a larger geographical region, or of similarly developed countries. We will refer to the small population as the *subpopulation* and the large population as the *reference* population. Although both unsystematic and systematic variability will be greater in the subpopulation it is reasonable to assume that the sub- and reference populations will share the same long-term trends in mortality decline, even in the presence of substantial deviations in current mortality levels. The alternative is diverging levels of mortality which in the long run seems highly unlikely for related populations. Wilson (2001); Wilmoth (1998) provide evidence for convergence in global mortality levels due to convergence of social and economic factors.

2.1.1. *Data.* We will assume that data consists of death counts, $\{D(t,x)\}$, and corresponding exposures, $\{E(t,x)\}$, for a range of years t and ages x. Data is assumed to be available for both the sub- and reference population (distinguished by subscript sub and ref, respectively), but not necessarily for the same ranges of years and ages. Data will typically also be gender specific, but it does not have to be. Since it is of no importance for the formulation of the model we will suppress a potential gender dependence in the notation.

$D(t,x)$ denotes the number of deaths occurring in calender year t among people aged $[x, x+1)$, and $E(t,x)$ 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*.

2.1.2. *Model structure.* From the death counts and exposures we can form the (crude) death rates

$$m(t,x) = \frac{D(t,x)}{E(t,x)}, \tag{1}$$

which are estimates of the underlying intensity, or force of mortality, $\bar{\mu}(t,x)$[1].

In order to proceed we assume that we have a family of intensity surfaces $H_\theta(t,x)$ parameterized by θ, and we consider the model where death counts are independent with

$$D_{\text{ref}}(t,x) \sim \text{Poisson}\left(\bar{\mu}_{\text{ref}}(t,x)E_{\text{ref}}(t,x)\right), \tag{2}$$

and $\bar{\mu}_{\text{ref}}(t,x) = H_\theta(t,x)$. Based on this model we find the maximum likelihood estimate (MLE) for θ, denoted by $\hat{\theta}$. In principle we could use a Lee-Carter specification of $\bar{\mu}$ in which case the model is the one proposed in Brouhns *et al.* (2002). However, by assumption the evolution of the reference population is smooth which allows us to get a good fit with a more parsimonious specification. In Section 2.2 we will develop a specific family of intensity surfaces, which will be shown to provide a very good fit to the reference data in our application. The use of a parametric model also offers insights into the dynamics of improvement rates over time.

The next step is to model the deviations of the subpopulation from the reference population. We will refer to the deviations as the spread and we propose to use a model of the form

$$D_{\text{sub}}(t,x) \sim \text{Poisson}\left(\bar{\mu}_{\text{sub}}(t,x)E_{\text{sub}}(t,x)\right), \tag{3}$$

with

$$\bar{\mu}_{\text{sub}}(t,x) = H_{\hat{\theta}}(t,x)\exp(y_t' r_x) \tag{4}$$

where $y_t' = (y_{0,t}, \ldots, y_{n,t})$ and $r_x' = (r_{0,x}, \ldots, r_{n,x})$ for some n. Again, this does in fact allow a Lee-Carter specification of the deviations. However, we will consider the situation in which the r's are fixed regressors and only the y's are estimated (by maximum likelihood estimation). Note, that in this case the estimates of y_t only depend on data from period t. In Section 2.3 we will propose a specific model with three regressors corresponding to level, slope and curvature of the spread.

The last step is to fit a time-series model for the multivariate time-series y_t. We will use a VAR(1)-model for which standard fitting routines exist. If the assumptions behind the modelling approach are fulfilled the time-series should not display drift but rather fluctuate around some level (which may be different from zero). In other words, we expect y_t to be *stationary*.

Forecasts are readily produced by combining trend forecasts with time-series forecasts of the spread. Assuming independence between trend and spread we have, with a slight abuse of notation[2], the following variance decomposition

$$\text{Var}(\log \bar{\mu}_{\text{sub}}(t,x)) = \text{Var}(\log H_{\hat{\theta}}(t,x)) + \text{Var}(y_t' r_x). \tag{5}$$

[1]We use $\bar{\mu}$ to indicate an intensity surface which is constant over calender years and over integer ages. We reserve the use of μ, which will later be used to denote a continuous intensity surface.

[2]In equation (5) $\hat{\theta}$ is considered as an estimator with a distribution rather than a fixed number.

We see that there are two sources of (systematic) variability: the trend and the spread. For most specifications of H_θ the variance will increase with the forecasting horizon. The variance of the spread, however, will only increase up to a given level under the assumed stationarity of y_t. The model does not forecast the mortality of the subpopulation to convergence in an absolute sense to that of the reference population, but the spread will be bounded (in probability).

In the following sections we develop a specific model which falls within the framework described above. The model will subsequently be used in an application to Danish mortality taking an international data set as reference population. With this application in mind the model has been dubbed SAINT as an abbreviation for Spread Adjusted InterNational Trend.

2.2. **Trend modelling.** A great number of functional forms have been suggested as models for adult mortality, see e.g. Chapter 2 of Gavrilov and Gavrilova (1991). Classical forms include the ones associated with the name of Gompertz

$$\mu(x) = \alpha \exp(\beta x), \tag{6}$$

and Makeham

$$\mu(x) = \alpha \exp(\beta x) + \gamma. \tag{7}$$

Both of these capture the approximate exponential increase in intensity observed for adult mortality. The Makeham form also includes an age-independent contribution which can interpreted as a rate of accidents. The additional term, referred to as background mortality, improves the fit at young ages.

Old age mortality, however, is generally overestimated by the assumed exponential increase. Empirical data typically shows decreasing acceleration in mortality at old ages, or even a late-life mortality plateau. A functional form that captures both the (approximate) exponential growth rate seen in adult mortality and the subsequent sub-exponential increase at old ages is the logistic family

$$\mu(x) = \frac{\alpha \exp(\beta x)}{1 + \alpha \exp(\beta x)} + \gamma. \tag{8}$$

This form has been proposed as the basis for mortality modelling by several authors, e.g. Cairns *et al.* (2006), Bongaarts (2005), Thatcher (1999), and it has been shown to fit empiric data very well in a number of applications.[3]

[3]Cairns *et al.* (2006) use the logistic form to describe one-year death probabilities q_x, while we use it as model for the underlying intensity. The two quantities are related by $q_x = 1 - \exp(-\int_{x-1}^{x} \mu(y)dt) \approx \mu(x)$.

2.2.1. *Selection and frailty.* Various theories have been proposed trying to explain why the increase in the force of mortality slows down at old ages, see e.g. Thatcher (1999) and references therein. In this paper we will focus on frailty theory as it provides a flexible and mathematically tractable framework for modelling mortality.

The theory assumes that the population is heterogeneous with each person having an individual level of susceptibility, or frailty. Frail individuals have a tendency to die earlier than more robust individuals and this selection causes the frailty composition of the cohort to gradually change over time. The continued concentration of robust individuals in effect slows down the mortality of the cohort and causes the cohort intensity to increase less rapidly than the individual intensities. The following example illustrates the idea.

Example 2.1 (Gamma-Makeham model)**.** *Assume that the ith person of a cohort has his own Makeham intensity:*

$$\mu(x; z_i) = z_i \alpha \exp(\beta x) + \gamma, \tag{9}$$

where z_i is an individual frailty parameter, while α, β and γ are shared by all persons in the cohort. Assume furthermore that Z follows a (scaled) Γ-distribution with mean 1 and variance σ^2. The force of mortality for the cohort then becomes

$$\mu(x) = \mathsf{E}[Z|x]\alpha \exp(\beta x) + \gamma = \frac{\alpha \exp(\beta x)}{1 + \sigma^2 \alpha(\exp(\beta x) - 1)/\beta} + \gamma, \tag{10}$$

where $\mathsf{E}[Z|x]$ denotes the conditional mean frailty of the cohort at age x. The cohort intensity in this model is of logistic form with an asymptotic value of $\beta/\sigma^2 + \gamma$ as x tends to infinity. Hence, although each individual intensity is exponentially increasing the selection mechanism is so strong that the cohort intensity levels off at a finite value.[4]

2.2.2. *The multiplicative frailty model.* The ideas of selection and frailty can be generalized to describe the evolution in mortality rates over time for a whole population. In the following we assume the existence of a smooth intensity surface, $\mu(t, x)$, which represents the instantaneous rate of dying for a person aged x at time t, i.e. the probability that the person will die between time t and $t + dt$ is approximately $\mu(t, x)dt$ for small dt.

We start by considering a general, multiplicative frailty model where the mortality intensity for an individual with frailty z has the form

$$\mu(t, x; z) = z\mu_s^I(t, x) + \gamma(t). \tag{11}$$

Hence, individual intensities have a senescent (age-dependent) component, $z\mu_s^I(t, x)$, and a background (age-independent) component, $\gamma(t)$. Frailty is assumed to affect the senescent component only and its effect is assumed to

[4]In fact, if $\sigma^2 > \beta/\alpha$ the level of heterogeneity is so large, and the selection effect thereby so strong, that the cohort intensity is decreasing with age! For $\sigma^2 = \beta/\alpha$ the cohort intensity is constant, while for smaller values of σ^2 the cohort intensity is increasing as expected.

be multiplicative. Thus z measures the excess (senescent) mortality relative to the mortality of an individual with frailty 1. This multiplicative structure is crucial for the following developments. Vaupel *et al.* (1979) consider the multiplicative frailty model for a single cohort and state results similar to ours.

Proposition 2.1. *Assuming (11) the population mortality surface is given by*

$$\mu(t,x) = \mathsf{E}[Z|t,x]\mu_s^I(t,x) + \gamma(t), \tag{12}$$

where $\mathsf{E}[Z|t,x]$ *denotes the conditional mean frailty at time* t *for persons of age* x*.*

We will denote the senescent component of the population intensity by $\mu_s(t,x)$, i.e. $\mu_s(t,x) = \mathsf{E}[Z|t,x]\mu_s^I(t,x)$. The result of Proposition 2.1 holds true regardless of the assumed frailty distribution at birth. However, in order to obtain an analytically tractable model we will assume that frailties at birth follow a scaled Γ-distribution with mean 1 and variance σ^2. Under this assumption the conditional frailty distributions are also Γ-distributed and explicit expressions for the conditional mean and variance can be derived. Let $Z|(t,x)$ denote the conditional frailty distribution at time t for persons of age x.

Proposition 2.2. *Assuming (11) and* $Z \sim \Gamma$ *with mean* 1 *and variance* σ^2 *then* $Z|(t,x) \sim \Gamma$ *with mean and variance given by*

$$\mathsf{E}[Z|t,x] = \left(1 + \sigma^2 I(t,x)\right)^{-1}, \tag{13}$$

$$\mathrm{Var}[Z|t,x] = \sigma^2\, \mathsf{E}[Z|t,x]^2, \tag{14}$$

where $I(t,x) = \int_0^x \mu_s^I(u+t-x,u)du$*.*

Proposition 2.2 characterizes how the frailty composition of a given birth-cohort changes over time. At early ages where the integrated intensity $I(t,x)$ is small the selection is modest and the conditional mean and variance are close to the unconditional values of 1 and σ^2. As the intensity increases so does $I(t,x)$ and the conditional mean and variance decrease towards 0. Thus, over time the frailty distribution gets more and more concentrated around 0.

The following proposition shows that the conditional mean frailty can also be expressed in terms of the senescent population mortality.

Proposition 2.3. *Under the assumptions of Proposition 2.2*

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

where $H(t,x) = \int_0^x \mu_s(u+t-x,u)du$*.*

As an immediate consequence of Proposition 2.3 we have the following inversion formula,

$$\mu_s^I(t,x) = \mu_s(t,x)\exp\left(\sigma^2 H(t,x)\right), \tag{16}$$

which allows us to recover the individual intensities from the population intensity and the level of heterogeneity, σ^2. The existence of such a formula implies that any population mortality surface can be described by a frailty model with a given level of heterogeneity.

When discussing improvement rates it is most illuminating to focus on the senescent part of the mortality. The background mortality component is primarily included for the purpose of improving the fit among young adults and, relative to the senescent part, its contribution to mortality is negligible for older age groups. Following the notation of Bongaarts (2005) we define the rate of improvement in senescent mortality as

$$\begin{aligned} \rho_s(t,x) &\triangleq -\frac{\partial \log \mu_s(t,x)}{\partial t} \\ &= -\frac{\partial \log \mathsf{E}[Z|t,x]}{\partial t} - \frac{\partial \log \mu_s^I(t,x)}{\partial t}. \end{aligned} \tag{17}$$

Generally, healthier conditions and other improvements in individual survival will mean that the contribution from the last term is positive. However, higher survival rates imply less selection and the average frailty of persons of age x will therefore increase to 1, the average frailty at birth, over time. Thus the contribution from the first term is negative. For old age groups with strong selection the changing frailty composition can substantially offset the general improvements but eventually the effect dies out and improvements occur.

To capture the idea that the mortality of an individual is affected by both accumulating and non-accumulating factors we will write the (baseline) individual intensity in the form

$$\mu_s^I(t,x) = \kappa(t,x) \exp\left(\int_0^x g(u+t-x,u)du \right). \tag{18}$$

We think of κ as representing the current level of treatment/health at time t for persons of age x, while the accumulating factor g represents the aging process. The idea is that $g(t,x)$ represents the increase in (log) mortality caused by aging at time t for persons of age x. Hence, to get the accumulated effect of aging one needs to integrate from age 0 at the time of birth, $t-x$, to the current age x at time t.

2.2.3. *Specification.* We will consider the following specialization of the model given by (11) and (18):

$$g(t,x) = g_1 + g_2(t-t_0) + g_3(x-x_0), \tag{19}$$

$$\kappa(t) = \exp(\kappa_1 + \kappa_2(t-t_0)), \tag{20}$$

$$\gamma(t) = \exp(\gamma_1 + \gamma_2(t-t_0)), \tag{21}$$

with $x_0 = 60$ and $t_0 = 2000$. The substraction of (year) 2000 in the specification of g, κ and γ and 60 in g is done for interpretability reasons only. Thus $g(2000, 60) = g_1$ is the "aging" of a 60-year old in year 2000, g_2 is

the additional aging across ages for each calender year, and g_3 is the additional biological aging for each year of age. Similarly, $\kappa(2000) = \kappa_1$ and $\gamma(2000) = \gamma_1$ while κ_2 and γ_2 give the annual rates of change. Notice, that κ depends only on time since the obvious "missing" term, $\kappa_3(x-x_0)$, is already present through g_1. The model has a total of 8 parameters; the 7 parameters appearing in the specification of g, κ and γ together with the variance of the frailty distribution, σ^2. As we will later demonstrate the model is able to capture the main characteristics of the data very well despite its parsimonious structure.

From a computational perspective it is convenient to think of the intensities as functions of birth year, rather than calender year, and age. By use of Propositions 2.1 and 2.2 we can write μ as

$$\mu(t,x) = \frac{K(t-x,x)}{1+\sigma^2\int_0^x K(t-x,y)dy} + \gamma(t), \tag{22}$$

where $K(b,x) = \mu_s^I(b+x,x)$. This representation highlights the fact that the integral in the denominator relates to a given birth-cohort.

For the model above we have

$$K(b,x) = \kappa(b)\exp\left((g(b,0)+\kappa_2)x + (g_2+g_3)x^2/2\right). \tag{23}$$

That is, $K(b,x)$ is log-quadratic in x (for fixed b). When $g_2+g_3<0$ the integral in the denominator can be expressed in terms of the cdf of a normal distribution, while this is not possible when the sum is positive. In either case, it is easy to evaluate the integral numerically.

The model allows us to derive the current and asymptotic improvement patterns in age-specific death rates.

Proposition 2.4. *Assume $\kappa_2 < 0$. If $\kappa_2 + g_2 x < 0$ then*

$$\rho_s(t,x) = -\frac{\partial \log \mathsf{E}[Z|t,x]}{\partial t} - (\kappa_2+g_2x) \to -(\kappa_2+g_2x) \text{ for } t\to\infty. \tag{24}$$

The conditions of the proposition imply that all age groups up to age x experience improvements. Note, however, that g_2 may be either positive or negative. Thus the model allows for (asymptotic) improvement rates in senescent mortality to be either increasing or decreasing with age. The presence of the first term means that improvement rates can be substantially lower for a long time before eventually approaching their long-run level. In extreme cases the first term may even dominate the second term, representing general improvements, in which case ASDR's will increase for a period before starting to decrease. Some people have indeed argued that this may happen in old-age mortality. However, we do not find support for increasing ASDR's in the data analyzed in this paper.

The model can be viewed as a generalization of the Gamma-Makeham model of Section 2.2.1. Indeed, that model is obtained by letting all of g, κ and γ be constant (if only g is constant we obtain the model proposed by

Vaupel (1999)). However, unlike the Gamma-Makeham model the cohort mortality profiles of our model will generally not have finite asymptotes.

Proposition 2.5. *Assume $\sigma^2 > 0$. The limit cohort mortality is given by*

$$\lim_{x\to\infty} \mu_s(b+x,x) = \begin{cases} \infty & \text{if } g_2+g_3>0, \\ \frac{g(b,0)+\kappa_2}{\sigma^2} & \text{if } g_2+g_3=0 \text{ and } g(b,0)+\kappa_2>0, \\ 0 & \text{else.} \end{cases}$$

The cohort mortality profile is the result of two opposite effects: the increase in individual mortality pushes the cohort mortality upwards, while the selection mechanism pushes it downwards. For Γ-distributed frailties and exponential individual intensities the two effects balance each other in such a way that an old-age mortality plateau occurs. This is the case in the Gamma-Makeham model and in the second case of Proposition 2.5. However, when individual intensities increase faster than exponential the individual effect dominates and the cohort mortality converges to infinity (although at a slower pace than the individual intensities). Conversely, for sub-exponential individual intensities the selection effect dominates and the cohort mortality goes to zero. These two situations correspond to the first and third case, respectively, of Proposition 2.5.[5] In the application later in the paper the estimates of g_2 and g_3 are both positive. Hence we find ourselves in the first case. That individual intensities increase faster than exponential was also found by Yashin and Iachine (1997).

2.2.4. *Estimation.* We next want to estimate the parameters of model (19)–(21) using the reference data. Since the intensity surface is a continuous function of time and age, while data is aggregated over calender years and one year age groups, we define, for integer values of t and x,

$$\bar{\mu}_{\text{ref}}(t,x) = \frac{1}{4}\left(\mu(t,x)+\mu(t,x+1)+\mu(t+1,x)+\mu(t+1,x+1)\right) \tag{25}$$

to represent the average intensity over the square $[t,t+1)\times[x,x+1)$ of the Lexis diagram.[6] We will use the Poisson-model in (2) with $\bar{\mu}_{\text{ref}}(t,x)$ given by (25) to find the MLE, $\hat{\theta}$, of the parameters $\theta = (\sigma, g_1, g_2, g_3, \kappa_1, \kappa_2, \gamma_1, \gamma_2)$. This is achieved by maximizing the log-likelihood function

$$\begin{aligned} l(\theta) &= \sum_{t,x} D_{\text{ref}}(t,x)\log(\bar{\mu}_{\text{ref}}(t,x)E_{\text{ref}}(t,x)) - \log(D_{\text{ref}}(t,x)!) - \bar{\mu}_{\text{ref}}(t,x)E_{\text{ref}}(t,x) \\ &= \sum_{t,x} D_{\text{ref}}(t,x)\log(\bar{\mu}_{\text{ref}}(t,x)) - \bar{\mu}_{\text{ref}}(t,x)E_{\text{ref}}(t,x) + \textit{constant}, \end{aligned}$$

[5]The third case of Proposition 2.5 is an extreme case of sub-exponential growth in which the individual intensities are in fact decreasing with age, at least from some age. However, the result holds for any sub-exponential intensity, e.g. polynomial.

[6]There are other possibilities for defining $\bar{\mu}_{\text{ref}}(t,x)$. For instance, $\bar{\mu}_{\text{ref}}(t,x) = \mu(t+1/2, x+1/2)$, or $\bar{\mu}_{\text{ref}}(t,x) = \int_0^1\int_0^1 \mu(t+s,x+u)dsdu$. If the exposure is uniform over the square one may argue in favor of the latter definition, but it is cumbersome to implement and unlikely to yield any substantial differences.

where the last term does not depend on θ and hence need not be included in the maximization. It is straightforward to implement the log-likelihood function and to maximize it by standard numeric optimization routines. We have used the `optim` method in the freeware statistical computing package R for our application.

Generally, maximum likelihood estimates are (under certain regularity conditions) asymptotically normally distributed with variance matrix given by the inverse Fisher information[7] evaluated at the true parameter. As an estimate of the variance-covariance matrix we will use

$$\widehat{\mathrm{Cov}}(\hat{\theta}) = -D^2_\theta l(\hat{\theta})^{-1}, \tag{26}$$

which can be computed numerically once $\hat{\theta}$ has been obtained. Using the variance estimates and the approximate normality (approximate) 95%-confidence intervals for the parameters can be computed as $\hat{\theta} \pm 1.96\widehat{\mathrm{Var}}(\hat{\theta})$, where $\widehat{\mathrm{Var}}(\hat{\theta})$ denotes the diagonal of $\widehat{\mathrm{Cov}}(\hat{\theta})$.

Bootstrapping constitutes an alternate approach to assessing the parameter uncertainty which does not rely on asymptotic properties, see e.g. Efron and Tibshirani (1993); Koissi *et al.* (2006). In short, the method consists of simulating a number of new data sets, i.e. new death counts, given the observed exposures and the estimated intensities and calculate the MLE for each data set. The resulting (bootstrap) distribution reflects the uncertainty in the parameter estimates. Although simple in theory the computational burden is in our case substantial as each maximization takes several minutes.

2.3. **Spread modelling.** The fundamental assumption behind the proposed method for modelling small population mortality is the existence of an underlying (smooth) mortality surface, the trend, around which the small population mortality evolves. In this section we focus on modelling and estimating the deviations of the small population mortality from the trend.

2.3.1. *Spread.* For given underlying trend, $\bar{\mu}_{\mathrm{ref}}$, we will assume that subpopulation death counts are independent and distributed according to

$$D_{\mathrm{sub}}(t,x) \sim \mathrm{Poisson}\left(\bar{\mu}_{\mathrm{sub}}(t,x)E_{\mathrm{sub}}(t,x)\right), \tag{27}$$

where

$$\bar{\mu}_{\mathrm{sub}}(t,x) = \bar{\mu}_{\mathrm{ref}}(t,x)\exp(y_t' r_x) \tag{28}$$

with $y_t' = (y_{0,t}, \ldots, y_{n,t})$ and $r_x' = (r_{0,x}, \ldots, r_{n,x})$ for some n. The spread between the mortality of the subpopulation and the trend is modelled by the last term in (28). The regressors $r_0, \ldots, r_n$ determine the possible age-profiles of the (log) spread, while $y_0, \ldots, y_n$ describe the evolution over time of the corresponding components of the spread. We will refer to the y's as the spread parameters.

[7]The Fisher information is defined as minus the expected value of the second derivative of the log-likelihood function, $\mathcal{I}(\theta) = -\mathsf{E}[D^2_\theta l(\theta)|\theta]$.

As opposed to the elaborate trend model we have chosen to use a rather simple log-linear parametrization of the spread. We do this for two reasons. First, assuming the trend model captures the main features of the mortality surface we expect there to be only little structure left in the spread. Introducing a complex functional form for the spread therefore seems fruitless. Second, a complicated spread model would to some extent counter the idea of the model. The spread is supposed to model only the random, but potentially time-persistent, fluctuations around the underlying mortality evolution.

Regarding the choice of dimensionality, n, we are faced with the usual trade-off. A high number of regressors can fit the spread evolution very precisely, but there is a risk of overfitting thereby impairing forecasting ability. Also, a high number of spread parameters are harder to model and will, typically, increase forecasting uncertainty. A low number of regressors, on the other hand, will fit the spreads less well and can be expected to capture only the overall shape. However, a low number of spread parameters are easier to model and, generally, provides more robust and less uncertain forecasts.

For a given number of regressors there are essentially two ways to proceed. Either, the regressors are specified directly and only the spread parameters are estimated, or both regressors and spread parameters are estimated simultaneously from the data. We prefer the former method due to ease of interpretability of the spread parameters and presumed better forecasting ability; although we recognize that the latter method provides a better (with-in sample) fit.

Specifically, we propose to parameterize the spread by the three regressors

$$r_{0,x} = 1, \tag{29}$$

$$r_{1,x} = (x - 60)/40, \tag{30}$$

$$r_{2,x} = (x^2 - 120x + 9160/3)/1000, \tag{31}$$

which describe, respectively, the level, slope and curvature of the spread. For ease of interpretability the regressors are chosen orthogonal and they are normalized to (about) unity at ages 20 and 100.[8] The number of regressors reflects a compromise between fit and ease of modelling which appears to work well in our application.

Due to the assumed independence of death counts the MLE of the spread parameters for year t depends only on data for that year. For each year

[8]In the application we use mortality data for ages 20 to 100, i.e. 81 one-year age groups. Seen as vectors the three regressors are orthogonal w.r.t. the usual inner product in $\mathbb{R}^{81}$. The regressors are normalized such that $r_{2,20} = -1$ and $r_{2,100} = 1$, while $r_{3,20} = r_{3,100} = 1.053$. If desired we can obtain $r_{3,20} = r_{3,100} = 1$ by changing the normalization constant from 1000 to 3160/3 in the definition of r_3.

of subpopulation data we obtain the MLE for y_t by maximizing the log-likelihood function

$$l(y_t) = \sum_x D_{\text{sub}}(t,x) y_t' r_x - \bar{\mu}_{\text{ref}}(t,x) \exp(y_t' r_x) E_{\text{sub}}(t,x) + \textit{constant},$$

where $\bar{\mu}_{\text{ref}}(t,x)$ is calculated with the maximum likelihood estimates from Section 2.2.4 inserted. Note that the parametric form of the underlying trend ensures that we can calculate $\bar{\mu}_{\text{ref}}(t,x)$ for all x and t. Thus the age and time windows for which we have data for the sub- and reference population need not coincide, or even overlap. In practice, of course, we expect there to be at least a partial overlap. For example, if the subpopulation is the current and former members of a specific pension scheme, or a specific occupational or ethnic group, we might have only a relatively short history of data, while we might have a considerably longer history of national data which we might want to use as reference data.

2.3.2. *Time dynamics.* The multivariate series of spread parameters describe the evolution in excess mortality in the subpopulation relative to the reference population. Over time we expect the two populations to experience similar improvements and we therefore believe the spread to be stationary rather than showing systematic drift. We also expect the spread to show time-persistence. If at a given point in time the mortality of the subpopulation is substantially higher or lower than the reference mortality we expect it to stay higher or lower for some time thereafter. Finally, we expect the spread parameters to be dependent rather than independent. The regressors are chosen to have a clear interpretation, but we do not expect, e.g. the level and the slope of the spread to develop independently of each other over time.

The simplest model meeting these requirements is the vector autoregressive (VAR) model which we will adopt as spread parameter model. Specifically we suggest to use the Gaussian VAR(1)-model

$$y_t = A y_{t-1} + \epsilon_t, \tag{32}$$

where A is a three by three matrix of autoregression parameters and the errors ϵ_t are three-dimensional i.i.d. normally distributed variates with mean zero and covariance matrix Ω, i.e. $\epsilon_t \sim N_3(0,\Omega)$.

By not including a mean term in the model we implicitly assume that the spread will converge to zero (in expectation) over time. We believe this is a natural condition to impose for the application to Danish data with an international reference data set considered in this paper. Indeed, it is hard to justify the opposite, that Danish mortality should deviate systematically from international levels indefinitely even if historic data were to suggest it. For other applications one may wish to include a mean term in the model and thereby allow for systematic deviations. Similarly, one may wish to consider more general VAR-models with additional lags to capture more complex time-dependence patterns.

The parameters A and Ω of model (32) can be estimated by the `ar` routine in R treating the time series of estimated spread parameters, y_t, as observed variables. The routine offers various estimation methods. It would have been in the spirit of this paper to use maximum likelihood estimation, but unfortunately this option is only implemented for univariate time series. Instead we use Yule-Walker estimation which obtains estimates by solving the Yule-Walker equations, cf. e.g. Brockwell and Davis (1991). In our application the estimated A defines a stationary time series, i.e. the modulus of A's eigenvalues is smaller than one, for both men and women. However, in general there is no guarantee for this. As with all statistical analysis where data contradicts modelling assumptions one will then have to propose a more suitable model, e.g. introduce a mean term or additional lags, and reiterate the analysis.

2.3.3. *Forecast.* Forecasting in the VAR-model (32) is based on the conditional distribution of the future values of the spread given the observed values. Assuming year T to be the last year of observation and h to be the forecasting horizon we need to find the conditional distribution of y_{T+h} given the observed values. Due to the Markov property of the VAR(1)-model this distribution depends on the last observed value, y_T, only. Expanding the data generating equation we obtain for $h \geq 1$

$$y_{T+h}|y_T \sim N(m_h, V_h),$$

where m_h and V_h are given by

$$m_h = A^h y_T, \quad V_h = \sum_{i=0}^{h-1} A^i \Omega (A^i)'.$$

From these expressions forecasted values for future spread parameters and corresponding two sided (pointwise) 95%-confidence intervals are easily obtained as

$$\mathrm{CI}_{95\%}(y_{T+h}) = m_h \pm 1.96\sqrt{\mathrm{diag}(V_h)}.$$

Note that due to stationarity the forecasting uncertainty will increase towards a finite limit as the forecasting horizon increases. Thus the deviations of the subpopulation from the reference population are bound (in probability) over time. By use of equation (28) we further have

$$\mathrm{CI}_{95\%}\left(\log \bar{\mu}_{\mathrm{sub}}(T+h,x)\right) = \log \bar{\mu}_{\mathrm{ref}}(T+h,x) + m_h' r_x \pm 1.96\sqrt{r_x' V_h r_x}.$$

The stated confidence intervals only reflect the stochasticity of the VAR-model itself without taking the parameter uncertainty into account. They are therefore, in a sense, the "narrowest" possible confidence intervals. Confidence intervals incorporating parameter uncertainty of both the trend and the spread model can be constructed by bootstrap, but we will not pursue that point further. Cairns (2000) also considers model uncertainty, i.e. the

uncertainty associated with determining the underlying model, and he discusses how all three types of uncertainty can be assessed coherently in a Bayesian framework.

We have concentrated on assessing the uncertainty of a single ASDR at a future point in time. Since the conditional distribution of the entire future $\{y_{T+1}, \ldots\}$ given y_T is readily available we can also derive simultaneous confidence intervals for any collection of ASDR's by the same method. In principle, it is therefore possible to derive analytic confidence intervals for any functional of the intensity surface. In practice, however, most quantities of interest, e.g. remaining life expectancy, are too complicated to allow analytic derivations. Instead it is necessary to resort to Monte Carlo methods to assess forecasting uncertainty of any but the simplest quantities. Fortunately, this is straightforward to implement. We simply simulate a large number of realizations from the VAR-model (32) and calculate the corresponding intensity surface by (28). For each surface we calculate the quantity of interest and we thereby obtain (samples from) the forecasting distribution. This will be illustrated in Section 3.4.

3. Application

To demonstrate the model in action we consider the case that gave rise to the name SAINT, namely Denmark as the (sub)population of interest and a basket of developed countries as the reference population. The model is applied to each sex separately.

3.1. **Data.** Data for this study originates from the Human Mortality Database,[9] which offers free access to updated records on death counts and exposure data for a long list of countries. The database is maintained by University of California, Berkeley, United States and Max Planck Institute for Demographics Research, Germany.

We will use both Danish data and a pooled international data set consisting of data for the following 19 countries: USA, Japan, West Germany, UK, France, Italy, Spain, Australia, Canada, Holland, Portugal, Austria, Belgium, Switzerland, Sweden, Norway, Denmark, Finland and Iceland. This set is chosen among the 34 countries represented in the Human Mortality Database because of their similarity to Denmark with respect to past and presumed future mortality. Table 6 in Appendix B contains a summary of the data.

The subsequent analysis uses data from the years 1933 to 2005 and ages 20 to 100. As far as the time dimension is concerned the cut points are determined by the availability of US data. Concerning the age span the analysis could in principle be based on all ages from 0 to 110, which are all available in the Human Mortality Database. However, since the prime focus is adult mortality and since the mortality pattern at young ages differs

[9] See www.mortality.org

markedly from adult mortality all ages below 20 have been excluded. For very high ages the quality of data is poor and sometimes based on disaggregated quantities and for this reason all ages above 100 have also been excluded.[10]

The international data set is constructed as the aggregate of the 19 national data sets. For each year from 1933 to 2005 and each age from 20 to 100 the international death count and international exposure consists of, respectively, the total death count and total exposure of those of the 19 countries for which data exists for that year. Measured in terms of death counts and exposures the international data set is more than 100 times larger than the Danish data set.

3.2. **Trend.** To illustrate the data we have plotted Danish and international female death rates for ages 40, 60 and 80 years in Figure 1. Compared to Denmark the international development in death rates has been quite stable with only slowly changing annual rates of improvement. The Danish mortality evolution, on the other hand, shows a much more erratic behavior with large year-to-year variation in improvement rates. The Danish level seems to follow that of the international community in the long run, but there are extended periods with substantial deviations. The most striking of these is the excess mortality of Danish females around age 60 which emerged in 1980, peaked more than a decade later and is still present today although less pronounced. From Figure 1 and similar plots it seems reasonable to think of Danish mortality rates as fluctuating around a stable international trend.

We have used the international data set to estimate the trend model described in Section 2.2.3. Table 1 contains maximum likelihood estimates of the eight parameters and corresponding two sided 95% confidence intervals. The narrow width of the confidence intervals reflects the fact that, relative to the amount of data, we have a very parsimonious model. Using a similar model Barbi (2003) reports standard errors of the same magnitude in an application to Italian data. The small standard errors indicate that the parameters are well determined, but this does not necessarily imply a good fit. The fit of the model can be assessed graphically on Figure 2 which shows international female mortality rates together with the estimated trend. Overall, it appears that the model does a remarkably good job at describing the data. There are appreciable deviations only for the very youngest and very highest ages. For now we settle with this informal graphical inspection of goodness-of-fit, but we will return to the issue more formally in Section 4.

The model has three parameters to describe three different types of improvement in mortality over time. The parameters g_2 and κ_2 affect the improvement in senescent mortality, while the reduction in background mortality is determined by γ_2. By Proposition 2.4 we know that the limiting rate of improvement in senescent mortality is given by $-(\kappa_2 + g_2 x)$ for ages x for

[10] In some countries and some years data for ages younger than 100 years is also based on disaggregated quantities, but we suspect this to be of minor importance.

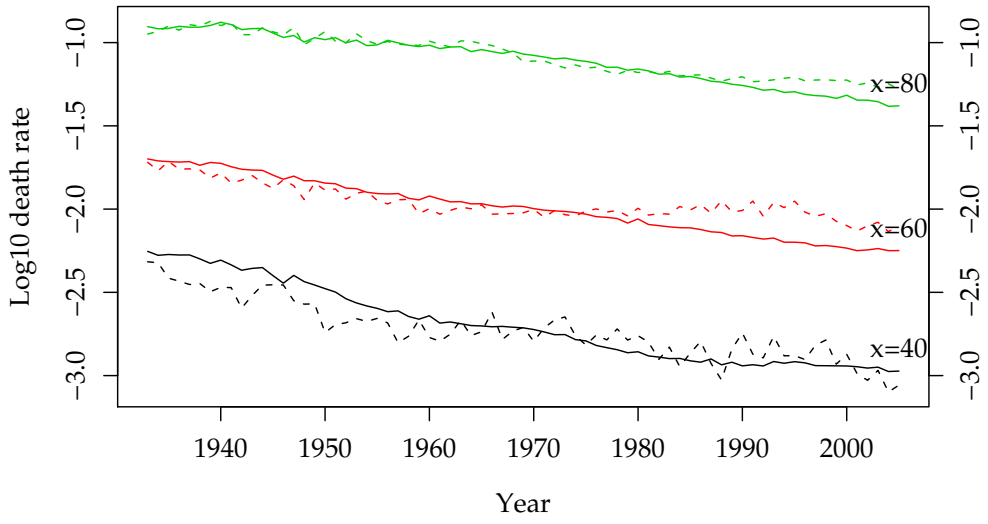

FIGURE 1. Danish (dashed line) and international (solid line) development in female log death rates from 1933 to 2005 for selected ages.

Parameter	**Women**		**Men**	
	Estimate	95%-CI	Estimate	95%-CI
σ	$4.2860 \cdot 10^{-1}$	$\pm 1.6 \cdot 10^{-4}$	$2.6243 \cdot 10^{-1}$	$\pm 2.3 \cdot 10^{-4}$
g_1	$9.8965 \cdot 10^{-2}$	$\pm 2.3 \cdot 10^{-6}$	$1.0551 \cdot 10^{-1}$	$\pm 2.2 \cdot 10^{-6}$
g_2	$4.7856 \cdot 10^{-6}$	$\pm 5.1 \cdot 10^{-8}$	$8.3744 \cdot 10^{-5}$	$\pm 3.7 \cdot 10^{-8}$
g_3	$1.3103 \cdot 10^{-3}$	$\pm 1.0 \cdot 10^{-7}$	$5.5903 \cdot 10^{-5}$	$\pm 8.6 \cdot 10^{-8}$
κ_1	$-8.7819 \cdot 10^{0}$	$\pm 1.7 \cdot 10^{-4}$	$-1.0576 \cdot 10^{1}$	$\pm 1.6 \cdot 10^{-4}$
κ_2	$-1.8510 \cdot 10^{-2}$	$\pm 5.8 \cdot 10^{-6}$	$-1.7827 \cdot 10^{-2}$	$\pm 5.0 \cdot 10^{-6}$
γ_1	$-1.1810 \cdot 10^{1}$	$\pm 1.9 \cdot 10^{-3}$	$-7.5222 \cdot 10^{0}$	$\pm 8.4 \cdot 10^{-4}$
γ_2	$-8.9038 \cdot 10^{-2}$	$\pm 3.2 \cdot 10^{-5}$	$-2.5005 \cdot 10^{-2}$	$\pm 2.0 \cdot 10^{-5}$

TABLE 1. Maximum likelihood estimates and 95% confidence intervals for the trend model given in Section 3.2. The estimation is based on international mortality data from 1933 to 2005 for ages 20 to 100 years.

which this quantity is positive.[11] Since g_2 is positive for both sexes the (limiting) rate of improvement is decreasing in age as expected. Due to frailty

[11]With the estimated parameter values this is satisfied up to age 213 for men and 3870 years for women, i.e. for all ages of practical relevance.

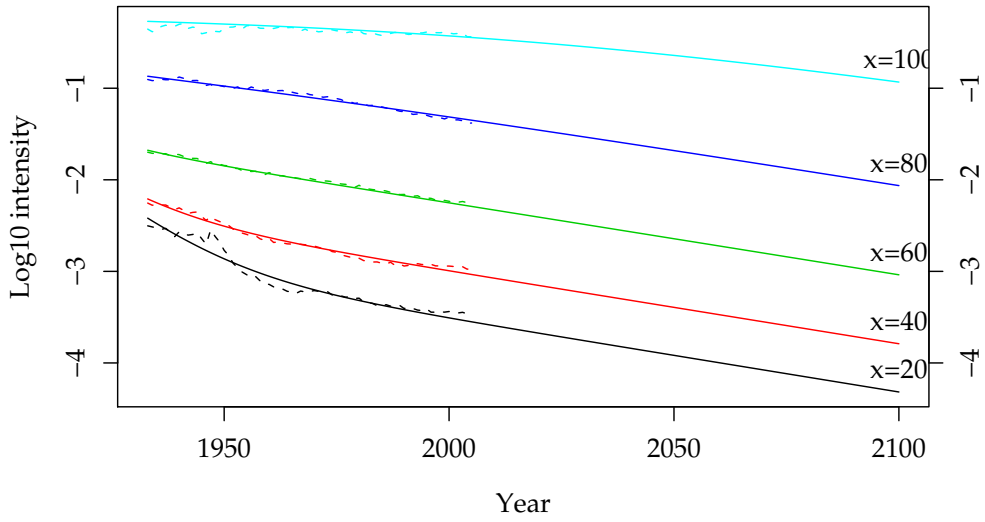

FIGURE 2. Historic development (dashed line) in international female log mortality from 1933 to 2005 for the age groups $20, 40, \ldots, 100$. Model estimate of trend with parameters given in Table 1 is superimposed (solid line).

we also have that the limiting improvement rate is achieved more slowly for higher ages than for lower ages. This further "steepens" the age-profile of improvement rates and causes it to change shape over time as illustrated in Figure 3. Note in particular how the improvement rate for 100-year-olds is projected to double over the next century. The projected increase in improvement rates can also be observed from the curved projection in Figure 2 for this age group (the effect is also present for the younger age groups but much less pronounced).

The value of κ_2 is estimated to about -1.8% for both women and men. The value of g_2, however, is almost 20 times larger for men than for women. This implies that whereas the limiting age-profile of improvement rates is almost flat for women the age-profile will remain steep for men. Thus 100 year-olds females will eventually experience the same annual rate of improvement as the younger age groups, while old men will continue to have lower improvement rates than younger men.

The improvement rate of background mortality is estimated to about 9% for women and about 2.5% for men. The large value for women is due to the dramatic decrease in mortality among the 20 to 30 year-olds in the beginning of the observation period, cf. Figure 2. However, since female background

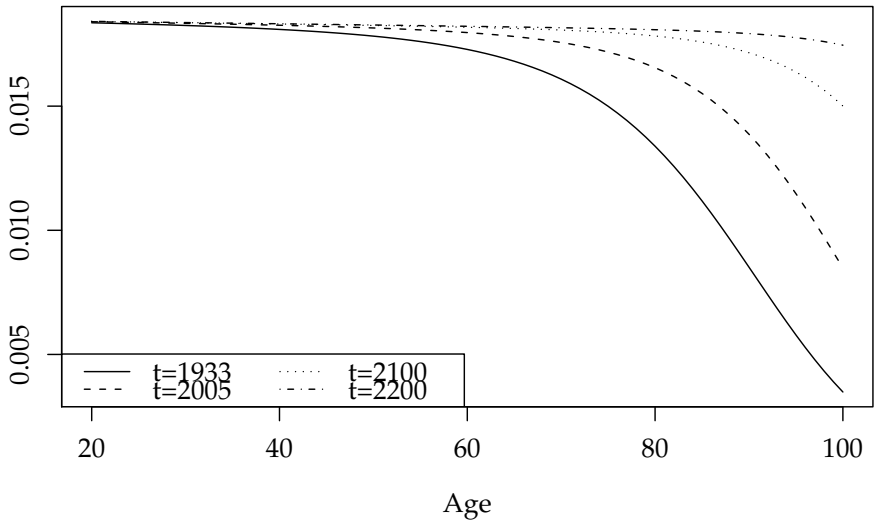

FIGURE 3. Rate of improvement, $\rho_s(t,x)$, in female senescent mortality. Based on parameter values from Table 1.

mortality is now negligible compared to senescent mortality further reductions will have virtually no effect on (total) mortality even at young ages. For men, on the other hand, background mortality is much higher (as can be seen by the difference in the values of γ_1) and future reductions will have a substantial effect on young age mortality.

The estimated level of heterogeneity, σ, is higher for women than for men. Since the σ parameter controls the "delay" in achieving the asymptotic rate of improvement, i.e. the first term in (24) of Proposition 2.4, this implies that the asymptotic rate is approached slower for women than for men. It is not clear why this is the case, but a gender difference of the same magnitude was found by Barbi (2003).

The remaining parameters control the age profile of (senescent) mortality. As expected the parameter for the overall level, κ_1, and the parameter for first-order age dependence, g_1, are both estimated to be smaller for women than for men. Only the second-order age dependence parameter, g_3, is higher, albeit still small, for women than for men. The parameter is positive for both sexes implying that aging is accelerating with age. Thus for all ages of practical relevance female mortality is estimated to be lower than male mortality, but for very advanced ages female mortality will in fact exceed male mortality.

3.3. **Spread.** We will apply the three factor spread model of Section 2.3 to describe the Danish fluctuations around the international level. The estimated spread series for women are shown in Figure 4, where the excess mortality of Danish women from around 1980 onwards is clearly visible. Note that simultaneously with the increase of the level the curvature has decreased. This means that only women around age 60 experience excess mortality while the mortality of very young and very old Danish women is similar to the international level.

In 2005 the estimated spread parameters were $(0.17, 0.06, -0.16)$ for women and $(-0.01, 0.15, 0.08)$ for men. This in fact implies an excess mortality of more than 25% for Danish women of age 60, but only 6% at age 100. Danish men, on the hand, are very much in line with the international level having an excess mortality of 3% at age 60.

The parameter estimates of the VAR(1)-model, which describes the dynamics of the spread series, are

$$A = \begin{pmatrix} 0.6861 & -0.1907 & -0.2739 \\ -0.1423 & 0.8724 & -0.1558 \\ -0.2422 & -0.1035 & 0.5179 \end{pmatrix}, \ \Omega = 10^{-3} \begin{pmatrix} 2.0449 & -0.7341 & 0.3012 \\ -0.7341 & 2.9779 & -0.7376 \\ 0.3012 & -0.7376 & 1.3278 \end{pmatrix},$$

for women and

$$A = \begin{pmatrix} 0.7885 & -0.1714 & -0.1283 \\ -0.2485 & 0.6387 & 0.0477 \\ -0.0650 & -0.0792 & 0.9130 \end{pmatrix}, \ \Omega = 10^{-3} \begin{pmatrix} 1.4840 & -0.3818 & 0.3646 \\ -0.3818 & 3.0056 & -1.1880 \\ 0.3646 & -1.1880 & 3.4158 \end{pmatrix},$$

for men. In both cases the A matrix give rise to stationary series. Note that diagonal and off-diagonal elements of A are of the same magnitude because of the high interdependence between the three spread components. Also the errors are highly correlated.

Figure 4 shows the mean forecast for the spread parameters for women with 95% confidence intervals. Due to stationarity all three spread components are forecasted to converge to zero. However, due to the structure of A the convergence is not necessarily monotone. The slope, for instance, starts out positive but is forecasted to become negative before approaching zero.

The width of the confidence intervals reflects the observed variation in the spread over the estimation period. The confidence intervals expand quite rapidly to their stationary values indicating that substantial deviations can build up or disappear in a matter of decades. The confidence intervals do not include parameter uncertainty, but only the uncertainty induced by the error term of the VAR-model. Incorporating parameter, or indeed model, uncertainty will most likely lead to even wider confidence intervals.

3.4. **Forecasting life expectancy.** Life expectancies are an intuitively appealing way to summarize a mortality surface. Letting $\bar{\mu}$ denote an intensity surface which is constant over Lexis squares, cf. Section 2.1.1, and letting t and x be integers we get the following approximation for the *cohort* mean

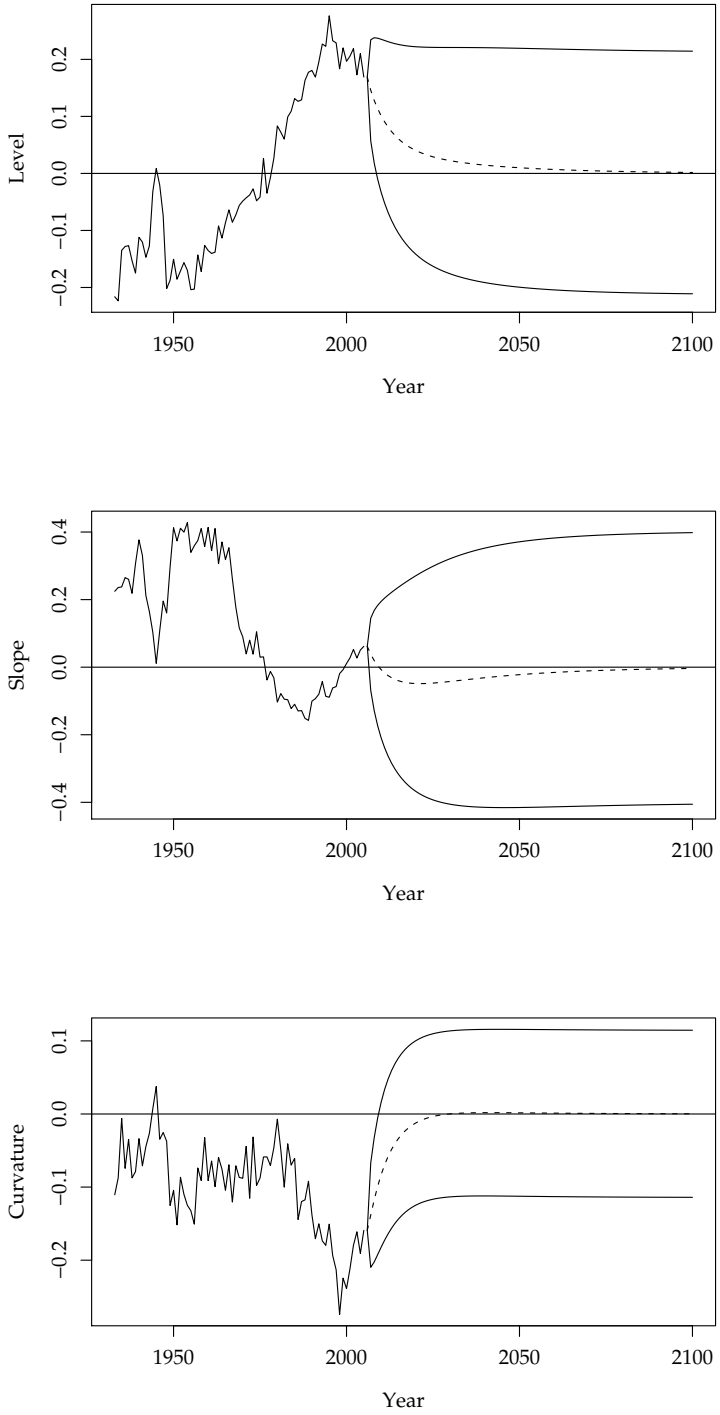

FIGURE 4. Estimated and forecasted spread parameters for Danish women with two sided pointwise 95% confidence intervals.

remaining life time of individuals born at $t-x$ (at time t)

$$\begin{aligned} e^{\mathrm{c}}(t,x) &= \mathsf{E}\left[T-x \middle| T \geq x\right] \\ &= \int_0^\infty \exp\left(-\int_0^s \bar{\mu}(t+v,x+v)dv\right) ds \\ &\approx \sum_{j=0}^{M} \exp\left(-\sum_{i=0}^{j-1} \bar{\mu}(t+i,x+i)\right) \frac{1-\exp\left(-\bar{\mu}(t+j,x+j)\right)}{\bar{\mu}(t+j,x+j)} \end{aligned} \tag{33}$$

for some large M. Similarly, the *period* mean remaining life time is

$$e^{\mathrm{p}}(t,x) \approx \sum_{j=0}^{M} \exp\left(-\sum_{i=0}^{j-1} \bar{\mu}(t,x+i)\right) \frac{1-\exp\left(-\bar{\mu}(t,x+j)\right)}{\bar{\mu}(t,x+j)}. \tag{34}$$

The cohort life expectancy is calculated from the ASDR's of a specific cohort, i.e. along a diagonal of the Lexis diagram, while the period life expectancy is calculated from the ASDR's at a given point in time, i.e. along a vertical line of the Lexis diagram. The cohort life expectancy represents the actual life expectancy of a cohort taking the future evolution of ASDR's into account. The period life expectancy, on the other hand, is the life expectancy assuming no future changes in ASDR's. For this reason the cohort life expectancy is (substantially) higher than the corresponding period life expectancy.

In Table 2 we have shown selected cohort and period life expectancies for women based on point estimates of the intensities (obtained from the mean forecast of the spread). Period life expectancies based on observed death rates for 2005 are also shown. We note that they correspond very well with the model estimates indicating a good fit of the model in the jump-off year.

The table contains period life expectancy forecasts up to year 2105 while cohort life expectancies are forecasted only up to year 2025. In principle, we can calculate cohort life expectancies for 2105 also. However, assuming a maximal age of 120 years this would require that we project ASDR's to year 2205. No matter how good a model, we cannot give credence to quantities based on projections 200 years into the future and we have therefore chosen to omit the numbers.

The current excess mortality for Danish women can be seen as lower period life expectancies in 2005. The Danish cohort life expectancies are also lower than the international levels but the differences are smaller due to future convergence of Danish rates to the international trend. After 20 years the differences between Danish and international life expectancies have virtually disappeared.

The forecasting uncertainty of complicated functionals such as life expectancy can be assessed by Monte Carlo methods as described in Section 2.3.3. As an illustration of this approach we show in Figure 5 the forecasting distribution for the cohort life expectancy of a 60-year-old Danish women in 2005, $e^{\mathrm{c}}(2005,60)$, based on 100,000 simulations. The empirical mean is 27.06 which is very close to the point estimate of 27.05 in Table 2.

	International				Denmark			
Year	Age				Age			
	20	60	70	80	20	60	70	80
2005	71.67	27.37	17.88	10.17	71.67	27.05	17.41	9.66
2025	74.96	30.27	20.36	11.98	74.96	30.26	20.34	11.96

	International				Denmark			
Year	Age				Age			
	20	60	70	80	20	60	70	80
2005	62.92	24.85	16.54	9.64	60.89	23.10	15.11	8.68
	(62.84)	(25.12)	(16.83)	(9.70)	(60.89)	(23.11)	(15.19)	(8.84)
2025	65.98	27.43	18.75	11.28	65.91	27.36	18.69	11.24
2045	68.94	30.02	21.04	13.08	68.94	30.02	21.04	13.08
2105	77.27	37.70	28.15	19.13	77.27	37.70	28.15	19.13

TABLE 2. Upper panel: Cohort remaining life expectancy in years for women. The numbers are based on model forecasts and calculated using (33) with $M = 120$. Lower panel: Period remaining life expectancy in years for women. The numbers are based on model forecasts and calculated using (34) with $M = 120$. The period remaining life expectancy based on observed death rates for year 2005 with $M = 110$ is shown in brackets.

Note that this need not necessarily be true in general for non-linear functionals. Confidence intervals can be obtained using either the empirical standard deviation of 0.32 and a normal approximation, or the percentiles of interest can be calculated directly from the sample.

4. GOODNESS-OF-FIT

Evaluation of a statistical model's performance is of uttermost importance, and hence we shall devote this section to investigating the fit of the SAINT model. To this end we evaluate the model's fit within-sample as well as out-of-sample. Having applications in mind we shall emphasize the latter. Our benchmark is the widespread "Poisson version" of the Lee-Carter (LC) model, cf. Brouhns *et al.* (2002). This model assumes that

$$D(t,x) \sim \text{Poisson}\left(\bar{\mu}(t,x)E(t,x)\right) \text{ with } \bar{\mu}(t,x) = \exp(\alpha_x + \beta_x\kappa_t), \tag{35}$$

where the parameters are subject to the constraints $\sum_t \kappa_t = 0$ and $\sum_x \beta_x = 1$ to ensure identifiability.

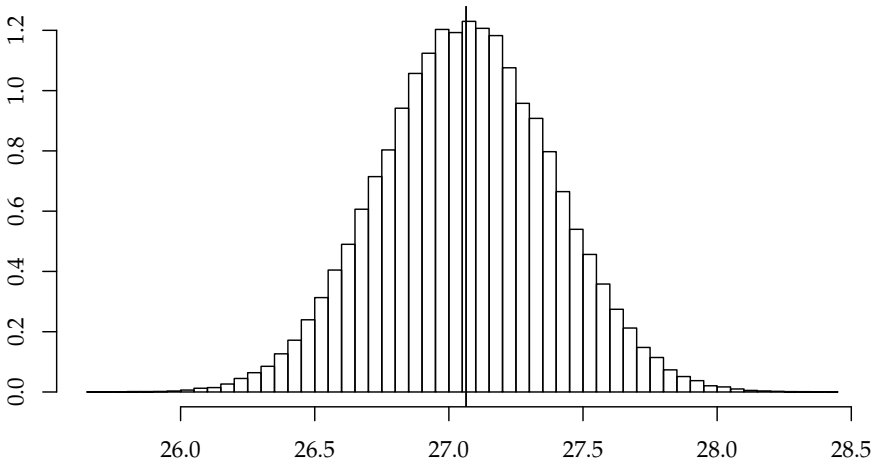

FIGURE 5. Forecasting distribution of the cohort remaining life expectancy $e^{c}(2005, 60)$ for Danish women. The mean and median (vertical line) are both 27.06, and the standard deviation is 0.32. Based on 100,000 simulations.

To assess the performance we measure the cellwise errors in death counts, i.e. the deviation from the expectation in the hypothesized Poisson distribution. We then sum these, their absolute values or their squares. For comparison and interpretation we normalize by the total number of deaths in the two former cases. Thus we consider

$$G_1 = \sum_{t,x} \left(D(t,x) - \bar{\mu}(t,x)E(t,x)\right) \Big/ \sum_{t,x} D(t,x)$$

$$= 1 - \sum_{t,x} \bar{\mu}(t,x)E(t,x) \Big/ \sum_{t,x} D(t,x)\,, \tag{36}$$

$$G_2 = \sum_{t,x} \left|D(t,x) - \bar{\mu}(t,x)E(t,x)\right| \Big/ \sum_{t,x} D(t,x)\,, \tag{37}$$

$$G_3 = \sum_{t,x} \left(D(t,x) - \bar{\mu}(t,x)E(t,x)\right)^2, \tag{38}$$

where $\bar{\mu}$ is either fitted or forecasted. The forecasted values for the SAINT model (LC model) are based on the mean forecast of the spread (κ-index).

Notice that G_1 is the weighted (by actual deaths) average of $1-\bar{\mu}(t,x)/m(t,x) \approx \log\left(m(t,x)/\bar{\mu}(t,x)\right)$. The weighted average of the latter therefore being a comparable measure of fit. In comparing some versions of the LC method

	Women		Men	
Estimation period	SAINT	Lee-Carter	SAINT	Lee-Carter
1933-1950	5.08	5.38	5.30	5.91
1933-1970	5.02	5.55	5.39	6.09
1933-1990	6.00	6.55	5.00	5.78
1933-2005	6.25	6.31	4.89	5.96

TABLE 3. Within-sample error measured by G_2 (as percentages). Danish data.

Booth *et al.* (2006) calculated the *unweighed* averages of $\log(m(t,x)/\bar{\mu}(t,x))$ and its absolute value. Modulo a log approximation this corresponds to weighing by $(\bar{\mu}(t,x)E(t,x))^{-1}$ in (36), which seems unreasonable considering the likelihood. Apart from these suggestions, most mortality models we have encountered base their evaluation of fit on graphical inspection.

G_2 and G_3 evaluate the fit in two different ways and we shall use both measures. As a supplement we shall also use G_1 to measure overall bias. Alternatively, one could evaluate how well key figures such as annuity values and remaining life expectancy match, but we do not pursue that here.

4.1. **Within-sample performance.** Within-sample both models perform well with an absolute relative error of about 6%. Table 3 reveals that there is not much to choose between the two, but it is encouraging that the more sparsely parameterized SAINT model fits at least as good over any subperiod considered. We choose to use G_2 because it is comparable across sexes and periods, but G_3 gives the same conclusion.

Recently, Dowd *et al.* (2008b) came up with some statistically testable suggestions to measure goodness-of-fit for mortality models, which we shall not go into—partly because we are somewhat skeptical of the underlying independence and dispersion assumptions.

4.2. **Out-of-sample performance.** The papers we are aware of are to a large extent silent regarding the out-of-sample performance of their respective models. Two exceptions are Booth *et al.* (2006) as mentioned above, and Dowd *et al.* (2008a), whose ideas are appealing. For sake of brevity we shall not perform any of their four suggested inspections in depth here.

In essence we believe that a good mortality forecast must fulfill two equally important criteria. First, the model should provide accurate forecasts over short and long horizons. And secondly, using different input data the mortality forecasts ought to be as little sensitive towards the choice of estimation period as possible, i.e. robust.

4.2.1. *Accuracy.* To evaluate the accuracy we will calculate G_3 for forecast horizons ranging from 10 to 35 years for two different, but overlapping, estimation periods. The results are displayed in Table 4 from which we conclude

	Forecast period length					
Estimation period	10		15		20	
	SAINT	LC	SAINT	LC	SAINT	LC
1933-1950	0.69	0.43	0.93	0.69	1.34	1.31
1933-1970	1.48	2.32	2.39	4.27	3.50	6.96
	Forecast period length					
Estimation period	25		30		35	
	SAINT	LC	SAINT	LC	SAINT	LC
1933-1950	2.43	2.81	4.26	5.34	6.44	8.58
1933-1970	4.82	9.70	6.22	13.0	7.27	15.7

TABLE 4. Out-of-sample error for different estimation periods and different forecast periods measured by G_3 (normalized by 10^6). Danish women.

	Forecast period length					
Estimation period	10		15		20	
	SAINT	LC	SAINT	LC	SAINT	LC
1933-1950	-7.97	-1.64	-6.84	-0.62	-6.36	-0.33
1933-1970	-4.86	-3.28	-3.64	-2.35	-2.10	-1.21
	Forecast period length					
Estimation period	25		30		35	
	SAINT	LC	SAINT	LC	SAINT	LC
1933-1950	-6.82	-1.04	-6.97	-1.51	-6.49	-1.39
1933-1970	+0.33	+0.80	+1.83	+1.93	+2.78	+2.53

TABLE 5. Out-of-sample error for different estimation periods and different forecast periods measured by G_1 (as percentages). Danish women.

(albeit based on limited evidence and overlapping estimation periods) that the LC method predicts slightly more accurately over short forecast periods, whereas on long horizons the SAINT model's performance is superior. Further analysis has indicated that the tipping point lies between about five and 15 years' forecast.

At the cost of potentially even worse long run forecasts it has been suggested to improve the short run accuracy by calibrating the Lee-Carter model to the latest observed death rates. This would likely reinforce the difference between the two models.

Very short term forecasts (not shown) are quite accurate in both cases—because of short term smoothness of data and the relatively dense parametrization. All conclusions above apply to men as well.

For the same estimation and forecast periods we have calculated G_1 as a measure of bias. The results are shown in Table 5. Note that negative values imply upward bias of death rates, i.e. projected death rates are higher than realized death rates. For the short estimation period the SAINT model has a higher bias than LC, while the bias of the two models is essentially the same for the long estimation period.

Examining the contributions of G_1 more thoroughly (numbers not shown) reveals that there is a seemingly systematical pattern in the errors of the SAINT model with most of the upward bias concentrated at ages below 40. Death rates for ages 40–60 are in fact slightly downward biased, while death rates for ages above 60 are essentially unbiased. Due to its structure the LC model suffers no such systematic age bias.

Dowd *et al.* (2008a) point out that most mortality forecasts are upward biased. We find the same, but of course this is not an intrinsic feature of the models.

4.2.2. *Robustness.* We will check robustness in two ways—by examining the stability of forecasts towards the inclusion of additional years in each end of the data window. This serves two distinct purposes. Adding extra years in the "left" side of the interval we examine the sensitivity towards the otherwise arbitrary choice of left end point of the input data. On the other hand adding years in the "right" side allows analysis of the desired feature that forecasts do not change substantially when the model is calibrated using new data. Of course these two tests are closely linked. For the sake of brevity we will provide graphical indications only but obviously a version of the G_3 measure, or something similar, should be considered as well to assess how close forecasts based on different data are.

For the former analysis see Figure 6. This graphical inspection clearly indicates that the SAINT model is more "backward robust". The particular evidence is based on two scenarios only, but in fact the conclusion applies to other ages and periods and to Danish men as well.

Finally, we consider the stability towards including new data. This is essentially no different from the preceding analysis, and the conclusion is repeated from above. Figure 7 compares mortality intensity forecasts at two key ages and suggests that the SAINT model is slightly more "forward robust". This conclusion is also representative across sexes, ages, and estimation periods.

We do not believe in the existence of an intrinsically optimal length for the sample period. Hence, we do not investigate this. Instead we have faith in the underlying model and use as much data as possible whenever it is deemed being of an acceptable quality.

As a closing remark we note that any full evaluation of the out-of-sample performance should take the entire fitted and forecasted distribution into account, cf. Dowd *et al.* (2008a). At first glance our model seems to provide reasonably wide distributions on both short and long horizons, thus nicely accompanying the reasonable forecasts. Presently we shall not, however, elaborate further on this.

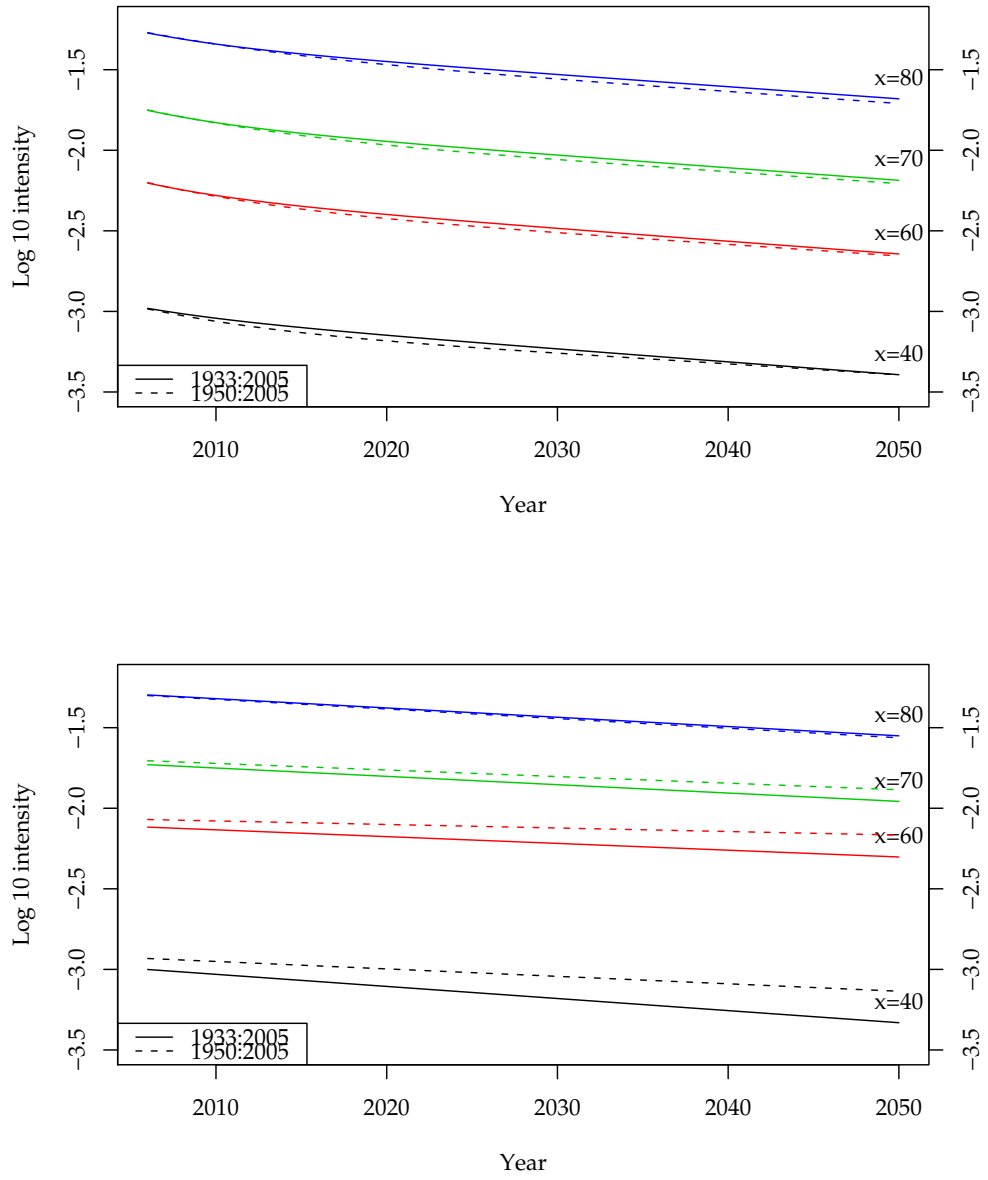

FIGURE 6. Mortality intensity forecasts based on two different estimation periods. Upper panel: SAINT model. Lower panel: Lee-Carter model. Danish women.

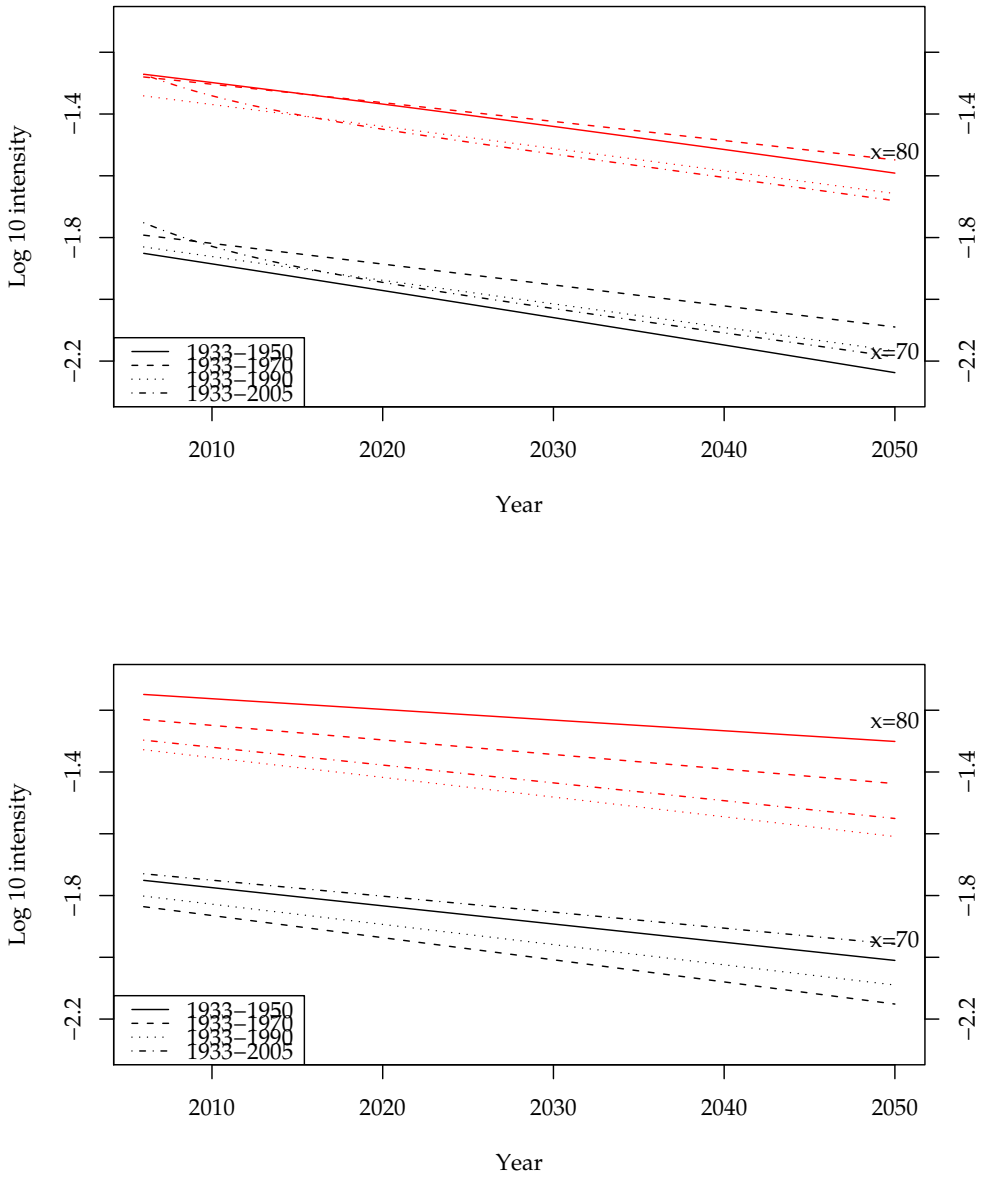

FIGURE 7. Mortality intensity forecasts based on four different estimation periods. Upper panel: SAINT model. Lower panel: Lee-Carter model. Danish women.

5. Concluding remarks

The mainstream in mortality modeling builds on linear time series of unobserved underlying random processes. This typically works very well when the population in question is sufficiently large that realised death rates are smooth, in particular over relatively short forecast horizons. Over longer horizons, and for small populations on the other hand the performance of these models is less convincing, and the estimates may be very sensitive to the choice of input data. We therefore developed a two-step approach—modeling first the mortality of a larger reference population, then the mortality spread between the two populations.

We have left the choice of reference population a subjective one. The reference population should be related to the population of interest as we have to believe that the two populations share the same long term trend. Observing this, we recommend to choose it as large as possible for best identification of the trend. The analysis obviously depends on the choice of reference population, but in this respect the choice of reference population is no different from the choice of estimation window, or indeed the choice of model.

In the presented model we have focused on forecasting a single population. However, the methodology can easily be extended to produce coherent, i.e. non-diverging, forecasts for a group of related populations by using the group as reference population and treat each population as a subpopulation of the group. Similarly, we could consider men and women as subpopulations of the same population rather than estimate separate models for each gender as done in the application. Using common and individual components to produce coherent forecasts for related populations in the Lee-Carter framework has been suggested by Li and Lee (2005).

We have used a two-stage estimation routine in which we first estimate the trend parameters and then estimate the spread parameters with the trend kept fixed. This approach can be justified when the reference population is substantially larger than the subpopulation, as in case of Danish and international data. For applications in which the reference and subpopulation are of comparable size one might consider to estimate the trend and the spread jointly. It is straightforward to write down the likelihood function so in principle this is possible, but it is numerically involved due to the large number of parameters involved.

The trend component of the SAINT model imposes structure on how mortality can evolve over time and across ages. The parametric form provides insight into the improvement patterns and it guarantees biologically plausible forecasts. Compared to the Lee-Carter model the structure of the SAINT model lead to more precise long-term forecasts at the price of higher bias. The higher bias was primarily for young ages which is not surprising as the focus of our modelling has been on old age mortality. The bias at young

ages could undoubtedly be reduced by more careful modelling of these age groups if so desired.

We have concentrated on the uncertainty generated by the stochastic term of the spread model, and only briefly mentioned the possibility of including parameter uncertainty. In the case of the trend, however, the latter source of uncertainty would be negligible since the trend parameters are so well-determined. Another possibility which might better reflect the uncertainty of the trend would be to introduce stochastic terms in the trend model also, thus treating both the trend and the spread as stochastic processes. This however is not straightforward since we wish to preserve the overall structure of the trend. We believe that constructing confidence intervals which properly reflects the intrinsic uncertainty of mortality projections is an important topic which calls for more attention than so far received.

Appendix A

A.1. **Background.** Recall that the density of the Γ-distribution with shape parameter $\lambda > 0$ and scale parameter $\beta > 0$ is given by

$$f(z) = \frac{\beta^{-\lambda}}{\Gamma(\lambda)} z^{\lambda-1} \exp\left(-z/\beta\right), \ (z \geq 0). \tag{39}$$

This distribution has been mean $\beta\lambda$ and variance $\beta^2\lambda$. Letting $\lambda = \beta^{-1} = \sigma^{-2}$ we obtain a Γ-distribution with mean 1 and variance σ^2.

The survival function, $\bar{F}(t,x)$, denotes the proportion of the cohort born at time $t-x$ still alive at time t (at age x). Similarly, the individual survival function, $\bar{F}(t,x;z)$, denotes the probability that a person with frailty z born at time $t-x$ is still alive at time t. If f denotes the density of the frailty distribution at birth we have

$$\bar{F}(t,x) = \int_0^\infty \bar{F}(t,x;z) f(z) dz, \tag{40}$$

while the conditional frailty density at time t for persons of age x is given by

$$f(z|t,x) = \frac{f(z)\bar{F}(t,x;z)}{\bar{F}(t,x)}. \tag{41}$$

The survival function can be expressed in terms of the force of mortality as

$$\bar{F}(t,x) = \exp\left(-\int_0^x \mu(u+t-x,u)du\right), \tag{42}$$

and, conversely,

$$\mu(t,x) = \left[-\frac{d}{d\delta}\log \bar{F}(t+\delta, x+\delta)\right]_{|\delta=0}. \tag{43}$$

The same relationships hold for $\bar{F}(t,x;z)$ and $\mu(t,x;z)$.

A.2. **Proofs.**

Proof of Example 2.1. The first equality in (10) follows from Proposition 2.1 with $\mu_s^I(t,x) = \alpha \exp(\beta x)$ and $\gamma(t) = \gamma$, and the second equality follows from Proposition 2.2. □

Proof of Proposition 2.1. By (43), (40), (41) and (11) the population mortality satisfies

$$\begin{aligned}
\mu(t,x) &= \frac{\left[-\frac{d}{d\delta}\bar{F}(t+\delta, x+\delta)\right]_{|\delta=0}}{\bar{F}(t,x)} \\
&= \frac{\int_0^\infty f(z)\left[-\frac{d}{d\delta}\bar{F}(t+\delta, x+\delta; z)\right]_{|\delta=0} dz}{\bar{F}(t,x)} \\
&= \frac{\int_0^\infty f(z)\bar{F}(t,x;z)\mu(t,x;z)dz}{\bar{F}(t,x)} \\
&= \int_0^\infty f(z|t,x)\left(z\mu_s^I(t,x) + \gamma(t)\right) dz \\
&= \mathsf{E}[Z|t,x]\mu_s^I(t,x) + \gamma(t).
\end{aligned}$$

□

Proof of Proposition 2.2. Using (11), the individual survival function can be written

$$\bar{F}(t,x;z) = \exp\left(-zI(t,x) - G(t,x)\right), \tag{44}$$

where $I(t,x) = \int_0^x \mu_s^I(u+t-x,u)du$ and $G(t,x) = \int_0^x \gamma(t-x+u)du$. Inserting (44) in (40) with f given by (39) for $\lambda = \beta^{-1} = \sigma^{-2}$ we get

$$\begin{aligned}
\bar{F}(t,x) &= \frac{\lambda^\lambda}{\Gamma(\lambda)} \int_0^\infty \exp\left(-z\left[\lambda + I(t,x)\right]\right) z^{\lambda-1} dz \exp\left(-G(t,x)\right) \\
&= \left(\frac{1}{1+\sigma^2 I(t,x)}\right)^{1/\sigma^2} \exp\left(-G(t,x)\right). \qquad (45)
\end{aligned}$$

Finally, inserting (44) and (45) in (41) with f as above we obtain

$$f(z|t,x) = \frac{\left(\sigma^{-2} + I(t,x)\right)^{\sigma^{-2}}}{\Gamma\left(\sigma^{-2}\right)} z^{\sigma^{-2}-1} e^{-z\left(\sigma^{-2}+I(t,x)\right)},$$

which we recognize as a Γ-density with $\lambda = \sigma^{-2}$ and $\beta^{-1} = \sigma^{-2} + I(t,x)$. □

Proof of Proposition 2.3. First note that in the notation of Proposition 2.2 we have

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) I(t,x) = \mu_s^I(t,x),$$

and thereby

$$\begin{aligned}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)\log \mathsf{E}[Z|t,x] &= -\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)\log\left(1+\sigma^2 I(t,x)\right)\\ &= -\frac{\sigma^2\mu_s^I(t,x)}{1+\sigma^2 I(t,x)}\\ &= -\sigma^2\mu_s^I(t,x)\,\mathsf{E}[Z|t,x]\\ &= -\sigma^2\mu_s(t,x),\end{aligned}$$

where the last equality follows from Proposition 2.1. Using $\mathsf{E}[Z|t-x,0]=1$ we then get

$$\begin{aligned}\log \mathsf{E}[Z|t,x] &= \int_0^x \left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}\right)\log \mathsf{E}[Z|t-x+u,u]du\\ &= -\sigma^2\int_0^x \mu_s(t-x+u,u)du.\end{aligned}$$

□

Proof of Proposition 2.4. The equality in (24) follows from (17) and the specification of μ_s^I. To show convergence we first use Proposition 2.2 to write

$$\frac{\partial \log \mathsf{E}[Z|t,x]}{\partial t} = -\frac{\sigma^2\frac{\partial}{\partial t}I(t,x)}{1+I(t,x)}, \tag{46}$$

where $I(t,x)=\int_0^x K(t-x,y)dy$. By (23) and dominated convergence we have

$$I(t,x)=\int_0^x \exp\left((\kappa_2+g_2 y)t\right)K(-x,y)dy \to 0 \text{ for } t\to\infty,$$

since $\kappa_2+g_2 y<0$ for all $0\le y\le x$ by assumption. Similarly,

$$\frac{\partial}{\partial t}I(t,x)=\int_0^x \exp\left((\kappa_2+g_2 y)t\right)(\kappa_2+g_2 y)K(-x,y)dy \to 0 \text{ for } t\to\infty,$$

and we conclude that (46) also converges to zero. □

Proof of Proposition 2.5. First case follows by applying l'Hôpital's rule to (22), second case is the Gamma-Makeham model, and third case is trivial. □

Appendix B

Country	Period	Male deaths	Male exposure	Female deaths	Female exposure
Australia	1933-2004	3656554	298774817	3050957	304038560
Austria	1947-2005	2362681	146948386	2544281	172938922
Belgium	1933-2005	3959202	237960436	3655934	254715105
Canada	1933-2005	5818067	501315204	4711059	510104624
Denmark	1933-2005	1736235	118573885	1651576	124503893
England & Wales	1933-2005	18790451	1155893702	18906239	1301265668
Finland	1933-2005	1649803	105805700	1493332	118540078
France	1933-2005	19554792	1197266420	18601741	1332188094
Iceland	1933-2005	50595	4531626	46223	4597037
Italy	1933-2004	17840263	1231244196	16616547	1360392086
Japan	1947-2005	22944373	2115001379	19852118	2274227259
Netherlands	1933-2005	3785909	304422103	3442716	316942482
Norway	1933-2005	1326283	94013584	1240955	98579501
Portugal	1940-2005	2859456	189334582	2769074	216359565
Spain	1933-2005	10815395	785643871	9940870	862709017
Sweden	1933-2005	2969115	202088902	2792398	209883920
Switzerland	1933-2005	1949244	145118601	1888688	159060151
USA	1933-2005	67513137	4756496475	58192508	5127612213
West Germany	1956-2005	16193960	1081937335	17307798	1227848814

Table 6. Summary of deaths and exposures for ages 20-100 over the period 1933-2005 for the countries included in the international data set. Source: The Human Mortality Database (www.mortality.org).

References

Andreev, K. F. (2002). *Evolution of the Danish Population from 1835 to 2000.* Monographs on Population Aging, 9. Odense University Press.

Barbi, E. (2003). Assessing the rate of ageing of the human population. Working Paper. Max Planck Institute for Demographic Research.

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

Booth, H., Hyndman, R. J., Tickie, L., and de Jong, P. (2006). Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research* **15**, 289–310.

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

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods, Second Edition.* Springer-Verlag, New York.

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**, 373–393.

Cairns, A. J., 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**, 687–718.

Cairns, A. J. G. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics* **27**, 313–330.

Currie, I. D., Durban, M., and Eilers, P. H. C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* **4**, 279–298.

de Jong, P. and Tickle, L. (2006). Extending Lee-Carter mortality forecasting. *Mathematical Population Studies* **13**, 1–18.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2008a). Backtesting stochastic mortality models: An ex-post evaluation of multi-period-ahead density forecasts. Pensions Institute Discussion paper PI-0803.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2008b). Evaluating the goodness of fit of stochastic mortality models. Pensions Institute Discussion paper PI-0802.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, New York.

Gavrilov, L. A. and Gavrilova, N. S. (1991). *The Biology of Life Span: A Quantitative Approach, ed. V. P. Skulachev.* Harwood Academic Publishers, Chur.

Jarner, S. F., Kryger, E. M., and Dengsøe, C. (2008). The evolution of death rates and life expectancy in Denmark. *Scandinavian Actuarial Journal* **108**, 147–173.

Koissi, M.-C., Shapiro, A. F., and Högnäs, G. (2006). Evaluating and extending the Lee-Carter model for mortality forecasting: Bootstrap confidence intervals. *Insurance: Mathematics and Economics* **38**, 1–20.

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

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

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

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

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

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

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

Vaupel, J. W. (1999). Discussion on "The Long-Term Pattern of Adult Mortality and the Highest Attained Age" by A. R. Thatcher. *Journal of the Royal Statistical Society. Series A* **162**, 31–32.

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

Wilmoth, J. R. (1998). Is the pace of Japanese mortality decline converging toward international trends? *Population and Development Review* **24**, 593–600.

Wilson, C. (2001). On the scale of global demographic convergence 1950-2000. *Population and Development Review* **27**, 155–171.

Yashin, A. I. and Iachine, I. A. (1997). How frailty models can be used for evaluating longevity limits: Taking advantage of an interdisciplinary approach. *Demography* **34**, 31–48.

The Danish Labour Market Supplementary Pension Fund, Kongens Vænge 8, DK-3400 Hillerød, Denmark