Statistical Estimation Techniques in Life and Disability Insurance—A Short Overview

Boualem Djehiche

Abstract This is a short introduction to some basic aspects of statistical estimation techniques known as graduation technique in life and disability insurance.

Keywords Life insurance · Disability insurance · Claims reserving · Mortality modeling · Thiele's equation

Mathematics Subject Classification 2010 90B30 · 60J75 · 62M10

1 Life Insurance

By life insurance policy or contract we mean any form of person insurance contract over a (long) period of time such as life or pension and disability or sickness coverage. In such products, premiums and benefits are typically contingent upon transitions of the policyholder between a number of states stated in the contract. Thereof the use of the powerful (semi)-Markov chain theory to carry out the valuation of insurance contracts and estimation of the underlying rates. We first give a short introduction to the basic constituents of a life insurance contract and related reserving. Then we single out the main parameters that control the evolution of the life insurance contract and focus on their statistical estimation. These parameters are the mortality rate and disability inception and recovery rates. Due to lack of space, the reader is referred to the list of references for an update of recent developments in claims reserving techniques for life and disability insurance. A detailed account for basic life insurance contracts can be found in the papers [11–15] by Norberg. A very short summary is displayed in Sects. 1.1–1.6, below.

B. Djehiche (✉)
Department of Mathematics, KTH Royal Institute of Technology,
SE-100 44, Stockholm, Sweden
e-mail: boualem@math.kth.se

M. Eddahbi et al. (eds.), *Statistical Methods and Applications in Insurance and Finance*, Springer Proceedings in Mathematics & Statistics 158,
DOI 10.1007/978-3-319-30417-5_5

1.1 A Markov Chain Model of a Life Insurance Contract

Let $E = \{0, 1, 2, \dots, m\}$ be the (finite) set of possible states of the policy. Starting at 0, the policy is assumed to be in one and only one state at each time. Let $X(t)$ denote the state of the policy at time $t \in [0, n]$. We assume that the process X is right-continuous with a finite number of jumps, with transition probability

$$p_{ij}(s,t) = P[X(t) = i | X(s) = j], \quad i, j \in E, \quad 0 \le s \le t \le n, \tag{1}$$

and transition intensity

$$\mu_{ij}(t) := \lim_{h \downarrow 0} \frac{p_{ij}(t, t+h)}{h}, \quad i \neq j. \tag{2}$$

The total transition intensity from state i at time t is

$$\mu_{i\cdot}(t) = \sum_{k:k\neq i} \mu_{ik}(t) \tag{3}$$

so that

$$p_{ii}(t, t+dt) = 1 - \mu_{i\cdot}(t)dt + o(t).$$

1.1.1 Basic Kolmogorov Equations

The transition probabilities $(p_{ij}(s,t),\ i, j \in E,\ 0 \le s \le t \le n)$ satisfy the following equations.

A. The Kolmogorov backward equation: for $s \le t$,

$$\begin{cases} \frac{\partial p_{ij}}{\partial s}(s,t) = \mu_{i\cdot}(s)p_{ij}(s,t) - \sum_{k:k\neq i} \mu_{ik}(s)p_{kj}(s,t), \\ p_{ij}(t,t) = \delta_{ij}. \end{cases} \tag{4}$$

B. The Kolmogorov forward equation: for $s \le t$,

$$\begin{cases} \frac{\partial p_{ij}}{\partial t}(s,t) = -p_{ij}(s,t)\mu_{j\cdot}(t) + \sum_{k:k\neq i} p_{ik}(s,t)\mu_{kj}(t), \\ p_{ij}(s,s) = \delta_{ij}. \end{cases} \tag{5}$$

C. The Chapman-Kolmogorov equation

$$p_{ik}(s,u) = \sum_{j\in E} p_{ij}(s,t)p_{jk}(t,u), \quad s \le t \le u. \tag{6}$$

The key parameter in this Markov chain framework is the *transition intensity* which is the object of our statistical inference study.

1.2 Examples

1.2.1 Single Life with One Cause of Death (One Absorbing State)

In this model $E = \{0, 1\}$, where state 0 = alive, state 1= dead (absorbing state). If T denotes the life length of a person with survival probability

$$\bar{F}(t) = P(T > t),$$

the Markov chain counts the number of deaths:

$$X(t) = \mathbb{1}_{\{T \leq t\}}, \quad t \in [0, n],$$

with transition probability

$$p_{00}(s, t) = \frac{\bar{F}(t)}{\bar{F}(s)} = e^{-\int_s^t \mu(u)du}.$$

μ is called *mortality intensity (rate or force)*. Its estimation from data is of central importance in life insurance Fig. 1

1.2.2 Single Life with *m* Causes of Death (*m* Absorbing States)

In this model $E = \{0, 1, \ldots, m\}$, where state 0 = alive, state j= dead with cause j (absorbing state). These absorbing states model different causes of death such as death by "car accident", "normal death" or "death caused by a disease" etc. Fig. 2.

The total mortality intensity is

$$\mu_{0\cdot}(t) := \mu(t) = \sum_{j=1}^{r} \mu_j(t), \tag{7}$$

where, $\mu_j(t) := \mu_{0j}(t)$ denotes the mortality rate for death with cause j. This is nothing but the transition intensity from state 0 (alive) to the absorbing state j.

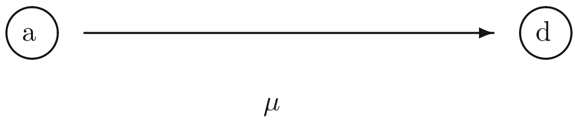


Fig. 1 Single life with one cause of death

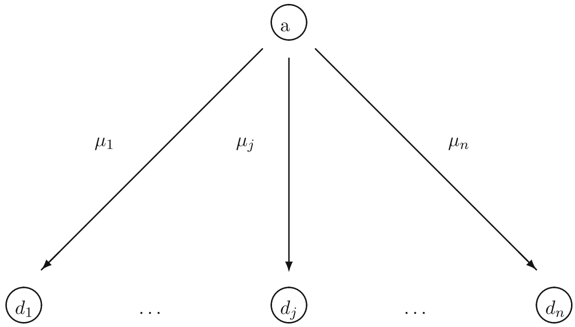


Fig. 2 Single life with m causes of death

The probability that an s years old person will die from cause j before age t is then

$$p_{0j}(s,t) = \int_s^t e^{-\int_s^u \mu(\tau)d\tau} \mu_j(u)du. \tag{8}$$

1.2.3 Disability, Recovery and Death

This model is widely used to analyze insurance contracts with payments depending on the state of the health of the insured. For example

- Sickness insurance that provides an annuity benefit during disability periods.
- Life insurance with premium waiver during disability.
- Pension with additional benefits to other members of the family.

The possible states are a = alive/active, i=invalid/unemployed, and d=dead/recovered or any other suitable labeling Fig. 3.

1.3 Payment Streams and Reserving Techniques

Let X be the Markov chain with intensities μ_{ij} associated with an insurance contract. Let

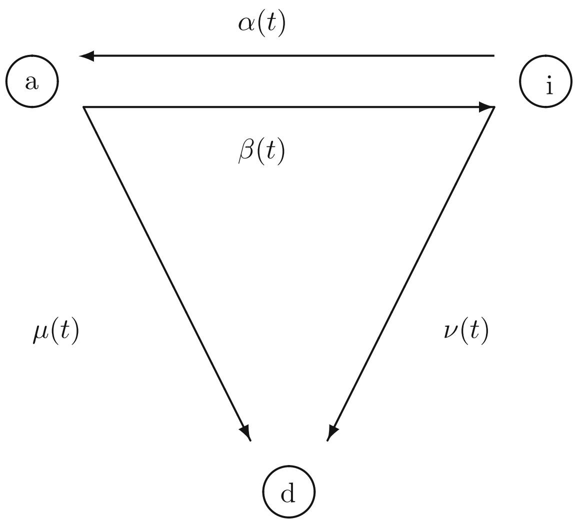
$$I_j(t) = 1\!\!1_{\{X(t)=j\}}, \quad t \in [0, T],$$

denote the indicator process of whether the policy is in state j or not, and

$$N_{ij}(t) = \#\{s\colon X(s^-) = i,\ X(s) = j,\ s \in (0,t]\}, \quad i \neq j,$$

denote the number of transitions from state i to state j during the time interval $(0, t]$.

Fig. 3 Three possible states of a life insurance contract



We have

$$dI_j(t) = dN_{\cdot j}(t) - dN_{j\cdot}(t), \tag{9}$$

where,

$$N_{\cdot j}(t) := \sum_{k;k\neq j} N_{kj}(t), \qquad N_{j\cdot}(t) := \sum_{k;k\neq j} N_{jk}(t).$$

We have, for $t \leq u$

$$\begin{aligned} E[I_j(u)|X(t) = i] &= p_{ij}(t, u), \\ E[dN_{jk}(u)|X(t) = i] &= p_{ij}(t, u)\mu_{jk}(u)du. \end{aligned} \tag{10}$$

A standard payment stream A (benefits less premiums) has usually the following form:

$$dA(t) := \sum_j \left(I_j(t)dA_j(t) + \sum_{k;k\neq j} a_{jk}(t)dN_{jk}(t) \right), \tag{11}$$

where,

$$dA_j(t) := a_j(t)dt + (A_j(t) - A_j(t^-)) = a_j(t)dt + \Delta A_j(t) \tag{12}$$

specifies the so-called *general life annuity payment* i.e. payments due during sojourn in state j. The payment $a_j(t)$ is the rate of a state-wise annuity payable continuously at time t, while the lump sum payment $\Delta A_j(t)$ is an endowment at time t. The annuity function A_j is usually assumed to have a finite number of discontinuity points $\{t_1, t_2, \dots, t_q\}$. The payments $a_{jk}(t)$ specify the so-called *general life assurance* i.e. amounts that are payable immediately upon transition from state j to state k.

1.4 Expected Present Values and Prospective Reserves

The liability at time t for which the insurer should provide a reserve (prospective reserve) is the present value of the payment streams (future benefits less premiums) A over the lifespan $[t, n]$ of the insurance contract:

$$V(t) = \int_t^n e^{-\int_t^s r(u)du} dA(s). \tag{13}$$

When the policy is in state i at time t, then, in view of Eq. (10), the state-wise *prospective reserve* is

$$\begin{aligned} V_i(t) &:= E[V(t)|X(t) = i] = \int_t^n e^{-\int_t^s r(u)du} E[dA(s)|X(t) = i] \\ &= \int_t^n e^{-\int_t^s r(u)du} \sum_j p_{ij}(t, s) \left(dA_j(t) + \sum_{k;k\neq j} a_{jk}(s)\mu_{jk}(s)ds \right), \end{aligned} \tag{14}$$

when r, a_j, a_{ik} are all deterministic function.
Written in differential form, V_j satisfies the following Feynman-Kac type formula known as **the backward Thiele's differential equation**:

$$\begin{cases} \frac{dV_i}{dt}(t) = (r(t) + \mu_{i\cdot}(t))V_i(t) - \sum_{j;j\neq i} \mu_{ij}(t)V_j(t) - a_i(t) - \sum_{j;j\neq i} a_{ij}(t)\mu_{ij}(t), \\ \qquad t \in (t_{p-1}, t_p), \quad p = 1, \dots, q, \\ \Delta V_j(t_p) = -\Delta A_j(t_p), \quad p = 1, 2, \dots, q, \quad i \in E, \\ V_j(n) = 0. \end{cases} \tag{15}$$

This equation admits an explicit solution only for a few uninteresting/trivial insurance contracts. In most cases it is solved using a numerical integration recipe. A fourth order "Runge-Kutta" procedure seems to work efficiently in almost all practical situations.

Thiele's equation can be recast in the following form "preferred by actuaries"

$$-a_i(t)dt = dV_i(t) - r(t)V_i(t)dt + \sum_{j;j\neq i} R_{ij}(t)\mu_{ij}(t)dt \tag{16}$$

where,

$$R_{ij}(t) = a_{ij}(t) + V_j(t) - V_i(t), \tag{17}$$

is the so-called "Sum-at-Risk" associated with a possible transition from state i to state j.

- The term $\sum_{j;j\neq i} R_{ij}(t)\mu_{ij}(t)dt$ is called the "risk premium" in $(t, t+dt)$.
- The term $dV_i(t) - r(t)V_i(t)dt$ is called the "savings premium" in $(t, t+dt)$.

1.5 The Equivalence Principle (aka Fairness Constraint)

The equivalence principle of insurance states that the expected present values of premiums and benefits should be equal. That is, roughly speaking, premiums and benefits should balance on the average. In our context this principle states that

$$V_0(0) = -A_0(0). \tag{18}$$

This condition imposes a constraint on the contractual payments a_j, A_j and a_{ij} to design a premium level for given benefits. Noting that $A_0(0^-) = 0$, we easily see that Eq. (18) is equivalent to

$$V_0(0^-) := E\left[\int_{0^-}^{n} e^{-\int_0^s r(u)du} dA(s)\right] = 0. \tag{19}$$

The state-wise prospective reserve $V(t)$ can be seen as the value function of a *singular* control problem subject to the fairness constraint, where the control parameter is the process $A(t)$.

1.6 First and Second Order Reserving Bases

The jump intensities μ_{ij} (purely actuarial parameters or liability driving parameter) and the discounting rate r which reflects the "expected return" of the investment portfolio (the main driver of the asset side) constitute the so-called *reserving basis*:

• **First order technical basis (prudent or conservative)**. This is a set of assumptions about the portfolio return (or just an interest rate that reflects the market value of the cash flow), r, the transition rates μ_{ij} (including mortality rates), costs and other relevant technical parameters etc. These assumptions are meant to yield premiums and reserves that include a high safety loading that hedges against worst case scenarios. The first order premiums and reserves are usually higher than experience based or historically observed values. This means that a systematic surplus is created by the company and, by law (which regulates mutual funds in some countries), it should be redistributed to the policyholder in terms of *bonuses* that are usually allocated but not distributed until the termination of the policy. Here we face a model risk!

• **Second order technical basis**: It is also called experience (or market) basis. It sets values of the parameters based on *realistic scenarios* collected based on the history of the policy. The company updates the reserves on a regular basis and adjusts for the parameters using the *bonus fund* created by applying the first order basis.

A typical example of adjustments to be made under the experience (market) basis is compensation for a possible non-equivalence of the first order payments i.e. $V_0(0^-) \neq 0$, i.e. the insurance company compensates for this by adding dividend payments D to the first order payments. D has usually the following form:

$$dD(t) := \sum_j \left(I_j(t) dD_j(t) + \sum_{k;k\neq j} \delta_{jk}(t) dN_{jk}(t) \right), \tag{20}$$

where,

$$dD_j(t) := \delta_j(t)dt + (D_j(t) - D_j(t^-)) = \delta_j(t)dt + \Delta D_j(t). \tag{21}$$

The coefficients $\delta_j, \Delta D_j$ and δ_{ij} are stochastic processes adapted to the "demographic-economic" history $\mathcal{F}$ with a more complex structure than the coefficients related to the payment processes A. The dividend process D is chosen (constrained) to attain the ultimate equivalence (fairness):

$$E\left[\int_{0^-}^{n} e^{-\int_0^s r(u)du} d(A+D)(s) \right] = 0. \tag{22}$$

In the Black and Scholes market model, the dividend payments are provided by an asset portfolio such as the following diffusion Y modulated by the jump process X:

$$\begin{aligned} &dY(t) = rY(t)dt + \sigma(t, X(t), Y(t))Y(t)dW(t) + d(C-D)(t), \\ &Y(0^-) = 0, \end{aligned} \tag{23}$$

where, C is the usual income (or contribution) process of the following form (similar to A and D):

$$dC(t) := \sum_j \left(I_j(t) dC_j(t) + \sum_{k;k\neq j} c_{jk}(t) dN_{jk}(t) \right), \tag{24}$$

$$dC_j(t) := c_j(t)dt + (C_j(t) - C_j(t^-)) = c_j(t)dt + \Delta C_j(t). \tag{25}$$

Assuming the coefficients $\delta_j(t)$, $\Delta D_j(t)$ and $\delta_{ij}(t)$ are functions of $(t, Y(t))$, the state-wise *prospective reserve* is

$$\begin{aligned} V_i(t,x) &:= E[V(t)|X(t)=i, Y(t)=x] \\ &= E\left[\int_t^n e^{-\int_t^s r(u)du} d(A+D)(s)|X(t)=i, Y(t)=x\right] \end{aligned} \tag{26}$$

satisfies a more complex "Thiele's" PDE (cf. [8, 16, 18]).

1.7 Graduation Techniques-Estimation of the Mortality Rates

We start with statistical inference of the mortality rate μ which is the only jump intensity in the simplest life insurance contract: Single life with one cause of death (one absorbing state) i.e. $E = \{0, 1\}$, where state 0 = alive, state 1= dead (absorbing state). The underlying Markov chain counts the number of deaths:

$$X(t) = 1\!\!1_{\{T \le t\}}, \quad t \in [0, n],$$

where, T denotes the life length of a person with survival probability

$$p_{00}(s,t) = \frac{\bar{F}(t)}{\bar{F}(s)} = e^{-\int_s^t \mu(u)du}, \qquad 0 \le s \le t \le n.$$

In actuarial practice one often considers the *remaining life length* T_x of an insured of age x. The corresponding survival probability over a time period of length $t \ge 0$ is

$$P(T_x > t) := P(T > x + t | T > x) = e^{-\int_x^{x+t} \mu(u)du} = e^{-\int_0^t \mu(x+u)du}. \tag{27}$$

In a more general framework where 'stochastic mortality' modeling can be incorporated, consider (the possibly random) force (or rate) of mortality $\mu(x, t)$ at t for individual aged x at time 0. Then, the *survival index*

$$S(x,t) := \exp\left(-\int_0^t \mu(x+s, s)ds\right)$$

is the probability of survival of an individual aged x during the time interval $[0, t]$, given the mortality force $\mu(x, s)$ i.e.

$$P(T_x > t) = E[S(x,t)].$$

In Eq. (27), $\quad \mu(x,t) = \mu(x+t)$.

The main goal of this section is to estimate the mortality force $\mu(x, s)$, given historical mortality data of a population of insured individuals.

1.7.1 An Age-Specific Model: Gompertz-Makeham Graduation Formula

This model captures the evolution of mortality in mutually exclusive age cohorts but disregards a possible common risk factor that links all cohorts together. Consider an insured population of ages $x_i,\ i = 1, 2, \ldots, n$. Let N_x denote the exposure i.e. the number of individuals of the same age x, and D_x denotes the number of individuals dead during the interval $(x, x + 1)$. Assuming that the remaining survival lengths of all individuals are *independent*, and the insured population is *homogeneous* in the sense that the survival probability of all individuals is the same. A stochastic model based on a "crude approximation" of the Binomial distribution by the Poisson distribution suggests that

$$D_{x_i} \sim \text{independent Poisson}(\mu_{x_i} N_{x_i}). \tag{28}$$

Then the mortality rate (or force) μ_{x_i} for a population of age $x_i,\ i = 1, 2, \ldots, n$ can be estimated by the so-called 'central or crude death rate'

$$\hat{\mu}_{x_i} = \frac{D_{x_i}}{N_{x_i}}, \quad i = 1, 2, \ldots, n. \tag{29}$$

Gompertz and later Makeham famous graduation formula suggests a mortality rate of the form

$$\mu_x := \alpha + \beta e^{\gamma x}, \tag{30}$$

where, the parameters α, β and γ which satisfy $\alpha + \beta > 0, \beta > 0$ and $\gamma \geq 0$ are estimated using the insured population data. When $\alpha = 0$ we get Gompertz mortality law. A fairly standard way to perform the parameter estimation is to use a weighted least squares method: minimize

$$Q = \sum_{i=0}^{n} w_{x_i} \left(\hat{\mu}_{x_i} - \alpha - \beta e^{\gamma x_i}\right)^2 \tag{31}$$

w.r.t. the parameters α, β and γ, where the weight is the inverse of the variance of $\hat{\mu}_{x_i}$:

$$w_{x_i} = \frac{N_{x_i}^2}{Var(D_{x_i})} = \frac{N_{x_i}^2}{N_{x_i}\hat{\mu}_{x_i}} = \frac{N_{x_i}}{\hat{\mu}_{x_i}}, \tag{32}$$

so that Q is approximately χ^2-distributed. In practice, one 'fixes' a value for γ 'based on experience' and finds the optimal values of α and β. In the Swedish life insurance business, there is a Central Mortality Committee that estimates these parameters to be used by insurance companies and pension funds. For example, in the so-called M90 investigation, the committee suggested that

$$\mu_x = \alpha + \beta e^{\gamma(x-f)},$$

where, the parameter f adjusts for mortality of females among the insured population. Values $f = 4$ or 5 years are used. For M90, $\alpha = 0.001$, $\beta = 0.000012$ and $\gamma = 0.044$.

1.7.2 Gompertz Graduation Formula with a View Towards GLM

Recall Gompertz' graduation formula:

$$\mu_x := \beta e^{\gamma x}, \tag{33}$$

or $\log \mu_x$, which is linear in age,

$$\log \mu_x = \log \beta + \log e^{\gamma x} := a + \gamma x.$$

This can be extended to a quadratic or a polynomial form

$$\log \mu_x = a + bx + cx^2, \quad \log \mu_x = a_0 + a_1 x + a_2 x^2 + \ldots + a_p x^p.$$

GLM means that we perform a regression of $\log \mu_x$ with respect to a basis

$$\{1, x\}, \quad \{1, x, x^2\}, \quad \{1, x, \ldots, x^p\},$$

or any other carefully chosen 'spline' basis $\{B_1(x), B_2(x), \ldots, B_p(x)\}$ such that

$$\mu_x = \sum_{j=1}^{p} B_j(x) a_j := P(a),$$

and estimate the coefficients $a_0, a_1, \ldots, a_p$ which maximize the *penalized log-likelihood function*:

$$L(a) - \frac{1}{2}\lambda P(a), \tag{34}$$

where, $L(a)$ is the log- likelihood of the model

$$D_{x_i} \sim \text{independent Poisson}(\mu_{x_i} N_{x_i}), \quad i = 1, \ldots, n, \tag{35}$$

and $\lambda > 0$ is a smoothing parameter.

A similar approach can be applied to obtain a smooth *year (or period) specific* mortality: maximize the *penalized log-likelihood function*

$$L(\theta) - \frac{1}{2}\lambda P(\theta), \tag{36}$$

where, $L(\theta)$ is the log- likelihood of the model

$$D_{t_i} \sim \text{independent Poisson}(\mu_{t_i} N_{t_i}), \quad t = t_{\min}, \ldots, t_{\max}, \tag{37}$$

and

$$P(\theta) = \sum_{j=1}^{p} B_j(t)\theta_j.$$

The smoothing parameter λ can be estimated using the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) or the Generalized Cross-Validation (GCV).

1.7.3 An Age-Period Model: Lee-Carter Graduation Formula

Lee and Carter [10] suggest a Gompertz type graduation formula for the full mortality rate $\mu(x, t)$:

$$\log \mu(x, t) := \alpha(x) + \beta(x)\kappa(t), \tag{38}$$

subject to the constraints

$$\sum_x \beta(x) = 1, \quad \sum_t \kappa(t) = 0, \tag{39}$$

fitting

$$\sum_{x,t} \left(\log \mu_{obs}(x, t) - \alpha(x) + \beta(x)\kappa(t)\right)^2.$$

This model captures the evolution of mortality in mutually exclusive age cohorts while at the same time includes a possible common risk factor (systemic risk) $k(t)$ that links all cohorts together over time. The parameters $a(x)$ and $b(x)$ are age-specific while $k(t)$ is time (period) dependent only and should capture the random period effect of the mortality rate. The risk factor $k(t)$ is usually modeled as a time series or a random walk with drift. Lee and Carter [10] suggest an ARIMA (discretized diffusion process) for κ of the form

$$k(t+1) = k(t) + a_1 + a_2\xi + \sigma z(t)$$

where, $z(t)$ is white noise and $\xi \in \{0, 1\}$ is a dummy variable that captures major outbreaks of disease leading to a huge mortality wave such as the 1918 worldwide flu outbreak or the 2008 earthquake in China etc. Statistical estimation of these parameters is usually performed w.r.t. each dimension: x and time (period) t. Here are some suggestions (see [2, 10], Currie, Richards and co-authors (2003–2012), [17] etc.).

- Given $\kappa(t) = \hat{\kappa}(t)$, fit a GLM with regressor $\hat{\kappa}$:

$$\log \mu(x,t) := \alpha(x) + \beta(x)\hat{\kappa}(t).$$

- Given $\alpha(x) = \hat{\alpha}(x), \quad \beta(x) = \hat{\beta}(x)$, fit a GLM with offset $\hat{\alpha}(x)$ and regressor $\hat{\beta}(x)$:

$$\log \mu(x,t) := \hat{\alpha}(x) + \hat{\beta}(x)\kappa(t).$$

- Perform a regression w.r.t. a 2-d spline basis $B_a(x) \otimes B_y(t)$ for age and time dimensions (x,t).

1.7.4 Building Blocks of the MLE for the Lee-Carter Model

Following [2], the MLE approach to the Lee-Carter model is based on the assumption that

$$\begin{aligned} D_{x,t} &\sim \text{Poisson}(\mu(x,t)N_{x,t}), \text{ where } \log \mu(x,t) := \alpha(x) + \beta(x)\kappa(t), \\ & x = x_{\min}, \ldots, x_{\max}, \quad t = t_{\min}, \ldots, t_{\max}. \end{aligned} \tag{40}$$

The parameters $\alpha(x)$, $\beta(x)$ and $\kappa(t)$ are estimated by maximizing the log-likelihood function

$$L(\alpha,\beta,\kappa) := \sum_{x,t} \big(D_{x,t}(\alpha(x) + \beta(x)\kappa(t)) - N_{x,t} \exp\left(\alpha(x) + \beta(x)\kappa(t)\right) \big) + C,$$

where, C contains all the terms that do not dependent on the parameters. The nonlinear term $\beta(x)\kappa(t)$ does not allow for a closed form of the maximizing parameters. One instead uses an iterative method such as the Newton-Raphson updating scheme (or any more efficient numerical optimization algorithm):

$$\theta^{(n+1)} = \theta^{(n)} - \frac{\partial L^{(n)}/\partial\theta}{\partial^2 L^{(n)}/\partial\theta^2},$$

which numerically solves $\partial L^{(n)}/\partial\theta = 0$.

$$
\begin{cases}
\hat{\alpha}_x^{(0)} = 0, \quad \hat{\beta}_x^{(0)} = 1, \quad \hat{\kappa}_t^{(0)} = 0, \\
\text{alternatively } \hat{\alpha}_x^{(0)} = \frac{1}{t_{\max}-t_{\min}+1} \sum_t \log(\hat{\mu}(x,t)), \; \hat{\beta}_x^{(0)} = \frac{1}{t_{\max}-t_{\min}+1}, \\
\hat{\kappa}_t^{(0)} = \sum_x \hat{\beta}_x^{(0)} \left(\log(\hat{\mu}(x,t)) - \hat{\alpha}_x^{(0)}\right), \\
\hat{D}_{x,t}^{(n)} = N_{x,t} \exp\,(\hat{\alpha}_x^{(n)} + \hat{\beta}^{(n)} \hat{\kappa}_t^{(n)}), \\
\hat{\alpha}_x^{(n+1)} = \hat{\alpha}_x^{(n)} - \frac{\sum_t (D_{x,t} - \hat{D}_{x,t}^{(n)})}{-\sum_t \hat{D}_{x,t}^{(n)}}, \quad \hat{\beta}_x^{(n+1)} = \hat{\beta}_x^{(n)}, \quad \hat{\kappa}_t^{(n+1)} = \hat{\kappa}_t^{(n)}, \\
\hat{\kappa}_t^{(n+2)} = \hat{\kappa}_t^{(n)} - \frac{\sum_t (D_{x,t} - \hat{D}_{x,t}^{(n+1)}) \hat{\beta}_x^{(n+1)}}{-\sum_t \hat{D}_{x,t}^{(n+1)} (\hat{\beta}_x^{(n+1)})^2}, \quad \hat{\alpha}_x^{(n+2)} = \hat{\alpha}_x^{(n+1)}, \; \hat{\beta}_x^{(n+2)} = \hat{\beta}_x^{(n+1)}, \\
\hat{\beta}_x^{(n+3)} = \hat{\beta}_x^{(n+2)} - \frac{\sum_t (D_{x,t} - \hat{D}_{x,t}^{(n+2)}) \hat{\kappa}^{n+2}(t)}{-\sum_t \hat{D}_{x,t}^{(n+2)} (\hat{\kappa}_t^{n+2})^2}, \quad \hat{\alpha}_x^{(n+3)} = \hat{\alpha}_x^{(n+2)}, \; \hat{\kappa}_t^{(n+3)} = \hat{\kappa}_t^{(n+2)}.
\end{cases}
$$

The parameters are standardized in each step of the iteration to satisfy the constraints

$$
\sum_x \beta(x) = 1, \quad \sum_t \kappa(t) = 0, \tag{41}
$$

by letting

$$
\hat{\alpha}_x^{(n+1)} = \hat{\alpha}_x^{(n)} + A\hat{\beta}_x^{(n)}, \quad \hat{\kappa}_x^{(n+1)} = (\hat{\kappa}_t^{(n)} - A)B, \quad \hat{\beta}_x^{(n+1)} = \hat{\beta}_x^{(n)}/B, \tag{42}
$$

where,

$$
A = \frac{1}{t_{\max} - t_{\min}} \sum_t \hat{\kappa}_t^{(n)}, \quad B = \sum_x \hat{\beta}_x^{(n)}. \tag{43}
$$

The estimated values of $\kappa(t), \;\; t = t_{\min}, \ldots, t_{\max}$ are used to fit it to a dynamical model. We mentioned above that Lee and Carter fit $\kappa(t)$ to an ARIMA model of the form

$$
k(t+1) = k(t) + a_1 + a_2\xi + \sigma z(t)
$$

where, $z(t)$ is white noise and $\xi \in \{0, 1\}$ is a dummy variable that captures major mortality changes.

This algorithm is illustrated by the Figs. 4, 5 and 6, applied to mortality data among Swedish insured (cf. Swedish Research Board for Actuarial Science [17]).

Mortality jumps, due to e.g. new life standards or medical development etc., are also important to capture in a mortality model, despite the serious difficulties to perform reliable estimation. Cox et al. [7] suggest two types of mortality jump events to the Lee-Carter model:

$$
\log \mu(x,t) := \alpha(x) + \beta(x)\kappa(t) - G(x,t) + H(x,t),
$$

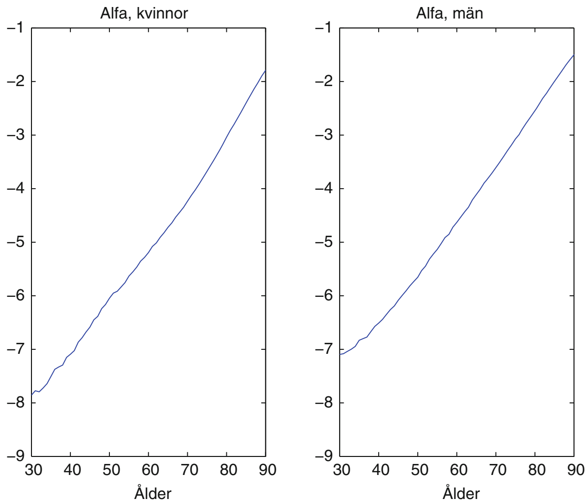


Fig. 4 The α_x parameter for ages 30–90 years (females and males). (From [17], reproduced with permission from Taylor and Francis Ltd.)

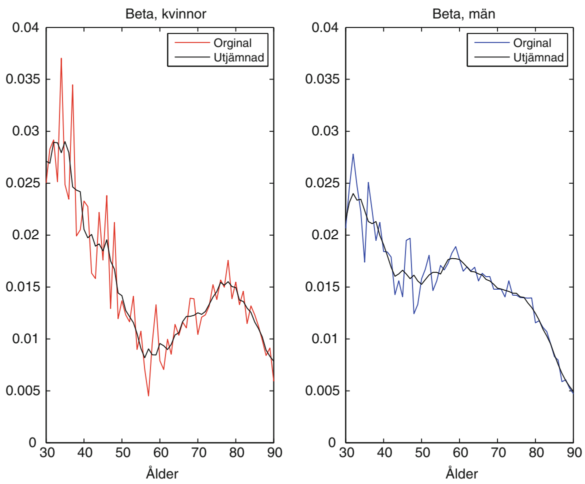


Fig. 5 Estimated and smoothed β_x parameter for ages 30–90 years (females and males). (From [17], reproduced with permission from Taylor and Francis Ltd.)

where

- $G(x, t)$ captures a *permanent longevity jump* and takes the form

$$G(x, t) := K(x, t) + D(x, t),$$

with

$$K(x, t) := \sum_{j=1}^{\infty} y_j A_j(x) \mathbb{1}_{\{t \geq \eta_j\}} = \text{Jump reduction component},$$

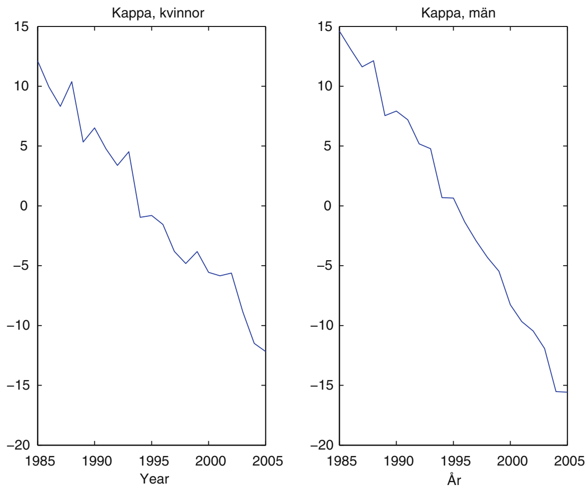


Fig. 6 Estimated and linearized $\kappa(t)$ parameter for data 1985–2005 (females and males). (From [17], reproduced with permission from Taylor and Francis Ltd.)

and

$$D(x,t) := \sum_{j=1}^{\infty} \zeta_i(t-\nu_i)F_i(x)e^{-\xi_i(t-\nu_i)}1\!\!1_{\{t\geq\nu_i\}} = \text{Trend reduction component.}$$

- $K(x,t)$ captures temporary adverse mortality jumps and takes the form

$$H(x,t) := \sum_{j=0}^{\infty} b_j B_j(x)e^{-\kappa_i(t-\tau_i)}1\!\!1_{\{t\geq\tau_j\}}.$$

1.8 An Age-Period-Cohort Model: Extending Lee-Carter Graduation Formula

The Lee-Carter model captures the age-period effect, but does not reflect the possible cohort effect (calender year-age $= t - x$). A simple model that would simultaneously capture the age-period-cohort effect is

$$\log\mu(x,t) := \alpha(x) + \kappa(t) + \gamma(t-x).$$

Renshaw and Haberman [9] suggested the following extension of the Lee-Carter model to capture the cohort effect (calender year-age $= t - x$):

$$\log\mu(x,t) := \beta_1(x) + \beta_2(x)\kappa(t) + \beta_3(x)\gamma(t-x). \tag{44}$$

A generalization of this mortality model for data divided into N components reads

$$\log \mu(x, t) := \sum_{j=1}^{N} \beta_j(x)\kappa_j(t)\gamma_j(t - x).$$

In a series of papers the Edinburgh teams including Currie, Richards and co-authors (2003–2012) and Cairns and co-authors (2006–2012) suggest other extensions and perform deep statistical analysis that seem tune the age-period-cohort effect when applied to mortality data from England and Wales, and USA.

1.9 An Infinite Dimensional Approach to Mortality Modeling

The mortality rate can be viewed as an (infinite dimensional) curve of (x, t). To capture the high level of uncertainty in projections of future mortality one is tempted to translate the "machinery" developed for "forward" interest rate yields such as "the HJM-model under the Musiela parametrization etc." to mortality rates. One is tempted to translated the calibration techniques of interest rate yield curves, to perform hopefully more accurate projections of future mortality (though with limited data points). Recent relevant references include [3–6, 19].

2 Disability Insurance

In the next sections we briefly describe a stochastic semi-Markov model for the development of disability inception and recovery rates and perform the corresponding statistical estimation. For more details see [1].

2.1 Disability Inception

Let $E_{x,t}$ denote the number of healthy individuals with age in $[x, x + 1)$ at the beginning of time period t, and let $D_{x,t}$ denote the number of individuals among $E_{x,t}$ with disability inception in the interval $[t, t + 1)$. In this section we model inception over time, $t = 0, 1, 2, \ldots$ and eventually estimate the underling parameters.

The Fig. 7 describes inception frequencies per 5-year age groups of females insured and a smoothed curve. This plot clearly shows that inception seems to be strongly time- and age-dependent. Below, we suggest a model of this behavior.

Assume $D_{x,t}$ is binomially distributed given $E_{x,t}$:

$$D_{x,t} \sim \text{Bin}(E_{x,t}, p_{x,t}) \tag{45}$$

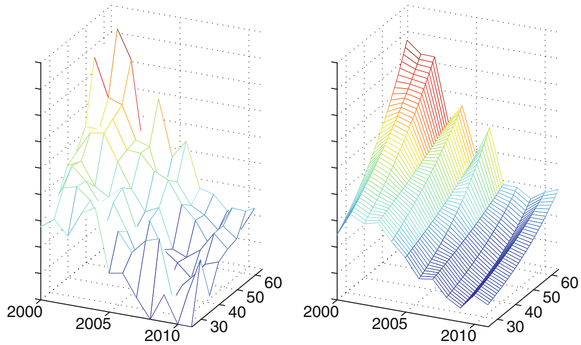


Fig. 7 *Left* Inception frequencies per 5-year age groups, females. *Right* Smoothed surface

where $p_{x,t}$ is the inception probability of an x-year-old. In order to reduce the dimensionality of the problem and achieve some level of smoothness, we use the logistic regression:

$$\operatorname{logit} p_{x,t} := \log\left(\frac{p_{x,t}}{1-p_{x,t}}\right) = \sum_{i=1}^{n} \nu_t^i \phi^i(x), \tag{46}$$

where $\phi^i(x)$ are age-dependent *basis functions*, and ν_t^i time-varying stochastic *risk factors* that we aim at estimating. Changing notation, $p_{\nu_t}(x) = p_{x,t}$, we invert the expression above, obtaining

$$p_{\nu_t}(x) = \frac{\exp\left(\sum_{i=1}^{n} \nu_t^i \phi^i(x)\right)}{1+\exp(\sum_{i=1}^{n} \nu_t^i \phi^i(x))}. \tag{47}$$

This guarantees that the probabilities $p_{\nu_t}(x) \in (0, 1)$.

Given historical values of $D_{x,t}$ and $E_{x,t}$, and a set of basis functions $\{\phi^i\}$, the log-likelihood function for yearly values of $\nu_t \in \mathbf{R}^n$ can be written

$$l(\nu_t) = \sum_{x\in X}\left[D_{x,t}\sum_{i=1}^{n} \nu_t^i \phi^i(x) - E_{x,t}\log\left(1+\exp\left\{\sum_{i=1}^{n} \nu_t^i \phi^i(x)\right\}\right)\right] + c_t. \tag{48}$$

If the basis functions are linearly independent it can be shown that $-l(\nu_t)$ is strictly convex. Thus it has a unique minimum. Minimizing $-l(\nu_t)$, using e.g. methods from numerical optimization, yields estimates of ν_t. The basis functions can be chosen by the user, according to some criteria. Desired properties of $p_{\nu_t}(\cdot)$, e.g. continuity or smoothness w.r.t. x, are achieved by choosing continuous or smooth $\phi^i(\cdot)$, by taking into account eventual population characteristics. Suitable choices of basis functions give the risk factors concrete interpretations. Alternatively, an optimal basis can be extracted from the data using functional principal component analysis. This approach yields better model fit, but harder to interpret results.

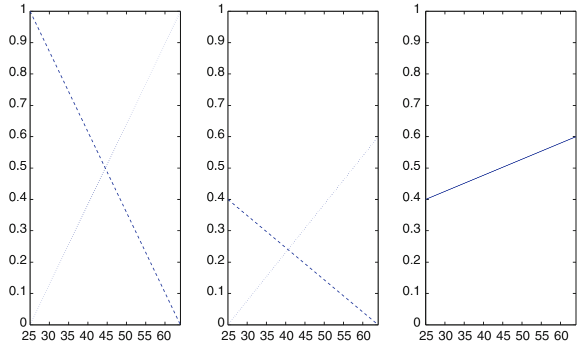


Fig. 8 *Left* Two basis functions. *Centre* Basis functions scaled with risk factor values 0.4 and 0.6. *Right* The resulting linear combination. *Note* $\phi^1(25) = \phi^2(64) = 1$, and $\phi^1(64) = \phi^2(25) = 0$

Consider the simple model

$$\operatorname{logit} p_{\nu_t}(x) = \nu_t^1 \phi^1(x) + \nu_t^2 \phi^2(x),$$

where the basis functions are linear on $x \in [25, 64]$:

$$\phi^1(x) = \frac{64 - x}{39} \quad , \quad \phi^2(x) = \frac{x - 25}{39}. \tag{49}$$

A linear combination of ϕ^1 and ϕ^2 is then also linear Fig. 8.
Under this model, the logistic inception probability of a 25-year old is given by

$$\operatorname{logit} p_{\nu_t}(25) = \nu_t^1 \phi^1(25) + \nu_t^2 \phi^2(25) = \nu_t^1.$$

Similarly, for a 64-year old we have $\operatorname{logit} p_{\nu_t}(64) = \nu_t^2$. An x-year old can be seen as a convex combination of a 25-year old and a 64-year old. Inception for the population is fully described by only ν_t^1 and ν_t^2.

2.2 Recovery from Disability

Recovery from disability is slightly more complicated. The probability of recovering from illness depends on the amount of time spent in the 'ill' state. This is known as the semi-Markov property. We extend the disability inception model above to the semi-Markov case, and apply it to recovery modeling.

Let $E_{x,d,t}$ denote the number of individuals with disability inception age in $[x, x + 1)$ and disability duration d at some point in the time period $[t, t + 1)$. Let $R_{x,d,t}$ denote the number of individuals among $E_{x,d,t}$ that recover during $[d, d + \Delta d)$ and $[t, t + 1)$. Assume $R_{x,d,t}$ is binomially distributed given $E_{x,d,t}$:

$$R_{x,d,t} \sim \operatorname{Bin}(E_{x,d,t}, p_{x,d,t}), \tag{50}$$

where, $p_{x,d,t}$ is the probability that an individual, with disability inception age in $[x, x+1)$ and disability duration d at some point in $[t, t+1)$, recovers during $[d, d+\Delta d)$.

We propose the following logistic regression model:

$$\text{logit} p_{\nu_t}(x, d) = \sum_{i=1}^{n} \phi^i(x) \sum_{j=1}^{k} \nu_t^{i,j} \psi^j(d), \tag{51}$$

where ϕ^i and ψ^j, are age and duration dependent basis functions, respectively, and $\nu_t^{i,j}$ are stochastic risk factors. This is the inception model Eq. (46), extended with one dimension. The likelihood also has the same structure as before. It is strict convexity if each of the sets of functions $\{\phi^i\}$ and $\{\psi^i\}$ are linearly independent. Again, we estimate ν_t using numerical optimization Fig. 9.

Consider the simple model

$$\text{logit}\, p_{\nu_t}(x, d) = \phi^1(x) \sum_{j=1}^{3} \nu_t^{1,j} \psi^j(d) + \phi^2(x) \sum_{j=1}^{3} \nu_t^{2,j} \psi^j(d)$$

where ϕ and ψ are given by:

$$\phi^1(x) = \tfrac{64-x}{39},\, \phi^2(x) = \tfrac{x-25}{39},\, \psi^1(d) = 1,\, \psi^2(d) = d,\, \psi^3(d) = \sqrt{d}.$$

Hence, the recovery probabilities for a 25-year old are given by

$$\text{logit}\, p_{\nu_t}(25, \cdot) = \phi^1(25) \sum_{j=1}^{3} \nu_t^{1,j} \psi^j(\cdot) + \phi^2(25) \sum_{j=1}^{3} \nu_t^{2,j} \psi^j(\cdot) = \sum_{j=1}^{3} \nu_t^{1,j} \psi^j(\cdot),$$

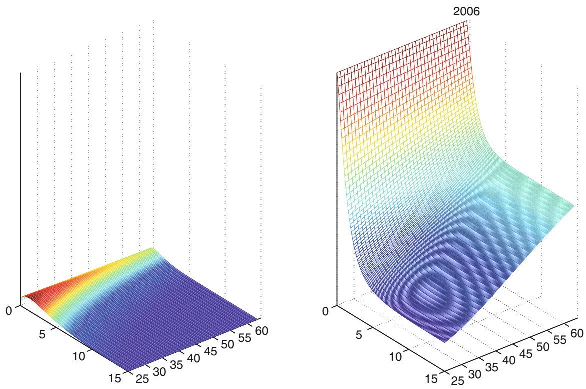


Fig. 9 *Left* Conditional recovery probabilities. *Right* Recovery surface, females, calendar year 2006

determined by $\nu_t^{1,1}$, $\nu_t^{1,2}$, $\nu_t^{1,3}$. Similarly, the recovery probabilities for a 64-year old determined by $\nu_t^{2,1}$, $\nu_t^{2,2}$, $\nu_t^{2,3}$. An x-year old can be seen as a convex combination of a 25-year old and a 64-year old. These considerations allow us to fully compute the probability that illness lasts longer than a given period. Let an x-year old's illness duration be the r.v. D. The probability that the illness lasts longer than d years is given by

$$\lambda(x, d) = P_{\nu_t}(D > d) = \prod_{n=0}^{d/\Delta d - 1} (1 - p_{\nu_t}(x, n\Delta d)).$$

This is analogous to survival curves.

References

1. Aro, H., Djehiche, B., Löfdahl, B.: Stochastic modelling of disability insurance in a multi-period framework. Scand. Actuar. J. (2013). http://dx.doi.org/10.1080/03461238.2013.779594
2. Brouhns, N., Denuit, D., Vermunt, J.K.: A Poisson log-bilinear regression approach to the construction of projected lifetables. Insurance: Mathematics and Economics **31**(3), 373–393 (2002)
3. Biffis, E.: Affine processes for dynamic mortality and actuarial valuation. Insur. Math. Econom. **37**(3), 443–468 (2005)
4. Biffis, E., Denuit, M.: Lee-Carter goes risk-neutral: An application to the Italian annuity market, pp. 33–53. Giornale dell'Istituto Italiano degli Attuari, LXIX (2006)
5. Biffis, E., Denuit, M., Devolder, P.: Stochastic mortality under measure changes. Scand. Actua. J. **4**, 284–311 (2010)
6. Biffis, E., Millosovitch, P.: A bidimensional approach to mortality risk. Decision in Economics and Finance **29**(2), 71–94 (2006)
7. Cox, S.H., Lin, Y., Pedersen, H.: Mortality risk modeling: Applications to insurance securitization. Insurance: Mathematics and Economics **46**(1), 242–253 (2010)
8. Fahrenwaldt, M.A.: Sensitivity of life insurance reserves via Markov semigroups (preprint) (2013)
9. Haberman, S., Renshaw, A.: On age-period-cohort parametric mortality rate projections. Insurance Math. Econom. **45**(2), 255–270 (2009)
10. Lee, R.D., Carter, L.: Modelling and forecasting the time series of US mortality. Journal of the American Statistical Association **87**, 659–671 (1992)
11. Norberg, R.: Reserves in life and pension insurance. Scand. Actuar. J. **1991**(1), 324 (1991)
12. Norberg, R.: Identities for present values of life insurance. Scand. Actuar. J. pp. 100–106 (1993)
13. Norberg, R.: A theory of bonus prognoses in life insurance. Finance Stoch. **3**(4), 373–390 (1999)
14. Norberg, R.: On bonus and bonus prognoses in life insurance. Scand. Actuar. J. **2001**(2), 126–147 (2001)
15. Norberg, R.: Basic life insurance mathematics. University of Copenhagen, Lecture Notes (2002)
16. Persson, S.V.: Stochastic interest rate in life insurance: The principle of equivalence revisited. Scandinavian Actuarial Journal **1998**(2), 97–112 (1998)
17. Samuelsson, E.: Mortality among Swedish insured. Scan. Actua. J. **2008**(2–3), 184–199 (2008). www.tandfonline.com
18. Steffensen, M.: Surplus-liked life insurance. Scandinavian Actuarial Journal **1**, 1–22 (2006)
19. Tappe, S., Weber, S.: Stochastic mortality models: An infinite-dimensional approach (preprint) (2013)