LIFE ANUITIES WITH STOCHASTIC SURVIVAL PROBABILITIES: A REVIEW

Michel Denuit
Institut de Statistique & Institut des Sciences Actuarielles
Université Catholique de Louvain
Louvain-la-Neuve, Belgium
michel.denuit@uclouvain.be

August 27, 2007

Abstract

An insurance company selling life annuities has to use projected life tables to describe the survival of policyholders. Such life tables are generated by stochastic processes governing the future path of mortality. To fix the ideas, the standard Lee-Carter model for mortality projection will be adopted here. In that context, the paper purposes to examine the consequences of working with random survival probabilities. Various stochastic inequalities are derived, showing that the risk borne by the annuity provider is increased compared to the classical independent case. Moreover, the type of dependence existing between the insured life times is carefully examined. The paper also deals with the computation of ruin probabilities and large portfolio approximations

Key words and phrases: Lee-Carter model, mortality projection, association, MTP2, stochastic orders, ruin probability, prudential life table, securitization.

1 Introduction and motivation

During the 20th century, the human mortality globally declined. These mortality improvements pose a challenge for the planning of public retirement systems as well as for the private life annuities business. When long-term living benefits are concerned, the calculation of expected present values (for pricing or reserving) requires an appropriate mortality projection in order to avoid underestimation of future costs. Actuaries have therefore to resort to life tables including a forecast of the future trends of mortality (the so-called projected life tables).

The composition of the population in the industrialized countries will change significantly in coming decades as the decline in fertility rates following the baby boom, coupled with increasing longevity, leads to an older population. This demographic shift will likely have significant implications for social security. In many industrialized countries, policy makers now wonder whether the programs should continue to be financed solely through the current pay-as-you-go structure or whether personal accounts or other innovations should be introduced. Possible social security reforms and the shift from defined benefit to defined contribution private pension plans are expected to increase demand for individual annuity products in the future. As demand for individual annuities increases, the need for risk management of the potential mortality improvements increases as the insurance companies write new individual annuity business.

Different approaches for building projected life tables have been developed so far; see e.g., Lee (2000), Pitacco (2004), Wong-Fupuy & Haberman (2004), Booth (2006), and Pitacco, Denuit, Haberman & Olivieri (2007) for a review. The most widely used model is the one proposed by Lee & Carter (1992). These authors suggested to describe the secular change in mortality as a function of a single time index. Precisely, the force of mortality at age x in calendar year t is of the form $\exp(\alpha_x + \beta_x\kappa_t)$, where the time-varying parameter κ_t reflects the general level of mortality and follows a stochastic process (usually, a random walk with drift). Section 2 offers a review of the Lee-Carter approach to mortality forecasting.

Section 3 is devoted to the type of dependence induced by the Lee-Carter model. The future lifetimes are all influenced by the same time index κ_t. Since the future path of this index is unknown and modelled as a stochastic process, the policyholders' lifetimes become dependent on each other. This is similar to the correlation arising in credit risk portfolios, where the dependence is induced by the common economic conditions. When the Lee-Carter model applies, life annuity present values are correlated random variables, contrarily to the standard actuarial assumptions: they are now conditionally independent given the future path of mortality. Consequently, the risk does not disappear as the size of the portfolio increases: there always remains some systematic risk, that cannot be diversified, whatever the number of policies.

Section 4 examines ruin probabilities for annuity portfolios. A simple formula to compute the ruin probability for life annuities portfolios is proposed, and we explain how the actual computations can be performed with the help of De Pril-Panjer recursive algorithms. We also use these results to determine a prudential life table that could be implemented by regulatory authorities, to protect annuitants against possible bankruptcy of annuity providers.

Section 5 is devoted to large portfolio approximations. In large portfolios, the risk borne by annuity providers (insurance companies or pension funds) is basically driven by the randomness in the future mortality rates. In Section 5, we provide accurate approximations for the quantiles of the conditional present value of the obligations contracted by the annuity provider, given the future mortality rates. Specifically, we derive approximations for the quantiles of the life annuity conditional expected present value. Note that this is the non diversifiable part of the insurance risk, that remains with the annuity provider whatever the size of the portfolio. When actuaries work with projected life tables, a new risk thus emerges: the risk that the mortality projections turn out to be erroneous, and that the annuitants live longer than predicted by the projected life tables. This is the so-called longevity risk, which has become a major issue for insurers and pension funds. In that respect, securitization of this risk could offer great opportunities for hedging, as discussed in the conclusion.

The results summarized in this review paper come from the following articles. Section 2 is based on Brouhns, Denuit & Vermunt (2002a,b), Brouhns, Denuit & Van Keilegom (2005), Cossette, Delwarde, Denuit, Guillot & Marceau (2007), Czado, Delwarde & Denuit (2005), Delwarde, Denuit & Eilers (2007), and Delwarde, Denuit & Partrat (2007). A general account of the statistical aspects of the Lee-Carter model for mortality forecasting in given in Denuit (2007c). Section 3 is based on Denuit & Frostig (2007a,c). Section 4 is based on Denuit & Frostig (2007b) and Frostig & Denuit (2007). Section 5 is based on Denuit & Dhaene (2007) and Denuit (2007a,b). We refer to these papers for extensive numerical illustrations, whereas here, we focus on methodological aspects. Many results presented in this paper call upon stochastic orders and dependence notions. We refer the reader to Denuit, Dhaene, Goovaerts & Kaas (2005) for an introduction to these topics, with applications in actuarial science.

2 Log-bilinear model for mortality forecasting

2.1 Notation

We analyze the changes in mortality as a function of both age x and calendar time t. Henceforth,

- $T_x(t)$ is the remaining life time of an individual aged x on January the first of year t; this individual will die at age $x + T_x(t)$ in year $t + T_x(t)$.
- $q_x(t)$ is the probability that an x-aged individual in calendar year t dies before reaching age $x + 1$, i.e. $q_x(t) = \Pr[T_x(t) \leq 1]$.
- $p_x(t) = 1 - q_x(t)$ is the probability that an x-aged individual in calendar year t reaches age $x + 1$, i.e. $p_x(t) = \Pr[T_x(t) > 1]$.
- $\mu_x(t)$ is the force of mortality at age x during calendar year t, that is,

$$\mu_x(t) = \lim_{\Delta \to 0} \frac{\Pr[x < T_0(t-x) \leq x + \Delta | T_0(t-x) > x]}{\Delta}.$$

Henceforth, we assume that the age-specific forces of mortality are constant within bands of age and time, but allowed to vary from one band to the next. Specifically, given any integer age x and calendar year t, it is supposed that

$$\mu_{x+\xi}(t+\tau) = \mu_x(t) \text{ for } 0 \leq \xi, \tau < 1. \tag{2.1}$$

Under (2.1), we have for integer age x and calendar year t that

$$p_x(t) = \exp(-\mu_x(t)). \tag{2.2}$$

2.2 Log-bilinear form for the forces of mortality

Actuaries have traditionally been calculating premiums and reserves using a deterministic mortality intensity. Here, the forces of mortality will be described by a stochastic process. Specifically, these forces are assumed to be of the form

$$\mu_x(t|\boldsymbol{\kappa}) = \exp(\alpha_x + \beta_x \kappa_t) \tag{2.3}$$

where the parameters β_x and κ_t are subject to constraints ensuring model identification (typically, the β_x's sum to 1 and the κ_t's are centered). Interpretation of the parameters involved in model (2.3) is quite simple. The value of α_x corresponds to the average of $\ln \mu_x(t)$ over time t so that $\exp \alpha_x$ is the general shape of the mortality schedule. The actual forces of mortality change according to an overall mortality index κ_t modulated by an age response β_x. The shape of the β_x profile tells which rates decline rapidly and which slowly over time in response of change in κ_t. The time factor κ_t is intrinsically viewed as a stochastic process and Box-Jenkins techniques are then used to model and forecast κ_t. Under the assumption (2.1) for the $\mu_x(t|\boldsymbol{\kappa})$'s, we have

$$p_x(t|\boldsymbol{\kappa}) = 1 - q_x(t|\boldsymbol{\kappa}) = \exp\big(-\mu_x(t|\boldsymbol{\kappa})\big).$$

Remark 2.1. Throughout the paper, we assume that all the β_x's are positive. This is typically the case when the parameters are estimated from empirical mortality data. Positive β_x's imply that the death rates are increasing in the κ_t's. If the κ_t's increase then the life lengths tend to shorten. To make this more precise, we need the hazard rate order that is defined as follows: if two non-negative random variables X and Y with hazard rate functions r and q, respectively, are such that $r(t) \geq q(t)$ for all $t \geq 0$, then X is said to be smaller than Y in the hazard rate order, which is denoted as $X \preceq_{\text{hr}} Y$. Recall that the hazard rate function is the ratio of the probability density function to the survival function. In life insurance mathematics, the hazard rate function is usually termed as force of mortality.

Denoting as F and G the respective distribution functions of X and Y, it is easy to verify that $X \preceq_{\text{hr}} Y$ holds if, and only if,

$$\frac{1-F(t+s)}{1-F(t)} \leq \frac{1-G(t+s)}{1-G(t)} \quad \text{for all} \quad s \geq 0 \quad \text{and all} \quad t. \tag{2.4}$$

When all the β_x's are positive, it is easy to see from (2.3) that the remaining lifetimes decrease in the $\preceq_{\text{hr}}$-sense as the κ_t's increase.

2.3 Survival probability

For any non-negative integer d, let ${}_dp_{x_0}(\boldsymbol{\kappa})$ be the d-year survival probability for an individual aged x_0 in year t_0 given the trajectory of the time index $\boldsymbol{\kappa}$. More specifically,

$$
{}_dp_{x_0}(\boldsymbol{\kappa}) = \Pr[T_{x_0}(t_0) > d|\boldsymbol{\kappa}],
$$

where $\boldsymbol{\kappa}$ stands for the random vector $(\kappa_{t_0}, \ldots, \kappa_{t_0+\omega-x_0})$ with ω being the ultimate age of the life table (precisely, ω is such that $p_\omega(t) = 0$ for every year t). Note that we drop the explicit reference to the calendar year t_0; by convention, we work with the cohort aged x_0 in year t_0, and follow the survival of this particular group of individuals.

For integer d, the d-year survival probability writes

$$
\begin{aligned}
{}_dp_{x_0}(\boldsymbol{\kappa}) &= \prod_{j=0}^{d-1} p_{x_0+j}(t_0 + j|\boldsymbol{\kappa}) \\
&= \exp\left(-\sum_{j=0}^{d-1} \mu_{x_0+j}(t_0 + j|\boldsymbol{\kappa})\right) \\
&= \exp\left(-\sum_{j=0}^{d-1} \exp\left(\alpha_{x_0+j} + \beta_{x_0+j}\kappa_{t_0+j}\right)\right).
\end{aligned}
$$

Note that ${}_dp_{x_0}(\boldsymbol{\kappa})$ is a random variable since it involves the random κ_{t_0+j}'s. In general, for any positive ξ, we have that

$$
{}_\xi p_{x_0}(\boldsymbol{\kappa}) = \Pr[T_{x_0}(t_0) > \xi|\boldsymbol{\kappa}] = \begin{cases} \exp\big(-\xi\mu_{x_0}(t_0|\boldsymbol{\kappa})\big) \text{ if } \xi \leq 1, \\ \exp\big(-(\xi - \lfloor\xi\rfloor)\mu_{x_0+\lfloor\xi\rfloor}(t_0 + \lfloor\xi\rfloor|\boldsymbol{\kappa})\big) \\ \qquad \prod_{k=0}^{\lfloor\xi\rfloor-1} \exp\big(-\mu_{x_0+k}(t_0 + k|\boldsymbol{\kappa})\big) \text{ if } \xi > 1, \end{cases} \tag{2.5}
$$

where $\lfloor\xi\rfloor$ is the integer part of the positive real number ξ.

2.4 Parameter estimation

The main statistical tool of Lee & Carter (1992) is least-squares estimation via singular value decomposition of the matrix of the log age-specific observed forces of mortality. This implicitly means that the errors are assumed to be homoskedastic, which is quite unrealistic: the logarithm of the observed force of mortality is much more variable at older ages than at younger ages because of the much smaller absolute number of deaths at older ages. Brouhns, Denuit & Vermunt (2002a,b) and Renshaw & Haberman (2003) implemented an alternative approach to mortality forecasting based on heteroskedastic Poisson error structures. They replaced ordinary least-squares regression with Poisson regression for the death counts.

The parameters α_x, β_x and κ_t are estimated by maximizing a Poisson log-likelihood. Specifically, let D_{xt} be the number of deaths recorded at age x during year t. These deaths originate from an exposure-to-risk ETR_{xt} (measured in individual-year). The estimated parameters maximize

$$L(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\kappa}) \;=\; \sum_t \sum_x \Big(D_{xt}(\alpha_x+\beta_x\kappa_t) - \mathrm{ETR}_{xt}\exp(\alpha_x+\beta_x\kappa_t)\Big) + \text{constant}.$$

The maximum likelihood estimates of α_x, β_x and κ_t are found with the help of a Newton-Raphson algorithm. Details of the fitting procedure can be found in Brouhns, Denuit & Vermunt (2002a).

The resulting estimated β_x's and κ_t's often exhibit an irregular pattern. This is problematic as these random variations propagate to the price list and reserves. Therefore, some smoothing is usually needed. Bayesian formulations assume some sort of smoothness of age and period effects in order to improve estimation and facilitate prediction. In order to implement this idea, Czado, Delwarde & Denuit (2005) resorted to a Bayesian model in which the prior portion imposes smoothness by relating the underlying mortality rates to each other over the Lexis plane. As a consequence, the rate estimate in each age-year square "borrows strength" from information in adjacent squares.

An alternative to the Bayesian approach is to penalize the Poisson log-likelihood $L(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\kappa})$. Specifically, Delwarde, Denuit & Eilers (2007) suggested to replace the Poisson log-likelihood with a penalized version of it $L(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\kappa}) - \frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{P}_\beta\boldsymbol{\beta}$ where

$$\boldsymbol{P}_\beta = \pi_\beta \boldsymbol{D}^T\boldsymbol{D} \text{ with } \boldsymbol{D} = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}$$

and π_β is the smoothing parameter. The term $\boldsymbol{\beta}^T\boldsymbol{P}_\beta\boldsymbol{\beta}$ penalizes irregular β_x's. The objective function can therefore be seen as a compromise between goodness-of-fit (first term $L(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\kappa})$) and smoothness of the β_x's (second term $\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{P}_\beta\boldsymbol{\beta}$). The penalty involves the sum of the squared second order differences of the β_x's, that is, the sum of the square of the second differences $\beta_{x+2} - 2\beta_{x+1} + \beta_x$. Second order differences penalize deviations from the linear trend. The trade off between fidelity to the data (governed by the sum of squared residuals) and smoothness (governed by the penalty term) is controlled by the smoothing parameters π_β. The larger the smoothing parameters the smoother the resulting fit. The choice of the smoothing parameters is crucial as we may obtain quite different fits by varying the smoothing parameters π_β. The choice of the optimal π_β is based on the observed data, using cross-validation.

The Poisson specification is not the only reasonable choice for mortality statistics. Cossette, Delwarde, Denuit, Guillot & Marceau (2007) proposed a Binomial regression model for estimating the parameters in logbilinear mortality projection models. The annual number D_{xt} of recorded deaths is then assumed to follow a Binomial distribution, with a death probability expressed in function of the force of mortality via $q_x(t|\boldsymbol{\kappa}) = 1 - \exp(-\mu_x(t|\boldsymbol{\kappa}))$. Delwarde, Denuit & Partrat (2007) suggested to take the D_{xt}'s Negative Binomial distributed to account for the heterogeneity with respect to mortality. See also Li, Hardy & Tan (2006).

2.5 Time index dynamics

To forecast, Lee & Carter (1992) assume that the α_x's and β_x's remain constant over time and forecast future values of κ_t using a standard univariate time series model. After testing several specifications, they found that a random walk with drift was the most appropriate model for their data. They made clear that other ARIMA models might be preferable for different data sets, but in practice the random walk with drift model for κ_t is used almost exclusively. According to this model, the κ_t's obey to

$$\kappa_t = \kappa_{t-1} + d + \xi_t, \tag{2.6}$$

where the ξ_t's are independent and Normally distributed with mean 0 and variance σ^2. In (2.6), d is known as the drift parameter. We will retain the model (2.6) throughout this paper. The next section explains why the dynamics (2.6) plays such a prominent role in the Lee-Carter approach.

Remark 2.2. It would be tempting to conclude that when the time index is described by a random walk with drift, increasing the volatility σ is dangerous for the annuity provider. Things are however more complicated, since the death rates are LogNormally distributed. Specifically, let us denote as $\preceq_{\text{cx}}$ the convex order (defined as $X \preceq_{\text{cx}} Y \Leftrightarrow \mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ for all the convex functions g such that the expectations exist) and as $\preceq_{\text{icx}}$ the increasing convex order (or stop-loss order, defined as $X \preceq_{\text{icx}} Y \Leftrightarrow \mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ for all the non-decreasing convex functions g such that the expectations exist). Switching from σ to $\sigma' > \sigma$ and denoting as κ_{t_0+k} and κ'_{t_0+k} the corresponding time indices, we have that $\kappa_{t_0+k} \preceq_{\text{cx}} \kappa'_{t_0+k}$ provided the drift parameter is left unchanged. This in turn gives $\mu_x(t_0+k|\boldsymbol{\kappa}) \preceq_{\text{icx}} \mu_x(t_0+k|\boldsymbol{\kappa}')$. Death rates become therefore "larger" and "more variable" when the volatility increases. If more variability causes trouble to the annuity provider, increasing the death rates is on the contrary beneficial. It is generally not possible to separate these two effects.

2.6 Selection of an optimal calibration period

Most actuarial studies base the projections on the mortality statistics relating to the years 1950 to present. The question then becomes why the post 1950 period better represents expectations for the future than does the post 1900 period. There are several justifications for the use of the second half of the 20th century. The pace of mortality decline was more even across all ages over the 1950-2000 period than over the 1900-2000 period. The quality of mortality data, particularly at older ages, for the 1900-1950 period is questionable. Infectious disease was an uncommon cause of death by 1950, while heart disease and cancer were the two most common causes, as they are today. This view seems to imply that the diseases affecting death rates from 1900 through 1950 are less applicable to expectations for the future than the dominant causes of death from 1950 through 2000.

Booth, Maindonald & Smith (2002) designed procedures for selecting an optimal calibration period which identified the longest period for which the estimated mortality index parameter κ_t was linear. Specifically, these authors seek to maximize the fit of the overall model by restricting the fitting period to maximize fit to the linearity assumption. The choice of the fitting period is based on the ratio of the mean deviances of the fit of the underlying Lee-Carter model to the overall linear fit. This ratio is computed by varying the starting year (but holding the jump-off year fixed) and the chosen fitting period is that for which the ratio is substantially smaller than for periods starting in previous years.

More specifically, Booth, Maindonald & Smith (2002) assume a priori that the trend in the adjusted $\widehat{\kappa}_t$'s is linear, based on the "universal pattern" of mortality decline identified by Tuljapurkar, Li & Boe (2000). When the $\widehat{\kappa}_t$'s depart from linearity, this assumption may be better met by appropriately restricting the fitting period. Restricting the fitting period to the longest recent period for which the adjusted $\widehat{\kappa}_t$'s do not deviate markedly from linearity has several advantages. Since systematic changes in the trend in $\widehat{\kappa}_t$ are avoided, the uncertainty in the forecast is reduced accordingly. Moreover, the β_x's are likely to better satisfy the assumption of time invariance. Finally, the estimate of the drift parameter more clearly reflects recent experience.

An ad-hoc procedure for selecting the optimal fitting period has been suggested in Denuit & Goderniaux (2005). The idea is to select this period in such a way that the series of the $\widehat{\kappa}_t$'s is best approximated by a straight line. To this end, the adjustment coefficient R^2 (which is the classical goodness-of-fit criterion in linear regression) is maximized (as a function of the number of observations included in the fit).

The restriction of the optimal fitting period clearly favors the random walk with drift model for the κ_t's. It also corresponds to a prudential approach, since the decline in the κ_t's usually tends to fasten after the the 1970's (where the optimal fitting period starts in most cases). In the numerical illustrations, the appropriate time series model can be selected on the basis of goodness-of-fit criteria including AIC and BIC.

2.7 Projecting the time index

We will assume in the remainder of this paper that the values $\kappa_1, \ldots, \kappa_{t_0}$ are known but that the κ_{t_0+k}'s, $k = 1, 2, \ldots$, are unknown and have to be projected from the random walk with drift model (2.6). To forecast the time index at time $t_0 + k$ with all data available up to t_0, we use

$$\kappa_{t_0+k} \quad = \quad \kappa_{t_0} + kd + \sum_{j=1}^{k} \xi_{t_0+j}. \tag{2.7}$$

The point estimate of the stochastic forecast is thus

$$\mathbb{E}[\kappa_{t_0+k}|\kappa_1, \ldots, \kappa_{t_0}] = \kappa_{t_0} + kd$$

which follows a straight line as a function of the forecast horizon k, with slope d. In the empirical studies performed on populations of industrialized countries, we always get a negative value for the estimated drift parameter d so that the κ_{t_0+k}'s follow a downward trend. Combined with positive β_x's, this downward trend expresses the decrease in the forces of mortality observed in industrialized countries.

The conditional variance of the forecast is

$$\mathbb{V}[\kappa_{t_0+k}|\kappa_1, \ldots, \kappa_{t_0}] = k\sigma^2.$$

Therefore, the conditional standard errors for the forecast increase with the square root of the distance to the forecast horizon k. Now, the covariance structure of the κ_{t_0+k}'s is given by

$$\mathbb{C}[\kappa_{t_0+k_1}, \kappa_{t_0+k_2}] = \sigma^2 \min\{k_1, k_2\}.$$

In this paper, we consider a group of n individuals, aged x_0 at time t_0. The random vector $\boldsymbol{\kappa}^T = (\kappa_{t_0+1}, \ldots, \kappa_{t_0+\omega-x_0})$ is Multivariate Normal with mean

$$\boldsymbol{m}^T = (\kappa_{t_0} + \theta, \ldots, \kappa_{t_0+\omega-x_0} + (\omega - x_0)\theta) \tag{2.8}$$

and variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & 2\sigma^2 & \cdots & 2\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2 & 2\sigma^2 & \cdots & (\omega - x_0)\sigma^2 \end{pmatrix}. \tag{2.9}$$

2.8 Prediction intervals and bootstrapping

Of course, the projection of the mortality itself is affected by uncertainty. Confidence intervals (for annuities and life expectancies) can be obtained by ignoring all the errors except those in forecasting the mortality index. According to Appendix B of Lee & Carter (1992), these errors dominate the others for annuities and expected remaining lifetimes. Because of the importance of appropriate measures of uncertainty in an actuarial context, Brouhns, Denuit & Vermunt (2002b) and Brouhns, Denuit & Vankeilegom (2005) derived confidence intervals taking into account all the sources of variability. The nonlinear nature of the quantities of interest makes an analytical approach not tractable and therefore Monte-Carlo simulation (or parametric bootstrap) as well as nonparametric bootstrap procedures are needed.

3 Type of dependence induced by the Lee-Carter model

3.1 Association of the insured lifetimes in the Lee-Carter framework

The concept of dependence called association has been introduced by Esary, Proschan & Walkup (1967). It is defined as follows. Random variables $X_1, X_2, \ldots, X_n$ (or the random vector $\boldsymbol{X}$) are said to be associated when

$$\mathbb{C}\left[\Psi_1(X_1, X_2, \cdots, X_n), \Psi_2(X_1, X_2, \cdots, X_n)\right] \geq 0 \tag{3.1}$$

for all non-decreasing functions Ψ_1 and Ψ_2 for which the covariances exist. Association has been first considered in actuarial science by Norberg (1989) who used it in order to investigate some alternatives to the independence assumption for multilife statuses in life insurance, as well as to quantify the consequences of a possible dependence on the amounts of premium relating to multilife insurance contracts. The intuitive meaning of association seems rather unclear. However, implicit in a conclusion that a set of random variables is associated is a wealth of inequalities, often of direct use in various problems.

It may appear impossible to check the condition (3.1) of association directly given a distribution function $F_{\boldsymbol{X}}$ for $\boldsymbol{X}$. Where association of a random vector can be established, it is usually done by making use of a stochastic representation of $\boldsymbol{X}$ or of the following properties (henceforth referred to as P1 to P5) that can be found in Esary, Proschan & Walkup (1967):

P1 If $\boldsymbol{X}$ is associated then any subset $(X_{i_1}, X_{i_2}, \ldots, X_{i_k})$ of $\boldsymbol{X}$ is associated.

P2 If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are associated and mutually independent then $(\boldsymbol{X}, \boldsymbol{Y})$ is associated.

P3 If $\boldsymbol{X}$ is associated and $\Psi_1, \Psi_2, \cdots, \Psi_k : \mathbb{R}^n \to \mathbb{R}$ are non-decreasing functions, then $\Psi_1(\boldsymbol{X})$, $\Psi_2(\boldsymbol{X})$, $\cdots$, $\Psi_k(\boldsymbol{X})$ are associated.

P4 Independent random variables $X_1, \ldots, X_n$ are associated (that is, inequality (3.1) is satisfied with independent random variables $X_1, \ldots, X_n$).

P5 Let $X_1, X_2, \ldots, X_n$ be independent random variables. Let us define the partial sums

$$S_i = \sum_{j=1}^{i} X_j, \;\; i = 1, 2, \ldots, n.$$

Then, considering P3-P4, the vector of the partial sums $(S_1, S_2, \ldots, S_n)$ is associated

The next example shows that the κ_t's obeying to a random walk with drift are associated.

Example 3.1. As mentioned above, we will assume in the remainder of this paper that the values $\kappa_1, \ldots, \kappa_{t_0}$ are known but that the κ_{t_0+k}'s, $k = 1, 2, \ldots$, are given by (2.7). Therefore, conditional upon κ_{t_0}, the κ_{t_0+k}'s appear as partial sums and are therefore associated by P5.

Let us prove, under mild conditions, that the Lee-Carter remaining lifetimes are associated. To this end, we need to recall the definition of stochastic dominance $\preceq_{\mathrm{d}}$. Given two random variables X and Y, $X \preceq_{\mathrm{d}} Y \Leftrightarrow \mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ for all the non-decreasing functions g, provided the expectations exist. It is easy to see that the inequality

$$\Pr[T_i > t | \boldsymbol{\kappa} = (k_1, \ldots, k_{\omega - x_0})] \geq \Pr[T_i > t | \boldsymbol{\kappa} = (k'_1, \ldots, k'_{\omega - x_0})]$$

holds for all t whenever $\boldsymbol{k} \leq \boldsymbol{k}'$ componentwise. This means that the T_i's decrease in the $\preceq_{\mathrm{d}}$-sense as $\boldsymbol{\kappa}$ increases (in fact, they even decrease in the $\preceq_{\mathrm{hr}}$-sense, which is stronger than $\preceq_{\mathrm{d}}$; see Remark 2.1). Considering the analysis conducted in Jogdeo (1978), we are now in a position to state the following result.

Property 3.2. *Let $T_1, \ldots, T_n$ be the remaining life times of n individuals aged x_0 at time t_0, distributed as $T_{x_0}(t_0)$, with common survival function given in* (2.5)*. If $\beta_x \geq 0$ for all the ages x, the lifetimes $T_1, \ldots, T_n$ are associated provided $\boldsymbol{\kappa}$ is associated.*

Note that the result stated in Property 3.2 also holds if $\beta_x \leq 0$ for all x; what really matters is that all the β_x's have same sign. In all the situations encountered in practice, the estimated β_x's have the same sign (usually positive) and $\boldsymbol{\kappa}$ is associated, so that the above result is very general. The association of the κ_{t_0+k}'s is then transmitted to the present values of insurance benefits, as shown in the next result. Henceforth, we denote as $v(s,t)$, $s < t$ the (deterministic) present value at time s of one monetary unit paid at time t. Obviously, $v(s,s) = 1$ whatever s.

Corollary 3.3. *Let $T_1, \ldots, T_n$ be as described in Property 3.2. If $\beta_x \geq 0$ for all x and if $\boldsymbol{\kappa}$ is associated then*

(i) the present values of pure endowments are associated, that is, the random variables

$$v(0,d)\mathbb{I}[T_1 > d], \ldots, v(0,d)\mathbb{I}[T_n > d]$$

are associated, where the indicator $\mathbb{I}[T_j > d] = 1$ if policyholder j survives up to time $t_0 + d$, and 0 otherwise;

(ii) the present values of life annuities are associated, that is, the random variables

$$\sum_{k=1}^{\lfloor T_1 \rfloor} v(0,k), \ldots, \sum_{k=1}^{\lfloor T_n \rfloor} v(0,k)$$

are associated, with the convention that the empty sum is zero;

(ii) the present values of d-year-deferred life annuities are associated, that is, the random variables

$$\mathbb{I}[T_1 > d] \sum_{k=d}^{\lfloor T_1 \rfloor} v(0,k), \ldots, \mathbb{I}[T_n > d] \sum_{k=d}^{\lfloor T_n \rfloor} v(0,k)$$

are associated.

These results are easily extended to policyholders with different ages $x_1, \ldots, x_n$ at time t_0.

3.2 Impact of association

The next result is a direct consequence of Corollary 1(a) in Christophides & Vaggelatou (2004). It allows us to compare the riskiness of a portfolio with lifetimes obeying to the Lee-Carter model, to the riskiness of another portfolio made of independent life times with the same marginal distributions, in the case of equal benefits. Recall the definition of the convex order $\preceq_{\text{cx}}$ from Remark 2.2.

Property 3.4. *Let $T_1, \ldots, T_n$ be as described in Property 3.2. Let us denote as $T_1^\perp, \ldots, T_n^\perp$ the independent version of the Lee-Carter lifetimes $T_1, \ldots, T_n$, that is, the $T_i^\perp$'s are indepen-*

dent, each $T_i^\perp$ being distributed as $T_{x_0}(t_0)$. Then,

$$\begin{aligned}
\sum_{i=1}^{n} v(0,d)\mathbb{I}[T_i^\perp > d] &\preceq_{cx} \sum_{i=1}^{n} v(0,d)\mathbb{I}[T_i > d] \\
\sum_{i=1}^{n} \sum_{k=1}^{\lfloor T_i^\perp \rfloor} v(0,k) &\preceq_{cx} \sum_{i=1}^{n} \sum_{k=1}^{\lfloor T_i \rfloor} v(0,k) \\
\sum_{i=1}^{n} \mathbb{I}[T_i^\perp > d] \sum_{k=d}^{\lfloor T_i^\perp \rfloor} v(0,k) &\preceq_{cx} \sum_{i=1}^{n} \mathbb{I}[T_i > d] \sum_{k=d}^{\lfloor T_i \rfloor} v(0,k).
\end{aligned}$$

3.3 Heterogeneity in the sum insured

Denuit & Frostig (2006) and Frostig & Denuit (2006) examined the impact of the portfolio heterogeneity on its riskiness (as measured by risk measures in agreement with $\preceq_{\text{d}}$ or $\preceq_{\text{icx}}$). The main finding was that the riskiness of the portfolio often increases with the degree of heterogeneity in the sum insured. The same phenomenon occurs in the Lee-Carter framework, as shown next.

To measure the degree of heterogeneity, we need a tool called majorization. Majorization aims to formalize the idea that the components of a vector $\boldsymbol{x}$ are "less spread out" or "more nearly equal" than the components of $\boldsymbol{y}$. It is a partial order defined on the positive orthant $\mathbb{R}_+^n$. For a vector $\boldsymbol{x} \in \mathbb{R}_+^n$ we denote its elements ranked in ascending order as $x_{(1:n)} \geq x_{(2:n)} \geq \ldots \geq x_{(n:n)}$. Thus $x_{(n:n)}$ is the smallest of the x_i's, while $x_{(1:n)}$ is the largest. Considering $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}_+^n$, $\boldsymbol{y}$ is said to majorize $\boldsymbol{x}$, which is denoted as $\boldsymbol{x} \preceq_{\text{maj}} \boldsymbol{y}$, if

$$\sum_{i=k}^{n} x_{(i:n)} \geq \sum_{i=k}^{n} y_{(i:n)} \text{ for } k = 1, 2, \ldots, n, \text{ and } \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i.$$

The inequality $\boldsymbol{x} \preceq_{\text{maj}} \boldsymbol{y}$ implies that, for a fixed sum, the y_i's are more diverse than the x_i's. To illustrate this point, note that $\overline{\boldsymbol{y}} \preceq_{\text{maj}} \boldsymbol{y}$ always holds, with

$$\overline{\boldsymbol{y}} = (\overline{y}, \overline{y}, \ldots, \overline{y}) \quad \text{where} \quad \overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{3.2}$$

and

$$\boldsymbol{x} \preceq_{\text{maj}} \left(\sum_{i=1}^{n} x_i, 0, \ldots, 0 \right). \tag{3.3}$$

A function $\varphi : \mathbb{R}^n \to \mathbb{R}$ agreeing with $\preceq_{\text{maj}}$, that is, such that $\boldsymbol{x} \preceq_{\text{maj}} \boldsymbol{y} \Rightarrow \varphi(\boldsymbol{x}) \leq \varphi(\boldsymbol{y})$, is called a Schur-increasing function. This function is called Schur-decreasing if $\boldsymbol{x} \preceq_{\text{maj}} \boldsymbol{y} \Rightarrow \varphi(\boldsymbol{y}) \leq \varphi(\boldsymbol{x})$.

The next result follows from the following property, that can be found in Marshall & Olkin (1979): If $Z_1, \cdots, Z_n$ are exchangeable random variables and $g : \mathbb{R}^n \to \mathbb{R}$ is a symmetric convex function, then the function φ defined by

$$\varphi(a_1, \cdots, a_n) = \mathbb{E}\left[g\big(a_1 Z_1, \cdots, a_n Z_n\big) \right]$$

is also symmetric and convex. Thus, φ is Schur-increasing. It suffices to note that the random variables $T_1, \ldots, T_n$ in Property 3.2 are exchangeable, so that we are in a position to invoke this property.

Property 3.5. *Let $T_1, \ldots, T_n$ be as described in Property 3.2. Let s_i be the sum insured for policy i, $i = 1, \ldots, n$. Then,*

(i) the more the s_i's are dispersed, the more the sum of the present values of pure endowments is dangerous (in the $\preceq_{cx}$-sense), that is,

$$\begin{aligned}\boldsymbol{s} \preceq_{maj} \boldsymbol{r} \quad \Rightarrow \quad & s_1 v(0,d)\mathbb{I}[T_1 > d] + \ldots + s_n v(0,d)\mathbb{I}[T_n > d] \\ & \preceq_{cx} r_1 v(0,d)\mathbb{I}[T_1 > d] + \ldots + r_n v(0,d)\mathbb{I}[T_n > d].\end{aligned}$$

(ii) the more the s_i's are dispersed, the more the sum of the present values of life annuities is dangerous, that is,

$$\boldsymbol{s} \preceq_{maj} \boldsymbol{r} \Rightarrow s_1 \sum_{k=1}^{\lfloor T_1 \rfloor} v(0,k) + \ldots + s_n \sum_{k=1}^{\lfloor T_n \rfloor} v(0,k) \preceq_{cx} r_1 \sum_{k=1}^{\lfloor T_1 \rfloor} v(0,k) + \ldots + r_n \sum_{k=1}^{\lfloor T_n \rfloor} v(0,k).$$

(ii) the more the s_i's are dispersed, the more the sum of the present values of d-year-deferred life annuities is dangerous, that is,

$$\begin{aligned}\boldsymbol{s} \preceq_{maj} \boldsymbol{r} \quad \Rightarrow \quad & s_1 \mathbb{I}[T_1 > d] \sum_{k=d}^{\lfloor T_1 \rfloor} v(0,k) + \ldots + s_n \mathbb{I}[T_n > d] \sum_{k=d}^{\lfloor T_n \rfloor} v(0,k) \\ & \preceq_{cx} r_1 \mathbb{I}[T_1 > d] \sum_{k=d}^{\lfloor T_1 \rfloor} v(0,k) + \ldots + r_n \mathbb{I}[T_n > d] \sum_{k=d}^{\lfloor T_n \rfloor} v(0,k).\end{aligned}$$

3.4 Increasingness of $\boldsymbol{\kappa}$ in the timing of deaths

The Lee-Carter setting falls in the scope of Application 4.5 in Shaked & Spizzichino (1998). The lifetimes $T_1, \ldots, T_n$ are not only associated but also weakened by failure. This positive dependence notion is stronger than association and has proven its usefulness in reliability theory. Translated in an actuarial context, it basically states that the death of a policyholder decreases the survival probability for the others (more formally, the conditional residual lives of the survivors at any time $t \geq 0$ are $\preceq_{\mathrm{d}}$-ordered, given two up-to-time-t realizations that are identical except that in one realization there is a death at time t whereas there is no death at that time in the other realization). This illustrates the type of positive dependence existing between the Lee-Carter lifetimes.

Let us first establish that $\boldsymbol{\kappa}$ obeying to a random walk with drift model is strongly positively dependent. Recall that a random vector $\boldsymbol{X}$ with probability density function f is MTP_2 if f is an MTP_2 function. Defining for $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$ the lattice operators $\vee$ and $\wedge$ as

$$\boldsymbol{x} \vee \boldsymbol{y} = \big(\max\{x_1, y_1\}, \ldots, \max\{x_n, y_n\}\big)$$

and

$$\boldsymbol{x} \wedge \boldsymbol{y} = \big(\min\{x_1, y_1\}, \ldots, \min\{x_n, y_n\}\big)$$

and assuming that the support of $\boldsymbol{X}$ is a lattice (that is, if $\boldsymbol{x}$ and $\boldsymbol{y}$ are in the support of $\boldsymbol{X}$, then so are $\boldsymbol{x} \vee \boldsymbol{y}$ and $\boldsymbol{x} \wedge \boldsymbol{y}$), this means that $\boldsymbol{X}$ is MTP$_2$ if the inequality

$$f(\boldsymbol{x})f(\boldsymbol{y}) \leq f(\boldsymbol{x} \vee \boldsymbol{y})f(\boldsymbol{x} \wedge \boldsymbol{y}) \tag{3.4}$$

holds for any $\boldsymbol{x}$, $\boldsymbol{y}$ in the support of $\boldsymbol{X}$. In the case of a bivariate density function, MTP$_2$ reduces to the standard TP$_2$.

If the dynamics of the κ_t's is given by (2.6) then $\boldsymbol{\kappa}$ is multivariate Normal, with mean vector $\boldsymbol{m}$ given in (2.8) and variance-covariance matrix $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{A}$ given in (2.9), where $\boldsymbol{A}$ is a square matrix of dimension $\omega - x_0$ with element ij given by $\min\{i, j\}$. We know from Tong (1990, Theorem 4.3.2.) that in the multivariate Normal case, MTP$_2$ occurs if, and only if, all the off-diagonal components of the inverse of the variance-covariance matrix are non-positive. It can be checked that the inverse of the matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A}^{-1} = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix},$$

so that $\boldsymbol{\kappa}$ obeying to (2.6) is indeed MTP$_2$.

Let us consider n policyholders, with future lifetimes $T_1, \ldots, T_n$ as in (3.2). The T_i's are independent given $\boldsymbol{\kappa}$, and have forces of mortality (2.3). Given a random vector $\boldsymbol{S}$ and an event A, let us denote as $\mathcal{L}(\boldsymbol{S}|A)$ the conditional distribution of $\boldsymbol{S}$ given A. Let $\boldsymbol{e} = (1, \ldots, 1)^T$ where the length depends on the context. For $J = \{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$, let $\boldsymbol{\tau}_J$ denote $(\tau_{i_1}, \ldots, \tau_{i_k})$ and let the complement of J be denoted as $\overline{J}$. Let h_t denote a realization of the T_i's up to time t, that is, an event of the form

$$h_t = \{\boldsymbol{T}_J = \boldsymbol{\tau}_J,\ \ \boldsymbol{T}_{\overline{J}} > t\boldsymbol{e}\} \tag{3.5}$$

where $\boldsymbol{0} \leq \boldsymbol{\tau}_J \leq t\boldsymbol{e}$. In words, h_t means that the policyholders in J died before t, at times $\tau_{i_1}, \ldots, \tau_{i_k}$, respectively, and that the policyholders in $\overline{J}$ are still alive at time t. Now, let us fix J, $i \in \overline{J}$, $t \geq 0$, $\boldsymbol{\tau}_J \leq t\boldsymbol{e}$ such that $(\boldsymbol{\tau}_J, \tau_i)$ is in the support of $(\boldsymbol{T}_J, T_i)$ and consider

$$h'_t = \{\boldsymbol{T}_J = \boldsymbol{\tau}_J,\ \ T_i = t \text{ and } \boldsymbol{T}_{\overline{J}\setminus\{i\}} > t\boldsymbol{e}\}. \tag{3.6}$$

In words, the policyholders in J died at times $\tau_{i_1}, \ldots, \tau_{i_k}$ under both h_t and h'_t. Those in $\overline{J}$ are still alive at time t under h_t whereas policyholder i dies at time t and those in $\overline{J} \setminus \{i\}$ outlive him under h'_t. Intuitively speaking, h'_t is a worse survival history than h_t so that we expect that $\boldsymbol{\kappa}$ becomes "larger" under h'_t relative to h_t. To formalize these ideas, we need the likelihood ratio order. Given two random variables X and Y with respective probability density functions f_X and f_Y, X is said to be smaller than Y in the likelihood ratio order, denoted as $X \preceq_{\text{lr}} Y$, if

$$f_X(u)f_Y(v) \geq f_X(v)f_Y(u) \text{ for all } u \leq v \in \mathbb{R}. \tag{3.7}$$

Considering (3.7), a ranking in the $\preceq_{\mathrm{lr}}$-sense can be given the following nice interpretation. Provided X and Y are independent (which can be assumed without loss of generality), the left-hand side of (3.7) can be regarded as the likelihood of the event "X is small and Y is large" whereas the right-hand side of this relation reads "X is large and Y is small". Then, (3.7) expresses the fact that the latter event is less likely to occur than the first one. This order is stronger than stochastic dominance, i.e. $X \preceq_{\mathrm{lr}} Y \Rightarrow X \preceq_{\mathrm{d}} Y$.

The multivariate version of $\preceq_{\mathrm{lr}}$ is defined by extending (3.7) to joint densities. More precisely, given two n-dimensional vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, with probability density functions $f_{\boldsymbol{X}}$ and $f_{\boldsymbol{Y}}$, respectively, $\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the likelihood ratio order, written as $\boldsymbol{X} \preceq_{\mathrm{lr}} \boldsymbol{Y}$, if the inequality

$$f_{\boldsymbol{X}}(\boldsymbol{x})f_{\boldsymbol{Y}}(\boldsymbol{y}) \leq f_{\boldsymbol{X}}(\boldsymbol{x} \wedge \boldsymbol{y})f_{\boldsymbol{Y}}(\boldsymbol{x} \vee \boldsymbol{y}) \tag{3.8}$$

hods for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$. The inequality in (3.8) defining multivariate $\preceq_{\mathrm{lr}}$ can be interpreted as (3.7). We also see from (3.4) that $\boldsymbol{X}$ is MTP$_2$ if $\boldsymbol{X} \preceq_{\mathrm{lr}} \boldsymbol{X}$ holds true.

We are now in a position to formalize the idea that $\boldsymbol{\kappa}$ gets larger under h'_t relative to h_t.

Property 3.6. *Consider the events h_t and h'_t defined in* (3.5) *and* (3.6), *and assume that $\boldsymbol{\kappa}$ is MTP$_2$. Then, we have $\mathcal{L}(\boldsymbol{\kappa}|h_t) \preceq_{lr} \mathcal{L}(\boldsymbol{\kappa}|h'_t)$.*

Let us now consider another history of the form

$$h_s = \{\boldsymbol{T}_J = \boldsymbol{\tau}_J,\ \boldsymbol{T}_{\overline{J}} > s\boldsymbol{e}\} \text{ with } s > t. \tag{3.9}$$

In words, this means that we do not modify the ages at death of the policyholders in J who died before t, but under h_s the policyholders in $\overline{J}$ are known to live longer than under h_t. We then have the following result.

Property 3.7. *Consider the events h_t and h'_t defined in* (3.5) *and* (3.9), *and assume that $\boldsymbol{\kappa}$ is MTP$_2$. Then, we have $\mathcal{L}(\boldsymbol{\kappa}|h_s) \preceq_{lr} \mathcal{L}(\boldsymbol{\kappa}|h_t)$.*

3.5 Decreasingness of the remaining lifetimes in the timing of deaths

We have seen above that the association of $\boldsymbol{\kappa}$ is transmitted to $\boldsymbol{T}$, as well as to future death rates and survival probabilities. Here, we aim to establish similar relationships for $\boldsymbol{T}$, inherited from the MTP$_2$ character of $\boldsymbol{\kappa}$. Specifically, we establish that the remaining lifetimes become larger if the survival history is more favorable. To this end, we need a multivariate extension of $\preceq_{\mathrm{d}}$. In case we want to compare random vectors, an intuitively acceptable strategy consists in transforming those vectors into random variables using increasing mappings, and then to compare the resulting outcomes with univariate stochastic orderings. This yields the following definition for multivariate $\preceq_{\mathrm{d}}$: given two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, $\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the stochastic dominance, written as $\boldsymbol{X} \preceq_{\mathrm{d}} \boldsymbol{Y}$, if $\Psi(\boldsymbol{X}) \preceq_{\mathrm{d}} \Psi(\boldsymbol{Y})$ for every non-decreasing function $\Psi : \mathbb{R}^n \to \mathbb{R}$. It can be shown that, as in the univariate case, $\boldsymbol{X} \preceq_{\mathrm{lr}} \boldsymbol{Y} \Rightarrow \boldsymbol{X} \preceq_{\mathrm{d}} \boldsymbol{Y}$.

We are now in a position to state the following result.

Property 3.8. *Assume that $\boldsymbol{\kappa}$ is MTP$_2$ and consider the events h_t and h'_t defined in* (3.5) *and* (3.6)*. Then, we have*

$$\mathcal{L}(\boldsymbol{T}_{\overline{\mathcal{J}}\setminus\{i\}} - t\boldsymbol{e}|h'_t) \preceq_d \mathcal{L}(\boldsymbol{T}_{\overline{\mathcal{J}}\setminus\{i\}} - t\boldsymbol{e}|h_t).$$

Since the financial obligations of the insurance company increase in the remaining lifetimes in the context of life annuities, this result gives the conditions under which the payments to annuitants increase in the $\preceq_{\mathrm{d}}$-sense.

4 Ruin probabilities and applications

4.1 Distribution of the survival indicators

Let us consider life times $T_1, \ldots, T_n$ as in Property 3.2. As above, let $\mathbb{I}[T_i > h]$ be the indicator of the event $T_i > h$, that is, $\mathbb{I}[T_i > h] = 1$ if policyholder i is still alive at time h, and 0 otherwise. Then, $L_t = \sum_{i=1}^n \mathbb{I}[T_i > t]$ denotes the (random) number of survivors at time t among the n annuitants. Defining D_t as the number of deaths recorded in period $(t-1, t)$, that is, $D_t = \sum_{i=1}^n \mathbb{I}[t-1 < T_i \leq t]$, we obviously have that $L_t = L_{t-1} - D_t$, $t = 1, 2, \ldots$.

Clearly, $I_j = \mathbb{I}[T_j > d]$ is Bernoulli distributed with mean

$$\mathbb{E}[I_j] = \Pr[I_j = 1] = \mathbb{E}\Big[{}_dp_{x_0}(\boldsymbol{\kappa})\Big] = \nu_1.$$

The survival indicators $I_1, \ldots, I_n$ are non-decreasing functions of the associated lifetimes $T_1, \ldots, T_n$ and are therefore associated (by property P3 of association). The vector $(I_1, \ldots, I_n)$ is exchangeable, that is, the I_j's are identically distributed and $(I_1, \ldots, I_n) =_d (I_{\pi(1)}, \ldots, I_{\pi(n)})$ for any permutation π of $\{1, \ldots, n\}$. For a survey about exchangeable Bernoulli random variables, we refer the reader to Madsen (1993).

Let us denote as $\nu_1, \nu_2, \nu_3, \ldots$ the moments of ${}_dp_{x_0}(\boldsymbol{\kappa})$, that is, $\nu_k = \mathbb{E}[({}_dp_{x_0}(\boldsymbol{\kappa}))^k]$, $k = 1, 2, 3, \ldots$. Then,

$$\Pr[I_1 = 1, \ldots, I_k = 1] \;=\; \mathbb{E}\Big[\Big({}_dp_{x_0}(\boldsymbol{\kappa})\Big)^k\Big] = \nu_k.$$

The distribution of the number of survivors at time $t_0 + d$ is given by

$$\Pr\left[\sum_{i=1}^n I_i = k\right] \;=\; \sum_{i=0}^{n-k} (-1)^i \frac{n!}{i!k!(n-k-i)!} \nu_{k+i}, \;\; k = 0, 1, \ldots, n.$$

The moments $\nu_1, \ldots, \nu_n$ of ${}_dp_{x_0}(\boldsymbol{\kappa})$ thus determine the distribution of the number of survivors L_d.

Let us now examine the dependence between the I_j's, as measured by their covariance. Specifically,

$$\mathbb{C}[I_i, I_j] = \Pr[I_i = 1, I_j = 1] - \nu_1^2 = \mathbb{V}\Big[{}_dp_{x_0}(\boldsymbol{\kappa})\Big]$$

and the Pearson's correlation coefficient for the pair (I_i, I_j), $i \neq j$, is given by

$$r[I_i, I_j] = \frac{\mathbb{V}\Big[{}_d p_{x_0}(\boldsymbol{\kappa})\Big]}{\nu_1 - \nu_1^2} \geq 0.$$

We thus see that our ignorance about ${}_d p_{x_0}(\boldsymbol{\kappa})$ drives the dependence between the I_j's. The more ${}_d p_{x_0}(\boldsymbol{\kappa})$ is uncertain, and has thus a large variance, the more the I_j's are correlated. If ${}_d p_{x_0}(\boldsymbol{\kappa})$ is constant then $r[I_i, I_j] = 0$ and the I_j's are independent (since zero correlation is equivalent to independence in the Bernoulli case).

From Section 3, we expect some positive dependence between the I_j's. Intuitively speaking, the fact that $I_k = 1$ suggests that ${}_d p_{x_0}(\boldsymbol{\kappa})$ is large, and this in turn increases the probability for $I_j = 1$, $j \neq k$. To make this point clear, let us compute for $j \neq k$

$$\begin{aligned}
\Pr[I_j = 1 | I_k = 1] &= \frac{\Pr[I_j = 1 \text{ and } I_k = 1]}{\Pr[I_k = 1]} = \frac{\mathbb{E}[I_j I_k]}{\Pr[I_k = 1]} \\
&= \frac{\mathbb{C}[I_j, I_k] + \Pr[I_j = 1]\Pr[I_k = 1]}{\Pr[I_k = 1]} \\
&= \Pr[I_j = 1] + \frac{\mathbb{C}[I_j, I_k]}{\Pr[I_k = 1]} \\
&= \Pr[I_j = 1] + \frac{\mathbb{V}\Big[{}_d p_{x_0}(\boldsymbol{\kappa})\Big]}{\mathbb{E}\Big[{}_d p_{x_0}(\boldsymbol{\kappa})\Big]}.
\end{aligned}$$

The knowledge that policyholder k survives increases the probability that policyholder j survives.

4.2 General formula for the ruin probability

Let us consider an insurance portfolio comprising n life annuity contracts, issued to policyholders with remaining life times $T_1, \ldots, T_n$ obeying to the assumptions of Property 3.2. The insurance company pays € 1 at the end of each year, provided the annuitant is still alive at that time.

Let U_t be the surplus of the insurance company at time t. Starting with an initial reserve $U_0 = w$, the surplus obeys to the dynamics

$$U_t = U_{t-1}(1 + r_t) - L_t, \quad t = 1, 2, \ldots, \tag{4.1}$$

where $L_0 = n$ by convention and r_t is the interest rate earned on the reserve during period t. The ruin probability at horizon k is defined as

$$\psi_k(w) = \Pr[U_t < 0 \text{ for some } t = 1, \ldots, k | U_0 = w];$$

$1 - \psi_k(w)$ measures the financial strength of the insurance company at horizon k, for an initial reserve amounting to w. The regulator wants that the insurer has enough capital w at its disposal to ensure $1 - \psi_k(w)$ be as large as possible for large k's. For instance, an acceptable limit is that w ensures $1 - \psi_k(w) \geq 99\%$ or $1 - \psi_k(w) \geq 99.5\%$. We thus work conditionally on the initial reserve w, for a given initial number n of annuitants aged x_0 at the origin.

As above, let us denote as $v(t_1, t_2)$ the (deterministic) present value at time t_1 of a payment of € 1 made at time t_2, $t_1 < t_2$, and as $u(t_1, t_2) = 1/v(t_1, t_2)$ the accumulated value at time t_2 of a payment of € 1 made at time t_1. Let $a_{\overline{k}|}$ be the present value of a sequence of unit payments made at times $1, 2, \ldots, k$, that is,

$$a_{\overline{k}|} = \sum_{t=1}^{k} v(0, t).$$

Note that the insurer gets at time t_0 all the annuities single premiums, so that ruin is impossible during the first few years. Precisely, define $k^\star$ as the largest integer k such that $a_{\overline{k^\star}|} \times n \leq w$. Then, $\psi_k(w) = 0$ if $k \leq k^\star$.

Let us define the following weighted sum of survivors

$$\begin{aligned} W_j &= \sum_{i=1}^{n} \sum_{h=1}^{j} u(h, j) \mathbb{I}[T_i > h] \\ &= \left(\sum_{i=1}^{n} a_{\overline{\min\{T_i, j\}}|} \right) u(0, j). \end{aligned}$$

It is the accumulated value at time j of all the payments made by the insurer to the n annuitants for the years 1 to j. For any horizon k, the ruin is avoided provided at each time j, W_j remains smaller than the accumulated value of the initial reserve w. Coming back to (4.1), we see that $U_t < 0 \Rightarrow U_{t+j} < 0$ for any $j = 1, 2, \ldots$ Therefore, the ruin probabilities can be expressed as

$$\psi_k(w) = \Pr\left[W_k > wu(0, k)\right] = \Pr\left[\sum_{i=1}^{n} a_{\overline{\min\{T_i, k\}}|} > w\right].$$

The formula for $\psi_k(w)$ involving only W_k, computing the ruin probability at horizon k then amounts to evaluate the distribution function of W_k. If the lifetimes are independent, this can be done using recursive algorithms designed for the individual model of risk theory. Here, the lifetimes are only conditionally independent, so that these algorithms only give the conditional ruin probabilities for some fixed $\boldsymbol{\kappa}$ sequence. These conditional probabilities must then be integrated with respect to $\boldsymbol{\kappa}$ to get the overall ruin probability. Since the application of the recursive algorithms can be rather time-consuming, this approach becomes rapidly prohibitive, except if the κ_t's are appropriately discretized (so that the discrete version of $\boldsymbol{\kappa}$ follows a binomial or trinomial tree). Hereafter, we describe several explicit formulas to approximate or bound the ruin probabilities.

4.3 Approximations and bounds based on mean and variance

For n large enough we have the following CLT-based approximation for the non-ruin probability:

$$\Pr\left[W_j \leq wu(0, j)\right] \approx \Phi\left(\frac{wu(0, j) - \mu_j}{\sigma_j}\right),$$

where Φ is the standard Normal distribution function, $\mu_j = \mathbb{E}[W_j]$ and $\sigma_j^2 = \mathbb{V}[W_j]$. We refer the reader to DENUIT & FROSTIG (2007c) for a derivation of the moments of W_j when the life times $T_1, \ldots, T_n$ conform to the Lee-Carter model.

Following KAAS & GOOVAERTS (1985) it is also possible to get bounds on the distribution function of W_j from the knowledge of its first moments. Specifically, it is easily seen that W_j is valued in $(0, b_j)$ with $b_j = nu(0, j)a_{\overline{j}|}$. Using the mean μ_j, the standard deviation σ_j and the upper bound b_j for W_j, we can find the bounds

$$M^{(\mu_j,\sigma_j,b_j)}(s) \leq F_{W_j}(s) \leq W^{(\mu_j,\sigma_j,b_j)}(s) \text{ for all } s \geq 0. \tag{4.2}$$

These bounds complement the CLT approximation by giving the range of possible values for the distribution function of W_j at a specific level s. Explicit expressions for these extremal distributions are provided in Table 1 in KAAS & GOOVAERTS (1985).

4.4 Approximations and bounds based on mean, variance and skewness

The approximation based on the CLT is often inaccurate for small values of n. This is why actuaries often resort to the NP approximation that takes into account the third moment. Instead of approximating the ratio $\frac{W_j-\mu_j}{\sigma_j}$ by a standard Normal random variable Z as with the Central Limit theorem, the Normal-Power approximation consists is using a linear combination $Z + b_1 Z^2 + b_2$ involving the Gamma distributed random variable Z^2. This allows to take into account the disymmetry of W_j. Let us denote as $\gamma_j = \frac{\mathbb{E}[(W_j-\mu_j)^3]}{\sigma_j^3}$ the skewness of W_j. The coefficients b_1 and b_2 are obtained by equating the three first moments of $\frac{W_j-\mu_j}{\sigma_j}$ and $Z + b_1 Z^2 + b_2$. The approximation

$$\Pr[W_j \leq w] \approx \Phi\left(-\frac{3}{\gamma_j} + \sqrt{\frac{9}{\gamma_j^2} + 1 + \frac{6}{\gamma_j}\frac{w-\mu_j}{\sigma_j}}\right)$$

is generally accurate provided $0 < \gamma_j < 1$ and $w \geq \mu_j + \sigma_j$.

Using the skewness $\gamma_j = \frac{\mathbb{E}[(W_j-\mu_j)^3]}{\sigma_j^3}$, tighter bounds $M^{(\mu_j,\sigma_j,\gamma_j,b_j)}$ and $W^{(\mu_j,\sigma_j,\gamma_j,b_j)}$, say, can be found such that

$$M^{(\mu_j,\sigma_j,\gamma_j,b_j)}(s) \leq F_{W_j}(s) \leq W^{(\mu_j,\sigma_j,\gamma_j,b_j)}(s) \text{ for all } s \geq 0. \tag{4.3}$$

Explicit expressions for these extremal distributions are provided in Table 3 in KAAS & GOOVAERTS (1985).

4.5 Prudential life table

4.5.1 Requirement

Now, let us determine a prudential life table as follows. Let us consider the cohort reaching age x_0 (typically, retirement age) in year t_0. For this cohort, we determine the prudential life table $\mu^{\text{prud}}_{x_0+k}$, $k = 1, 2, \ldots$, in order to satisfy

$$\Pr[\mu_{x_0+k}(t_0 + k|\boldsymbol{\kappa}) \leq \mu^{\text{prud}}_{x_0+k} \text{ for some } k = 1, 2, \ldots] \leq \epsilon_{\text{mort}}$$

for some probability level ϵ_{mort} small enough (motivation for the choice of ϵ_{mort} will be given in Section 4.5.3). This is equivalent to requiring that

$$\Pr[\exp(\alpha_{x_0+k}+\beta_{x_0+k}\kappa_{t_0+k})\geq \mu^{\mathrm{prud}}_{x_0+k} \text{ for all } k=1,2,\ldots]\geq 1-\epsilon_{\mathrm{mort}}.$$

In order to find the $\mu^{\mathrm{prud}}_{x_0+k}$'s, we express them as a percentage π of a set of reference forces of mortality $\mu^{\mathrm{ref}}_{x_0+k}$, i.e. $\mu^{\mathrm{prud}}_{x_0+k}=\pi\mu^{\mathrm{ref}}_{x_0+k}$. Then, the value of π comes from the constraint

$$\Pr\left[\kappa_{t_0+k}\geq \frac{\ln\left(\pi\mu^{\mathrm{ref}}_{x_0+k}\right)-\alpha_{x_0+k}}{\beta_{x_0+k}} \text{ for all } k=1,2,\ldots\right]=1-\epsilon_{\mathrm{mort}}.$$

Note that the reduction of death rates by a constant factor π is in line with the proportional hazard transform approach to measure risk that has been proposed by Wang (1995).

4.5.2 Reference life table

The set of the $\mu^{\mathrm{ref}}_{x_0+k}$'s can be the latest available population life table, for instance. Here, we take for the $\mu^{\mathrm{ref}}_{x_0+k}$'s the exponential of the point estimates of the κ_{t_0+k}'s, that is,

$$\mu^{\mathrm{ref}}_{x_0+k}=\exp(\alpha_{x_0+k}+\beta_{x_0+k}(\kappa_{t_0}+k\theta)). \tag{4.4}$$

We thus require that

$$\Pr[\exp(\alpha_{x_0+k}+\beta_{x_0+k}\kappa_{t_0+k})\geq \pi\exp(\alpha_{x_0+k}+\beta_{x_0+k}(\kappa_{t_0}+k\theta)) \text{ for all } k=1,2,\ldots]\geq 1-\epsilon_{\mathrm{mort}}$$

$$\Leftrightarrow \Pr[\beta_{x_0+k}(\kappa_{t_0+k}-(\kappa_{t_0}+k\theta))\geq \ln\pi \text{ for all } k=1,2,\ldots]\geq 1-\epsilon_{\mathrm{mort}}.$$

The value of $\ln\pi$ can then be determined as a quantile of the random vector

$$\Big(\beta_{x_0+1}\big(\kappa_{t_0+1}-(\kappa_{t_0}+\theta)\big),\ldots\beta_{\omega}\big(\kappa_{t_0+\omega-x_0}-(\kappa_{t_0}+(\omega-x_0)\theta)\big)\Big)$$

that is multivariate Normal with $\mathbf{0}$ mean and variance-covariance matrix

$$\widetilde{\boldsymbol{\Sigma}}=\left(\begin{array}{cccc}\sigma^2\beta^2_{x_0+1} & \sigma^2\beta_{x_0+1}\beta_{x_0+2} & \cdots & \sigma^2\beta_{x_0+1}\beta_{\omega}\\ \sigma^2\beta_{x_0+1}\beta_{x_0+2} & 2\sigma^2\beta^2_{x_0+2} & \cdots & 2\sigma^2\beta_{x_0+2}\beta_{\omega}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma^2\beta_{x_0+1}\beta_{\omega} & 2\sigma^2\beta_{x_0+2}\beta_{\omega} & \cdots & (\omega-x_0)\sigma^2\beta^2_{\omega}\end{array}\right).$$

4.5.3 Property of the prudential life table

Let $\phi_k(w|\boldsymbol{\kappa})$ denote the non-ruin probability conditional on $\boldsymbol{\kappa}$. Clearly, $\phi_k(w)=\mathbb{E}[\phi_k(w|\boldsymbol{\kappa})]$. Let us denote as $\phi_k^{\mathrm{prud}}(w)$ the non-ruin probability computed with the life table $\mu^{\mathrm{prud}}_{x_0+k}$, that is, computed assuming the mutual independence between the life times with the set $\{q^{\mathrm{prud}}_{x_0+k},\ k=0,1,\ldots\}$ of the one-year death probabilities given by

$$q^{\mathrm{prud}}_{x_0+k} \;=\; 1-\exp\Big(-\pi\exp\big(\alpha_{x_0+k}+\beta_{x_0+k}(\kappa_{t_0}+k\theta)\big)\Big) \text{ for } k=0,1,\ldots$$

We then have

$$\Pr[\phi_k(w|\boldsymbol{\kappa}) \geq \phi_k^{\text{prud}}(w)] \geq 1 - \epsilon_{\text{mort}}$$

since $\phi_k(w|\boldsymbol{\kappa})$ is increasing in the death rates $\mu_x(t|\boldsymbol{\kappa})$. Hence,

$$\phi_k(w) \geq (1 - \epsilon_{\text{mort}})\phi_k^{\text{prud}}(w)$$

so that

$$\phi_k(w) \geq (1 - \epsilon_{\text{mort}})(1 - \epsilon_{\text{solv}})$$

if w is determined so that $\phi_k^{\text{prud}}(w) \geq 1 - \epsilon_{\text{solv}}$ for some acceptable insolvency probability ϵ_{solv}. Taking $\epsilon_{\text{mort}} = \epsilon_{\text{solv}} = 1\%$ gives a ruin probability of at most 1.99%.

4.5.4 Panjer algorithm for the ruin probability

As demonstrated above, computing the ruin probability amounts to evaluate the distribution function of the W_j's, and this amounts to compute the distribution function of the sum of the $a_{\overline{\min\{T_i,j\}}|}$'s. In general, we must account for the dependence induced among the T_i's by the unknown life table (i.e., by the $\boldsymbol{\kappa}$ random vector). However, if we switch to the prudential life table, this dependence disappears since the life table becomes deterministic. In this case, computing the distribution function of W_j thus amounts to compute the distribution function of a sum of n independent and identically distributed random variables $a_{\overline{\min\{T_1,j\}}|}, \ldots, a_{\overline{\min\{T_n,j\}}|}$. Now, the probability distribution of $\sum_{i=1}^n a_{\overline{\min\{T_i,j\}}|}$ can be obtained as follows. Clearly, $a_{\overline{\min\{T_i,j\}}|}$ is valued in $\{0, a_{\overline{1}|}, \ldots, a_{\overline{j}|}\}$ and has probability distribution

$$\begin{aligned}
\Pr[a_{\overline{\min\{T_i,j\}}|} = 0] &= q_{x_0}^{\text{prud}} \\
\Pr[a_{\overline{\min\{T_i,j\}}|} = a_{\overline{\ell}|}] &= p_{x_0}^{\text{prud}} \ldots p_{x_0+\ell-1}^{\text{prud}} q_{x_0+\ell}^{\text{prud}} \text{ for } \ell = 1, \ldots, j-1 \\
\Pr[a_{\overline{\min\{T_i,j\}}|} = a_{\overline{j}|}] &= p_{x_0}^{\text{prud}} \ldots p_{x_0+j-1}^{\text{prud}}.
\end{aligned}$$

Let X_i be $a_{\overline{\min\{T_i,j\}}|}$ that has been appropriately discretized. Here, we keep the original probability mass at the origin, and round the other values in the support of $a_{\overline{\min\{T_i,j\}}|}$ to the least upper integer. This ensures that the ruin probability computed in this way will be at least as large as the exact ruin probability. The probability mass function p_X of the X_i's has support $\{0, 1, \ldots, \lceil a_{\overline{j}|} \rceil\}$, with $p_X(0) > 0$ (since the probability mass $q_{x_0}^{\text{prud}}$ of $a_{\overline{\min\{T_i,j\}}|}$ at the origin is kept unchanged). De Pril (1985) developed a simple recursion giving the n-fold convolution of p_X directly in terms of p_X. This substantially reduces the number of required operations. Specifically, the probability mass function of the sum $S = \sum_{i=1}^n X_i$ can be computed from the following recursive formula:

$$p_S(s) = \frac{1}{p_X(0)} \sum_{\eta=1}^{s} \left(\frac{n+1}{s}\eta - 1 \right) p_X(\eta) p_S(s-\eta), \quad s = 1, 2, \ldots,$$

starting from $p_S(0) = \big(p_X(0)\big)^n$. This recurrence relation is a particular case of Panjer recursion formula in the compound Binomial case. It is known to be numerically unstable so that particular care is needed when performing the computations. Backward and forward computations are often needed to reach a given numerical accuracy. For more details about these issues, we refer the reader to Panjer & Wang (1993).

5 Large portfolios approximations

5.1 Life annuity net single premium

Assume that each policyholder has a life annuity contract that pays € 1 at the end of each year provided he or she is still alive at that time. Given $\boldsymbol{\kappa}$, the present value of the benefits for policyholder i aged x writes

$$a_x(\boldsymbol{\kappa}) = \sum_{d=1}^{\omega-x} v(0,d)\,{}_dp_x(\boldsymbol{\kappa}).$$

Let us consider a portfolio of n life annuity contracts. The policyholders are all aged x_0 in year t_0, when the annuity contracts are issued. We assume that the future lifetimes $T_1, T_2, \ldots, T_n$ of the n policyholders are as described in Property 3.2. By the conditional law of large numbers, we have that

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} a_{\overline{T_i}|} = a_x(\boldsymbol{\kappa}) \text{ almost surely.}$$

Note that in the last equation, $\boldsymbol{\kappa}$ is present in the left-hand side since it influences the common distribution of the T_i's.

For large values of n, the following approximation can thus be used:

$$\Pr\left[\sum_{i=1}^{n} a_{\overline{T_i}|} > w\right] \approx \Pr\left[a_x(\boldsymbol{\kappa}) > \frac{w}{n}\right].$$

This comes from the fact that $\frac{1}{n}\sum_{i=1}^{n} a_{\overline{T_i}|}$ converges to $a_{x_0}(\boldsymbol{\kappa})$ with probability 1. One can then determine the prudent life table by setting the approximation equal to some acceptable probability level ϵ.

Note that, despite the positive dependence existing between the Lee-Carter lifetimes, there is still some diversification effect in the portfolio. Indeed, invoking Theorem 4 in Arnold & Villasenor (1986) allows us to write

$$\frac{1}{n+1}\sum_{i=1}^{n+1} a_{\overline{T_i}|} \preceq_{\text{cx}} \frac{1}{n}\sum_{i=1}^{n} a_{\overline{T_i}|}$$

so that less capital is needed as the size of the portfolio increases.

5.2 Comonotonic approximation to $a_{x_0}(\boldsymbol{\kappa})$

5.2.1 Definition of comonotonicity

Comonotonicity corresponds to perfect positive dependence: all the random variables can be written as non-decreasing transformations of the same underlying random variable. They thus "move in the same direction", are "common monotonic", hence the name. Precisely, a random vector $(X_1, \ldots, X_d)$ is said to be comonotonic if, and only if, there exist a random variable Z and non-decreasing functions $t_1, t_2, \ldots, t_d$, such that $(X_1, \ldots, X_d)$ is distributed as $\big(t_1(Z), t_2(Z), \ldots, t_d(Z)\big)$. Equivalently, $(X_1, \ldots, X_d)$ is comonotonic if it is distributed as $\big(t_1(Z), t_2(Z), \ldots, t_d(Z)\big)$ with all the t_i's non-increasing. A detailed account of comonotonicity can be found in Dhaene et al. (2002a,b).

5.2.2 Quantile functions

Given a distribution function F_X for a random variable X, its inverse F_X^{-1} is generally defined as

$$F_X^{-1}(\epsilon) = \inf\left\{x \in \mathbb{R} \mid F_X(x) \geq \epsilon\right\},$$

for $0 \leq \epsilon \leq 1$. Given some probability level ϵ, $F_X^{-1}(\epsilon)$ is the ϵth quantile of X.

Now, assume that F_X is continuous (this assumption will be satisfied in all the applications contained in this paper). The quantiles of the random variables X and $t(X)$, for a continuous non-decreasing function t, are related as follows: for any $0 < \epsilon < 1$

$$F_{t(X)}^{-1}(\epsilon) = t\left(F_X^{-1}(\epsilon)\right). \tag{5.1}$$

If t is continuous and non-increasing then, for any $0 < \epsilon < 1$

$$F_{t(X)}^{-1}(\epsilon) = t\left(F_X^{-1}(1-\epsilon)\right). \tag{5.2}$$

5.2.3 Additivity of quantile functions for sums of comonotonic random variables

It easy to see that the sum of comonotonic random variables are quantile-additive. Having a comonotonic random vector $(X_1, \ldots, X_d)$ distributed as $\big(t_1(Z), \ldots, t_d(Z)\big)$ with all the t_i's non-decreasing, it is easily seen that $\sum_{i=1}^d X_i$ is distributed as $t(Z)$ with $t = t_1 + \ldots + t_d$ non-decreasing. Hence, the quantile function of $\sum_{i=1}^d X_i$ is obtained as in (5.1) by applying t to the quantile function of Z, that is

$$F_{X_1+\ldots+X_d}^{-1}(\epsilon) = \sum_{i=1}^{d} t_i(F_Z^{-1}(\epsilon)).$$

Now, since (5.1) ensures that $F_{X_i}^{-1}(\epsilon) = t_i(F_Z^{-1}(\epsilon))$ for $i = 1, \ldots, d$, we have thus established that if $(X_1, \ldots, X_d)$ is comonotonic then

$$F_{X_1+\ldots+X_d}^{-1}(\epsilon) = \sum_{i=1}^{d} F_{X_i}^{-1}(\epsilon).$$

5.2.4 Comonotonic approximations

Consider the vector $(X_1, \ldots, X_d)$ of correlated random variables. Assume that we are interested in the distribution of $S_d = \sum_{i=1}^d X_i$. The determination of the distribution function of this sum is in general difficult (and requires either numerical integration or simulation).

If the X_i's are strongly correlated, then they can be approximated by their comonotonic version $(F_{X_1}^{-1}(Z), \ldots, F_{X_d}^{-1}(Z))$ where Z is uniformly distributed on the unit interval. The idea is thus to approximate S_d by $S_d^u = \sum_{i=1}^d F_{X_i}^{-1}(Z)$. Now, the quantile function of S_d^u is obtained directly from the sum of the marginal quantile functions.

In particular, if $X_i = t_i(Z_i)$, where the t_i's are monotonic functions (all decreasing or all increasing) and where the Z_i's are identically distributed and strongly correlated, then we can approximate S_d by $S_d^u = \sum_{i=1}^d t_i(Z)$, where Z is distributed as the Z_i's.

Another approach to approximate the distribution of S_d consists in approximating this sum with $S_d^l = \mathbb{E}[S_d|\Lambda_d]$, for some random variable Λ_d "close" to S_d. The approximation $S_d^l = \sum_{i=1}^d \mathbb{E}[X_i|\Lambda_d]$ can be regarded as a sum of comonotonic random variables provided $\mathbb{E}[X_i|\Lambda_d]$ is a non-decreasing function of Λ_d. The additivity of the quantile function can then be invoked again, to get the quantile function of S_d^l without effort.

The random variables S_d^u and S_d^l have extremal properties (hence, the superscripts "u" for "upper" and "l" for "lower"). Broadly speaking, S_d^u is the most variable sum of terms with the same marginal distributions as the terms involved in S_d, whereas S_d^l is less variable than S_d. Formally, the stochastic inequality

$$S_d^l \preceq_{\text{cx}} S_d \preceq_{\text{cx}} S_d^u \tag{5.3}$$

holds true. In particular, we have the equality $\mathbb{E}[S_d^l] = \mathbb{E}[S_d] = \mathbb{E}[S_d^u]$ for the expectations and the inequality $\mathbb{V}[S_d^l] \leq \mathbb{V}[S_d] \leq \mathbb{V}[S_d^u]$ for the variances.

5.2.5 Comonotonic approximation for the random cohort survival probabilities

Let us now define

$$S_d = \sum_{j=0}^{d-1} \exp\left(\alpha_{x+j} + \beta_{x+j}\kappa_{t_0+j}\right) = \sum_{j=0}^{d-1} \delta_j \exp(Z_j), \tag{5.4}$$

where $\delta_j = \exp(\alpha_{x+j}) > 0$ and $Z_j = \beta_{x+j}\kappa_{t_0+j}$. The d-year survival probability ${}_dp_x(\boldsymbol{\kappa})$ is then equal to $\exp(-S_d)$. Conditional upon κ_{t_0}, we have that Z_j is Normally distributed with mean μ_j and variance σ_j^2 given by

$$\mu_j = \beta_{x+j}(\kappa_{t_0} + j\theta) \text{ and } \sigma_j^2 = (\beta_{x+j})^2 j\sigma^2, \tag{5.5}$$

with the convention that a Normally distributed random variable with zero variance is constantly equal to its mean. The random variable S_d appears as a linear combination of correlated LogNormal random variables. The random d-year survival probability is the exponential of this linear combination.

Being driven by the κ_t's, the terms in S_d are certainly strongly positively dependent. Therefore, as explained in Section 3.4, we could think of approximating S_d by a sum of perfectly dependent random variables, with the same marginal distributions, that is,

$$S_d \approx S_d^u = \sum_{j=0}^{d-1} \delta_j \exp(\mu_j + \sigma_j Z)$$

where Z obeys to the standard Normal distribution. Since S_d^u is a sum of comonotonic random variables, its quantile function is additive. Considering (5.1), the quantile function

$F^{-1}_{S^u_d}$ of S^u_d is given by

$$F^{-1}_{S^u_d}(\epsilon) = \sum_{j=0}^{d-1} \delta_j \exp\Big(\mu_j + \sigma_j \Phi^{-1}(\epsilon)\Big), \tag{5.6}$$

where Φ^{-1} is the quantile function of the standard Normal distribution.

Another approximation of S_d is $S^l_d = \mathbb{E}[S_d|\Lambda_d]$, where Λ_d is taken as a first-order approximation of S_d, that is,

$$\Lambda_d = \sum_{j=0}^{d-1} \delta_j \exp(\mu_j) Z_j.$$

It is expected that S_d and S^l_d be "close" to each other. A straightforward computation then gives

$$S^l_d = \sum_{j=0}^{d-1} \delta_j \exp\Big(\mu_j + r_j(d)\sigma_j Z + \frac{1}{2}(1-(r_j(d))^2)\sigma_j^2\Big)$$

where $r_i(d)$, $i = 0, \ldots, d-1$, is the correlation coefficient between Λ_d and Z_i, that is,

$$r_i(d) = \frac{\mathbb{C}[Z_i, \Lambda_d]}{\sigma_i \sigma_{\Lambda_d}} = \frac{\sum_{j=0}^{d-1} \delta_j \exp(\mu_j) \mathbb{C}[Z_i, Z_j]}{\sigma_i \sqrt{\sum_{j=0}^{d-1} \sum_{k=0}^{d-1} \delta_j \delta_k \exp(\mu_j + \mu_k) \beta_{x+j} \beta_{x+k} \min\{j,k\} \sigma^2}}$$

where $\mathbb{C}[Z_i, Z_j] = \beta_{x+i}\beta_{x+j} \min\{i,j\}\sigma^2$.

In the application we have in mind, β_{x+i} and β_{x+j} typically have the same sign so that all the $r_i(d)$'s are non-negative. This makes the S^l_d's sums of comonotonic random variables and allows us to take advantage of the quantile additivity. Specifically, the quantile function of S^l_d is then given by

$$F^{-1}_{S^l_d}(\epsilon) = \sum_{j=0}^{d-1} \delta_j \exp\Big(\mu_j + r_j(d)\sigma_j \Phi^{-1}(\epsilon) + \frac{1}{2}(1-(r_j(d))^2)\sigma_j^2\Big). \tag{5.7}$$

5.3 Life annuity conditional expected present value

An analytical computation of the distribution of $a_x(\boldsymbol{\kappa})$ seems to be out of reach. From the approximations S^u_d and S^l_d derived for S_d satisfying (5.3), we get the following approximations for $a_x(\boldsymbol{\kappa})$:

$$a_x(\boldsymbol{\kappa}) \approx \sum_{d \geq 1} \exp(-S^u_d) v(0,d)$$

or

$$a_x(\boldsymbol{\kappa}) \approx \sum_{d \geq 1} \exp(-S^l_d) v(0,d).$$

Since the S^u_d's are sums of comonotonic random variables, their quantile functions are additive. Moreover, the ϵth quantile of $\exp(-S^u_d)$ is $\exp(-F^{-1}_{S^u_d}(1-\epsilon))$. This provides the following approximations for the the quantile function $F^{-1}_{a_x(\boldsymbol{\kappa})}$ of $a_x(\boldsymbol{\kappa})$:

$$F^{-1}_{a_x(\boldsymbol{\kappa})}(\epsilon) \approx \sum_{d \geq 1} \exp\Big(-F^{-1}_{S^u_d}(1-\epsilon)\Big) v(0,d) \tag{5.8}$$

where $F^{-1}_{S^u_d}$ is given in (5.6). Now, assuming that the S^l_d's are comonotonic, we get

$$F^{-1}_{a_x(\boldsymbol{\kappa})}(\epsilon) \approx \sum_{d\geq 1} \exp\Big(- F^{-1}_{S^l_d}(1-\epsilon)\Big) v(0,d) \tag{5.9}$$

where $F^{-1}_{S^l_d}$ is given by (5.7).

6 Conclusion

This paper discusses some of the consequences of the use of projected life tables. In the Lee-Carter framework, a single time index drives mortality forecasts. This index obeys to a stochastic process (often, a Gaussian random walk with drift). The randomness in the future trajectory of this process induces dependence between policyholders' remaining lifetimes, that are no more independent but conditionally independent, given the time index.

From the Lee-Carter specification, we have established that the remaining lifetimes are associated. Also, present values of pure endowments and life annuities are associated. This allows us to use the convex order to compare the riskiness of the situation with respect to independence. Large portfolio approximations are also obtained. In that respect, this paper examined the distribution of the life annuity conditional expected present value, given future mortality rates. This random variable can be seen as the residual, non diversifiable risk that remains with the annuity provider whatever the size of the portfolio. Comonotonic approximations are derived, which are easily tractable. The accuracy of the approximations makes them suitable for practical evaluations.

Since no one can accurately predict the future, risk management of mortality and longevity is an indispensable part in the annuity providers operations. Life annuity contracts typically run for several decades so that a life table which may seem to be on the safe side at the beginning of the contract might well turn out not to be so. Moreover, contrarily to financial assets (that can be very volatile), changes in forces of mortality slowly occur and pose a long term, but permanent, problem. Reinsurance treaties covering longevity risk are usually expensive and many life insurance companies are reluctant to buy long-term reinsurance coverage (because of substantial credit risk). Survivor bonds have coupon payments that depend on the proportion of the population surviving to particular ages. These bonds provide a very good hedge against mortality improvement risk: if annuitants live longer, the insurance companies would then make annuity payments for longer periods, but they would also receive greater offsetting coupon payments on their survivor bonds asset positions. See Denuit, Devolder & Goderniaux (2007) for more details, as well as Biffis & Denuit (2006), Biffis, Denuit & Devolder (2005), and Devolder & Denuit (2006).

The results derived in this paper are also useful to design risk transfer mechanisms. The life annuity conditional expected present value could be an appropriate index of longevity for the insurance market. Mortality derivatives can then define their payoffs with respect to this index.

Of course, the Lee-Carter model is not the only mortality projection method. Since the early 1900's, the evolution over time of graduated mortality curves is popular for the purpose of extrapolation. One classical procedure is based on the projection of parameters.

Specifically, a parametric model is fitted to the mortality experienced during each calendar year over a selected period. Trend curves are then fitted to the progression in the estimated parameters to derive life tables in future calendar years. This approach of course heavily relies on the appropriateness of the retained parametric models. Moreover, the estimated parameters are often strongly dependent so that univariate extrapolations may be misleading. Failing to take into account interdepencies among parameters may lead to implausible forecasts or inaccurate forecast prediction intervals.

Renshaw, Haberman & Hatzopoulos (1996) suggested a modelling structure in the framework of Generalized Linear Models, which incorporates both the age variation in mortality and the underlying trends in the mortality rate. This approach has been used in Sithole, Haberman & Verrall (2000) to investigate mortality trends for immediate annuitants and life office pensioners. Such regression models give a very accurate in-sample fit. However, a main disadvantage of this deterministic trend approach is that the accurate in-sample fit is translated into quite small prediction intervals, when extrapolated out of sample. However, such accurate predictions do not seem to be realistic.

The log-bilinear approach shares close similarities with principal component analysis. Hyndman & Ullah (2007) extended the principal components approach by adopting a functional data paradigm combined with nonparametric smoothing (penalized regression splines) and robust statistics. Univariate time series are then fitted to each component coefficient (or level parameter). The Lee-Carter method appears to be a particular case of this general approach.

Age-period-cohort models in demography study variations in mortality rates along three dimensions: age, year (or period), and cohort. Within this framework, “cohort” refers to an individual’s year of birth. If there is a significant cohort effect then the Lee-Carter method smooths it out and the resulting forecasts might become unrealistic. As pointed out by Sunamoto (2005), the Lee-Carter model drives the cohort effect out to the residual term. A time series modelling of the residuals can then be performed to account for the cohort effects present in the data. Renshaw & Haberman (2006) do not follow this route, but enrich the Lee-Carter model with additional effects as follows: the force of mortality is decomposed into

$$\mu_x(t) = \exp(\alpha_x + \beta_x\kappa_t + \gamma_x\lambda_c)$$

with an extra pair of bilinear terms $\gamma_x\lambda_c$ to represent additional cohort effects.

Empirical analyses suggest that $\ln q_x(t)/p_x(t)$ is reasonably linear in x for fixed t. Cairns, Blake & Dowd (2006) assumed that $q_x(t)$ is governed by the following two-factor Perks stochastic process:

$$q_x(t) = \frac{\exp(\kappa_t^{[1]} + \kappa_t^{[2]}x)}{1 + \exp(\kappa_t^{[1]} + \kappa_t^{[2]}x)}$$

where $\boldsymbol{\kappa}_t = (\kappa_t^{[1]}, \kappa_t^{[2]})^T$ obeys to a random walk with drift:

$$\boldsymbol{\kappa}_{t+1} = \boldsymbol{\kappa}_t + \boldsymbol{\mu} + \boldsymbol{C}\boldsymbol{Z}(t+1)$$

where $\boldsymbol{\mu}$ is a constant 2×1 vector of drift parameters, $\boldsymbol{C}$ is the constant 2×2 lower triangular matrix reflecting volatilities and correlations (specifically, $\boldsymbol{C}$ is the Choleski square root matrix of the covariance matrix), and $Z(t)$ is a 2×1 vector of independent standard Normal variables. This model can be extended to incorporate cohort effects.

Acknowledgements

The Author acknowledges the financial support of the *Communauté française de Belgique* under contract "Projet d'Actions de Recherche Concertées" ARC 04/09-320, as well as the financial support of the *Banque Nationale de Belgique* under grant "Risk measures and Economic capital".

References

[1] Arnold, B. C., & Villasenor, J. A. (1986). Lorenz ordering of means and medians. Statistics and Probability Letters 4, 47-49.

[2] Artzner, Ph., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent risk measures. Mathematical Finance 9, 203-228.

[3] Biffis, E., & Denuit, M. (2006). Lee-Carter goes risk-neutral: An application to the Italian annuity market. Giornale dell'Istituto Italiano degli Attuari 69, 1-21.

[4] Biffis, E., Denuit, M., & Devolder, P. (2005). Stochastic mortality under measure changes. Working Paper 05-14, Institut des Sciences Actuarielles, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

[5] Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. International Journal of Forecasting 22, 547-581.

[6] Booth, H., Maindonald, J. & Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. Population Studies 56, 325-336.

[7] Brouhns, N., Denuit, M., & Van Keilegom, I. (2005). Bootstrapping the Poisson log-bilinear model for mortality projection. Scandinavian Actuarial Journal, 212-224.

[8] Brouhns, N., Denuit, M., & Vermunt, J.K. (2002a). A Poisson log-bilinear approach to the construction of projected lifetables. Insurance : Mathematics and Economics 31, 373-393.

[9] Brouhns, N., Denuit, M., & Vermunt, J.K. (2002b). Measuring the longevity risk in mortality projections. Bulletin of the Swiss Association of Actuaries, 105-130.

[10] Cairns, A.J.G., Blake, D., & Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. Journal of Risk and Insurance 73, 687-718.

[11] Christofides, T. C., & Vaggelatou, E. (2004). A connection between supermodular ordering and positive/negative association. Journal of Multivariate Analysis 88, 138-151.

[12] Cossette, H., Delwarde, A., Denuit, M., Guillot, F. & Marceau, E. (2007). Pension plan valuation and dynamic mortality tables. North Americal Actuarial Journal 11, 1-34.

[13] Czado, C., Delwarde, A., & Denuit, M. (2005). Bayesian Poisson log-bilinear mortality projections. Insurance: Mathematics & Economics 36, 260-284.

[14] Delwarde, A., Denuit, M., & Eilers, P. (2007). Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach. Statistical Modelling 7, 29-48.

[15] Delwarde, A., Denuit, M., & Partrat, Ch. (2007). Negative Binomial version of the Lee-Carter model for mortality forecasting. Applied Stochastic Models in Business and Industry, in press.

[16] Denuit, M. (2007a). Distribution of the random future life expectancies in log-bilinear mortality projection models. Lifetime Data Analysis, in press.

[17] Denuit, M. (2007b). Comonotonic approximations to quantiles of life annuity conditional expected present values. Working Paper 07-03, Institut des Sciences Actuarielles, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

[18] Denuit, M. (2007c). Forecasting mortality: Applications and examples of log-bilinear models. In Pitacco, E., Denuit, M., Haberman, S., & Olivieri, A. "Modelling Longevity Dynamics for Pensions and Annuity Business", forthcoming.

[19] Denuit, M., Devolder, P., & Goderniaux, A.-C. (2007). Securitization of longevity risk: Pricing survivor bonds with Wang transform in the Lee-Carter framework. Journal of Risk and Insurance 74, 87-113.

[20] Denuit, M., & Dhaene, J. (2007). Comonotonic bounds on the survival probabilities in the Lee-Carter model for mortality projections. Computational and Applied Mathematics 203, 169-176.

[21] Denuit, M., Dhaene, J., Goovaerts, M.J., & Kaas, R. (2005). Actuarial Theory for Dependent Risks: Measures, Orders and Models. Wiley, New York.

[22] Denuit, M., & Frostig, E. (2006). Heterogeneity and the need for economic capital in the individual model. Scandinavian Actuarial Journal 2006(1), 42-66.

[23] Denuit, M., & Frostig, E. (2007a). Association and heterogeneity of insured lifetimes in the Lee-Carter framework. Scandinavian Actuarial Journal 107, 1-19.

[24] Denuit, M., & Frostig, E. (2007b). Prudential rules for life annuity pricing and reserving. Working Paper 07-02, Institut des Sciences Actuarielles, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

[25] Denuit, M., & Frostig, E. (2007c). Life insurance mathematics with random life tables. Working Paper 07-07, Institut des Sciences Actuarielles, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

[26] Denuit, M., & Goderniaux, A.-C. (2005). Closing and projecting lifetables using log-linear models. Bulletin of the Swiss Association of Actuaries, 29-49.

[27] De Pril, N. (1985). Recursions for convolutions of arithmetic distributions. ASTIN Bulletin 15, 135-139.

[28] Devolder, P., & Denuit, M. (2006). Continuous time stochastic mortality and securitization of longevity risk. Working Paper 06-02, Institut des Sciences Actuarielles, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

[29] Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R., & Vyncke, D. (2002a). The concept of comonotonicity in actuarial science and finance: Theory. Insurance: Mathematics & Economics 31, 3-33.

[30] Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R., & Vyncke, D. (2002b). The concept of comonotonicity in actuarial science and finance: Applications. Insurance: Mathematics & Economics 31, 133-161.

[31] Esary, J. D., Proschan, F., & Walkup, D. W. (1967). Association of random variables with applications. Annals of Mathematical Statistics 38, 1466-1474.

[32] Frostig, E., & Denuit, M. (2006). Monotonicity results for portfolios with heterogeneous claims arrival processes. Insurance: Mathematics & Economics 38, 484-494.

[33] Frostig, E., & Denuit, M. (2007). Ruin probabilities and optimal capital allocation for heterogeneous life annuity portfolios. Working Paper 07-01, Institut des Sciences Actuarielles, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

[34] Hyndman, R.J., & Ullah, Md. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. Computational Statistics and Data Analysis 51, 4942-4956.

[35] Jogdeo, K. (1978). On a probability bound of Marshall and Olkin. Annals of Statistics 6, 232-234.

[36] Kaas, R., & Goovaerts, M.J. (1985). Bounds on distribution functions under integral constraints. Bulletin de l'Association Royale des Actuaires Belges – Koninklijke Vereniging der Belgische Actuarissen 79, 45-60.

[37] Lee, R.D. (2000). The Lee-Carter method of forecasting mortality, with various extensions and applications. North American Actuarial Journal 4, 80-93.

[38] Lee, R.D. & Carter, L. (1992). Modelling and forecasting the time series of US mortality. Journal of the American Statistical Association 87, 659-671.

[39] Li, S.H., Hardy, M.R., & Tan, K.S. (2006). Uncertainty in mortality forecasting: An extension of the classical Lee-Carter approach. Technical Report, University of Waterloo.

[40] Madsen, R.W. (1993). Generalized binomial distribution. Communications in Statistics - Theory and Methods 22, 3065-3086.

[41] Marshall, A. W., & Olkin, I. (1979). Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.

[42] Norberg, R. (1989). Actuarial analysis of dependent lives. Bulletin of the Swiss Association of Actuaries, 243-254.

[43] Panjer, H.H., & Wang, S. (1993). On the stability of recursive formulas. ASTIN Bulletin 23, 227-258.

[44] Pitacco, E. (2004). Survival models in a dynamic context: A survey. Insurance: Mathematics & Economics 35, 279-298.

[45] Pitacco, E., Denuit, M., Haberman, S., & Olivieri, A., Eds (2007). Modelling Longevity Dynamics for Pensions and Annuity Business. Forthcoming.

[46] Renshaw, A.E. & Haberman, S. (2003). Lee-Carter mortality forecasting with age specific enhancement. Insurance: Mathematics & Economics 33, 255-272.

[47] Renshaw, A.E. & Haberman, S. (2006). A cohort-based extension of the Lee-Carter model for mortality reduction factors. Insurance: Mathematics & Economics 38, 556-570.

[48] Renshaw, A.E., Haberman, S. & Hatzopoulos, P. (1996). Recent mortality trends in U.K. male assured lives. British Actuarial Journal 2, 449-477.

[49] Shaked, M., & Spizzichino, F. (1998). Positive dependence properties of conditionally independent random lifetimes. Mathematics of Operations Research 23, 944-959.

[50] Sithole, T.Z., Haberman, S. & Verrall, R.J. (2000). An investigation into parametric models for mortality projections, with applications to immediate annuitants and life office pensioners' data. Insurance: Mathematics & Economics 27, 285-312.

[51] Sunamoto, N. (2005). Cohort effect structure in the Lee-Carter residual term. Technical Report.

[52] Tong, Y.L. (1990). The Multivariate Normal Distribution. Springer-Verlag, New York.

[53] Tuljapurkar, S., Li, N. & Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. Nature 405, 789-792.

[54] Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazard transforms. Insurance: Mathematics & Economics 17, 43-54.

[55] Wang, S. (2000). A class of distortion operators for pricing financial and insurance risks. Journal of Risk and Insurance 67, 15-36.

[56] Wong-Fupuy, C., & Haberman, S. (2004). Projecting mortality trends: recent developments in the United Kingdom and the United States. North American Actuarial Journal 8, 56-83.