THRESHOLD LIFE TABLES AND THEIR APPLICATIONS*

Johnny Siu-Hang Li,† Mary R. Hardy,‡ and Ken Seng Tan§

ABSTRACT

The rapid emergence of centenarians has highlighted the importance of survival probabilities at extreme ages and has motivated actuaries to look for alternative ways to close of life tables in place of assigning a death probability of 1 at an arbitrarily chosen age. Using the asymptotic results of modern extreme value theory, we propose a model, which we call the threshold life table, to extrapolate survival distributions to extreme ages and to determine the appropriate end point of a life table. By combining the threshold life table with the Lee-Carter model for stochastic mortality forecasting, we consider applications to the valuation of a life annuity portfolio and to the prediction of the highest attained age. We illustrate the theoretical results using U.S., Canadian, and Japanese mortality data.

1. INTRODUCTION

Since World War II, a significant contribution to overall mortality improvement in developed countries has arisen from the extension of life among the old rather than improving mortality among the young. The steady improvement of old-age mortality has led to a rapid emergence of supercentenarians. The number of supercentenarians in the International Database of Longevity has been increasing exponentially since the mid-1970s (Robine and Vaupel 2002). Although this tendency may be subject to an upward bias because of the possible omission of earlier cases and unanticipated age exaggerations, the growth of the supercentenarian population is indisputable, posing great challenges to actuaries as well as planners of age-based entitlement programs.

From an actuarial viewpoint, the emergence of supercentenarians gives rise to low-frequency, high-severity losses in cash flows that are contingent on survival. Prime examples include life annuity portfolios and defined-benefit pension plans. Reverse mortgages (also known as home income plans in Britain) are another example. Roughly speaking, a reverse mortgage is a loan that enables senior homeowners to convert part of their equity in their home into a stable income stream without having to sell their home or take on a new regular mortgage payment. The loan is paid off by the proceeds of the sale of the property when the homeowner either dies or moves out. A prolonged lifetime may result in a severe loss because of the lengthened income stream and the reduced present value of the property.

A reliable model of old-age mortality is therefore crucial in the valuation of these and similar businesses. Unfortunately, problems associated with the available data make the modeling difficult. Specifically, the number of deaths and exposures-to-risk dwindle at advanced ages, leading to large sampling errors and highly volatile crude death rates. It is also well known that ages at death are often misreported. Today's supercentenarians were, of course, born more than 100 years ago, when record keeping was not as exhaustive as it is today. Huge effort is required to validate their ages at death (see, e.g., Bernard and Vaupel 1999; Bourbeau and Desjardins 2002), and it is often impossible to confirm

* Professors Hardy and Tan acknowledge the support of the Natural Science and Engineering Research Council of Canada.

† Johnny Siu-Hang Li, ASA, PhD, is Assistant Professor in the Department of Statistics & Actuarial Science at the University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada. shli@uwaterloo.ca.

‡ Mary R. Hardy, FSA, CERA, FIA, PhD, is CIBC Professor of Financial Risk Management in the Department of Statistics & Actuarial Science at the University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada. mrhardy@waterloo.ca.

§ Ken Seng Tan, ASA, CERA, PhD, is Associate Professor and Canada Research Chair in the Department of Statistics & Actuarial Science at the University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada, and Cheung Kong Scholar at the China Institute for Actuarial Science, Central University of Finance and Economics, Bejing, P.R. China, kstan@uwaterloo.ca.

or disprove the accuracy in the absence of documentary evidence. Without reliable measurements of mortality at advanced ages, actuaries sometimes close the life table at an arbitrarily chosen age, say, 100, by setting the probability of death at that particular age as 1. The effect of an early closure on the aggregate loss may be insubstantial in terms of expectation, but certainly not in terms of variance and other risk measures, such as value-at-risk, as the odds of an extreme loss are largely determined by the tail of the survival distribution.

Recent developments in extreme value theory (EVT) may offer an attractive solution to the problem. The major contribution of this study—the threshold life table—is a systematic integration of EVT to the parametric modeling of mortality. More specifically, we use the asymptotic distribution of the exceedances over threshold to model the survival distribution beyond a particular age, the threshold age, which is chosen to ensure that the tail of the fitted distribution is consistent with the parametric graduation for earlier ages. This strategy allows us to extrapolate a survival distribution to extreme ages without the need for accurate mortality data for supercentenarians, and to determine statistically the appropriate end point of a life table.

The threshold life table is further extended to a dynamic version by using the Lee-Carter model for stochastic mortality forecasting. The dynamic version takes account of both mortality risk, which arises from the randomness of age at death, and longevity risk, which originates from the uncertainty in the trend of mortality improvement. We use the dynamic version to study the joint impact of mortality and longevity risk on the risk measures of a life annuity portfolio and on the highest attained age, which is commonly referred to as omega or ω in the actuarial literature.

The rest of this paper is organized as follows: Section 2 defines the notation and describes the data used throughout our discussion. Section 3 provides a brief review of previous work on the modeling of old-age mortality. Section 4 defines the threshold life table and details how the model parameters and the threshold age may be estimated. Section 5 considers applications to risk measures and the highest attained age. Section 6 concludes.

2. NOTATION AND DATA

The following notation is used throughout this paper:

- X: the age at death random variable, assumed to be continuous
- $f(x)$: the probability density function of X
- $F(x)$: the distribution function of X
- $S(x) = 1 - F(x)$: the survival function
- $\mu(x) = f(x)/(1 - F(x))$: the force of mortality
- d_x: the number of deaths between ages x and $x + 1$
- E_x: the number of exposures-to-risk between ages x and $x + 1$; in practice, E_x is approximated by the midyear population at age x
- l_x: the number of survivors to age x
- $m_x = d_x/E_x$: the central rate of death at age x
- $q_x = d_x/l_x$: the probability of death between ages x and $x + 1$, conditioning on survival to age x.

The subscript t may be appended to d_x, E_x, l_x, m_x, and q_x when they are considered to be varying with time; for example, $m_{x,t}$ refers to the central rate of death at age x and in calender year t. For detailed interpretations of the above notation, we refer readers to Chapter 3 of Bowers et al. (1997).

We illustrate the theoretical results using U.S., Canadian, and Japanese mortality data. Historical death counts and midyear population estimates (proxy for E_x) are provided by the Human Mortality Database (2006).

3. PREVIOUS RESEARCH ON MODELING OLD-AGE MORTALITY

Over the past few decades, actuaries and demographers have suggested various methods to model old-age mortality. Below we summarize those that are commonly used.

3.1 Cubic Polynomial Extrapolation

In analyzing Canadian individual insurance mortality experience, Panjer and Russo (1992) and Panjer and Tan (1995) used a cubic polynomial to extrapolate the survival distribution beyond age 100. The estimation consists of three stages.

In stage 1, graduated death probabilities, denoted by q_x^{grad}, were obtained by the Whittaker-Henderson method (see London 1985). In stage 2, Makeham's Second Law was fitted to the graduated mortality rates by minimizing the weighted least squares,

$$WLS = \sum_{x=z}^{99} E_x(\mu_x^{grad} - a - bx - cd^x)^2, \tag{3.1}$$

subject to the constraint that $\mu_z^{grad} = a - bz - cd^z$, where $\mu_x^{grad} = -\ln(1 - q_x^{grad})$, and z equals 40 and 44 for males and females, respectively. In the final stage, the mortality function beyond age 99 was assumed to follow a cubic polynomial of the form

$$q_{x+99}^{grad} = ax^3 + bx^2 + cx + d, \tag{3.2}$$

where $x \geq 99$. The parameters of the function can be derived using the following constraints: $q_{99}^{poly} = q_{99}^{grad}$, $q_{99}'^{poly} = q_{99}'^{grad}$, $q_{99}''^{poly} = q_{99}''^{grad}$, and $q_{105}^{poly} = 1$.

3.2 The Heligman-Pollard Model

Heligman and Pollard (1980) proposed the following eight-parameter formula to describe the mortality of the entire life span:

$$\frac{q_x}{1 - q_x} = A^{(x+B)^C} + D \exp(-E[\ln x - \ln F]^2) + GH^x, \tag{3.3}$$

where $A^{(x+B)^C}$, $D \exp(-E[\ln x - \ln F]^2)$ and GH^x represent early childhood mortality, accidental mortality, and senescent mortality, respectively. In addition, the component GH^x may be interpreted as the discrete version of the Gompertz law of mortality.

Although the estimation of model parameters can be undertaken relatively easily, by minimizing the weighted sum of squared errors, formulated in a similar manner as equation (3.1), there is strong empirical evidence that old-age mortality is not Gompertzian (see, e.g., Olshansky and Carnes 1997). Hence, the model may not be ideal for old-age mortality analysis.

3.3 The Coale-Kisker Method

The Coale-Kisker method (Coale and Guo 1989; Coale and Kisker 1990) has been widely applied to the mortality data of various developed countries. The method assumes that mortality rates increase at a varying rate instead of a constant rate as the Gompertz curve assumes. Let us suppose that we begin the Coale-Kisker extrapolation at age 85. We define

$$k(x) = k(x - 1) - R, \qquad x \geq 84, \tag{3.4}$$

where $k(x) = \ln(m_x/m_{x-1})$, R is a constant. Extending the formula up to $x = 110$ and summing, we have

$$k(85) + \cdots + k(110) = 26k(84) - (1 + 2 + \cdots + 26)R. \tag{3.5}$$

Solving for R, we obtain

$$R = \frac{26k(84) + \ln(m_{84}) - \ln(m_{110})}{351}. \tag{3.6}$$

To determine R, Coale and Kisker (1990) assumed that $m_{110} = 1.0$ for males, based on the fact that there are almost no survivors at ages greater than 110. They also assumed that $m_{110} = 0.8$ for females so as to avoid imposing a crossover of male and female mortality at age 110.

3.4 The Relational Model of Mortality

The relational model (Himes, Preston, and Coudran 1994) consists of a ''standard'' life table—for ages 45 to 99—calibrated from 82 different mortality schedules observed in several low-mortality countries. The ''standard'' is made useful at higher ages by fitting a straight line to the logits of the death rates from age 80:

$$\text{logit}(m_x^s) = \alpha + \beta x, \tag{3.7}$$

where $\text{logit}(m_x^s) = \ln(m_x^s/1 - m_x^s)$, and m_x^s denotes the ''standard'' central rate of death at age x. Equation (3.7) can be used to produce the ''standard'' death rates for ages 100 and beyond.

Then we can extrapolate any life table by relating it to the extended ''standard'' schedule through the following logit regression:

$$\text{logit}(m_x) = \delta + \gamma\text{logit}(m_x^s), \tag{3.8}$$

where δ and γ are the regression coefficients, and m_x are the central death rates in the life table that we intend to extrapolate.

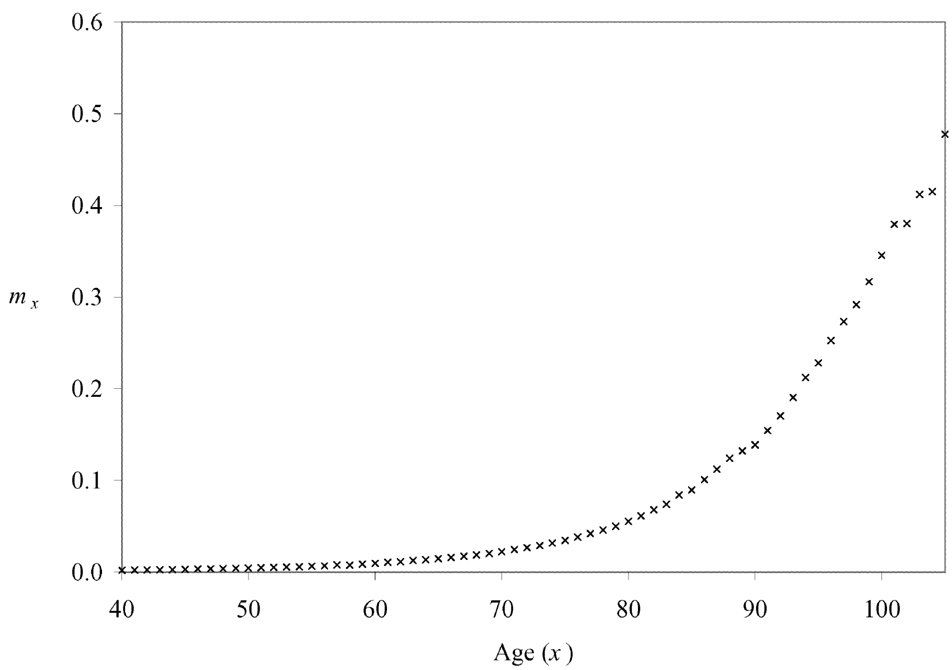
4. THE THRESHOLD LIFE TABLE

4.1 Motivation

We begin with an exploratory graphical analysis of a typical pattern of the central rate of death. Figure 1 suggests that the rate of increase after age 85 is much more rapid than before. The modeling of an age pattern of mortality is usually done in a piecewise approach. First, a parametric graduation is

Figure 1

Central Death Rates from Age 40, the U.S. Population, 2004



applied to the death rates at earlier ages to smooth out irregularities and to obtain a summary of the age pattern by a few parameters. Then the graduated rates are extrapolated to advanced ages, perhaps by using one of the four methodologies we reviewed in the last section. These methods, however, require an assumption on either the age at which the life table is closed (the cubic polynomial extrapolation and the Coale-Kisker method) or the trajectory of old-age mortality, without any scientific justification (the relational model and the Heligman-Pollard model). Furthermore, all these methods demand a subjective decision on the age at which the extrapolation begins.

The threshold life table is also a piecewise approach. It retains the parametric graduation for earlier ages, but the mathematical extrapolation is replaced by a statistical distribution, justified by extreme value theory. The fitted statistical distribution determines the appropriate end point of the life table, if any. Moreover, the age at which the modeling switches from a parametric graduation to an extreme value distribution can be chosen in a statistical way without the need of any subjective decision.

4.2 Model Specification

Let $Z = Y - d|Y > d$ be the conditional exceedances over the threshold d. The Balkema–de Haan–Pickands Theorem (Balkema and de Haan 1994) states that, under some regulatory conditions, the limiting distribution of Z is a Generalized Pareto as d approaches the right-hand-end support of Y. This important result in EVT provides a theoretical justification of the threshold life table, which is defined as follows:

for $x \leq N$,

$$F(x) = 1 - \exp\left(-\frac{B}{\ln C}(C^x - 1)\right); \tag{4.1}$$

for $x > N$,

$$F(x) = 1 - p\left(1 + \gamma\left(\frac{x - N}{\theta}\right)\right)^{-1/\gamma}, \tag{4.2}$$

where $p = S(N)$, and N is known as the *threshold age*. In other words, we assume that the survival distribution is Gompertzian before the threshold age, and the exceedances over the threshold age follow a Generalized Pareto distribution, according to the Balkema–de Haan–Pickands Theorem. To ensure that F is a proper distribution function, we require $B > 0$, $C > 1$, and $\theta > 0$. It is noteworthy that such a construction guarantees that F is continuous at the threshold age, but that the smoothness of $F(x)$ during the transition requires a careful choice of the threshold age.

In many applications of the Generalized Pareto distribution, the threshold d can be determined by the Hill estimator (Hill 1975; Mason 1982), which is defined as follows:

$$\hat{\alpha}(k) = \left(\frac{1}{k}\sum_{i=1}^{k} \ln \frac{X_{i,n}}{X_{k+1,n}}\right)^{-1}, \tag{4.3}$$

where $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$ is a random sample of size n arranged from largest to smallest. Suppose that the plot of $\hat{\alpha}(k)$ is flat when $k = j$. Then $X_{j+1,n}$ may be taken as an estimate of the threshold d.

Unfortunately the Hill estimator is not applicable in the selection of the threshold age, partly because of the fact that mortality data are highly censored, as we know only the number of deaths in different age intervals but not the precise age at death for each individual. The Hill estimator is not designed for censored data. Moreover, the use of Hill's plot requires eyeballing, and this is infeasible in the simulation studies that we conduct in later sections. Finally, Hill's plot focuses only on the tail of the distribution, and this contradicts our objective of producing a model suitable for the entire life span.

4.3 Estimation of Model Parameters

We propose two methods of choosing the threshold age. In both, the threshold age is chosen in the process of parameter estimation to ensure that the tail (the Generalized Pareto part) and the body (the Gompertz part) are consistent with each other. In the following discussion, we assume that the coverage of the mortality data reaches the open age group 100 and above, and that the mortality data from age 65 are used.

- **METHOD 1: MAXIMUM LIKELIHOOD ESTIMATION**

The likelihood function for the threshold life table can be written as

$$L(B, C, \gamma, \theta; N) = \left[\prod_{x=65}^{99} \left(\frac{S(x) - S(x+1)}{S(65)} \right)^{d_x} \right] \left(\frac{S(100)}{S(65)} \right)^{l_{100}}. \tag{4.4}$$

For age $x = 65, \ldots, 99$, the data are interval-censored, which means the contribution to the likelihood function is $S(x) - S(x + 1)$ per individual. Similarly, for age 100, the data are right-censored, which gives $S(100)$ as the contribution per individual. Each term in equation (4.4) is divided by $S(65)$ to acknowledge the left-truncation at age 65. We can easily show that the logarithm of equation (4.4) can be decomposed as the sum of two components, $l_1(B, C; N) + l_2(\gamma, \theta; N)$, where

$$l_1(B, C; n) = \sum_{x=65}^{N-1} [d_x \ln(S(x) - S(x+1))] + l_N \ln(S(N)) - l_{65} \ln(s(65)), \tag{4.5}$$

where $S(x) = \exp(-B/\ln C\ (C^x - 1))$, and

$$l_2(\gamma, \theta; N) = \sum_{x=N}^{99} \left[d_x \ln \left(\frac{S(x)}{S(N)} - \frac{S(x+1)}{S(N)} \right) \right] + l_{100} \ln \left(\frac{S(100)}{S(N)} \right), \tag{4.6}$$

where $S(x)/S(N) = (1 + \gamma\ (x - N/\theta))^{-1/\gamma}$.

This observation suggests that, for fixed N, the Gompertz part and the Generalized Pareto part can be estimated separately by maximizing l_1 and l_2, respectively. The choice of N relies on the maximization of the profile log-likelihood, which is defined by

$$l_p(N) = l(\hat{B}(N), \hat{C}(N), \hat{\gamma}(N), \hat{\theta}(N); N), \tag{4.7}$$

where $l = \ln(L)$; $\hat{B}(N)$, $\hat{C}(N)$, $\hat{\gamma}(N)$, and $\hat{\theta}(N)$ are the maximum likelihood estimate of B, C, γ, and θ for fixed N, respectively. The process of threshold age and model parameter estimation can be summarized by the following algorithm:

1. For $N = 98$,
   a. Find the values of B and C that maximize l_1
   b. Find the values of γ and θ that maximize l_2
   c. Compute the value of the profile log-likelihood, l_p
2. Repeat step 1 for $N = 97, 96, \ldots, 85$
3. Find the value of N that gives the maximum profile log-likelihood.

The value of N obtained in step 3 is the optimal threshold age. The maximum likelihood estimates of B, C, γ, and θ under the optimal threshold age are the ultimate model parameter estimates.

In fitting the threshold life table to period (static) mortality experience, the required values of d_x and l_x can be computed by constructing a hypothetical cohort using the series of static q_x. Let us assume an arbitrary number, say, 100,000, for the size of radix (l_0). Then d_x and l_x can be computed by the following recursive relations: $d_x = l_x q_x$ and $l_{x+1} = l_x - d_x$. We can easily show that the maximum likelihood parameter estimates are independent of the size of the radix chosen.

- **METHOD 2: WEIGHTED LEAST SQUARES ESTIMATION**

If the mortality data are given in terms of central death rates, then the method of weighted least squares can be readily applied. Let us rewrite the threshold life table as a function of the force of mortality:

for $x \leq N$,

$$\ln(\mu(x)) = \ln(B) + (\ln C)x; \tag{4.8}$$

for $x > N$,

$$\ln(\mu(x)) = -\ln(\theta) - \ln\left(1 + \gamma\left(\frac{x - N}{\theta}\right)\right). \tag{4.9}$$

Under the assumption of constant force of mortality for fractional ages, m_x may be considered the best estimate of $\mu(x)$. This relation suggests that we may estimate the model parameters by minimizing the sum of squared errors, *SSE,* which is defined as follows:

$$SSE = SSE_1 + SSE_2, \tag{4.10}$$

where

$$SSE_1 = \sum_{x=65}^{N} E_x\,(\ln(m_x) - \ln(B) - (\ln(C))x)^2, \tag{4.11}$$

and

$$SSE_2 = \sum_{x=N+1}^{99} E_x\left(\ln(m_x) + \ln(\theta) + \ln\left(1 + \gamma\left(\frac{x - N}{\theta}\right)\right)\right)^2. \tag{4.12}$$

The squared errors are weighted by the number of exposures-to-risk to allow for a better fit for ages with larger risk exposures.[1] Similar to maximum likelihood estimation, the Gompertz part and the Generalized Pareto part can be estimated separately given a fixed value of N. The estimation process can be summarized by the following algorithm:

1. For $N = 98$,
   a. Find the values of B and C that minimize SSE_1
   b. Find the values of γ and θ that minimize SSE_2
   c. Compute the value of the total sum of squared errors, *SSE*
2. Repeat step 1 for $N = 97, 96, \ldots, 85$
3. Find the value of N that gives the minimum *SSE*.

The value of N obtained in step 3 is the optimal threshold age, and the estimates of B, C, γ, and θ under the optimal threshold age are the ultimate model parameter estimates.

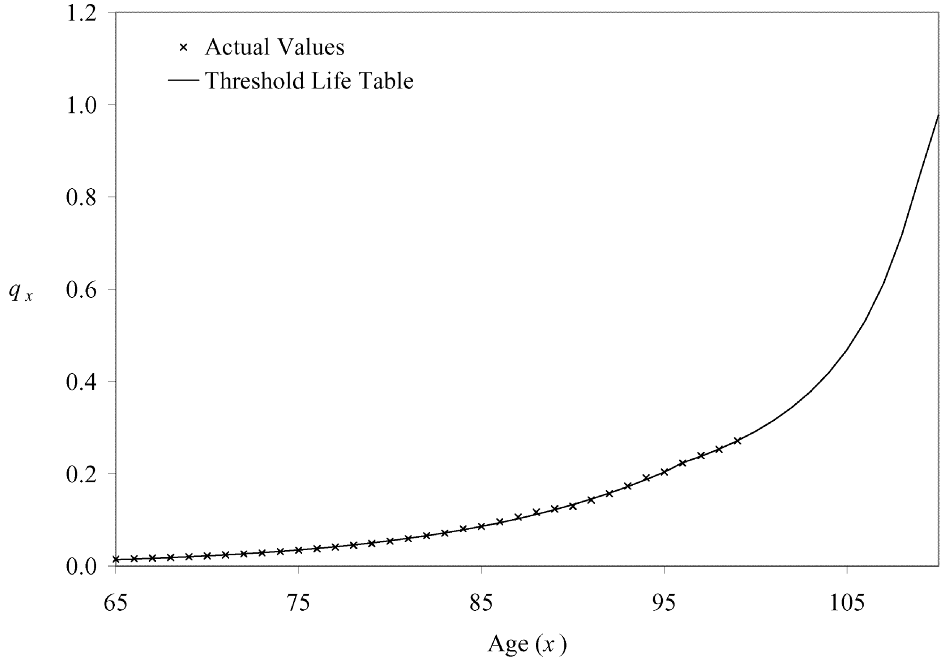
4.4 Empirical Results

Figure 2 depicts the fitted threshold life tables for the U.S., Canadian, and Japanese populations. Although the model specification does not guarantee smoothness at the threshold age, we observe that the body and tail of the fitted distribution join with each other smoothly. This result arises because the proposed methods of threshold age and model parameter estimation emphasize the overall goodness-of-fit. We also observe an excellent fit over the entire life span. Beyond the threshold age, the death probability increases progressively to 1 at the end point, which is determined statistically. We provide a further discussion on the determination of end point in Section 5.1, which deals with the prediction of the highest attained age.

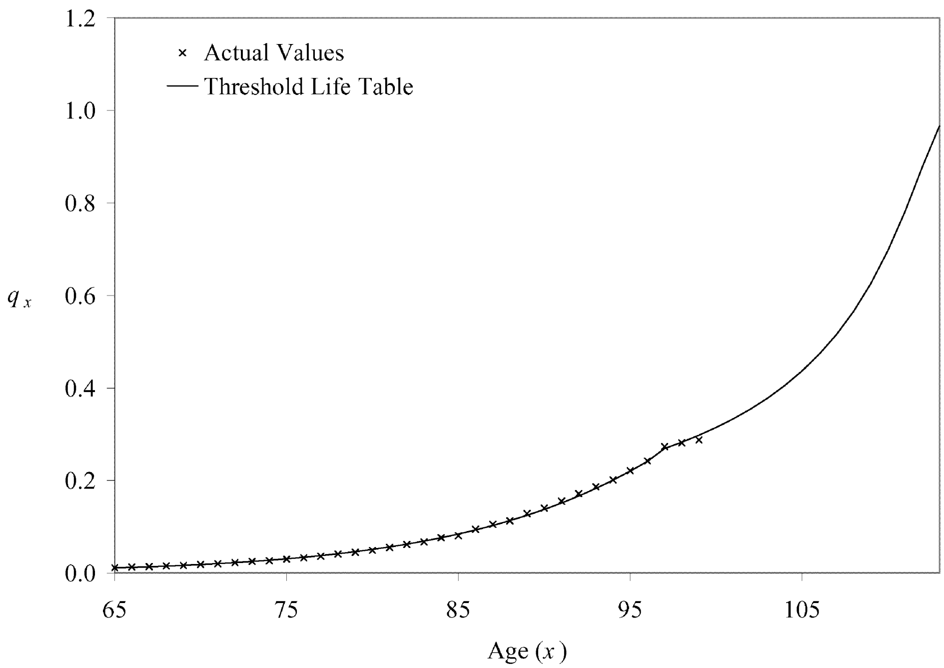
[1] An equal weight of 1 yields a similar result.

Figure 2

Threshold Life Tables Fitted to Static Mortality Data for the U.S., Canadian, and Japanese Populations in 2004

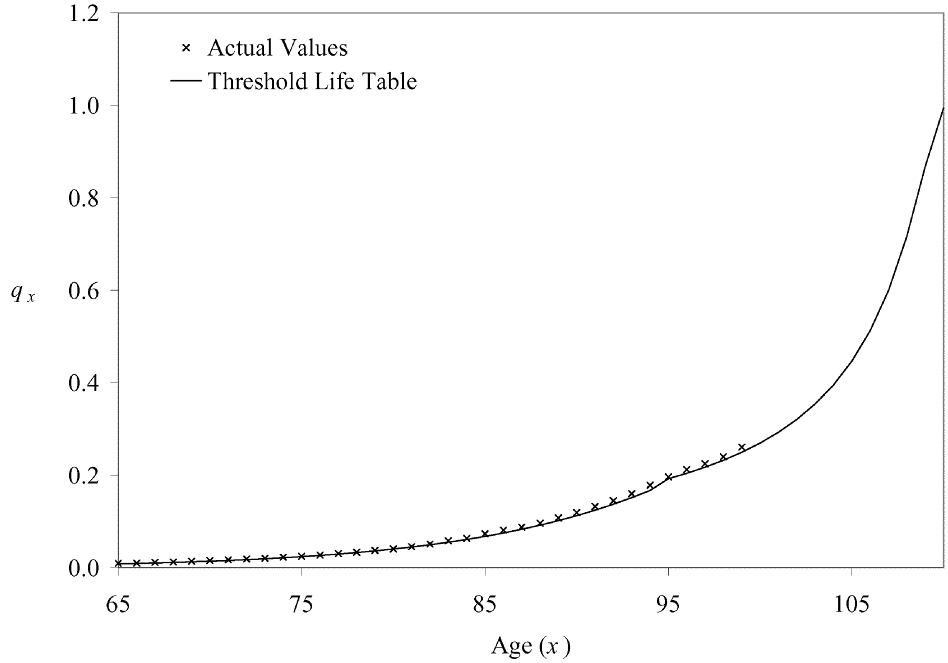


United States, $N = 96$



Canada, $N = 97$

Figure 2
(continued)



Japan, $N = 95$

5. APPLICATIONS

5.1 The Highest Attained Age

Suppose that n members of a birth cohort survive to age x_0; then, when all these n members have died, there will be a highest value, say, M_n, among their ages at death; and prior to realization, M_n can be regarded as a random variable with a certain probability distribution.

5.1.1 Previous Work on Modeling the Highest Attained Age

Thatcher (1999) modeled the highest attained age by using classical extreme value theory. Specifically, Thatcher assumed that the age-at-death random variable follows a three-parameter logistic distribution, that is,

$$\mu(x) = \frac{\alpha \exp(\beta x)}{1 + \alpha \exp(\beta x)} + \gamma. \tag{5.1}$$

According to classical extreme value theory, the distribution function of the highest attained age, M_n, can be expressed as

$$F_{M_n}(x) = \left(1 - \exp\left(-\int_{x_0}^{x} \mu(t)\, dt\right)\right)^n. \tag{5.2}$$

Combining equations (5.1) and (5.2), Thatcher obtained the following closed-form expression for the distribution function of M_n:

$$F_{M_n}(x) = \left[1 - \exp\left(\gamma(x_0 - x) + \beta^{-1} \ln\left(\frac{1 + \alpha \exp(\beta x_0)}{1 + \alpha \exp(\beta x)}\right)\right)\right]^n. \tag{5.3}$$

The choice of x_0 is arbitrary, and n can be approximated by the midyear population estimate for the cohort at age x_0.

Li and Chan (2007) extended the work of Thatcher (1999) by considering the impact of mortality improvement on the highest attained age. First, they extrapolated historical period death rates to age 150 by the relational model proposed by Himes, Preston, and Condran (1994). Given this matrix of extrapolated death rates, they obtained a Lee-Carter mortality projection from which cohort life tables for different birth cohorts were computed. Finally, they derived the distribution function of omega for each cohort using classical extreme value theory. The model they used is identical to that specified by equation (5.2). However, the distribution function of M_n must be calculated numerically because the extrapolation of death rates was performed in a nonparametric manner.

Han (2005) modeled the exceedance $Z = X - N|X > N$ over the threshold age N by a Generalized Pareto distribution. Han fitted Generalized Pareto distributions to period life tables for the U.S. population from 1901 to 1999, resulting in a trend of estimated omega. This trend was further modeled by a linear regression from which omegas in the future may be estimated. In Han's method the threshold ages were chosen in a completely arbitrary manner, leading to possible inconsistency with the age pattern at earlier ages. Also, the application to historical period life tables seems inappropriate since the highest attained age makes sense only if it is referenced to a particular birth cohort. Furthermore, the use of a linear regression in the prediction of future omegas has totally ignored the impact of future changes in the age pattern of mortality on the highest attained age.

Watts, Dupois, and Jones (2006) modeled the highest attained age directly by using the Generalized Extreme Value (GEV) distribution; that is,

$$F_{M_n}(x) = \exp\left(-\left(1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right)^{-1/\xi}\right), \tag{5.4}$$

where μ is the location parameter, σ is the scale parameter, and ξ is the shape parameter. The GEV distribution was fitted to the 10 highest ages at death in each year from 1980 to 2000. To acknowledge the fact that mortality data are not time homogeneous, they assumed that μ and the logarithm of σ were linear functions of time, leading to the following likelihood function:

$$L = \prod_{t=1980}^{2000} \left\{\exp\left(-\left(1 + \xi\left(\frac{z_t^{(10)} - \mu(t)}{\sigma(t)}\right)\right)^{-1/\xi}\right) \times \prod_{k=1}^{10} \sigma(t)^{-1}\left[1 + \xi\left(\frac{z_t^{(k)} - \mu(t)}{\sigma(t)}\right)\right]^{-1/\xi - 1}\right\}, \tag{5.5}$$

where $z_t^{(k)}$ is the kth largest age at death in year t, $\mu(t) = \mu_0 + \mu_1 t$, and $\sigma(t) = \exp(\sigma_0 + \sigma_1 t)$. Model parameters were estimated by maximizing L. The distribution of M_n in year s was obtained by substituting μ and σ in equation (5.4) by $\mu(s)$ and $\sigma(s)$, respectively. It is noteworthy that in assuming $\mu(s)$ is a linear function of time, the model implicitly preset a linear trajectory for the highest attained age. Note also that this direct approach requires precise data about the highest age at death, which is heavily subject to the problem of misreporting. Watts, Dupuis, and Jones (2006) showed that the right-hand-end support of the distribution of M_n can change from infinite to finite on removal of an outlier.

5.1.2 Our Method

Recall that in the threshold life table the exceedance $Y = X - N|X > N$ over the threshold age N is assumed to follow a Generalized Pareto distribution with the following specification:

$$F_Y(y) = 1 - \left(1 + \gamma \frac{y}{\theta}\right)^{-1/\gamma}, \tag{5.6}$$

where $y > 0$. The value of γ determines the precise type of distribution. If $\gamma > 0$, then Y follows a Pareto distribution; if $\gamma = 0$, then Y follows an exponential distribution; if $\gamma < 0$, then Y follows a beta distribution, which has a finite right-hand-end support, given by $-\theta/\gamma$. Equivalently speaking, when parameter γ is significantly less than zero (which we observe when the model is fitted to mortality data for many developed countries), the theoretical end point of the threshold life table is $N - \theta/\gamma$, which is strictly greater than N since $\theta > 0$ and $\gamma < 0$.

We let Y_i, $i = 1, 2, \ldots, n$, be a sequence of exceedances over the threshold age, and assume that Y_i follows the Generalized Pareto distribution specified by equation (5.6) independently. Also, we define $M_n = \max\{Y_i, i = 1, 2, \ldots, n\}$. If $\gamma < 0$, then the distribution function of M_n can be expressed as

$$F_{M_n}(y) = \begin{cases} 0, & y < 0 \\ (F_Y(y))^n, & 0 \le y < -\theta/\gamma. \\ 1, & y \ge -\theta/\gamma \end{cases} \tag{5.7}$$

It follows that when n tends to infinity, the distribution of M_n degenerates at $-\theta/\gamma$. Mathematically,

$$\lim_{n\to\infty} F_{M_n}(y) = \begin{cases} 0, & y < -\theta/\gamma \\ 1, & y \ge -\theta/\gamma. \end{cases} \tag{5.8}$$

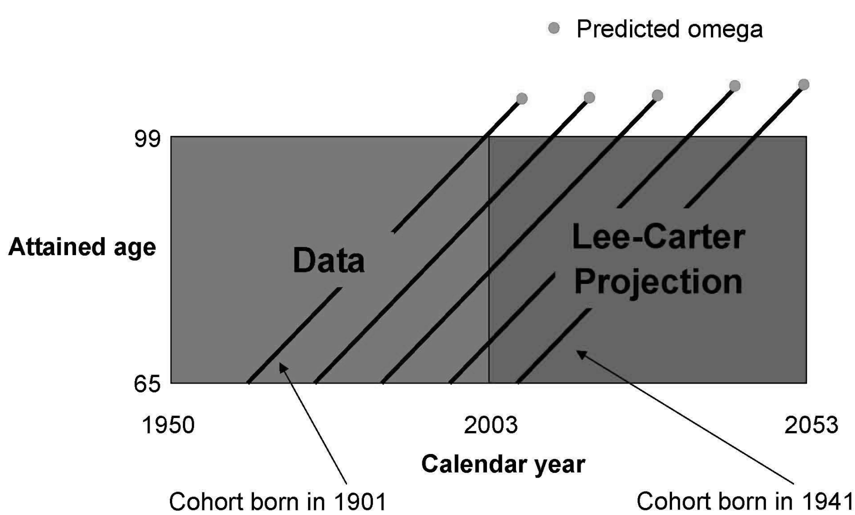
In other words, if the number of survivors (n) at the threshold age is sufficiently large, the highest attainable age of these survivors (and hence all members of the cohort) converges (in probability) to $\omega = N - \theta/\gamma$, the theoretical end point of the threshold life table.

Note that the variability in the estimate of ω is inherited from those in the estimates of θ and γ. Hence, the asymptotic variance of $\hat{\omega}$, the estimate of ω, can be computed by using the delta method, which can be summarized by the following equation:

$$\mathrm{Var}(\hat{\omega}) = \begin{pmatrix} \frac{\partial \omega}{\partial \gamma} & \frac{\partial \omega}{\partial \theta} \end{pmatrix} (I(\gamma, \theta))^{-1} \begin{pmatrix} \frac{\partial \omega}{\partial \gamma} \\ \frac{\partial \omega}{\partial \theta} \end{pmatrix}, \tag{5.9}$$

Figure 3

Graphical Illustration of the Methods for Predicting the Highest Attained Age



where $I(\gamma, \theta)$ is the information matrix for estimating γ and θ. Assuming normality holds, the approximate 95% confidence interval for ω can be written as

$$(\hat{\omega} - 1.96\sqrt{\mathrm{Var}(\hat{\omega})}, \hat{\omega} + 1.96\sqrt{\mathrm{Var}(\hat{\omega})}). \tag{5.10}$$

For cohorts born in earlier years, q_x for $x = 65, \ldots, 99$ can be obtained from the available data directly. However, for cohorts born in later years, we require an estimate of future death rates; for example, the prediction of omega for the cohort born in 1931 requires $q_{65,1996}, q_{66,1997}, \ldots, q_{99,2030}$. The values of $q_{74,2005}, \ldots, q_{99,2030}$ must be estimated from a mortality projection, based on the historical death rates. Here we use the Lee and Carter (1992) methodology to obtain a mortality projection from which cohort life tables for years of birth 1901 to 1941 are derived. Then for each birth cohort we model the death probabilities by a threshold life table whose end point is the predicted omega for that particular cohort. Finally, we compute the confidence intervals using equations (5.9) and (5.10). Figure 3 graphically illustrates our methodology.

Readers should note that there exists uncertainty in projecting future death rates. Continuing from the previous example, the actual and projected values of $q_{74,2005}, \ldots, q_{99,2030}$ may not be identical. We can take account of the uncertainty involved in estimating $q_{74,2005}, \ldots, q_{99,2030}$, by bootstrapping the Lee-Carter model (Brouhns, Denuit, and Van Keilegom 2005). The extended procedure for estimating the probability distribution of the highest attained age can be summarized as follows:

1. Generate N future mortality scenarios by bootstrapping the Lee-Carter model
2. For each scenario,
   a. Derive the cohort death rates for the birth cohort we are interested in
   b. Estimate the highest attained age for the cohort by fitting a threshold life table to the cohort death rates
3. Obtain an empirical distribution of omega for the cohort.

5.1.3 Data Issues

Raw data should be preferred over abridged rates in modeling old-age mortality. Although raw death counts for the U.S. and Canadian populations are quite exhaustive, raw population counts are highly right-censored; for example, raw populations counts (by single year of age) for Canadians are available up to age 89 only. According to the Human Mortality Database documentation (Wilmoth et al. 2005),

Figure 4

Illustration of the Extinct Cohorts Method Using a Lexis Diagram

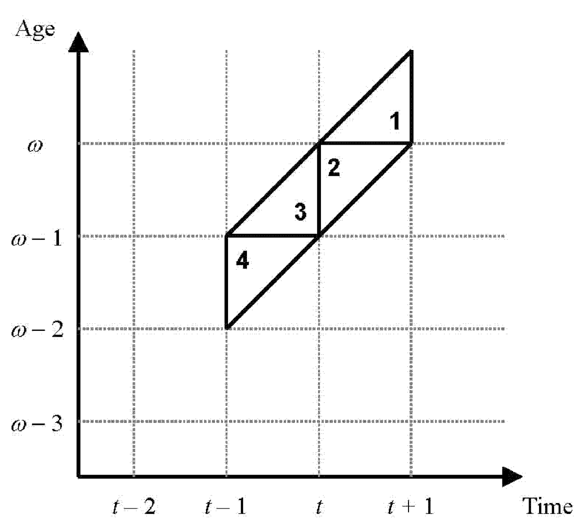


Table 1
Predicted Omegas for Various Cohorts of the Japanese Population

Year of Birth	Attained Age in 2007	Predicted Omega
1901	106	113.4477 (109.4321, 115.3740)
1911	96	114.6435 (109.0398, 120.1731)
1921	86	115.0247 (107.8065, 122.2430)
1931	76	119.9964 (110.0398, 132.1508)
1941	66	121.0953 (107.0769, 132.9158)

most of the unknown population counts beyond age 89 are estimated by using the extinct cohort method. With this method, the population size, $P\ (x,\ t)$, for a cohort at age x last birthday at the beginning of year t is estimated by summing all future deaths for the cohort. Mathematically,

$$P(x, t) = \sum_{i=0}^{\omega-x-1} [D_U(x + i, t + i) + D_L(x + i + 1, t + i)], \tag{5.11}$$

where $D_L(x, t)$ is the number of lower-triangle deaths recorded among those aged $[x, x + 1)$ in year t, $D_U(x, t)$ is the number of upper-triangle deaths recorded among those age $[x, x + 1)$ in year t, and ω is the assumed limiting age of the cohort. In Figure 4 we illustrate the extinct cohort method. $P(\omega - 1, t)$ is the sum of the death counts in triangles 1 and 2, while $P(\omega - 2, t - 1)$ is the sum of the death counts in triangles 1 to 4.

Although the contamination of data has little effect on the projection of future life expectancies, its impact on the estimation of the highest attained age can be significant. Also, as the extinct cohort method requires an assumption about the limiting age, the appropriateness of using the data derived by the extinct cohort method to predict the highest attained age is highly questionable. Therefore, in this section we apply only the theoretical results to the Japanese population for which both raw death and population counts (by single year of age) are available up to and including age 99.

5.1.4 Empirical Results

In Table 1 we show the mean and interval estimates of omegas for various birth cohorts. Here the interval estimates take no account of the uncertainty involved in the Lee-Carter mortality projection. Moving down the column, we observe that mortality improvement has a positive effect on the highest attained age. Over the period of analysis, the average increase in the predicted highest attained age is approximately 0.15 year per annum. Also, we observe that the width of the confidence interval increases with the year of birth. This trend is because the continual improvement in human mortality allows more people to survive to the open age group (100 and above), leading to more uncertainty about the tail of the survival distribution.

In Table 2 we evaluate how the consideration of the uncertainty involved in the Lee-Carter projection may affect the estimation of the highest attained age. We observe that, while the mean estimates of

Table 2
Predicted Omega for Cohort Born in 1941, the Japanese Population, with and without Considering the Uncertainty in the Lee-Carter Projection

Uncertainty in Mortality Improvement	No	Yes
The best estimate	121.0953	119.4668
Standard error	6.0309	15.7407
95% confidence interval	(107.0769, 132.9158)	(108.5603, 170.0000)

omega are similar, the standard error becomes significantly larger if the uncertainty involved in the Lee-Carter projection is taken into account. Note that the confidence intervals in Table 2 are obtained by different methods. The one that makes no allowance for uncertainty in mortality improvement is based on the asymptotic normality of the maximum likelihood estimates (eqs. (5.9) and (5.10)), whereas the other is obtained from the percentiles of a simulated distribution, which is permitted to be skewed.

5.2 Risk Measures

We consider a simple portfolio of a single-premium life annuity contract issued to a life aged 65 in 2004. In the following numerical illustration, we use the following assumptions:

1. Mortality follows the Japanese population experience
2. The interest rate is fixed at 5% per annum effective
3. There is only one premium (the net single premium) payable at age 65
4. No allowance is made for profit and expenses.

We evaluate the present value random variable by four different methods. The first two consider mortality risk only, and the other two take account of both mortality and longevity risk.

- **METHOD 1**

  We use the static death rates in 2004 and close the life table at age 100; that is, we let $q_{100} = 1$. To simulate the age at death (curtate future lifetime, $K(65)$), we generate z from the Uniform (0, 1) distribution; then we can obtain the simulated age at death by the following equation:

$$K(65) = \sup\left\{y : \prod_{x=65}^{y} (1 - q_x) > z\right\}, \tag{5.12}$$

  where q_x is the death probability from the static survival distribution. The present value of the portfolio can be expressed as

$$PV = \frac{1 - 1.05^{-K(65)}}{0.05/1.05}. \tag{5.13}$$

  We repeat the simulation 10,000 times to obtain an empirical loss distribution from which the mean, variance, value-at-risk (VaR), and conditional tail expectation (CTE) of the loss are calculated.

Figure 5
Distinction between the Four Methods of Valuation

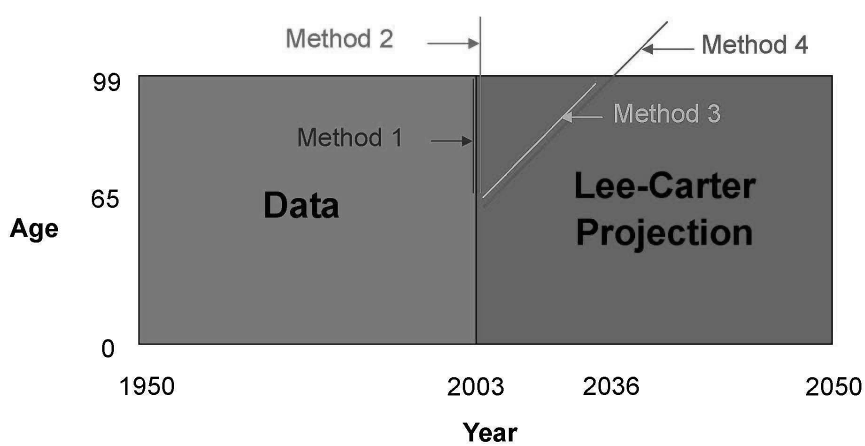


Table 3
Estimated Mean, Variance, 95% VaR, and 95% CTE of the Loss Using Different Methods

Method	Mean	Variance	95% VaR	95% CTE
1	−0.0316	36.777	8.3301	8.7125
2	−0.0112	40.950	8.7809	10.014
3	−0.0749	40.190	8.7228	10.228
4	0.0533	43.114	9.0421	11.532

- **METHOD 2**

  We use the static death rates in 2004 and model these death rates by a threshold life table, which extends the rates to extreme ages. The evaluation of the loss is similar to that in method 1 except that the values of q_x in equation (5.12) are taken from the fitted threshold life table.

- **METHOD 3**

  We use the cohort death rates for the cohort born in 1939 (aged 65 in 2004). To take longevity risk into account, we simulate 10,000 Lee-Carter mortality scenarios, based on the historical mortality data from 1950 to 2004, using the parametric bootstrapping method proposed by Brouhns, Denuit, and Van Keilegom (2005). For each of these scenarios, we compute the projected cohort death rates $(q_{66,2005}, \ldots, q_{99,2038})$. Here we do not extrapolate the rates and assume that $q_{100,2039} = 1$. Given the projected cohort death rates in each mortality scenario, we generate an empirical loss distribution from which we derive the risk measures for the loss.

- **METHOD 4**

  We model the cohort death rates in method 3 by a threshold life table, which extrapolates the rates up to a certain extreme age. The generation of the required empirical loss distribution is based on the extended cohort death rates.

Figure 5 provides a graphical illustration of the four methods. In Table 3 we observe that different methods yield significantly different risk measures. The difference between the risk measures obtained by methods 1 and 2 indicates that the consideration of extreme-age mortality has revealed a large part of variability that is neglected if the life table is closed too early. The comparison between methods 3 and 4 suggests that the effect of extreme-age mortality is amplified by the uncertainty about future mortality.

It is noteworthy that if the number of contracts in the portfolio is sufficiently large, the variance of the loss per contract will tend to zero in methods 1 and 2 but not in methods 3 and 4, because methods 3 and 4 have taken account of the nondiversifiable longevity risk.

6. CONCLUDING REMARKS

The tail behavior of the lifetime survival function has long been a controversial issue. There are three schools of thought on the subject. The first argues that q_x may reach 1 at a finite age (e.g., Vincent 1951). The second is that q_x may tend to 1 asymptotically as age increases (e.g., Gompertz 1825; Heligman and Pollard 1980). And the third view suggests that q_x may converge to a limit that is strictly less than 1 as age increases (e.g., Thatcher 1999). In fact, these three possibilities are provided for in our model. In the threshold life table, the form of the tail is determined by the model parameters and, hence, the data. We have demonstrated that if $\gamma < 0$, then $q_x = 1$ when $x = N - \theta/\gamma$. We can also easily show that if $\gamma > 0$, then q_x tends asymptotically to 1, and that if $\gamma = 0$, then the tail is exponential, implying that q_x tends to a limit that is less than 1.

Although mortality data among the oldest-old are crucial to the analysis of the highest attained age, the question on how much we should rely on raw data remains open. Unverified ages at death obtained from government agencies are often exaggerated, leading to a positive bias in the predicted distribution of the highest attained age. Watts, Dupuis, and Jones (2006) provided an excellent illustration. On the other hand, the use of verified mortality data is not a plausible alternative, as the number of validated supercentenarians in Japan is too scanty for statistical inference of any kind.[2] Until sufficient validated data among the oldest-old is available, a minimal extrapolation, either mathematical or statistical, is apparently the only feasible solution.

The accuracy of the predicted omegas is highly dependent on the validity of the assumptions used. The omegas predicted in this study would be accurate if the underlying survival distribution belongs to the peaks-over-threshold domain of attraction.[3] Similarly, the omegas predicted by Li and Chan (2007), for example, would be accurate if the relational model (see Section 3.4) they used is appropriate. These conditions, unfortunately, cannot be verified with the available data.

Finally, readers should be cautious about the model risk entailed in the analysis. The Lee-Carter mortality forecast relies on the long-term stability of the mortality index, which is highly aggregated. Consequently the model could have smoothed out the unusually rapid decline of old age mortality rates since the 1970s, and this may affect the forecast of the highest attained age.

7. ACKNOWLEDGMENTS

The authors wish to thank Doug Andrews and Harry Panjer for their invaluable suggestions. This work was supported by a PhD grant from the Casualty Actuarial Society and Society of Actuaries.

REFERENCES

BALKEMA, AUGUST A., AND LAURENS DE HAAN. 1974. Residual Life Times at Great Age. *Annals of Probability* 2: 792–804.

BERNARD, JEUNE, AND JAMES W. VAUPEL. 1999. *Validation of Exceptional Longevity*. Odense, Denmark: Odense University Press.

BOURBEAU, ROBERT, AND BERTRAND DESJARDINS. 2002. Dealing with Problems in Data Quality for the Measurement of Mortality at Advanced Ages in Canada. *North American Actuarial Journal* 6(3): 1–13.

BOWERS, NEWTON L., JR., HANS U. GERBER, JAMES C. HICKMAN, DONALD A. JONES, AND CECIL J. NESBITT. 1997. *Actuarial Mathematics*. 2nd ed. Schaumburg, IL: Society of Actuaries.

BROUHNS, NATACHA, MICHEL DENUIT, AND INGRID VAN KEILEGOM. 2005. Bootstrapping the Poison Log-Bilinear Model for Mortality Forecasting. *Scandinavian Actuarial Journal* 3: 212–24.

COALE, ANSLEY, AND GUANG GUO. 1989. Revised Regional Model Life Tables at Very Low Mortality. *Population Index* 55: 614–43.

COALE, ANSLEY, AND ELLEN E. KISKER. 1990. Defects in Data on Old-Age Mortality in the United States: New Procedures for Calculating Mortality Schedules and Life Tables at the Highest Ages. *Asian and Pacific Population Forum* 4: 1–31.

GOMPERTZ, BENJAMIN. 1825. On the Nature of the Function Expressive of the Law of Human Mortality and on a New Mode of Determining Life Contingencies. *Royal Society of London, Philosophical Transactions, Series A* 115: 513–85.

HAN, ZHONGXIAN. 2005. Living to 100 and Beyond: An Extreme Value Study. In *Living to 100 and Beyond Symposium Monograph*. http://www.soa.org/ccm/content/research-publications/library-publications/monographs/life-monographs/living-to-100-and-beyond-monograph/.

HELIGMAN, LARRY, AND JOHN H. POLLARD. 1980. The Age Pattern of Mortality. *Journal of the Institute of Actuaries* 107: 437–55.

HILL, BRUCE M. 1975. A Simple Approach to Inference about the Tail of a Distribution. *Annals of Statistics* 3: 1163–74.

HIMES, CHRISTINE L., SAMUEL H. PRESTON, AND GRETCHEN A. CONDRAN. 1994. A Relational Model of Mortality at Older Ages in Low Mortality Countries. *Population Studies* 48: 269–91.

HUMAN MORTALITY DATABASE. 2006. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). www.mortality.org or www.humanmortality.de. Accessed May 15, 2006.

[2] See Robine and Vaupel (2002) for the number of verified cases of supercentenarians in different countries.

[3] A distribution function F belongs to the peaks-over-threshold domain of attraction if and only if $n\bar{F}(a_n x + b_n) \rightarrow -\ln(W(x))$ as $n \rightarrow \infty$, for all x, where W is the distribution function of a Generalized Pareto random variable, for some sequences of real numbers $\{a_n\}$ and $\{b_n\}$ (see Embrechts et al. 1997). This assumption is not overly restrictive and is satisfied by many commonly encountered distributions, including normal, lognormal, gamma, Pareto, Gompertz, and Weibull (see Zelterman 1993).

LEE, RONALD D., AND LAWRENCE R. CARTER. 1992. Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association* 87: 659–71.

LI, SIU-HANG, AND WAI-SUM CHAN. 2007. The Lee-Carter Model for Forecasting Mortality, Revisited. *North American Actuarial Journal* 7(1): 72–91.

LONDON, RICHARD L. 1985. *Graduation: The Revision of Estimates*. Winsted, CT: ACTEX Publications.

MASON, DAVID M. 1982. Laws of Large Numbers for Sums of Extreme Values. *Annals of Probability* 10: 754–64.

OLSHANSKY, S. JAY, AND BRUCE A. CARNES. 1997. Ever since Gompertz. *Demography* 34: 1–15.

PANJER, HARRY H., AND GIUSEPPE RUSSO. 1992. Parametric Graduation of Canadian Individual Insurance Mortality Experience: 1982–1988. *Proceedings of the Canadian Institute of Actuaries* 23: 378–449.

PANJER, HARRY H., AND KEN SENG TAN. 1995. *Graduation of Canadian Individual Insurance Mortality Experience: 1986–1992*. Ottawa: Canadian Institute of Actuaries.

ROBINE, JEAN-MARIE, AND JAMES W. VAUPEL. 2002. Emergence of Supercentenarians in Low-Mortality Countries. *North American Actuarial Journal* 6(3): 54–63.

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

VINCENT, PAUL. 1951. La Mortalité des Vieillards. *Population* 6: 181–204.

WATTS, KATHRYN A., DEBBIE J. DUPUIS, AND BRUCE L. JONES. 2006. An Extreme Value Analysis of Advanced Age Mortality Data. *North American Actuarial Journal* 10(4): 162–78.

WILMOTH, JOHN R., EVGUENI ANDREEV, DMITRI A. JDANOV, AND DANA A. GLEI. 2005. Methods Protocol for the Human Mortality Database. www.mortality.org.

Discussions on this paper can be submitted until October 1, 2008. The authors reserve the right to reply to any discussion. Please see the Submission Guidelines for Authors on the inside back cover for instructions on the submission of discussions.