UNCERTAINTY IN MORTALITY FORECASTING: AN EXTENSION TO THE CLASSICAL LEE-CARTER APPROACH

BY

JOHNNY SIU-HANG LI, MARY R. HARDY AND KEN SENG TAN

ABSTRACT

Traditionally, actuaries have modeled mortality improvement using deterministic reduction factors, with little consideration of the associated uncertainty. As mortality improvement has become an increasingly significant source of financial risk, it has become important to measure the uncertainty in the forecasts. Probabilistic confidence intervals provided by the widely accepted Lee-Carter model are known to be excessively narrow, due primarily to the rigid structure of the model. In this paper, we relax the model structure by considering individual differences (heterogeneity) in each age-period cell. The proposed extension not only provides a better goodness-of-fit based on standard model selection criteria, but also ensures more conservative interval forecasts of central death rates and hence can better reflect the uncertainty entailed. We illustrate the results using US and Canadian mortality data.

KEYWORDS

Confidence interval; heterogeneity; Lee-Carter; maximum likelihood estimation; parametric bootstrap.

1. INTRODUCTION

The estimation of uncertainty involved in mortality forecasts is of critical importance to the management of longevity (mortality improvement) risk. Unlike traditional diversifiable mortality risk, that is, the random variation around a fixed mortality probability, longevity risk affects the entire portfolio and thus cannot be mitigated by selling a large number of policies. On the other hand, there is a lack of mortality derivative securities that can be used for hedging longevity risk. Therefore, risk capital is often required to cushion against longevity risk, and such capital is, of course, determined by measures of uncertainty associated with mortality projections. The computation of measures of uncertainty requires a stochastic methodology; examples include the Lee-Carter model (Lee and Carter, 1992), the P-splines regression (Currie

Astin Bulletin 39(1), 137-164. doi: 10.2143/AST.39.1.2038060

et al., 2004) and the parameterized time-series approach (McNown and Rogers, 1989).

Since its introduction in 1992, the Lee-Carter model has been widely used in diverse actuarial and demographic applications. For instance, it has been used as the basis of stochastic forecasts of the finances of the U.S. social security system and other aspects of the U.S. federal budget (see Congressional Budget Office of the United States, 1998). It has also been successfully applied to populations in Scandinavia (Li and Chan, 2005) and the G7 countries (Tuljapurkar *et al.*, 2000). An advantage of the Lee-Carter model is that the number of parameters is small relative to other stochastic mortality models. The parsimonious model structure gives constraints to the behavior of future death rates, resulting in a stable age pattern of mortality in the projections. This effectively prevents mortality crossovers, that is, a non-monotonicity in adulthood and senescent mortality over age[1], and various anti-intuitive behaviors that may be encountered in some other stochastic approaches.

Nevertheless, the stringent model structure has been seen to generate overly narrow confidence intervals (see Lee and Miller, 2001). This narrowness may result in underestimation of the risk of more extreme outcomes, defeating the original purpose of moving on to a stochastic framework. This phenomenon is an example of model risk (see Cairns, 2000), which can be reduced by relaxing the model structure. One way to relax the Lee-Carter model structure is to consider extra bilinear terms (see, e.g., Renshaw and Haberman, 2003; Booth *et al.*, 2002). Although the extra bilinear terms can explain a portion of temporal variance that is not captured by the original model, the time-varying components in the extra bilinear terms are typically highly non-linear, making forecasting more complicated[2]. Another way to relax the model structure is to introduce a dispersion parameter. Renshaw and Haberman (2006) and Delward *et al.* (2007) introduced a single (non-age-specific) parameter to the original model, aiming at a better goodness-of-fit. However, they made no attempt to consider the impact of the additional parameter on the width of prediction intervals. Furthermore, as we demonstrate in Section 4, the variation of dispersion over age is highly significant, implying that the use of a single dispersion parameter is insufficient.

Our objective in this paper is to explore the feasibility of extending the model from a different angle, by considering individual heterogeneity at the cell level. In more detail, the implementation of any mortality model requires the division of the Lexis plane into cells. The original Lee-Carter model allows death rates to vary between cells, but not within each cell; that is, individuals within a single cell are assumed to be homogeneous. However, researchers have

[1] In principle, parameter b_x could be negative for some ages, indicating that mortality at those ages tends to rise when falling at other ages. In this situation, which might happen when data are sparse, a crossover may occur in a mortality projection. However, this situation does not seem to occur when the model is applied to national populations and large pension plans from which we have data with an adequate number of exposures.

[2] Section 2 gives a further discussion on this problem.

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

noted that individuals at the same age may differ substantially in their endowment for longevity, and that individual differences are important to population-based mortality studies (see, e.g., Hougaard, 1984; Vaupel *et al.*, 1979). The primary contribution of this paper is to relax the model structure by incorporating individual differences into the Lee-Carter model. To justify our contribution, the proposed extension must satisfy the following criteria.

1. *Provision of wider confidence intervals.* Taking account of additional variations, the interval forecasts should encompass a broader range of probable outcomes.
2. *Improvement of goodness-of-fit.* The improvement of fit should be significant enough that the introduction of additional parameters is worthwhile. We base our conclusions on standard model selection criteria.
3. *Retention of the appealing features in the original version.* Relaxation of the model structure should be done carefully so that criteria 1 and 2 can be satisfied without distorting the stability in the projected age patterns and the linearity in the time-varying component.

The rest of this paper is organized as follows: Section 2 provides a brief review of the Lee-Carter methodology; Section 3 specifies our proposed extension and provides theoretical justifications for it; this section also details how mean and interval forecasts of death rates and other demographic quantities can be computed using our proposed extension; Section 4 compares the performance of our proposed extension with the original model, using US and Canadian mortality data; finally Section 5 concludes the paper.

2. The Lee-Carter Model

The Lee-Carter model describes the central rate of death ($m_{x,t}$) at age x and time t by three series of parameters, $\{a_x\}$, $\{b_x\}$, and $\{k_t\}$, in the following way:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}, \tag{1}$$

where a_x is an age-specific parameter that indicates the average level of mortality at age x; $\{k_t\}$ is a time-varying parameter that signifies the general speed of mortality improvement; b_x is another age-specific parameter that characterizes the sensitivity of $\ln(m_{x,t})$ to changes in k_t; and $\varepsilon_{x,t}$ is the error term that captures all remaining variations.

Mortality forecasting is performed in two stages. In the first stage, we estimate $\{a_x\}$, $\{b_x\}$, and $\{k_t\}$ using historical mortality data. Note that all parameters on the right hand side of equation (1) are unobservable. As a result, we cannot fit the model by simple methods like the ordinary least squares. To solve this problem, researchers have proposed a few alternative approaches including the method of singular value decomposition considered by Lee and

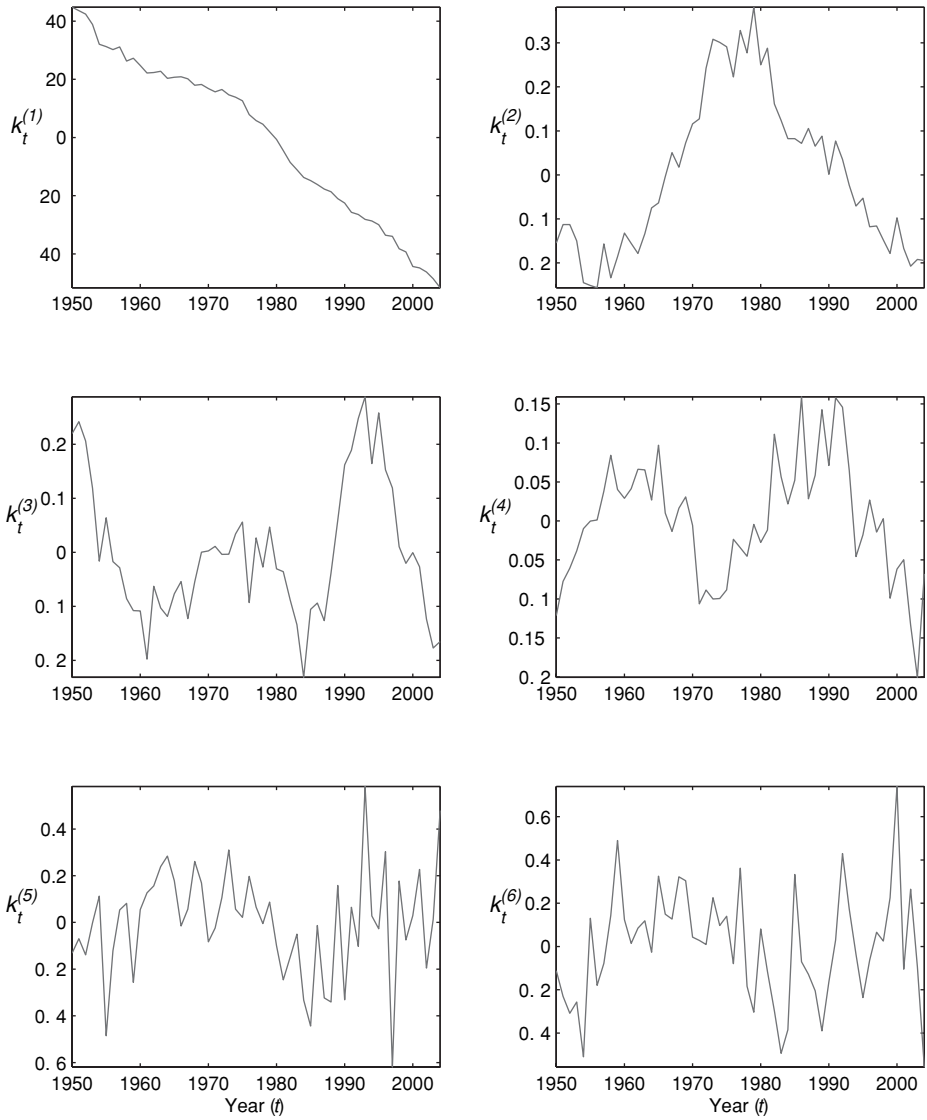


FIGURE 1: The i^{th} time-varying component, $k_t^{(i)}$, in the augmented Lee-Carter model fitted to Canadian mortality data.

Carter (1992), the method of maximum likelihood estimation implemented by Wilmoth (1993) and Brouhns *et al.* (2002), and the method of generalized linear models employed by Renshaw and Haberman (2006).

In the second stage, fitted values of k_t are modeled by an autoregressive integrated moving average (ARIMA) process, determined by the Box and Jenkins (1976) approach. Then we extrapolate k_t through the fitted ARIMA model to

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

obtain a forecast of future death rates. Let T be the forecast origin and $\hat{k}_{T+s}$ be the s-period ahead forecast of k_t. The s-period ahead forecast of $m_{x,t}$ is given by

$$\hat{m}_{x,T+s} = e^{\hat{a}_x + \hat{b}_x \hat{k}_{T+s}}. \tag{2}$$

In recent years, researcher have extended the original Lee-Carter model by introducing extra bilinear terms. Their extension generalizes equation (1) to

$$\ln(m_{x,t}) = a_x + \sum_{i=1}^{p} b_x^{(i)} k_t^{(i)} + \varepsilon_{x,t}. \tag{3}$$

For instance, Renshaw and Haberman (2003) considered $p = 2$ and Booth *et al.* (2002) considered $p = 5$. Although the extra bilinear terms can improve the goodness-of-fit to the historical data, they make forecasting more complicated because the additional time-varying components $k_t^{(i)}$, $i = 2, 3, \ldots, p$, are highly non-linear (see Figure 1), which may not be well handled by ARIMA processes.

3. The Proposed Extension

In the original Lee-Carter model, it is assumed that the observed number of deaths is a realization of the Poisson distribution with mean equal to the expected number of deaths under the Lee-Carter model; that is,

$$\Pr[D_{x,t} = y] = \frac{\lambda_{x,t}^{y} e^{-\lambda_{x,t}}}{y!}, \tag{4}$$

$\lambda_{x,t} = \mathrm{E}(D_{x,t}) = E_{x,t} e^{a_x + b_x k_t}$; $D_{x,t}$ and $E_{x,t}$ are the number of deaths and exposures-to-risk at age x and time t, respectively.

There are several drawbacks associated with the original setting. In assuming Poisson models, we are imposing the mean-variance equality restriction on the random variable $D_{x,t}$. In practice, however, the variance of $D_{x,t}$ can be greater than the mean. This situation is referred to as overdispersion. McCullagh and Nelder (1989) pointed out that overdispersion is commonplace and therefore it should be assumed to be present to some extent unless it is shown to be absent. If overdispersion exists, the analysis of data using a single parameter distribution such as Poisson will result in overestimating the degree of precision. In other words, in mortality forecasting, the interval forecasts of death rates and other demographic quantities under the assumption of Poisson death counts will be overly narrow. Cox (1983) also pointed out that there is a possible loss of efficiency if a single parameter distribution is used when overdispersion exists.

Motivated by the mean-variance equality restriction in the original Lee-Carter model, we consider the modeling distribution for the number of deaths. In the original model, it is assumed individuals in each age-period cell are

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

homogeneous and have the same probability of death. However, other than age and time, there are various factors affecting human mortality, for example, ethnicity, education, occupation, marital status and obesity (Brown and McDaid, 2003). These factors may divide the age-period cells into clusters, so that individuals in the same cell but different clusters will have different likelihood of death. The presence of clustering will not only violate the assumption of homogeneity, but also induce extra variation that is not reflected in the interval forecasts based on the original Lee-Carter model.

To account for the possibility of clustering, we segregate each age-period cell into smaller clusters of equal size. The number of clusters N_x in an age-period cell is age-specific and is assumed to be non-random. That is, the i^{th} cluster will have $E_{x,t}/N_x$ exposures-to-risk and $D_{x,t}(i)$ deaths, where $i = 1, 2, \ldots, N_x$. The total number of deaths, $D_{x,t}$, in an age-period cell is therefore given by

$$D_{x,t} = \sum_{i=1}^{N_x} D_{x,t}(i). \tag{5}$$

We assume that $D_{x,t}(i)$ and $D_{x,t}(j)$ are independent for all $i \neq j$, and that

$$D_{x,t}(i) \,|\, z_x(i) \sim \text{Poisson}\left(z_x(i) \frac{\lambda_{x,t}}{N_x} \right), \tag{6}$$

where $z_x(i)$ is an age-specific random variable that accounts for heterogeneity of individuals. We further assume that $\mathrm{E}(z_x) = 1$, which implies that on average, the clusters have the same mortality level as the cell to which they belong.

Note that $z_x(i)$ varies from cluster to cluster. When $z_x(i) > 1$, individuals in cluster i are more frail than the overall, and similarly when $0 < z_x(i) < 1$, individuals in cluster i are less frail. Although any distribution with a positive support can be a candidate for modeling $z_x(i)$, here we assume a Gamma distribution for mathematical tractability. In fact, Gamma distributions are often utilized in modeling heterogeneity (see, e.g., Hougaard, 1984; Vaupel *et al.*, 1979; Wang and Brown, 1998).

Assuming that $z_x(i)$ follows a Gamma distribution with $\mathrm{E}[z_x(i)] = 1$ and $\mathrm{Var}[z_x(i)] = \iota_x$, where $\iota_x > 0$, we can easily show that the probability mass distribution for $D_{x,t}(i)$ is given by

$$\Pr[D_{x,t}(i) = y] = \frac{\Gamma\left(y + \iota_x^{-1}\right)}{y!\,\Gamma\left(\iota_x^{-1}\right)} p^{\iota_x^{-1}} (1-p)^y, \tag{7}$$

and that the moment generating function for $D_{x,t}(i)$ can be expressed as

$$M_{D_{x,t}(i)}(z) = \left(\frac{p}{1-(1-p)e^z} \right)^{\iota_x^{-1}}, \tag{8}$$

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

where $p = \frac{\iota_x^{-1}}{\lambda_{x,t}/N_x + \iota_x^{-1}}$.

Given that $D_{x,t} = \sum_{i=1}^{N_x} D_{x,t}(i)$ and that $D_{x,t}(i)$ and $D_{x,t}(j)$ are independent for all $i \neq j$, we have the following moment generating function for $D_{x,t}$:

$$M_{D_{x,t}}(z) = \left(\frac{p}{1-(1-p)e^z}\right)^{N_x \iota_x^{-1}}, \tag{9}$$

which immediately implies that $D_{x,t}$ follows the Negative Binomial distribution with the following probability mass function:

$$\Pr[D_{x,t} = y] = \frac{\Gamma\left(y + \alpha_x^{-1}\right)}{y!\,\Gamma\left(\alpha_x^{-1}\right)} \left(\frac{\lambda_{x,t}}{\lambda_{x,t} + \alpha_x^{-1}}\right)^y \left(\frac{\alpha_x^{-1}}{\lambda_{x,t} + \alpha_x^{-1}}\right)^{\alpha_x^{-1}}, \tag{10}$$

where $\alpha_x = \iota_x / N_x$.

Summing up, the introduction of Gamma-distributed heterogeneity in age-period cells is equivalent to the assumption that $D_{x,t}$ follows a Negative Binomial distribution instead of a Poisson one. Note that

$$\mathrm{E}(D_{x,t}) = \lambda_{x,t} = E_{x,t} e^{a_x + b_x k_t}, \tag{11}$$

which means our generalization still complies with the Lee-Carter specification. More importantly, we have

$$\mathrm{Var}(D_{x,t}) = \mathrm{E}(D_{x,t}) + \alpha_x [\mathrm{E}(D_{x,t})]^2, \tag{12}$$

which means that our generalization explicitly allows for overdispersion. The dispersion parameters ($\{\alpha_x\}$) are age-specific, allowing for different degrees of overdispersion at different ages. As the mean-variance equality is relaxed, measures of uncertainty obtained from the proposed extension can capture a large part of the variation that is ignored in the original model. Note that the limiting case $\alpha_x \to 0$ yields a Poisson distribution.

Furthermore, the proposed extension gives an additional series of parameters to the model without altering the structure specified by equation (1). As a result, the desirable features of the original model are preserved.

We may use maximum likelihood to estimate the model parameters. The log-likelihood function is given by

$$l = \sum_{x,t} \left\{ \left[\sum_{i=0}^{D_{x,t}-1} \ln\left(\frac{1+\alpha_x i}{\alpha_x}\right) \right] + D_{x,t} \ln\left(\alpha_x \lambda_{x,t}\right) - \left(D_{x,t} + \alpha_x^{-1}\right) \ln\left(1 + \alpha_x \lambda_{x,t}\right) \right\} + c, \tag{13}$$

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

where c is a constant that is free of a_x, b_x, k_t, and α_x. Note that the model is overparameterized. For example, if $\{a_x\}$, $\{b_x\}$, $\{k_t\}$, and $\{\alpha_x\}$ form one set of parameters for the model, then $\{a_x\}$, $\{b_x/A\}$, $\{Ak_t\}$, and $\{\alpha_x\}$ will be an exactly equivalent set for any constant A. In most applications of the model including the pioneering work of Lee and Carter (1992) for the US population, a unique representation is obtained by setting the sum of k_t and b_x to zero and one, respectively. The Newton's procedure for maximizing the log-likelihood function subject to the constraints for parameter uniqueness is provided in Appendix 1.

However, if we set $\sum_t k_t = 0$ and $\sum_x b_x = 1$, the model will not fit the age-specific mortality data exactly at the forecast origin; that is, $m_{x,T}$ is not necessarily equal to $e^{a_x + b_x k_T}$. This situation would inevitably lead to error, which would be especially important in the early years of forecast. Bell (1997) and Lee and Miller (2001) noted that the error at the forecast origin caused significant bias in the forecasts for the first decade. This problem can be resolved by letting $a_x = m_{x,T}$ and $k_T = 0$, thereby fitting the age-specific mortality data at the forecast origin exactly. While the correction will not alter the log-likelihood function, it will require some modifications to the Newton's procedure for maximizing the log-likelihood[3]. We integrate the forecast origin correction into our proposed extension. The revised Newton's procedure is given in Appendix 2.

Having estimated the model parameters, we can fit an ARIMA process to $\{k_t\}$ and extrapolate k_t to the future. The mean forecasts of future central death rates can be obtained by using equation (2).

The computation of interval forecasts is not straightforward. Here we extend the parametric bootstrap procedure proposed by Brouhns *et al.* (2005) to form the following algorithm for computing interval forecasts based on our proposed extension.

1. Simulate N realizations from the Negative Binomial distribution specified by equation (10).
2. For each of these N realizations:
   (a) re-estimate $\{a_x\}$, $\{b_x\}$, and $\{k_t\}$ using maximum likelihood;
   (b) specify a new ARIMA process for the re-estimated $\{k_t\}$;
   (c) simulate future values of k_t, that is, k_{T+s}, $s = 1, 2, \ldots$, using the newly specified ARIMA process;
   (d) compute future values of $m_{x,t}$, that is, $m_{x,T+s}$, $s = 1, 2, \ldots$, using the re-estimated $\{a_x\}$ and $\{b_x\}$, and the simulated future values of k_t.
3. Step (2) gives an empirical distribution of $m_{x,T+s}$ for all x and $s = 1, 2, \ldots$. The 2.5th and the 97.5th percentiles of the empirical distribution respectively gives the lower and upper limit of the 95% interval forecast of $m_{x,T+s}$.

[3] When we apply the correction, parameter a_x is treated as a constant rather than a free parameter. Therefore, even though the log-likelihood function remains unchanged, the resulting log-likelihood value will be lower due to the additional constraints imposed on the model parameters.

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

It is noteworthy that the above algorithm allows both sampling fluctuations in the model parameters and stochastic error in the forecast of k_t be included in the interval forecasts. In addition, the algorithm does not require the assumption of normality and this makes an asymmetric confidence interval possible. Furthermore, the algorithm allows us to obtain interval forecasts for other variables such as life expectancies and annuity pure premiums. In more detail, the values of $m_{x,T+s}$, $s = 1, 2, \ldots$, in Step (2(d)) can be used to compute empirical distributions and hence confidence intervals for life expectancies and annuity values.

4. Two Examples

4.1. The data

In this section, we evaluate the performance of our proposed extension using US and Canadian mortality data. These data have a wide range of applications, particularly in the area of social security finances (see, e.g., Congressional Budget Office of the United States, 1998; Lee and Anderson, 2005).

In recent years, mortality derivatives have become increasingly popular. Most of these securities, such as JP Morgan's 'Q-forward,' are linked to the mortality of national populations. We believe that our illustrations, which are based on US and Canadian data, would be useful for users of these securities to assess the potential hedging error involved in a hedging exercise.

For both populations, the number of deaths ($D_{x,t}$) and exposures-to-risk ($E_{x,t}$) by single year of age from 0 to 99 are provided by the Human Mortality Database (2006). As in Tuljapurkar *et al.* (2000), we consider the post-war period from year 1950 to 2004.

4.2. Parameter estimates

We fit both the original model and our proposed extension (with the forecast original correction) to the historical data. Figures 2 and 3 display the parameter estimates for the United States and Canadian populations, respectively. Note that $\{a_x\}$ from both models are identical as we have imposed the constraint $a_x = m_{x,T}$ in the forecast origin correction. Our proposed extension has a minor effect on $\{b_x\}$ and $\{k_t\}$. Nevertheless, the long-term stability in the time-varying component $\{k_t\}$ is well preserved.

Next, we analyze the series of dispersion parameters ($\{\alpha_x\}$) that we introduced to the original model. It is clear that the values of α_x are not uniform over age. This observation has two important implications: (1) the degree of heterogeneity (individual differences) varies significantly with age; (2) it is insufficient to handle over-dispersion in the Lee-Carter model with a single (non-age-specific) dispersion parameter.

We observe a few peaks in the estimates of $\{\alpha_x\}$ for both populations. Note that the higher the value of α_x is, the more individual differences at age x. The

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

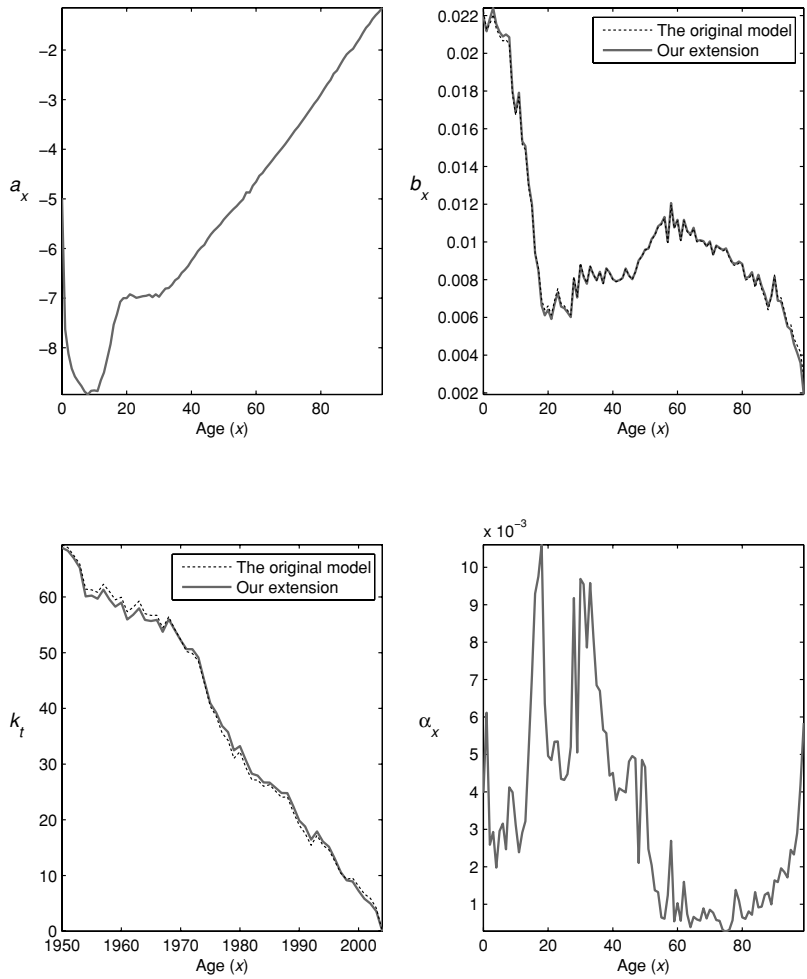


FIGURE 2: Estimates of parameters a_x, b_x, k_t, and α_x for the United States population.

first peak at age 0 to 1 may be attributed to the fact that infant mortality is divided into many categories[4]. In addition, there exist other factors, for example, birth weight, that have been proven to be influential to mortality in infancy and childhood (Prager, 1994). The second peak, which is centered at age 20, can be explained by the mortality differentials due to occupational differences

[4] Infant mortality is divided into three main categories: (1) perinatal mortality only includes deaths between the foetal viability (28 weeks gestation) and the end of the seventh day after delivery; (2) neonatal mortality only includes deaths in the first 28 days of life; (3) post-neonatal death only includes deaths after 28 days of life but before one year. These three mortality rates are significantly different (see, e.g., World Health Organization, 2006).

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

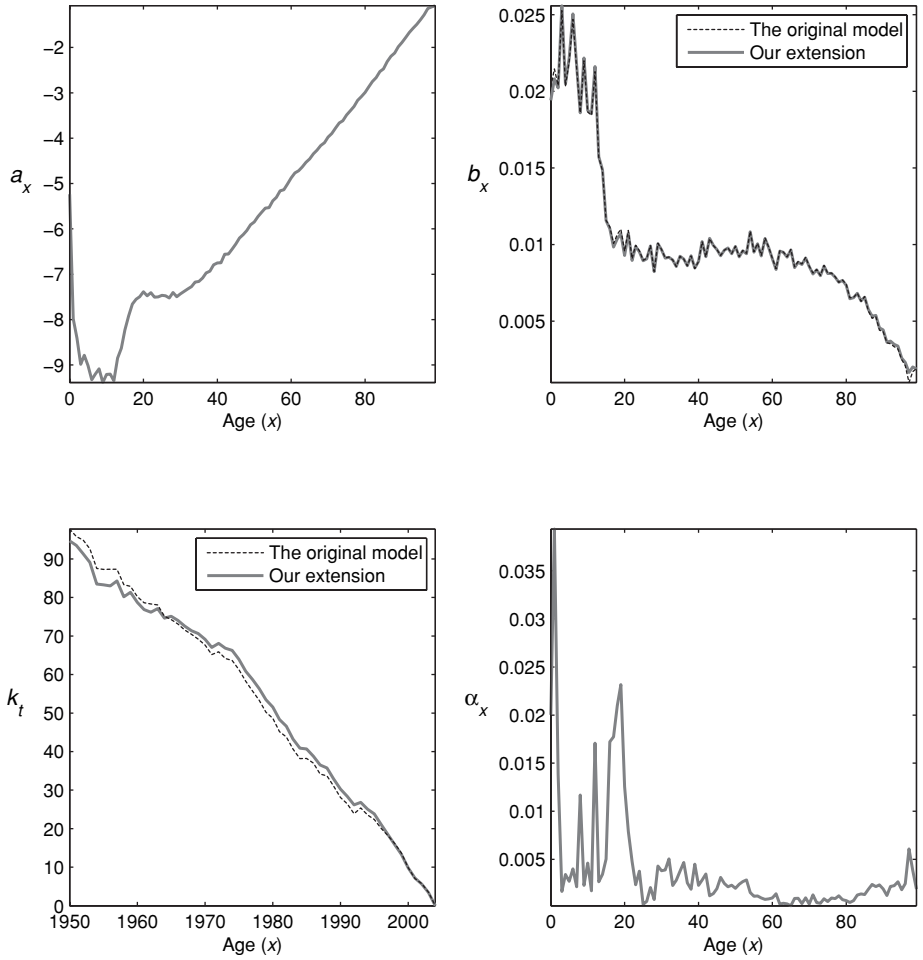


FIGURE 3: Estimates of parameters a_x, b_x, k_t, and α_x for the Canadian population.

(Burnett *et al.*, 1997). The final peak at around age 95 can be justified by the classification of the elderly by their activity of daily living (ADL) limitations (Kassner and Jackson, 1998). There is empirical evidence for the positive relationship between mortality and the number of ADL limitations (Pritchard, 2006).

4.3. Goodness-of-fit

Here we give, on the basis of the log-likelihood, a formal evaluation of the goodness-of-fit. The log-likelihood function of our proposed extension is given in equation (13), while that of the original model can be found in Brouhns *et al.*

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

(2002). Table 1 compares the log-likelihood of the two models. The significant increase in log-likelihood suggests that our proposed extension provides a better fit to the historical data. However, under the principal of parsimony, we should make use of the least possible number of parameters for adequate representations, and it is therefore inappropriate to base the conclusion only on the increase in log-likelihood, as we have introduced an additional series of parameters ($\{\alpha_x\}$). To account for the extra parameters, we can use the following model selection criteria.

1. Akaike Information Criterion (AIC) (Akaike, 1974), defined by $l - j$, where l is the log-likelihood and j is the number of parameters. AIC takes account of the increase in the number of parameters. Models with a higher value of AIC are more preferable.
2. Schwarz-Bayes Criterion (SBC) (Schwarz, 1978), defined by $l - 0.5j\ln(n)$, where n is the number of observations. The intuitions of SBC and AIC are similar, but SBC adjusts for sample size. Again, we prefer models with a higher SBC.
3. Likelihood-ratio test (LRT) (see, e.g., Klugman *et al.* 2004). The null hypothesis of LRT is that there is no significant improvement in the more complex model. Let l_1 and l_2 be the log-likelihood of the original model and our proposed extension, respectively. The test statistic is $2(l_2 - l_1)$. Under the null hypothesis, the test statistic has a chi-square distribution, with degrees of freedom equal to the number of additional parameters.

The values of AIC and SBC (see Table 1) indicates that our proposed extension gives a better fit than the original model even after penalizing for the extra parameters. The p-values of the LRT for both populations are less than 10^{-6},

TABLE 1

COMPARISON OF SELECTION INFORMATION FOR THE ORIGINAL MODEL AND OUR PROPOSED EXTENSION.

	The original model	**Our proposed extension**
	The United States	
Number of Parameters	255	355
Log-likelihood	–125,980	–41,886
AIC	–126,230	–42,241
SBC	–127,080	–43,415
	Canada	
Number of Parameters	255	355
Log-likelihood	–40,730	–29,734
AIC	–40,985	–30,089
SBC	–40,828	–31,263

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

providing further evidence for the improved goodness-of-fit. Although the new model specification introduces more flexibility, the projected age patterns of mortality remains stable, as shown in Figure 4. From Figure 4 we also observe that the proposed extension makes little change to the central mortality projections. However, the resulting interval forecasts, which we will show in Section 4.4, are substantially different.

4.4. Interval forecasts

Given the parameter estimates, we compute forecasts of future central death rates using the parametric bootstrap procedure discussed in Section 3. In Figures 5 and 6, we present the results for some representative ages. Note that the forecast origin correction has resulted in an exact fit in year 2004 when the forecasts were made.

The interval forecasts from our proposed extension are significantly wider than that from the original model. The increase in width varies by age, and depends on the degree of overdispersion in the historical data. For Canada, the increase ranges from 25% to 92%; and for the United States, it ranges from 18% to more than 180%. This observation agrees with our assertion that the mean-variance equality restriction in the original model has lead to understating the variations.

The increase in width has important implications to insurers. To illustrate, let us suppose that an insurer requires an estimate of m_{30} in 2014 (10 year from the forecast origin) for pricing a term-life insurance product. Based on the US mortality data, the original model gives the insurer a mean estimate of 0.000834 and a 95% confidence interval of (0.000633, 0.001060). The difference $0.001060 - 0.000834 = 0.000226$ can be considered a probabilistic margin for adverse deviation. If the original model is correct, this margin ensures that there remains no more than 2.5% chance that the mortality is underestimated. However, under our proposed extension, which models the uncertainty involved more realistically, the margin of 0.000226 is far less prudent: the chance that the mortality is underestimated is 24%, which is almost 10 times the original probability!

Readers are reminded the limitations of our extension. First, even though the model structure is relaxed, model risk still exists. For instance, our extension has not taken into account the possibility of outliers and structural changes in the time-varying component, $\{k_t\}$. Therefore, forecasters are still required to monitor the experience and re-calibrate the model from time to time. Second, the statistical tests can only tell us that, given the data we consider, our proposed extension gives a better fit than the original model. They do not indicate the model provides an adequate fit to the data, and do not imply our proposed extension outperforms other types of mortality models. We refer the interested readers to Cairns et al. (2009) for a comparison of the goodness-of-fit provided by various stochastic mortality models.

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

4.5. Backtesting

A way to test a projection model is to carry out backtesting, that is, to consider what results would have been produced if the model had been used in the past[5]. We use this approach to test the original model and our proposed extension, particularly on the interval forecasts they generate. Using this approach, we rebase the mortality projections only on the data as at 1994 (i.e., 10 years before the original forecast origin), starting the forecasts in 1995. Then we compare the "forecasts" with the actual values.

We consider three measures in the backtesting exercise; they are (1) the expectancy of life at birth ($\mathring{e}_0$), (2) the pure premium of a unit benefit whole life insurance sold to a life-aged-30 (A_{30}), and (3) the actuarial present value of a whole life annuity due of \$1 sold to a life-aged-60 ($\ddot{a}_{60}$). Measure (1) is often used as a convenient summary of the mortality of a population, while measures (2) and (3) have a wide range of actuarial applications. In Figures 7 to 9 we display the results for both populations.

The confidence level of 95% means that there is a 95% probability that the actual value is covered by the confidence interval. In other words, it is expected that, in each of the diagrams in Figures 7 to 9, there is no more than one true value lying outside the interval. However, in the original Lee-Carter forecasts of A_{30} and $\ddot{a}_{60}$ for the Canadian population, the coverage of actual values is less than 60%. In the forecast of A_{30} for the United States population, there are also two true values lying outside the interval based on the original model. The empirical results provide further evidence that the original Lee-Carter model tends to understate the uncertainty involved in estimating future mortality.

On the other hand, for all three measures we considered, the 95% probability bounds based on our proposed extension contain all actual values. The empirical results support the use of our proposed extension, particularly in applications for which the uncertainty associated with the central projections is important.

[5] We refer readers to Dowd *et al.* (2008) for a discussion on backtesting stochastic mortality models.

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

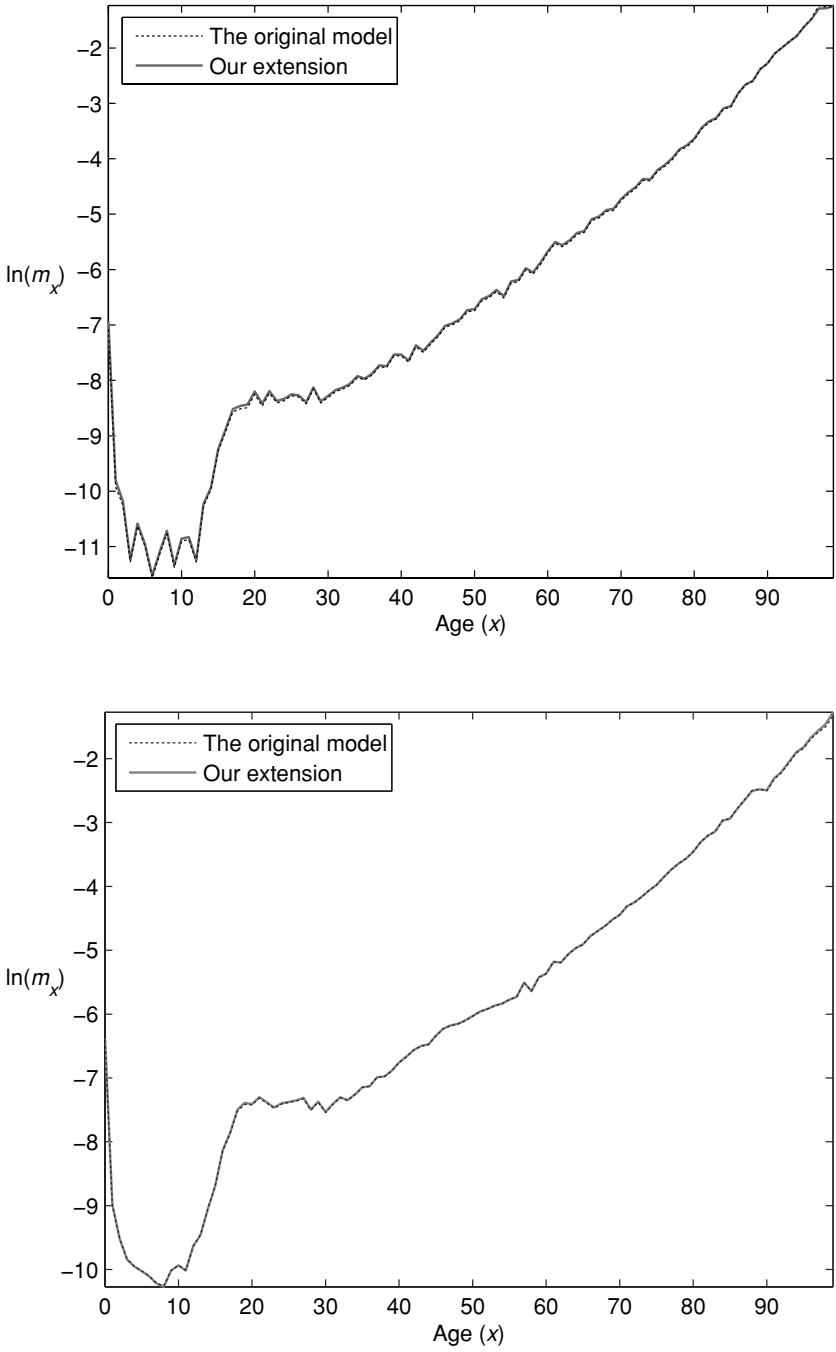


FIGURE 4: Projected age patterns of mortality in 2054 for the United States and Canadian populations.

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

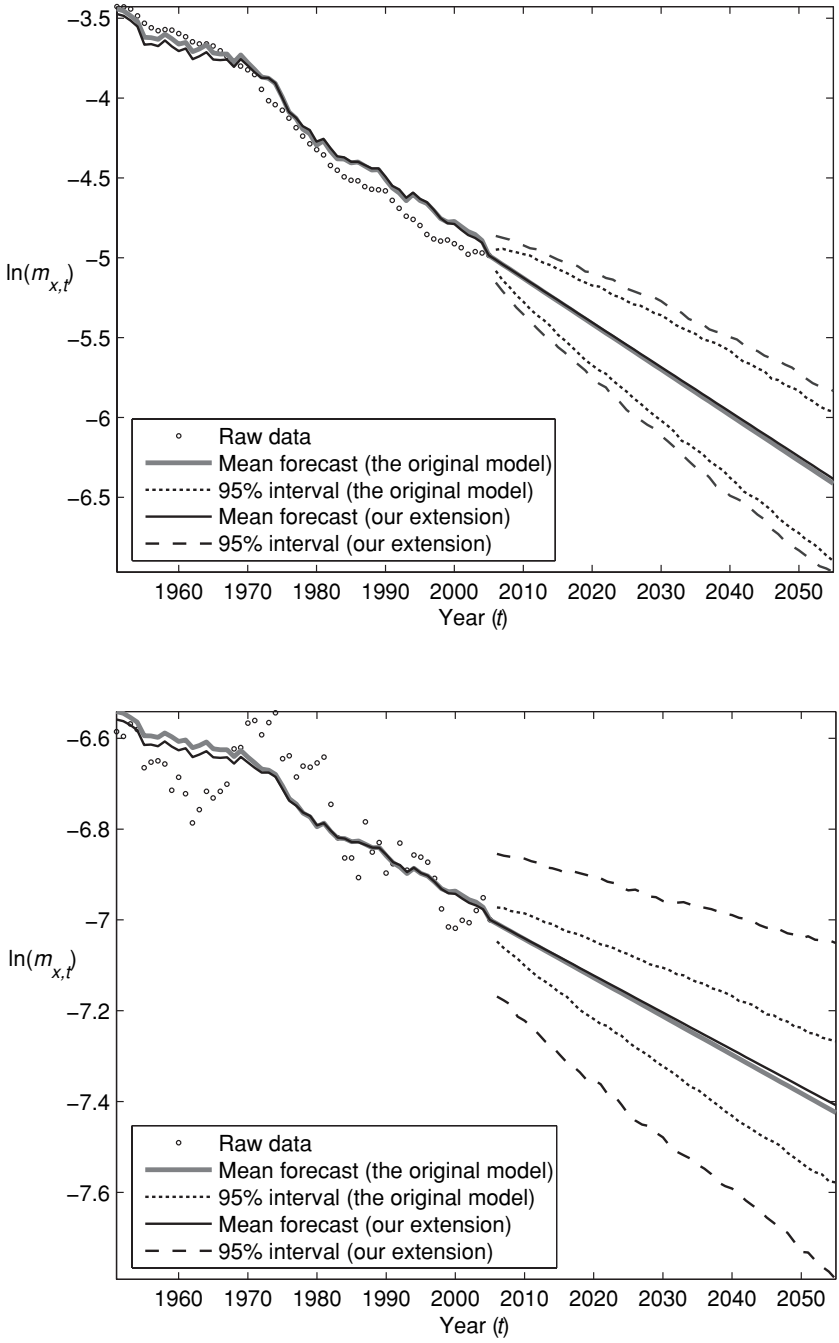
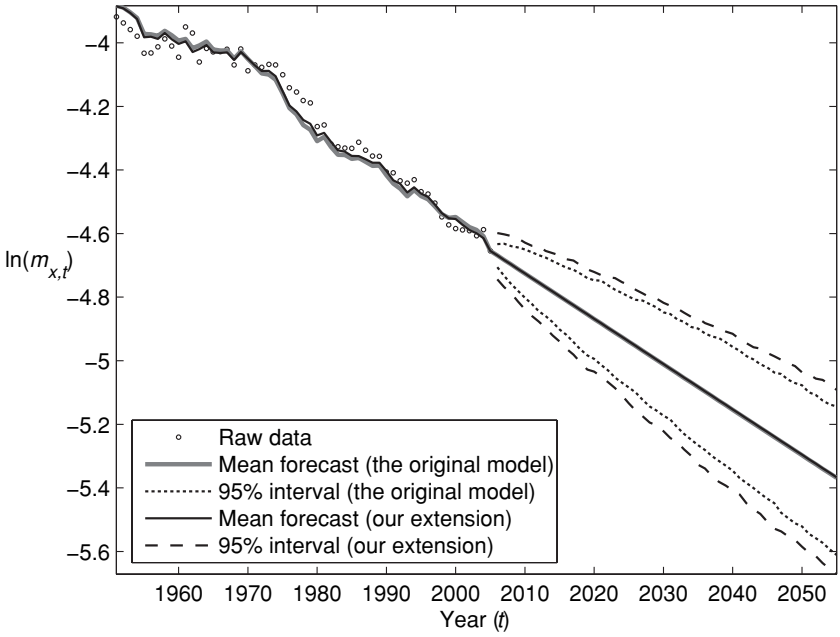


FIGURE 5: The fit (1950-2004) and projection (2005-2054) of central death rates at representative ages, the United States population.

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press



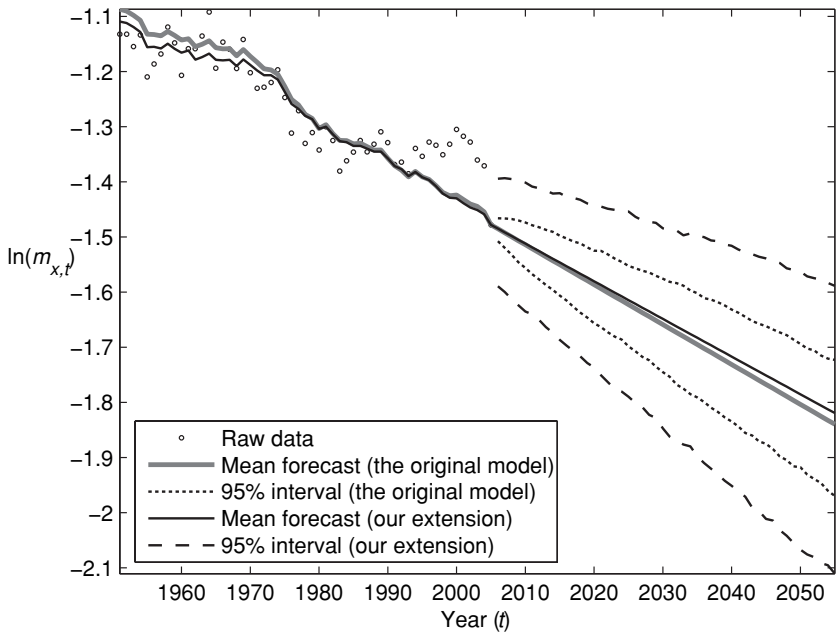


FIGURE 5: (cont'd).

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

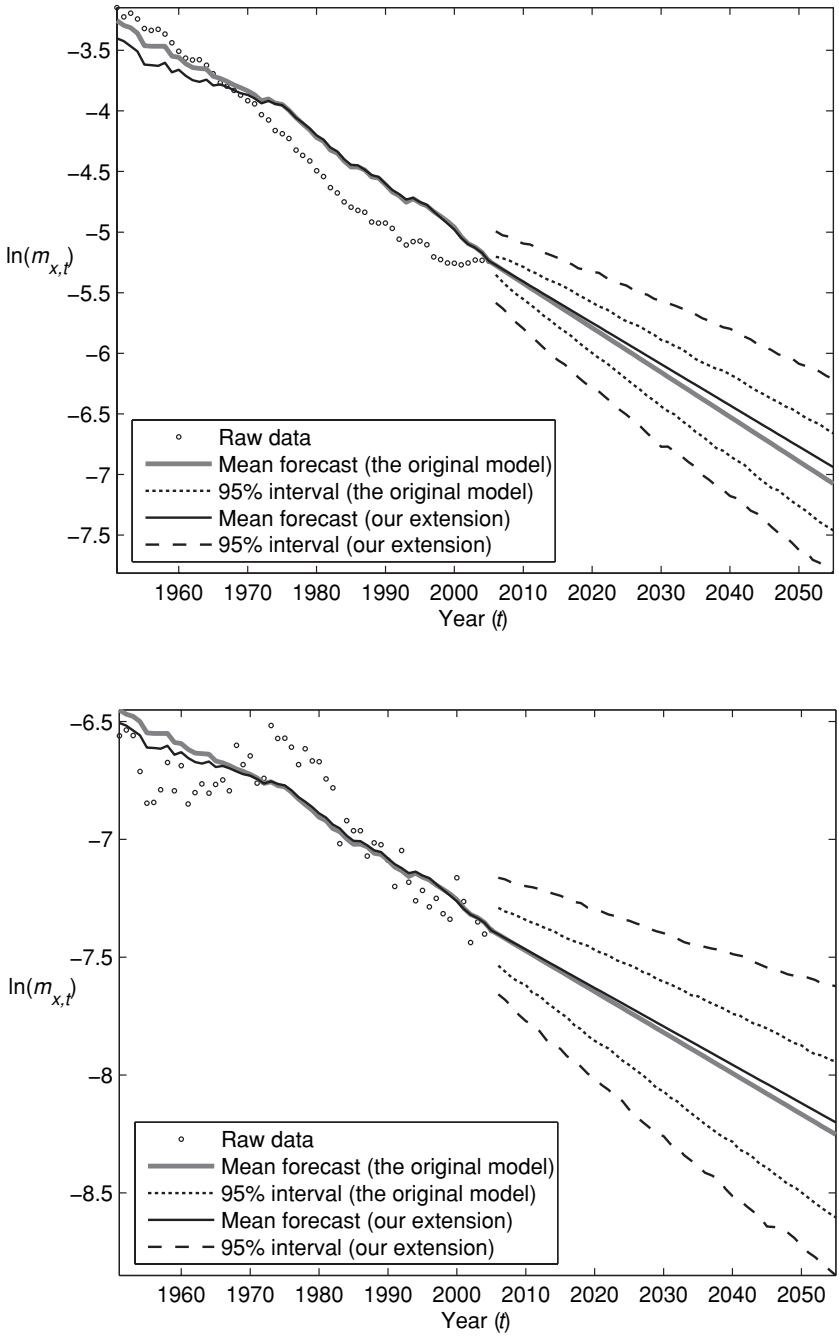


FIGURE 6: The fit (1950-2004) and projection (2005-2054) of central death rates at representative ages, the Canadian population.

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

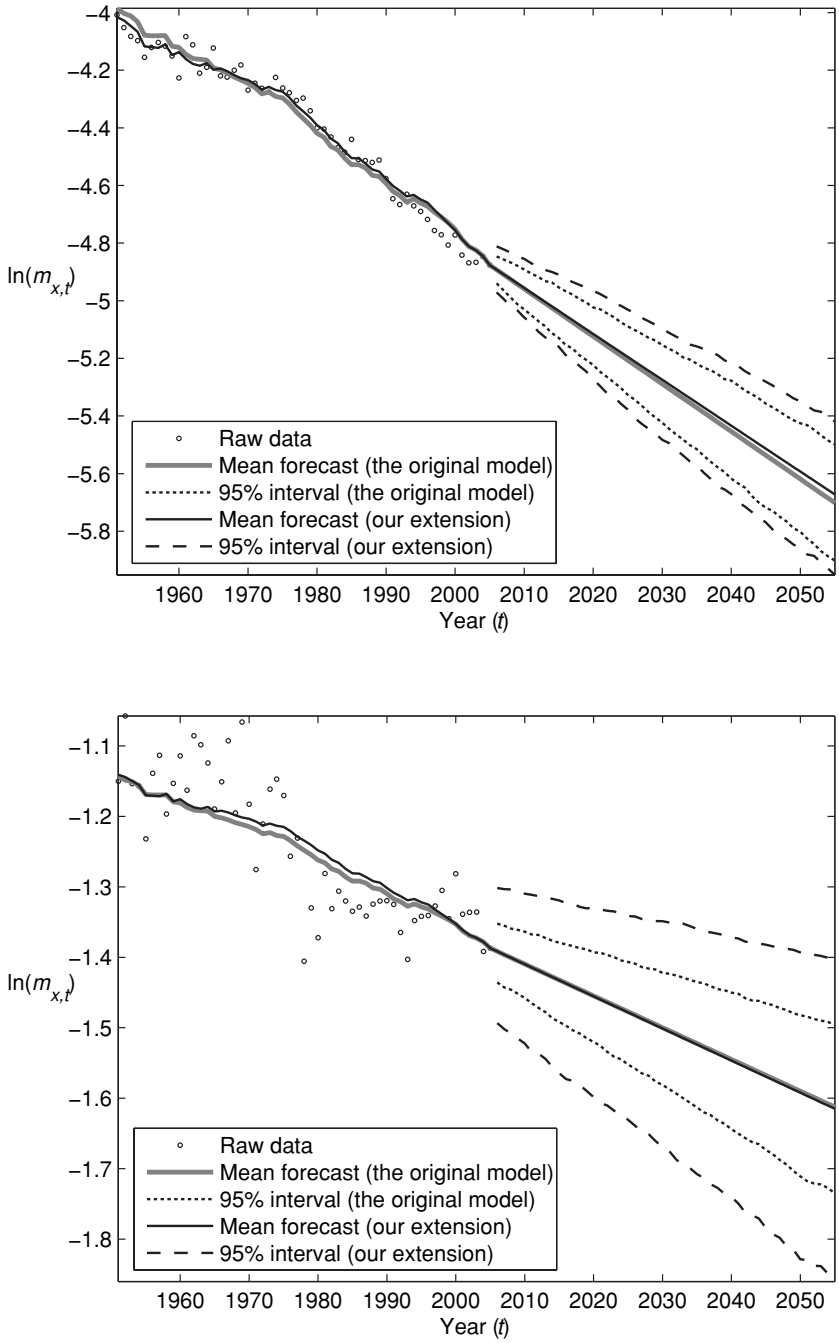


FIGURE 6: (cont'd).

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

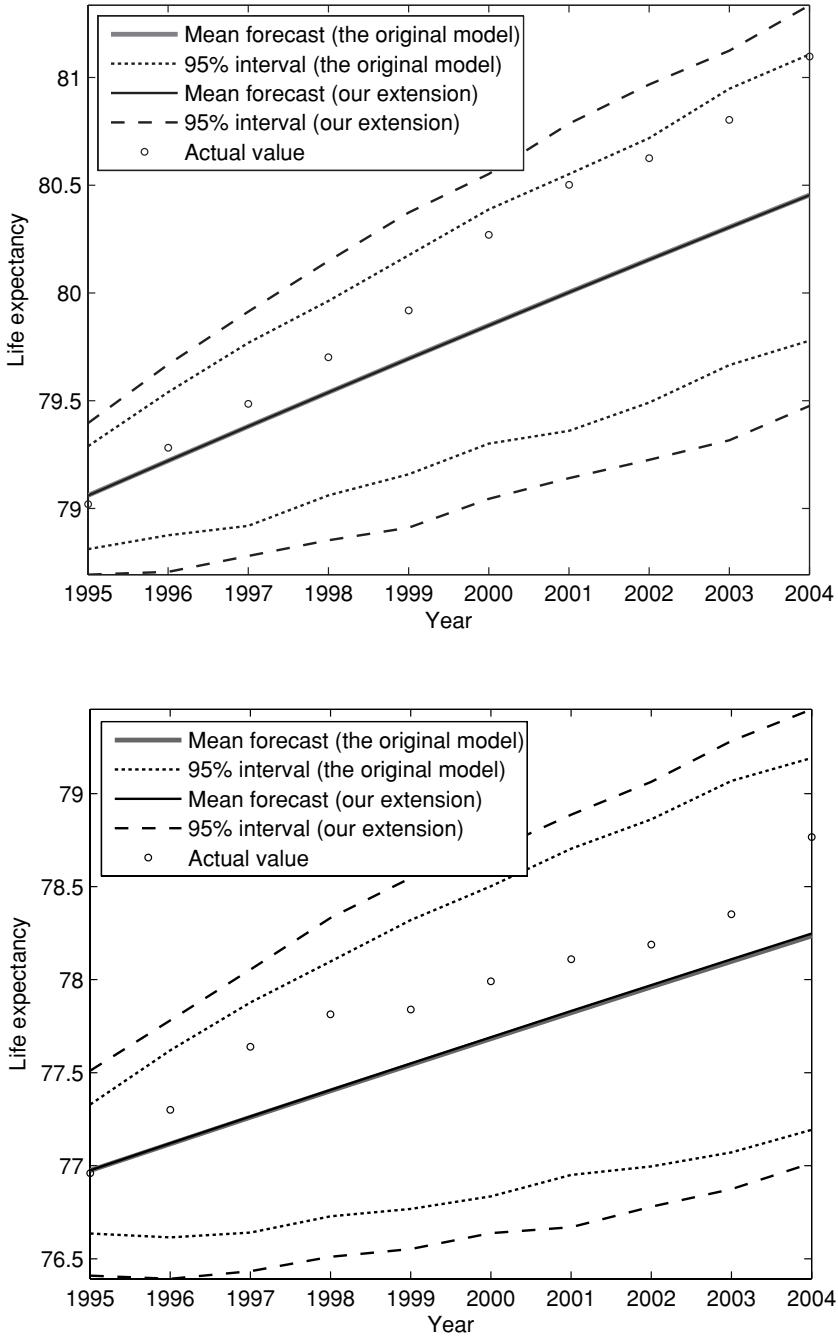


FIGURE 7: Projected life expectancy at birth ($\mathring{e}_0$), 1995-2004.

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

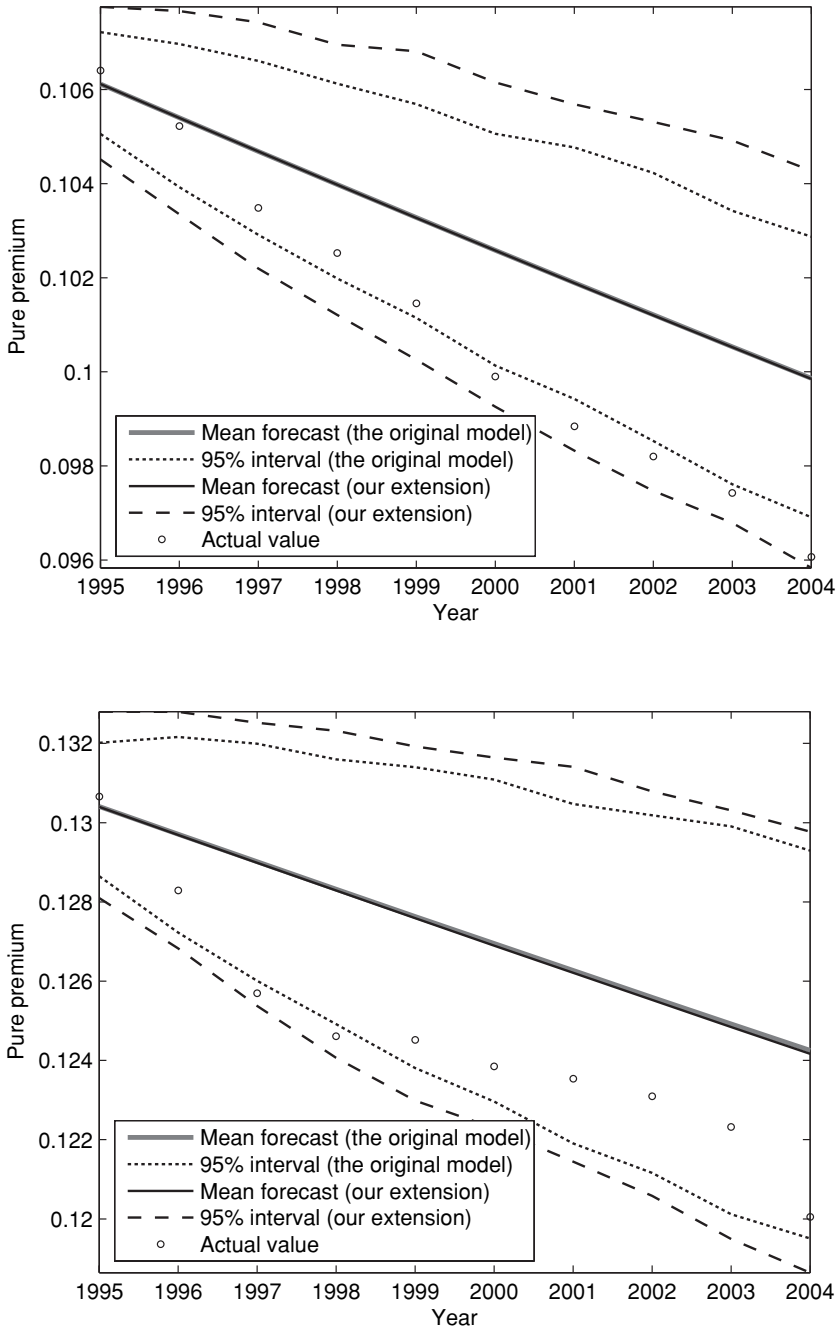


FIGURE 8: Projected pure premium of a unit-benefit whole life insurance sold to a life-aged-30 (A_{30}), 1995-2004.

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

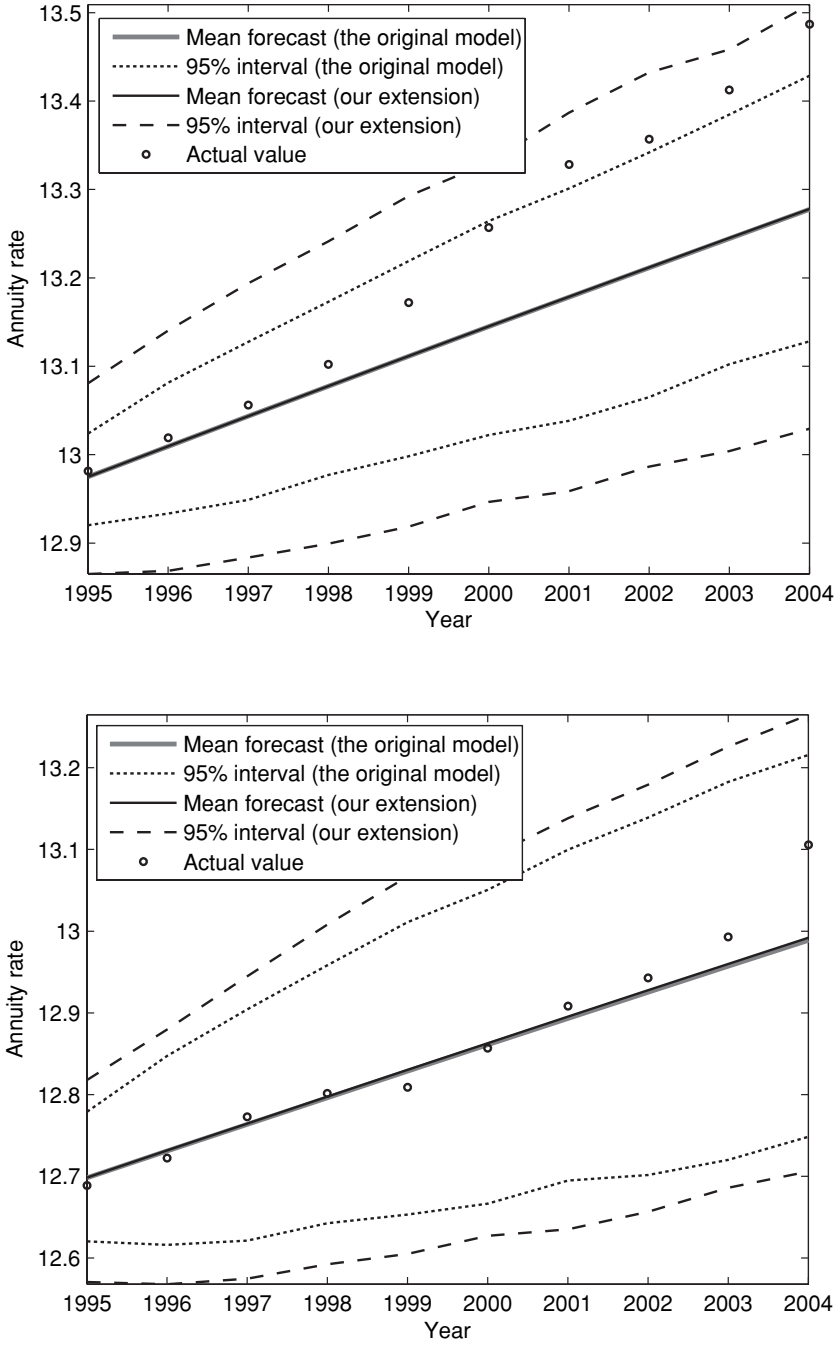


FIGURE 9: Projected actuarial present value of a whole life annuity due of \$1 sold to a life-aged-60 ($\ddot{a}_{60}$), 1995-2004.

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

5. Concluding Remarks

We have demonstrated that confidence intervals from the original Lee-Carter model are not sufficiently prudent, which can result in underestimating the capital required for cushioning against longevity risk. We addressed this issue by adapting the original Lee-Carter model for a better measurement of uncertainty.

The adaptation incorporates the uncertainty that arises from individual differences other than age and year of birth into the modeling of future mortality dynamics. The adapted model yields significantly wider (more conservative) confidence intervals, taking account of the uncertainty that is ignored in the original version due to the assumption of homogeneity of individuals in each age-period cell. In the back-testing exercise, we found that the adapted model gives a more realistic measurement of the variability of mortality rates in the past decade.

From a technical viewpoint, the adapted model allows for overdispersion through the age-specific dispersion parameters, $\{\alpha_x\}$. The additional parameters are justified for several reasons. First, the adapted model gives a better goodness-of-fit to historical data even if the increase in number of parameters is penalized. Second, since the new parameters do not alter the main model specification, the adaptation preserves the long-term stability of $\{k_t\}$ and age-patterns of mortality. Third, the dispersion parameters, which originate from the probability distribution function for modeling individual differences, have an intuitive interpretation. They convey invaluable information about the relative levels of heterogeneity at different ages.

Overdispersion in the Lee-Carter model has been considered previously by Renshaw and Haberman (2005) and Delward *et al.* (2007). Renshaw and Haberman set $\mathrm{Var}(D_{x,t}) = \phi \mathrm{E}(D_{x,t})$, where ϕ is a non-age-specific scale parameter, while Delward *et al.* use a Negative Binomial distribution with a scale parameter κ. Renshaw and Haberman's method suffers from the problem that the relationship between $\mathrm{E}(D_{x,t})$, $\mathrm{Var}(D_{x,t})$ and the probability function of $D_{x,t}$ is internally inconsistent. It is noteworthy that in both methods, the allowance of overdispersion entirely depends on a single parameter. However, as shown in Section 4.2, overdispersion in the Lee-Carter model is far from being constant over age. It is clear that age-specific parameters are required for adequate modeling. Furthermore, both Renshaw and Haberman (2005) and Delward *et al.* (2007) have made no attempt to consider uncertainty, which is our primary concern.

For mathematical convenience, we used a single parameter Gamma distribution for modeling the unobserved heterogeneity. Although we can replace the Gamma distribution with any continuous distribution that has a positive support, the mixture of distributions may not be carried out analytically. In this case, the maximum likelihood estimates of the model parameters may have to be determined by a combination of numerical integrations and the EM algorithm (see Brillinger, 1986).

In all versions of the Lee-Carter model, the modeling proceeds in two steps: the model parameters a_x, b_x, k_t, and α_x (if applicable) are estimated; then an ARIMA process is fitted to parameter k_t for extrapolation. Czado

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

et al. (2005) pointed out that this two-step procedure may give rise to incoherence. They proposed using the Markov Chain Monte Carlo (MCMC) method, which combines the two steps and thus may lead to desirable smooth variations over the Lexis plane. It would be interesting to incorporate our proposed extension with the MCMC method for a deeper understanding of the uncertainty associated with the parameter estimates and forecasts.

Acknowledgements

The authors are grateful to the editor and two anonymous referees for their invaluable comments. They would also like to acknowledge the support of the Natural Science and Engineering Research Council of Canada.

References

Akaike, H. (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, **AC-19**, 716-723.

Bell, W.R. (1997) Comparing and Assessing Time Series Methods for Forecasting Age Specific Demographic Rates. *Journal of Official Statistics*, **13**, 279-303.

Booth, H., Maindonald, J. and Smith, L. (2002) Applying Lee-Carter under Conditions of Variable Mortality Decline. *Population Studies*, **56**, 325-336.

Box, G.E.P. and Jenkins, G.M. (1976) *Time Series Analysis Forecasting and Control. 2nd ed.*, San Francisco: Holden-Day.

Brillinger, D.R. (1986) A Biometrics Invited Paper with Discussion: The Natural Variability of Vital Rates and Associated Statistics. *Biometrics*, **42**, 693-734.

Brouhns, N., Denuit, M. and Vermunt, J.K. (2002) A Poisson Log-bilinear Regression Approach to the Construction of Projected Lifetables. *Insurance: Mathematics and Economics*, **31**, 373-393.

Brouhns, N, Denuit, M. and Keilegom, I.V. (2005) Bootstrapping the Poisson Log-bilinear Model for Mortality Forecasting. *Scandinavian Actuarial Journal*, **31**, 212-224.

Brown, R.L. and McDaid, J. (2003) Factors Affecting Retirement Mortality. *North American Actuarial Journal*, **2**, 24-43.

Burnett, C., Maurer, J. and Dosemeci, M. Mortality by Occupation, Industry, and Cause of Death: 24 Reporting States (1984-1988). Cincinnati: National Institute for Occupational Safety and Health.

Cairns, A.J.G. (2000) A Discussion of Parameter and Model Uncertainty in Insurance. *Insurance: Mathematics and Economics*, **27**, 313-330.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., and Balevich, I. (2009) A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, to appear.

Congressional Budget Office of the United States (1998) *Long-Term Budgetary Pressures and Policy Options.* Washington, D.C.: U.S. Government Printing Office.

Cox, D.R. (1983) Some Remarks on Overdispersion. *Biometrika*, **70**, 269-274.

Currie I. D., Durban, M. and Eilers, P. H. C. (2004) Smoothing and Forecasting Mortality Rates. *Statistical Modelling*, **4**, 279-298.

Czado, C., Delwarde, A., Denuit, M. (2005) Bayesian Poisson Log-bilinear Mortality Projections. *Insurance: Mathematics and Economics*, **36**, 260-284.

Delwarde, A., Denuit, M. and Partrat, C. (2007) Negative Binomial Version of the Lee-Carter Model for Mortality Forecasting. *Applied Stochastic Models in Business and Industry*, in press.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2008) Backtesting Stochastic Mortality Models: An Ex-Post Evaluation of Multi-Period-

Ahead Density Forecasts. The Pensions Institute Discussion Paper PI-0803. Available at http://www.pensions-institute.org/papers.html.

HOUGAARD, P. (1984) Life Table Methods for Heterogeneous Populations: Distributions Describing the Heterogeneity. *Biometrika*, **71**, 75-83.

HUMAN MORTALITY DATABASE. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.human-mortality.de (data downloaded on 25 May 2005).

KASSNER, E. and JACKSON, B. (1998) *Determining Comparable Levels of Functional Disability.* Washington, DC: American Association of Retired Persons.

KLUGMAN, S.A., PANJER, H.H., and WILLMOT, G.E. (2004) *Loss Models: from Data to Decisions. 2nd edition.* New York: Wiley.

LEE, R. and ANDERSON, M. (2005) Stochastic Infinite Horizon Forecasts for US Social Security Finances. *National Institute Economic Review*, **194**, 82-93.

LEE, R. and MILLER, T. (2001) Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. *Demography*, **28**, 537-549.

LEE, R. and CARTER, L. (1992) Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, **87**, 659-671.

LI, S.H. and CHAN, W.S. (2005) Outlier Analysis and Mortality Forecasting: the United Kingdom and Scandinavian countries. *Scandinavian Actuarial Journal*, **31**, 187-211.

MCCULLAGH, P. and NELDER, J.A. (1989) *Generalized Linear Models, 2nd edition*, Chapman and Hall.

MCNOWN, R. and ROGERS, A. (1989) Forecasting Mortality: A Parameterized Time Series Approach. *Demography*, **26**, 645-650.

PRAGER, K. (1994) *Infant mortality by birthweight and other characteristics: United States, 1985 birth cohort.* National Center for Health Statistics. Vital Health Statistics Series 20, No. 4., DHHS Pub. No. (PHS) 94-1852.

PRITCHARD, D.J. (2006) Modeling Disability in Long-Term Care Insurance. *North American Actuarial Journal*, **10**, 48-75.

RENSHAW, A.E. and HABERMAN, S. (2003) Lee-Carter Mortality Forecasting with Age-specific Enhancement. *Insurance: Mathematics and Economics*, **33**, 255-272.

RENSHAW, A.E. and HABERMAN, S. (2006) A Cohort-based Extension to the Lee-Carter Model for Mortality Reduction Factors. *Insurance: Mathematics and Economics*, **38**, 556-570.

ROSS, S.M. (2003) *Simulation, 3rd edition.* San Diego, California: Academic Press.

SCHWARZ, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics*, **6**, 461-464.

TULJAPURKAR, S., LI, N. and BOE, C. (2000) A universal pattern of mortality decline in the G7 countries. *Nature*, **405**, 789-792.

VAUPEL, J.W., MANTON, K.G. and STALLARD, E. (1979) The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, **16**, 439-454.

WANG, S.S. and BROWN, R.L. (1998) A Frailty Model for Projection of Human Mortality Improvement. *Journal of Actuarial Practice*, **6**, 221-241.

WORLD HEALTH ORGANIZATION (2006) *Neonatal and Perinatal Mortality: Country, Regional and Global Estimates.* Geneva: World Health Organization.

WILMOTH, J.R. (1993) Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change. Technical report. Department of Demography. University of California, Berkeley.

APPENDIX

APPENDIX 1. An algorithm for estimating model parameters without the forecast origin correction

Let $a_x^{(v)}$, $b_x^{(v)}$, $k_t^{(v)}$, and $\alpha_x^{(v)}$ be the estimates of a_x, b_x, k_t, and α_x in the v^{th} iteration, respectively. We define the following functions:

$$\hat{D}_{x,t}^{(v)} = E_{x,t} \exp\left(a_x^{(v)} + b_x^{(v)} k_t^{(v)}\right);$$

$$f_{x,t}^{(v)} = \sum_{i=0}^{D_{x,t}-1} \left(\frac{i}{1+\alpha_x^{(v)} i} - \frac{1}{\alpha_x^{(v)}}\right);$$

$$g_{x,t}^{(v)} = \hat{D}_{x,t}^{(v)} \left(D_{x,t} + \frac{1}{\alpha_x^{(v)}}\right)\left(1+\alpha_x^{(v)} \hat{D}_{x,t}^{(v)}\right)^{-1};$$

$$h_{x,t}^{(v)} = \sum_{i=0}^{D_{x,t}-1} \left(\frac{-i^2}{\left(1+\alpha_x^{(v)} i\right)^2} - \frac{1}{\alpha_x^{(v)^2}}\right);$$

$$r_{x,t}^{(v)} = \frac{\ln\left(1+\alpha_x^{(v+3)} \hat{D}_{x,t}^{(v+3)}\right)}{\alpha_x^{(v+3)^2}}.$$

The Newton's procedure for estimating the model parameters in our proposed extension (without the forecast origin correction) is as follows.

- Update on $\{a_x\}$
  For all x,

$$a_x^{(v+1)} = a_x^{(v)} - \frac{\sum_t \left(D_{x,t} - \alpha_x^{(v)} g_{x,t}^{(v)}\right)}{\sum_t \left(-\alpha_x^{(v)} g_{x,t}^{(v)}\right) / \left(1+\alpha_x^{(v)} \hat{D}_{x,t}^{(v)}\right)},$$

  $b_x^{(v+1)} = b_x^{(v)}$, and $\alpha_x^{(v+1)} = \alpha_x^{(v)}$. For all t, $k_t^{(v+1)} = k_t^{(v)}$.

- Update on $\{b_x\}$
  For all x,

$$\tilde{b}_x^{(v+2)} = b_x^{(v+1)} - \frac{\sum_t \left(D_{x,t} k_t^{(v+1)} - \alpha_x^{(v+1)} k_t^{(v+1)} g_{x,t}^{(v+1)}\right)}{\sum_t \left(-\alpha_x^{(v+1)} k_t^{(v+1)^2} g_{x,t}^{(v+1)}\right) / \left(1+\alpha_x^{(v+1)} \hat{D}_{x,t}^{(v+1)}\right)},$$

$$b_x^{(v+2)} = \frac{\tilde{b}_x^{(v+2)}}{\sum_x \tilde{b}_x^{(v+2)}},$$

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

$a_x^{(v+2)} = a_x^{(v+1)}$, and $\alpha_x^{(v+2)} = \alpha_x^{(v+1)}$. For all t, $k_t^{(v+2)} = k_t^{(v+1)}$.

- Update on $\{k_t\}$
  For all t,

$$\tilde{k}_t^{(v+3)} = k_t^{(v+2)} - \frac{\sum_x \left(D_{x,t} b_x^{(v+2)} - \alpha_x^{(v+2)} b_x^{(v+2)} g_{x,t}^{(v+2)} \right)}{\sum_x \left(-\alpha_x^{(v+2)} b_x^{(v+2)^2} g_{x,t}^{(v+2)} \right) / \left(1 + \alpha_x^{(v+2)} \hat{D}_{x,t}^{(v+2)} \right)},$$

$$k_t^{(v+3)} = \tilde{k}_t^{(v+3)} - \frac{1}{T} \sum_t \tilde{k}_t^{(v+3)}.$$

  For all x, $a_x^{(v+3)} = a_x^{(v+2)}$, $b_x^{(v+3)} = b_x^{(v+2)}$, and $\alpha_x^{(v+3)} = \alpha_x^{(v+2)}$.

- Update on $\{\alpha_x\}$
  For all x,

$$\alpha_x^{(v+4)} = \alpha_x^{(v+3)} - \frac{\sum_t \left\{ f_{x,t}^{(v+3)} - g_{x_t}^{(v+3)} + r_{x,t}^{(v+3)} + \frac{D_{x,t}}{\alpha_x^{(v+3)}} \right\}}{\sum_t \left\{ h_{x,t}^{(v+3)} - \frac{D_{x,t} + 2 r_{x,t}^{(v+3)} \alpha_x^{(v+3)}}{\alpha_x^{(v+3)^2}} + \left(\frac{2}{\alpha_x^{(v+3)^2}} + g_{x,t}^{(v+3)} \right) \left(\frac{\hat{D}_{x,t}^{(v+3)}}{1 + \alpha_x^{(v+3)} \hat{D}_{x,t}^{(v+3)}} \right) \right\}},$$

  $a_x^{(v+4)} = a_x^{(v+3)}$, and $b_x^{(v+4)} = b_x^{(v+3)}$. For all t, $k_t^{(v+4)} = k_t^{(v+3)}$.

The iteration stops when the change in the log-likelihood function (equation (13)) is sufficiently small, say 10^{-6}. The starting values can be arbitrary, but a faster convergence can be achieved if they are set to the Poisson maximum likelihood estimates.

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press

Appendix 2. An algorithm for estimating model parameters with the forecast origin correction

To correct the potential error at the forecast origin, the procedure for updating $\{a_x\}$ and $\{k_t\}$ in Appendix 1 should be changed as follows.

- Update on $\{a_x\}$

  For all x, $a_x^{(v+1)} = D_{x,T}/E_{x,T}$ (i.e., we keep a_x unchanged), $b_x^{(v+1)} = b_x^{(v)}$, and $\alpha_x^{(v+1)} = \alpha_x^{(v)}$. For all t, $k_t^{(v+1)} = k_t^{(v)}$.

- Update on $\{k_t\}$

  For all t,

$$\tilde{k}_t^{(v+3)} = k_t^{(v+2)} - \frac{\sum_x \left(D_{x,t} b_x^{(v+2)} - \alpha_x^{(v+2)} b_x^{(v+2)} g_{x,t}^{(v+2)} \right)}{\sum_x \left(-\alpha_x^{(v+2)} b_x^{(v+2)^2} g_{x,t}^{(v+2)} \right) / \left(1 + \alpha_x^{(v+2)} \hat{D}_{x,t}^{(v+2)} \right)},$$

$$k_t^{(v+3)} = \tilde{k}_t^{(v+3)} - \tilde{k}_T^{(v+3)}.$$

  For all x, $a_x^{(v+3)} = a_x^{(v+2)}$, $b_x^{(v+3)} = b_x^{(v+2)}$, and $\alpha_x^{(v+3)} = \alpha_x^{(v+2)}$.

Johnny Siu-Hang Li
Department of Statistics and Actuarial Science
University of Waterloo
Ontario, Canada
E-Mail: shli@uwaterloo.ca

Mary R. Hardy
Department of Statistics and Actuarial Science
University of Waterloo
Ontario, Canada
E-Mail: mrhardy@uwaterloo.ca

Ken Seng Tan
Department of Statistics and Actuarial Science
University of Waterloo
Ontario, Canada
E-Mail: kstan@uwaterloo.ca

https://doi.org/10.2143/AST.39.1.2038060 Published online by Cambridge University Press