Stochastic Mortality Models with Applications in Financial Risk Management

by

Siu Hang Li

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2007

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Siu Hang Li

Abstract

In product pricing and reserving, actuaries are often required to make predictions of future death rates. In the past, this has been performed by using deterministic improvement scales that give only a single mortality trajectory. However, there is enormous likelihood that future death rates will turn out to be different from the projected ones, and so a better assessment of longevity risk would be one that consists of both a mean estimate and a measure of uncertainty. Such assessment can be performed using a stochastic mortality model, which is the core of this thesis.

The Lee-Carter model is one of the most popular stochastic mortality models. While it does an excellent job in mean forecasting, it has been criticized for providing overly narrow prediction intervals that may have underestimated uncertainty. This thesis mitigates this problem by relaxing the assumption on the distribution of death counts. We found that the generalization from Poisson to negative binomial is equivalent to allowing gamma heterogeneity within each age-period cells. The proposed extension gives not only a better fit, but also a more conservative prediction interval that may reflect better the uncertainty entailed.

The proposed extension is then applied to the construction of mortality improvement scales for Canadian insured lives. Given that the insured lives data series are too short for a direct Lee-Carter projection, we build an extra relational model that could borrow strengths from the Canadian population data, which covers a far longer period. The resultant scales consist of explicit measures of uncertainty.

The prediction of the tail of a survival distribution requires a special treatment due to the lack of high quality old-age mortality data. We utilize the asymptotic results in modern extreme value theory to extrapolate death probabilities to the advanced ages, and to statistically determine the age at which the life table should be closed. Such technique is further integrated with the Lee-Carter model to produce a stochastic analysis of old-age mortality, and a prediction of the highest attained age for various cohorts.

The mortality models we considered are further applied to the valuation of mortality-related financial products. In particular we investigate the no-negative-equity-guarantee that is offered in most fixed-repayment lifetime mortgages in Britain. The valuation of such guarantee requires a simultaneous consideration of both longevity and house price inflation

risk. We found that house price returns can be well described by an ARMA-EGARCH time-series process. Under an ARMA-EGARCH process, however, the Black-Scholes formula no longer applies. We derive our own pricing formula based on the conditional Esscher transformation. Finally, we propose some possible hedging and capital reserving strategies for managing the risks associated with the guarantee.

Acknowledgements

I would like to express my deepest gratitude to my advisors and dear friends, Mary Hardy and Ken Seng Tan, for their support, guidance and unwavering encouragement over the years. I am also extremely grateful to Phelim Boyle for his generosity to share with me his research ideas, particularly on the topic of equity release mechanisms.

Special thanks to other members of my thesis committee – Robert Brown, Samual Cox and Justin Wan – for their valuable participation and insights, to Wai-Sum Chan for introducing me to the field of mortality studies, to Doug Andrews and Harry Panjer for helpful comments on several chapters, and to Andrew Ng for insightful discussions on many topics.

Thanks to the Institute of Quantitative Finance and Insurance, the Government of Ontario, the Society of Actuaries and the Casualty Actuarial Society for providing me with the financial means to complete this thesis.

Finally, my heart goes to my family for their love and care, and to my teddy bear, Owee, for always being at my side and helping me deal with life's ups and downs.

Contents

Chapter 1

Introduction

1.1 Background

Human mortality has experienced remarkable improvement in the past century. For instance, in Canada, the life expectancy at birth has increased from about 56 years in 1921 to the current level of about 79 years[1], and its rise continues. While the enormous improvement in life expectancy is usually viewed as one of the greatest achievements of modern civilization, unanticipated mortality improvement can cause insurance companies and providers of mortality-related financial products to suffer huge losses. To protect financial institutions from the threats of reduction in mortality, actuaries often utilize life tables including a forecast of future trends in survival probabilities. However, the production of such tables is not easy.

A renowned demographer described the challenge of forecasting mortality as "a bumpy road to Shangri-La"[2]. This is because the demographic future of any human population is a result of complex and only partially understood mechanisms, and is highly uncertain. Recently, actuaries and demographers have been understandably concerned about error in their predictions, and part of their response is a new wave of work that is focused on the forecasting of uncertainty in mortality projections. Such work aims to forecast a range of possible outcomes along with associated probabilities, instead of a single prediction that

[1]Source: Human mortality database (2005) and own calculations.

[2]See Tuljapurkar (2005).

will almost surely be wrong.

The forecasting of uncertainty in mortality improvement is especially important to actuarial risk management. More specifically, longevity risk resembles investment risk in that it is non-diversifiable, since any change in the overall mortality level is likely to affect all policyholders of the same cohort. As a result, it cannot be mitigated by the traditional insurance mechanism of selling large number of policies. However, it is different in that there is a lack of mortality-linked securities that could possibly be used for hedging. In other words, capital is often required to cushion longevity risk, and such capital is, of course, determined by measures of uncertainty associated with mortality projections.

Tuljarpurkar (1997) suggested that methods for forecasting uncertainty can be segregated into two categories, namely, static and dynamic. In a static approach, the forecaster makes assertions about demographic characteristics, such as life expectancy at birth, and the trajectories along which those characteristics will change between the start and end. These trajectories are then assigned a probability, usually determined subjectively. A shortcoming of the static approach is that the dynamics of getting from "here" to "there" are largely ignored. In contrast, dynamic forecasts employ a stochastic model that is fitted to historical data. The resulting models have uncertainty embedded within them, as reflected in historical change, and yield trajectories in the form of sample paths, which are particularly valuable in assessing financial liabilities that are sensitive to the timing and pattern of demographic change. The application of such models in financial risk management has recently been highly regarded by leading actuarial organizations (see, e.g., Continuous Mortality Investigation Bureau, 2004).

1.2 Previous Mortality Improvement Scales

For ease of use and interpretation, actuaries sometimes prefer using simple formulae that summarize the entire projection of future mortality rates. Such formulae are often referred to as mortality improvement scales, and they work with the base mortality rates in a multiplicative manner, that is,

$$q_{x,t} = q_{x,0}\mathrm{IS}(x,t), \tag{1.1}$$

where $q_{x,0}$ and $q_{x,t}$ are the probabilities that a person who is aged x dies within a year, at

time 0 and at time t, respectively; and $\mathrm{IS}(x,t)$ is the improvement scale for projecting $q_{x,t}$. Below we summarize the improvement scales that have been widely used in Britain and North America.

- *The Society of Actuaries 1994 Group Annuity Mortality Table.*

  The Society of Actuaries Group Annuity Valuation Table Task Force (1995) developed a table that it recommended as suitable for a Group Annuity Reserve Valuation Standard. In addition, the task force recommended the following improvement scale for projecting future mortality rates of group annuitants:

  $$\mathrm{IS}(x,s) = (1 - AA_x)^s, \tag{1.2}$$

  where AA_x are age-dependent parameters, whose numerical values can be found in the Society of Actuaries Group Annuity Valuation Table Task Force (1995, p.892).

- *The Society of Actuaries 2001 Valuation Basic Experience Table.*

  In 2002, the Society of Actuaries and the Academy of Actuaries developed the 2001 CSO tables, which were intended to replace the 1980 CSO Table in the existing statutory valuation structure (American Academy of Actuaries, 2002). The 2001 CSO tables consist of six tables − for each sex, there are separate tables for nonsmoker, smoker, and composite nonsmoker/smoker. Each table has values for a 25-year select period and for ultimate ages.

  The entire work was divided into two pieces: the construction of the 2001 valuation basic experience table (VBT, done by the SoA Task Force), and the development of the loads (done by the Academy's Task Force). A key step in the construction of the VBT was to project the mortality experience underlying the 1990-95 Basic Mortality Tables to year 2001, the date at which the valuation table would be released. Given the information from the 1985-90 and 1990-95 Basic Mortality Tables, and from various non-life insurance sources, the SoA Task Force derived a mortality improvement scale, which can be expressed as follows:

for males:

$$\mathrm{IS}(x,s) = \begin{cases} 0, & x < 45 \\ \left(1 - \frac{0.01(x-45)}{10}\right)^s, & 45 \le x < 55 \\ 0.99^s, & 55 \le x < 80 \\ \left(1 - \frac{0.01(90-x)}{5}\right)^s, & 80 \le x < 90 \\ 0, & x > 90 \end{cases} ; \tag{1.3}$$

for females:

$$\mathrm{IS}(x,s) = \begin{cases} 0, & x < 45 \\ \left(1 - \frac{0.005(x-45)}{10}\right)^s, & 45 \le x < 55 \\ 0.995^s, & 55 \le x < 85 \\ \left(1 - \frac{0.005(90-x)}{5}\right)^s, & 85 \le x < 90 \\ 0, & x > 90 \end{cases} . \tag{1.4}$$

- *The Institute of Actuaries "80" and "92" Series Mortality Table.*

  In the United Kingdom, the Continuous Mortality Investigation Bureau (CMIB) of the Institute of Actuaries has periodically been considering future improvements in mortality for annuitants and pensioners. In two recent sets of published CMIB tables, namely the "80" Series (CMIB, 1990) and the "92" Series (CMIB, 1999), the projected mortality values are estimated by multiplying death rates in the base table by the improvement scale, which can be written as

$$\mathrm{IS}(x,t) = \alpha(x) + [1 - \alpha(x)][1 - f_n(x)]^{\frac{t}{n}}. \tag{1.5}$$

  In other words, the improvement scale is characterized by two age-dependent parameters, $\alpha(x)$ and $f_n(x)$. Parameter $\alpha(x)$ is the limiting value of the improvement scale as t tends to infinity, and $f_n(x)$ is the percentage of the total improvement $(1-\alpha(x))$ that is assumed to occur in n years.

  For the "80" Series, n was fixed at 20 and $f_{20}(x)$ at 0.6 for all ages, $\alpha(x)$ was assumed to be linearly increasing for $x = 60, ..., 110$, and fixed at 0.5 and 1 for $x < 60$ and

$x > 110$, respectively. In other words, the form of $\alpha(x)$ can be expressed as

$$\alpha(x) = \begin{cases} 0.5, & x < 60 \\ \dfrac{x-10}{100}, & 60 \leq x \leq 110 \\ 1, & x > 110 \end{cases} . \tag{1.6}$$

The "92" Series shows important differences in the parameters assumed. The value for n remains fixed at 20, but $f_{20}(x)$ varies linearly from 0.55 to 0.29 between ages 60 and 110. In addition, $\alpha(x)$ takes the value of 0.13 instead of 0.5 under 60; that is,

$$\alpha(x) = \begin{cases} 0.13, & x < 60 \\ 1 + 0.87\dfrac{x-110}{50}, & 60 \leq x \leq 110 \\ 1, & x > 110 \end{cases} , \tag{1.7}$$

$$f_{20}(x) = \begin{cases} 0.55, & x < 60 \\ \dfrac{0.55(110-x) + 0.29(x-60)}{50}, & 60 \leq x \leq 110 \\ 0.29, & x > 110 \end{cases} . \tag{1.8}$$

In all these improvement scales, values assumed for the parameters have tended to result from qualitative analysis of past trends and expert opinion, which is not generally regarded as an exact science. More importantly, these scales project future mortality rates in a purely deterministic manner, and give no measure of uncertainly at all.

It has been found that the rate of improvement in male pensioner mortality has been significantly faster than anticipated in the CMIB "92" scale, resulting in pension plan providers experiencing losses. The need for a more prudent improvement scale is pressing, and the CMIB (2002) offered three different scales as an alternative to the "92" scale. Although, in some sense, the range between these scales indicates uncertainty, no probabilities are associated with the alternative outcomes, and so it is difficult to interpret, employ and evaluate them. A better improvement scale would be one that is calibrated on a scientific ground, and provides explicit measures of uncertainty.

1.3 Objectives and Outline of the Thesis

This thesis explores the stochastic models for forecasting uncertainty in future mortality rates. In particular, we research on how the weaknesses of the existing models can be overcome, and how they may be turned into useful applications in financial risk management.

In Chapter 2, we focus on the celebrated Lee-Carter model, which has been extensively used for a wide range of purposes such as population projection (Booth and Tickle, 2003), and assessment of retirement income adequacy (Chia and Tsui, 2003). However, it has been criticized for its provision of overly narrow prediction intervals, which may lead to underestimating uncertainty entailed in mortality projections. This motivates us to relax the model structure so that the underlying uncertainty can be more realistically reflected in the forecast. The relaxation has to be performed carefully since an arbitrary modification of the model structure may lead the resultant mortality forecast to behave erroneously.

In Chapter 3, we applied the proposed variant of the Lee-Carter model to the derivation of mortality improvement scales for Canadian insured lives. The derivation is hampered by the lack of data about the insured lives. Specifically, the insured lives experience is too short to give a prudent statistical projection. To overcome this problem, we develop a joint model that utilizes mortality data of both the general population and insured lives. Such a joint model gives improvement scales that include explicit measures of uncertainty, which are particularly useful for actuaries to access longevity risk quantitatively.

In Chapter 4, we focus on the uncertainty of the tail of survival distributions. This is not considered in the previous chapters, partly because of the scarcity and unreliability of mortality data at the advanced ages − old-age mortality data are often right-censored, and are available only in small numbers, leading to significant sampling error. This chapter utilizes modern extreme value theory to extrapolate survival distributions to the higher ages, and to determine an appropriate end point of a life table. The technique is further integrated with the stochastic mortality model in Chapter 2 to analyze the uncertainty of future old-age death rates, and to estimate the probability distribution of the highest attained age.

In Chapter 5, we apply the stochastic mortality models in previous chapters to the risk management for mortality-related financial products. In particular, we investigate the no-negative-equity-guarantee (NNEG) in equity release mechanisms offered in Britain. In

a typical equity release plan, the customer (homeowner) is advanced a certain amount of cash, and the loan and interest are repaid from the property sale proceeds when the borrower dies, or moves into long-term care. The NNEG guarantees that the amount repayable is no greater than the property sale proceeds. This means that the guarantee may be viewed as an option that involves a myriad of demographic and financial risks. In this chapter, we attempt to value the NNEG, and to mitigate the risks entailed in offering such a guarantee.

Finally, in Chapter 6, we conclude the thesis with several suggestions for further research.

Chapter 2

Modeling Uncertainties in Mortality Forecasting

2.1 Introduction

In recent years, mortality has improved considerably faster than had been predicted, resulting in unforeseen mortality losses for annuity liabilities (see, e.g., CMIB, 2004). As a result, there has been a growing consensus about the need for reliable, scientific models for mortality that can be used to measure not only the expected improvement in mortality over the next few decades, but also the associated uncertainty in the projections. Several solutions have been proposed including the Lee-Carter model (Lee and Carter, 1992), the P-splines regression (Currie et al., 2004) and the parameterized time-series approach (McNown and Rogers, 1989).

The Lee-Carter model may be considered the current gold standard. 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 Australia (Booth et al., 2002), Scandinavia (Li and Chan, 2005a), and the G7 countries (Tuljapurkar et al., 2000). The model has several appealing features. First, the number of parameters is small relative to that of other stochastic mortality models. Secondly, the model parameters can be interpreted rather easily. Thirdly, 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 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, e.g., Lee (2000)). This narrowness may result in underestimation of the risk of more extreme outcomes, and this may defeat the original purpose of moving on to a stochastic framework. This phenomenon is an example of model risk (see, e.g., Cairns, 2000). Reduction of model risk can be achieved by relaxing the model structure to obtain a more general class of models. So far, relaxation of the Lee-Carter structure has been confined in the inclusion of additional bilinear terms, from two (Renshaw and Haberman, 2003) to five (Booth et al., 2002). Although these augmented versions can explain a higher proportion of temporal variance than the original model, the additional time-varying components are typically highly non-linear, which makes time-series forecasting harder.

The objective of this chapter 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. For example, given mortality data for years 1951 to 2000, the forecaster may divide the Lexis plane into 5000 cells for ages 0 to 99 and for the ex ante period of 50 years. The original Lee-Carter and other conventional methods allow 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 noted that individuals may differ substantially in their capacity for longevity, and that individual differences are important in population-based mortality studies (see, e.g., Hougaard, 1984; Vaupel et al., 1979). The primary contribution of this chapter is the incorporation of heterogeneity into the Lee-Carter model by introducing an unobserved variate for individual differences in each age-period cell. To justify our contribution, the proposed extension must satisfy the following criteria.

1. *Provision of wider confidence intervals.* Taking account of individual 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 so that the introduction of additional parameters is worthwhile.

3. *Retention of the appealing features in the original version.* Introduction of additional parameters should be done carefully so that criteria 1 and 2 can be satisfied without making any undesirable distortion in the projected age patterns. Linearity in the time-varying component of the model should also be retained.

The rest of this chapter is organized as follows. In Section 2, we provide a brief review of current methodologies of fitting the Lee-Carter model, with emphasis on the properties of the confidence intervals obtained under each method. In Section 3, we present our proposed extension. Technical details on model parameter estimation and forecasting are also given. In Section 4, we compare the performance of our proposed generalizations with the older models, using Canadian population mortality data. Section 5 concludes this chapter.

2.2 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, namely $\{a_x\}$, $\{b_x\}$ and $\{k_t\}$, in the following way:

$$\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t}, \tag{2.1}$$

where a_x gives the average level of mortality at each age; the time-varying component, k_t, sometimes referred to as the mortality index, signifies the general speed of mortality improvement, and it is worked in a multiplicative manner with another age-specific component b_x that characterizes the sensitivity to k_t at different ages; the error term $\epsilon_{x,t}$ captures all the remaining variations. The variance of $\epsilon_{x,t}$ is assumed to be constant in Lee and Carter (1992). Such assumption is relaxed in some variants of the model.

Equation (2.1) assumes that mortality improvement is additive to the "base" age pattern of mortality, represented by a_x. As a result, even though mortality at different ages is allowed to improve at various paces, determined by b_x, the age pattern in the distant future is not easily distorted.

The mortality forecasting in Lee-Carter is performed in two stages. In the first stage, we estimate parameters a_x, b_x and k_t by historical mortality data. 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. Finally, we extrapolate k_t through the fitted ARIMA model to obtain a forecast of future death rates.

Note that parameters on the right hand side of equation (2.1) are unobservable. Hence, we are not able to fit the model by simple methods like the ordinary least squares. To solve the problem, researchers have proposed a few alternative approaches, which may be roughly categorized as non-likelihood-based and likelihood-based. An in-depth understanding of methods in both categories is important as the differences in their fundamental assumptions often lead to dissimilar forecasting results. In the following, we give detailed descriptions on model parameter estimation and interval forecasting under each approach.

2.2.1 Non-likelihood-based Methods

In non-likelihood-based methods, we are not required to specify any probability distribution during model parameter estimation. Examples in this category include the method of Singular Value Decomposition (SVD) proposed in Lee and Carter (1992) and the method of weighed least squares (WLS) suggested later in Wilmoth (1993).

In the method of SVD, we set $a_x = \frac{1}{T}\sum_{t=0}^{T} \ln(m_{x,t})$, where T is the number of periods in the time-series mortality data, and apply SVD to the matrix of $\{\ln(m_{x,t}) - a_x\}$. The first left and right singular vectors give a tentative estimate of b_x and k_t, respectively. To satisfy the constraints for parameter uniqueness, the estimates of b_x and k_t are normalized so that they sum to one and zero, respectively.

As the method of SVD is purely a mathematical approximation, the fitted and actual number of deaths may not be the same. To reconcile the fitted and observed number of deaths, we are required to make an ad hoc adjustment to k_t. Having performed the re-estimation, we can model k_t by an appropriate ARIMA process and then obtain the mortality forecast by extrapolation.

Let $\hat{k}_{T+s}$ be the s-period ahead forecast of k_t. Then, the s-period ahead forecast of $m_{x,t}$ is given by

$$\hat{m}_{x,T+s} = \exp(\hat{a}_x + \hat{k}_{T+s}\hat{b}_x). \tag{2.2}$$

Assuming that the model specification is correct, the true value of $\ln(m_{x,T+s})$ can be

expressed as

$$\ln(m_{x,T+s}) = (\hat{a}_x + \alpha_x) + (\hat{k}_{T+s} + u_{T+s})(\hat{b}_x + \beta_x) + \epsilon_{x,T+s}, \tag{2.3}$$

where α_x and β_x are the error in estimating parameters a_x and b_x, respectively, and u_{T+s} is the stochastic error in the s-period ahead forecast of k_t. By combining equations (2.2) and (2.3), we have the following expression for the error in the forecast of $\ln(m_{x,t})$:

$$E_{x,T+s} = \alpha_x + \epsilon_{x,T+s} + (\hat{b}_x + \beta_x)u_{T+s} + \beta_x\hat{k}_{T+s}. \tag{2.4}$$

Unfortunately, there is no analytic solution to the covariance between the error terms and the parameter estimates. Hence, in the computation of variance of the forecast error, Lee and Carter (1992) assumed independence between all the terms on the right hand side of equation (2.4). Under this assumption, the variance of the forecast error can be expressed as

$$\sigma^2_{E,x,T+s} = \sigma^2_{\alpha,x} + \sigma^2_{\epsilon,x,T+S} + (\hat{b}^2_x + \sigma^2_{\beta,x})\sigma^2_{k,T+s} + \sigma^2_{\beta,x}\hat{k}^2_{T+s}, \tag{2.5}$$

where $\sigma^2_{\alpha,x}$ and $\sigma^2_{\beta,x}$ correspond to the variance of the error in estimating a_x and b_x, respectively, $\sigma^2_{\epsilon,x,T+s}$ is the variance of $\epsilon_{x,T+s}$, and $\sigma^2_{k,T+s}$ is the stochastic error in the s-step ahead forecast of k_t. We can estimate $\sigma^2_{\alpha,x}$ by the dividing the sample variance of $\ln(m_{x,t})$, for $t = 1, 2, ..., T$, by T, $\sigma^2_{\epsilon,x,T+s}$ by the variance of the error in fitting age x within the sample period, and $\sigma^2_{\beta,x}$ by a small-scaled bootstrap (see Appendix B, Lee and Carter, 1992). The form of $\sigma^2_{k,T+S}$ depends on the order of the ARIMA process.

To generate the interval forecast, we are required to make the additional assumption that normality holds. Under this assumption, we have the following expression for the approximate 95% point-wise interval forecast of $\ln(m_{x,t})$:

$$\ln(m_{x,T+s}) = \ln(\hat{m}_{x,T+s}) \pm 1.96\sigma_{E,x,T+s}. \tag{2.6}$$

In most cases, the original Lee-Carter (rank-1 SVD approximation) gives a good fit in terms of the proportion of temporal variance explained[1]. To further improve the fit,

[1]The proportion of temporal variance explained by a rank-p SVD approximation is measured by $\sum_{i=1}^{p} s_i^2 / \sum_{i=1}^{n} s_i^2$, where s_i is the i^{th} singular value, and n is the rank of the matrix $\{\ln(m_{x,t}) - a_x\}$.

researchers later considered using a rank-p SVD approximation. For example, Renshaw and Haberman (2003) considered $p = 2$, and Booth et al. (2002) considered $p = 5$. Under a rank-p SVD approximation, equation (2.1) is generalized to

$$\ln(m_{x,t}) = a_x + \sum_{i=1}^{p} b_x^{(i)} k_t^{(i)} + \epsilon_{x,t}. \tag{2.7}$$

Given an increased number of parameters, a rank-p ($p > 1$) SVD approximation can obviously explain a higher proportion of temporal variance. However, it makes forecasting more complicated because (1) the time-varying component in the higher order terms are often non-linear, making it difficult to be handled by an ARIMA process, and (2) analytic solutions to the covariance between errors in estimating the additional parameters are not available, so that stronger assumptions are required in computing the interval forecast.

In WLS method, the model parameters a_x, b_x and k_t are derived by minimizing

$$\sum_{x,t} w_{x,t}(\ln(m_{x,t}) - a_x - b_x k_t)^2, \tag{2.8}$$

where $w_{x,t}$ can be taken as the reciprocal of the number of exposures-to-risk at age x and time t. Interval forecasting in WLS is similar to that in SVD, and therefore we do not restate the details. Note that interval forecasting under WLS also requires the normality assumption.

2.2.2 Likelihood-Based Methods

In likelihood-based methods, we have to specify a probability distribution for the death counts. Examples in this category include the method of maximum likelihood estimation (MLE) considered by Wilmoth (1993) and implemented later by Brouhns et al. (2002), and the method of generalized linear models (GLM) employed by Renshaw and Haberman (2006). All these methods assume that the observed number of deaths is a realization of a Poisson distribution with mean equal to the expected number of deaths under the Lee-Carter model; that is,

$$D_{x,t} \sim \text{Poisson}(E_{x,t} \exp(a_x + b_x k_t)), \tag{2.9}$$

where $D_{x,t}$ and $E_{x,t}$ are the number of deaths and exposures-to-risk at age x and time t, respectively.

In the method of MLE, we obtain estimates of the model parameters by maximizing the log-likelihood function, which is given by

$$l_p = \sum_{x,t} \left(D_{x,t}(a_x + b_x k_t) - E_{x,t}(\exp(a_x + b_x k_t)\right) + c, \tag{2.10}$$

where c is a constant that is free of the model parameters. The maximization can be accomplished via a standard Newton's method.

By differentiating both sides of equation (2.10), we can immediately see that the observed and fitted number of deaths over time are equal when the algorithm converges. This implies that the ad hoc re-estimation of k_t required for the SVD method is not required here.

Having estimated the model parameters, we can project future values of k_t by a properly identified ARIMA process. The s-period ahead forecast of $m_{x,t}$ is again given by equation (2.2). The interval forecast of $m_{x,t}$ cannot be computed analytically, but it may be obtained by using a parametric bootstrap (Brouhns et al., 2005), which we now summarize.

1. Simulate N realizations from the Poisson distribution with mean equal to the fitted number of deaths under the Lee-Carter model. In the illustration presented in later sections, we use the transformed rejection method (Hörmann, 1993) to generate the Poisson random numbers.

2. For each of these N realizations:

   (a) re-estimate the model parameters a_x, b_x and k_t using MLE;

   (b) specify a new ARIMA process for the re-estimated k_t;

   (c) compute future values of $m_{x,t}$ using the re-estimated a_x and b_x, and the simulated future values of k_t under the newly specified ARIMA process.

3. Step (2) gives an empirical distribution of $m_{x,T+s}$ for all x and s. The 2.5^{th} and the 97.5^{th} percentiles of the empirical distribution respectively gives the lower and upper limit of the 95% interval forecast of $m_{x,T+s}$.

It is noteworthy that the above algorithm allows both the sampling fluctuation in the model parameters and the stochastic error in the projection of k_t to be included in the interval forecast. Furthermore, the algorithm does not require the assumption of normality and this makes an asymmetric confidence interval possible.

Koissi et al. (2005) also proposed a very similar method, known as residual bootstrapping, to compute the interval forecast of $m_{x,t}$. The details of this method are omitted in this review.

In the method of GLM, we use the log-link in modeling the "responses" $D_{x,t}$. The linear model can be written as

$$\ln(D_{x,t}) = \ln(E_{x,t}) + a_x + b_x k_t, \tag{2.11}$$

where $\ln(E_{x,t})$ is the offset. As usual, parameters a_x, b_x and k_t are unobservable and have to be estimated by the following algorithm.

1. Set arbitrary starting values of b_x.
2. Treat b_x as a known covariate; estimate a_x and k_t.
3. Treat a_x and k_t as known covariates; estimate b_x.
4. Repeat steps (2) and (3) until the deviance converges.

Steps (2) and (3) can be performed easily using S-plus. Methods of GLM and MLE give the same parameter estimates if the same constraints for parameter uniqueness are chosen. Mean forecasting of $m_{x,t}$ is identical to that in MLE, and interval forecasting can also be achieved by bootstrapping.

2.2.3 Problems Associated with the Existing Approaches

The non-likelihood-based methods are purely mathematical. In particular, in SVD, values of k_t obtained from the first right singular vector are often illegitimate and thus require further adjustment through an ad hoc procedure. Interval forecasting is quite straightforward, but it requires very strong assumptions. First, even though an analytic formula for the variance of forecast error of $\ln(m_{x,t})$ is available, co-variances between the error terms and

the model parameter estimates are ignored to avoid the mathematical complexities. This would inevitably deflate the variance of the forecast error and make the interval forecast narrower than it should be. Second, as the methods specify no probabilistic distribution during model parameter estimation, normality has to be assumed in the computation of the interval forecast. Inter alia, the assumption of normality rules out the possibility of an asymmetric interval forecast.

The existing likelihood-based methods assume that the observed number of deaths is a realization of a Poisson distribution. In this setting, interval forecasts and standard errors of parameters can be estimated by bootstrapping, although it is computationally intensive. The bootstrap explicitly reflects the statistical dependency between the model parameters in the interval forecast, and it also allows a simultaneous consideration of parameter uncertainty and the stochastic uncertainty arises from the process of k_t. Moreover, we can obtain asymmetric interval forecasts as they are based on the percentiles of the simulated empirical distribution rather than normality. The interval widths are however close to that obtained by non-likelihood-based methods since there is no change in the effective number of parameters. Finally, the likelihood-based methods may give a marginally better fit as the ad hoc re-estimation of k_t is not required.

Nevertheless, there are several drawbacks associated with the existing likelihood-based methods. In assuming Poisson models, we are imposing the mean-variance equality restriction on the variable $D_{x,t}$. In practice, however, the variance can be greater than the mean, and this situation is commonly known 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. Cox (1983) also pointed out that there is a possible loss of efficiency if a single parameter distribution is used when overdispersion exists. If overdispersion exists, the assumption of Poisson death counts will lead us to overestimating the degree of precision of the mortality projection.

2.3 The Proposed Extension

In this section we attempt to loosen the original model structure carefully so that restrictions on the movement of future death rates can be reduced without relinquishing the

desirable properties we stated in Section 2.1.

Motivated by the mean-variance equality restriction in the original likelihood-based estimation methods, we begin with the probability distribution assumed for the number of deaths. Recall that in the original likelihood-based methods, we assume that $D_{x,t}$ follows a Poisson distribution, and that individuals within each age-period cell are homogeneous. 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 exposures-to-risk in each age-period cell into clusters that differ in their 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 computed by the previous methods.

To account for the possibility of clustering, we segregate each age-period cell into $N_{x,t}$ clusters of equal size, where $N_{x,t}$ is assumed to be non-random. This implies that the i^{th} cluster will have $E_{x,t}/N_{x,t}$ exposures-to-risk and $D_{x,t}(i)$ deaths, where $i = 1, ..., N_{x,t}$. The total number of deaths, $D_{x,t}$, can therefore be expressed as

$$D_{x,t} = D_{x,t}(1) + D_{x,t}(2) + ... + D_{x,t}(N_{x,t}). \tag{2.12}$$

We assume that $D_{x,t}(i)$s are independently distributed, and

$$D_{x,t}(i)|z_x(i) \sim \text{Poisson}\left(z_x(i)\frac{E_{x,t}}{N_x}\exp(a_x + b_x k_t)\right), \tag{2.13}$$

where $z_x(i)$ is a random variable accounting for the heterogeneity of individuals. Note that $z_x(i)$ is age-specific and that it 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 probability distribution with a positive support can be a candidate distribution for $z_x(i)$, we assume here that $z_x(i)$ follows a gamma distribution for tractability:

$$z_x(i) \sim \text{Gamma}(\iota_x^{-1}, \iota_x). \tag{2.14}$$

In fact, gamma distributions are often utilized in modeling heterogeneity; for example, see Hougaard (1984), Vaupel et al. (1979), and Wang and Brown (1998). We follow the usual parameterization of the gamma distribution with $\text{E}[z_x(i)] = 1$ and $\text{Var}[z_x(i)] = \iota_x$,

where $\iota_x > 0$. It can be easily shown that

$$\mathrm{E}[D_{x,t}(i)] = \frac{E_{x,t}}{N_{x,t}} \exp(a_x + b_x k_t), \tag{2.15}$$

which will be denoted as $\mu_{x,t}$. Also, we have

$$\mathrm{E}[D_{x,t}] = E_{x,t} \exp(a_x + b_x k_t), \tag{2.16}$$

which means our extension still complies with the Lee-Carter specification. Furthermore, we can show that the unconditional distribution of $D_{x,t}(i)$ can be written as

$$\Pr[D_{x,t}(i) = y] = \frac{\Gamma(y + \iota_x^{-1})}{y!\Gamma(\iota_x^{-1})} \left(\frac{\mu_{x,t}}{\mu_{x,t} + \iota_x^{-1}}\right)^y \left(\frac{\iota_x^{-1}}{\mu_{x,t} + \iota_x^{-1}}\right)^{\iota_x^{-1}}. \tag{2.17}$$

By letting $\alpha_x = \iota_x / N_x$, equation (2.17) immediately leads to

$$\Pr[D_{x,t} = y] = \frac{\Gamma(y + \alpha_x^{-1})}{y!\Gamma(\alpha_x^{-1})} \left(\frac{N_x \mu_{x,t}}{N_x \mu_{x,t} + \alpha_x^{-1}}\right)^y \left(\frac{\alpha_x^{-1}}{N_x \mu_{x,t} + \alpha_x^{-1}}\right)^{\alpha_x^{-1}}. \tag{2.18}$$

This means that $D_{x,t}$ follows a negative binomial distribution unconditionally.

It is interesting to note that the negative binomial distribution has another statistical interpretation. In more detail, we may assume that

$$D_{x,t} | w_x \sim \text{Poisson}(w_x E_{x,t} \exp(a_x + b_x k_t)), \tag{2.19}$$

where w_x is a random variable that captures part of the variations not explained by the Poisson distribution, such as the random error in recording the age at death. If we assume further that

$$w_x \sim \text{Gamma}(\alpha_x^{-1}, \alpha_x), \tag{2.20}$$

then the unconditional distribution of $D_{x,t}$ is the same as that specified by equation (2.18).

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. More importantly, we have

$$\mathrm{Var}[D_{x,t}] = \mathrm{E}[D_{x,t}] + \alpha_x \{\mathrm{E}[D_{x,t}]\}^2, \tag{2.21}$$

which means that the extension explicitly allows for overdispersion via the dispersion parameters α_x's. In other words, the assumption of the mean-variance equality is not required, and consequently measures of uncertainty under 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 additional parameters (α_xs) to the model without altering the structure specified by equation (2.1). This introduces more flexibility without affecting the desirable features such as the linearity in k_t and the stability of the age-pattern of mortality over time.

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

$$\begin{aligned} l_{nb} &= \sum_{x,t} \left\{ \left[\sum_{i=0}^{D_{x,t}-1} \ln\left(\frac{1+\alpha_x i}{\alpha_x}\right) \right] + D_{x,t} \ln(\alpha_x \tilde{D}_{x,t}) \right. \\ &\qquad \left. -(D_{x,t} + \alpha_x^{-1}) \ln(1+\alpha_x \tilde{D}_{x,t}) \right\} + c, \end{aligned} \tag{2.22}$$

where $\tilde{D}_{x,t} = E_{x,t} \exp(a_x + b_x k_t)$, and c is a constant that is free of a_x, b_x, k_t and α_x.

We may maximize l_{nb} by a standard Newton's procedure as shown below.

$$\begin{aligned} a_x^{(v+1)} &= a_x^{(v)} - \frac{\sum_t (D_{x,t} - \alpha_x^{(v)} g_{x,t}^{(v)})}{\sum_t (-\alpha_x^{(v)} g_{x,t}^{(v)})/(1+\alpha_x^{(v)} \hat{D}_{x,t}^{(v)})}, \\ b_x^{(v+1)} &= b_x^{(v)}, \\ k_t^{(v+1)} &= k_t^{(v)}, \\ \alpha_x^{(v+1)} &= \alpha_x^{(v)}; \end{aligned} \tag{2.23}$$

$$\begin{aligned} b_x^{(v+2)} &= b_x^{(v+1)} - \frac{\sum_t (D_{x,t} k_t^{(v+1)} - \alpha_x^{(v+1)} k_t^{(v+1)} g_{x,t}^{(v+1)})}{\sum_t (-\alpha_x^{(v+1)} k_t^{(v+1)^2} g_{x,t}^{(v+1)})/(1+\alpha_x^{(v+1)} \hat{D}_{x,t}^{(v+1)})}, \\ a_x^{(v+2)} &= a_x^{(v+1)}, \\ k_t^{(v+2)} &= k_t^{(v+1)}, \\ \alpha_x^{(v+2)} &= \alpha_x^{(v+1)}; \end{aligned} \tag{2.24}$$

$$
\begin{aligned}
k_t^{(v+3)} &= k_t^{(v+2)} - \frac{\sum_x (D_{x,t} b_x^{(v+2)} - \alpha_x^{(v+2)} b_x^{(v+2)} g_{x,t}^{(v+2)})}{\sum_x (-\alpha_x^{(v+2)} b_x^{(v+2)^2} g_{x,t}^{(v+2)})/(1+\alpha_x^{(v+2)} \hat{D}_{x,t}^{(v+2)})}, \\
a_x^{(v+3)} &= a_x^{(v+2)}, \\
b_x^{(v+3)} &= b_x^{(v+2)}, \\
\alpha_x^{(v+3)} &= \alpha_x^{(v+2)};
\end{aligned} \tag{2.25}
$$

$$
\begin{aligned}
\alpha_x^{(v+4)} &= \alpha_x^{(v+3)} - \frac{\sum_t \left\{ f_{x,t}^{(v+3)} - g_{x_t}^{(v+3)} + r_{x,t}^{(v+3)} + \frac{D_{x,t}}{\alpha_x^{(v+3)}} \right\}}{\sum_t \left\{ h_{x,t}^{(v+3)} - \frac{D_{x,t} + 2 r_{x,t}^{(v+3)} \alpha_x^{(v+3)}}{\alpha_x^{(v+3)^2}} + \left(\frac{2}{\alpha_x^{(v+3)^2}} + g_{x,t}^{(v+3)} \right) \left(\frac{\hat{D_{x,t}}^{(v+3)}}{1+\alpha_x^{(v+3)} \hat{D}_{x,t}^{(v+3)}} \right) \right\}}, \\
a_x^{(v+4)} &= a_x^{(v+3)}, \\
b_x^{(v+4)} &= b_x^{(v+3)}, \\
k_t^{(v+4)} &= k_t^{(v+3)};
\end{aligned} \tag{2.26}
$$

where $\hat{D}_{x,t}^{(v)} = E_{x,t} \exp(a_x^{(v)} + b_x^{(v)} k_t^{(v)})$; and $f_{x,t}^{(v)}$, $g_{x,t}^{(v)}$, $h_{x,t}^{(v)}$ and $r_{x,t}^{(v)}$ are defined respectively by

$$
f_{x,t}^{(v)} = \sum_{i=0}^{D_{x,t}-1} \left(\frac{i}{1+\alpha_x^{(v)} i} - \frac{1}{\alpha_x^{(v)}} \right), \tag{2.27}
$$

$$
g_{x,t}^{(v)} = \hat{D}_{x,t}^{(v)} \left(D_{x,t} + \frac{1}{\alpha_x^{(v)}} \right) \left(1 + \alpha_x^{(v)} \hat{D}_{x,t}^{(v)} \right)^{-1}, \tag{2.28}
$$

$$
h_{x,t}^{(v)} = \sum_{i=0}^{D_{x,t}-1} \left(\frac{-i^2}{(1+\alpha_x^{(v)} i)^2} - \frac{1}{\alpha_x^{(v)^2}} \right), \tag{2.29}
$$

$$
r_{x,t}^{(v)} = \frac{1}{\alpha_x^{(v+3)^2}} \ln(1 + \alpha_x^{(v+3)} \hat{D}_{x,t}^{(v+3)}). \tag{2.30}
$$

As with the standard Lee-Carter, we normalize b_x and k_t to sum to unity and zero, respectively. The iteration stops when the change in the log-likelihood function is sufficiently small, say 10^{-6}. Usually, the rate of convergence is slower than that in the Poisson-based

MLE. A simple approach for achieving faster convergence is using the SVD or Poisson ML estimates as the starting values. Finally, we can use parametric bootstrapping to obtain interval forecasts of $m_{x,t}$. The required generation of negative binomial random numbers is fairly easy (see Ross, 2002).

Renshaw and Haberman (2005) proposed another method to handle over-dispersion in the Lee-Carter model. They assume that $D_{x,t}$ follows the Poisson distribution given by equation (2.9). They keep $\mathrm{E}[D_{x,t}] = E_{x,t}\exp(a_x + b_x k_t)$ as before, but they set $\mathrm{Var}[D_{x,t}] = \phi\mathrm{E}[D_{x,t}]$ to relax the mean-variance equality. The scale parameter ϕ can be estimated by the ratio of the residual deviance and its degree of freedom, given by $(X-1)(T-2)$, where X is the total number of age groups. The advantage of this method is that we can circumvent the assumption of a probability distribution for the unobserved heterogeneity. Nevertheless, there are several problems associated with this approach. First, the relationship between $\mathrm{E}[D_{x,t}]$, $\mathrm{Var}[D_{x,t}]$ and the probability function of $D_{x,t}$ is internally inconsistent. Second, as the dispersion parameter is not involved in the estimating equations, ML estimates of model parameters remain unchanged. In other words, the goodness-of-fit is not improved for certain. Third, as only a single dispersion parameter is used, the percentage increase in $\mathrm{Var}[D_{x,t}]$ is the same for all ages. This is unrealistic as the increase should be age-specific. For instance, variations at the advanced ages tend to be higher due to the inaccuracy of reported age at death and the sampling error when the number of deaths and exposures-to-risk are small. Fourth, the dispersion parameter may not be integrated into the bootstrapping of interval forecasts.

2.4 An Example

This section evaluates the performance of our proposed extension using Canadian population mortality data, provided by the Human Mortality Database (HMD). The required data, that is, $E_{x,t}$ and $D_{x,t}$, are given for $x = 0, 1, ..., 99$, and $t = 1921, 1922, ..., 2001$. We use the Coale-Kisker method (Coale and Guo, 1989) to smooth the death rates beyond age 85. The issue of modeling old-age mortality will be discussed in more detail in Chapter 4.

Through a graphical analysis of residuals, Koissi et al. (2005) have shown that the method of Poisson-based MLE provides a better fit than the methods of SVD and WLS. We

shall therefore focus on the comparison between the Poisson-based MLE and our proposed generalization.

2.4.1 Parameter Estimates

Figures 2.1 and 2.2 show that the influence of the extension on the base age pattern a_x and the relative speed of improvement b_x is minor. Figure 2.3 suggests that the linearity in the time-varying component is preserved.

Figure 2.4 shows that the estimates of α_x are especially high at the advanced ages, indicating a high level of heterogeneity within cells. For example, the elderly may be classified by their activities of daily living (ADL) limitations (Kassner and Jackson, 1998), which may be closely related to their mortality. From another point of view, this observation is also consistent with the empirical fact that vital statistics at the extreme ages are subject to significant recording and sampling errors (see, e.g., Bourbeau and Desjardins, 2002).

2.4.2 Goodness-of-fit

Table 2.1 gives a formal evaluation of the goodness-of-fit. The significantly higher log-likelihood suggests that our proposed extension provides a better fit. Nevertheless, 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 change in the log-likelihood, given that we have introduced additional parameters (α_x's). To account for the extra parameters, we can use the following model selection criteria:

1. The Akaike Information Criterion (AIC) (Akaike, 1974), defined by $l - j$, where l is the log-likelihood and j is the number of parameters. The AIC merits the increase in log-likelihood, but penalizes for the introduction of additional 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. Again, we prefer models with a higher SBC.

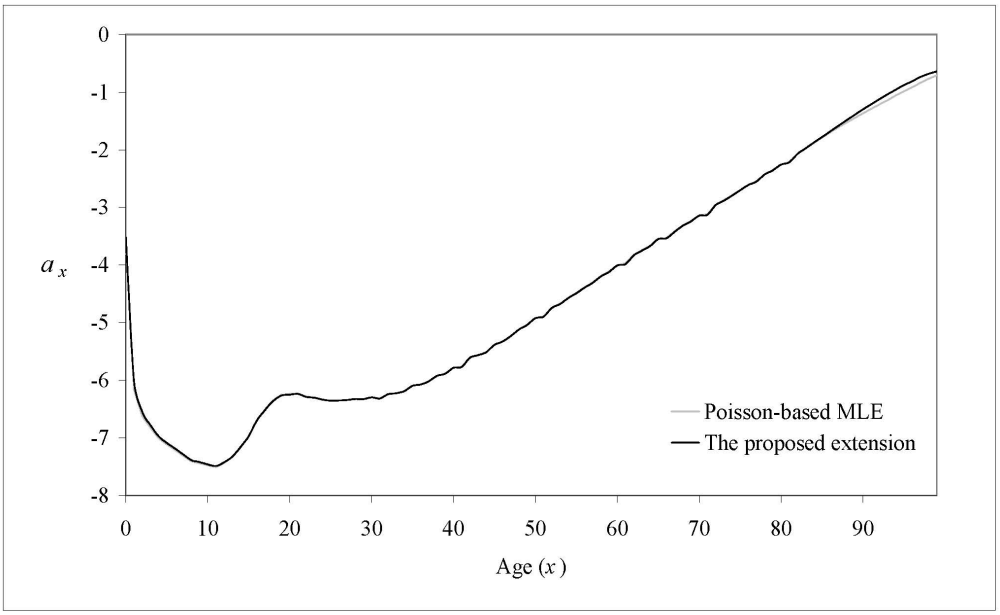
3. Likelihood-ratio test (LRT) (see Klugman et al. 2004). The null hypothesis of LRT is that there is no significant improvement in the more complex model. Let l_1 be the log-likelihood of under Poisson MLE, and l_2 be the log-likelihood under our proposed extension. The test statistic is $2(l_2 - l_1)$. 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 the AIC and the SBC are presented in Table 2.1. The p-values of the LRTs are less than 10^{-6}. All three criteria support the use of our proposed extension. Furthermore, as shown in Figures 2.5 and 2.6, the increased flexibility seems capable of correcting the under-fit manifested near the forecast origin, particularly at the young ages. However, the relaxed model structure has no discernible harm to the projected age pattern of mortality, as shown in Figure 2.7.

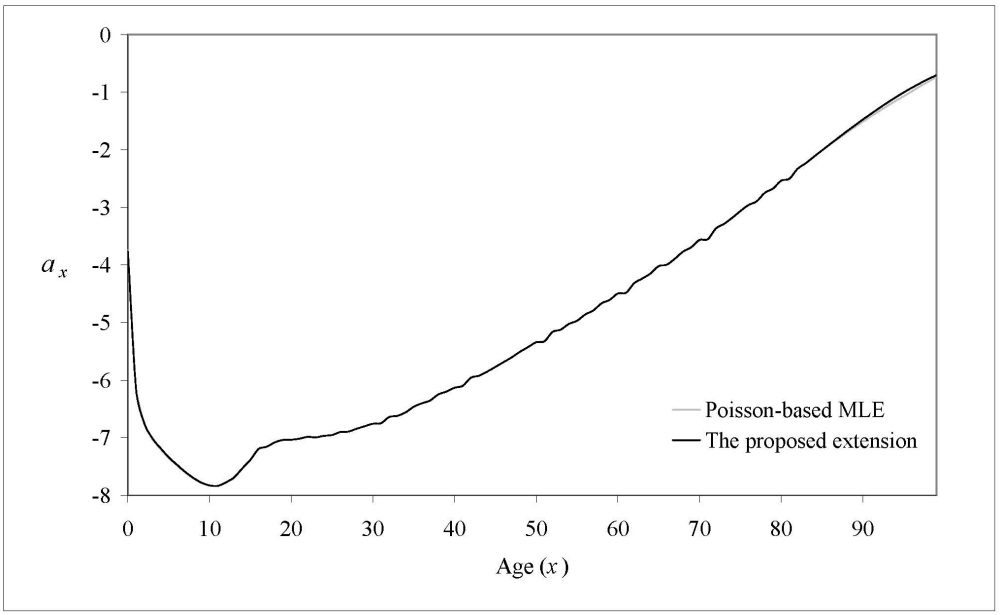
2.4.3 Interval Forecasts

Finally, we obtain the interval forecasts by parametric bootstrapping, and the results are presented graphically in Figures 2.5 and 2.6. Notice that the percentage increase in the interval width is proportional to α_x. On average, the increase in width at the younger ages (0-40) is around 6 percent and that at the higher ages (75 and over) is over 100 percent. This agrees with our previous assertion that the mean-variance equality restriction in the Poisson MLE has lead us to understating the variations, mostly at the higher ages.

Readers should bear in mind that no matter how the structure is relaxed, model risk still exists. The proposed extension has not taken into account the possibility of future structural changes, and it has not eliminated the need for model parameter re-calibration when future rates are realized.

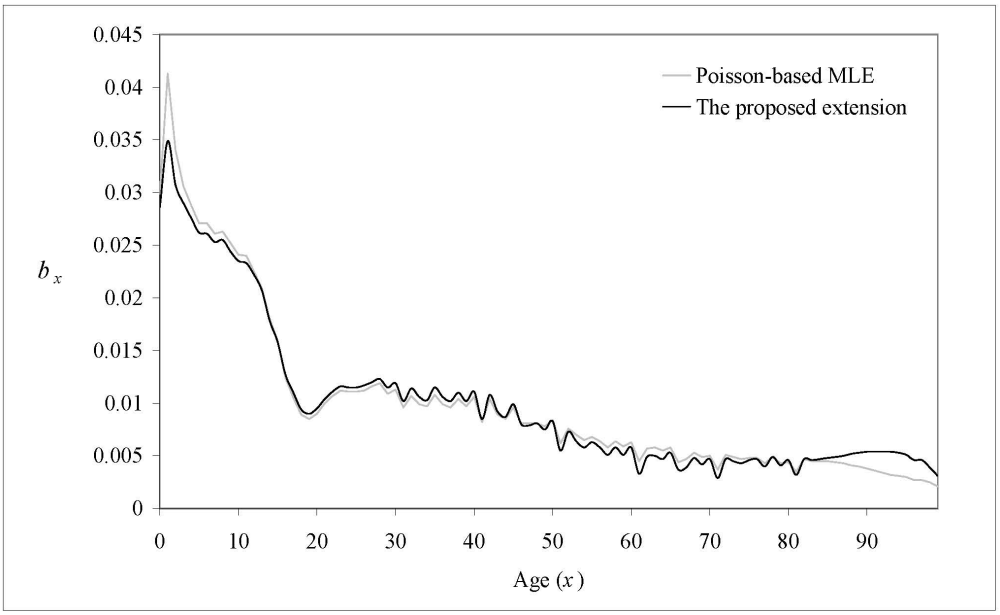


Male

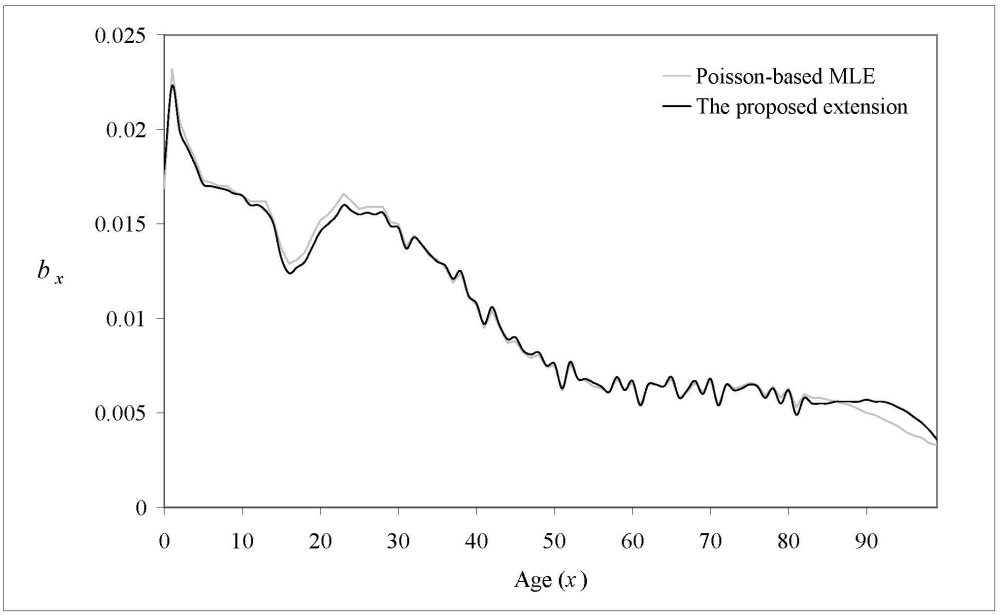


Female

Fig. 2.1. Estimate of parameter a_x.

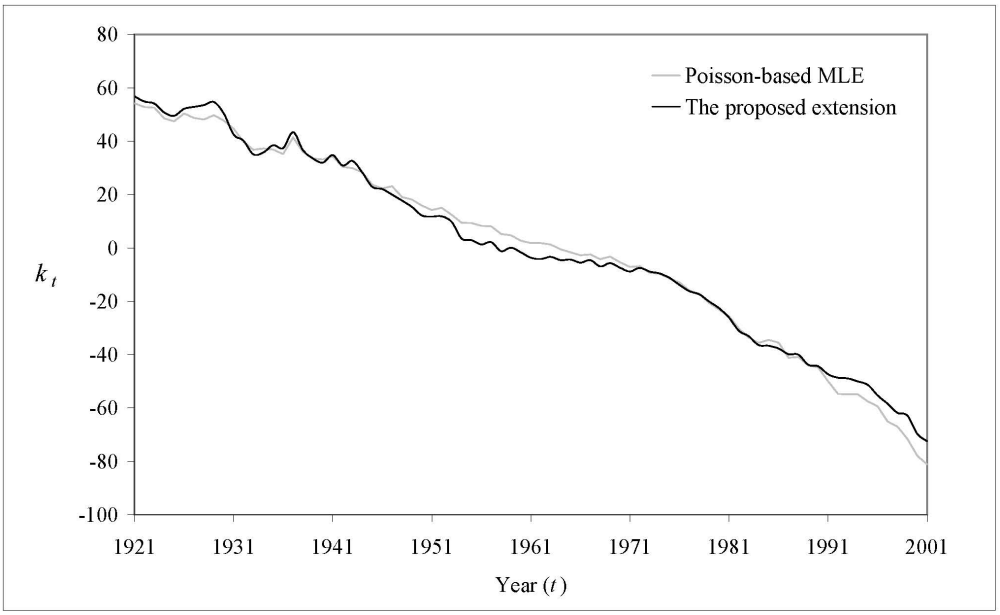


Male

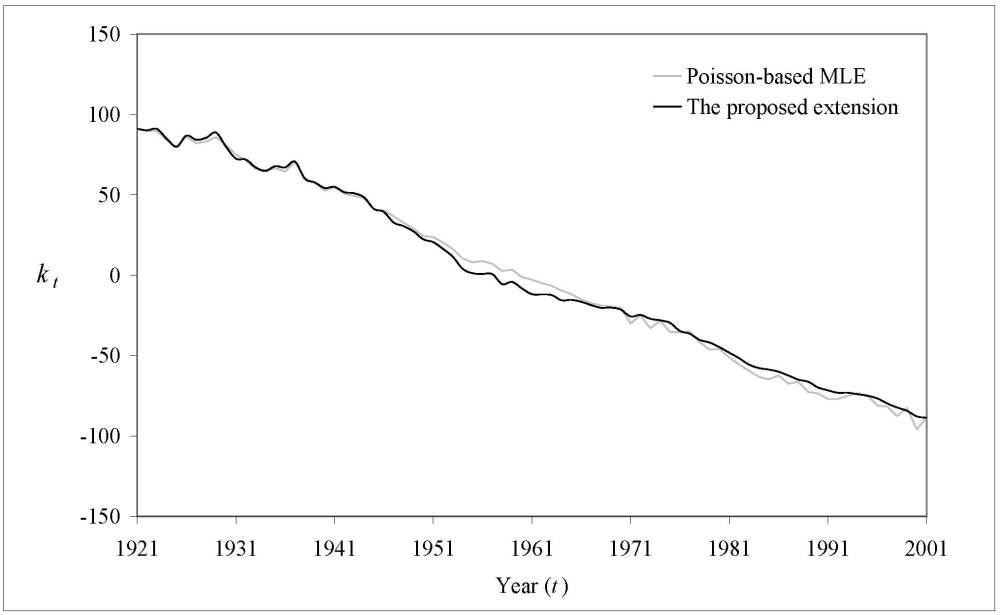


Female

Fig. 2.2. Estimate of parameter b_x.

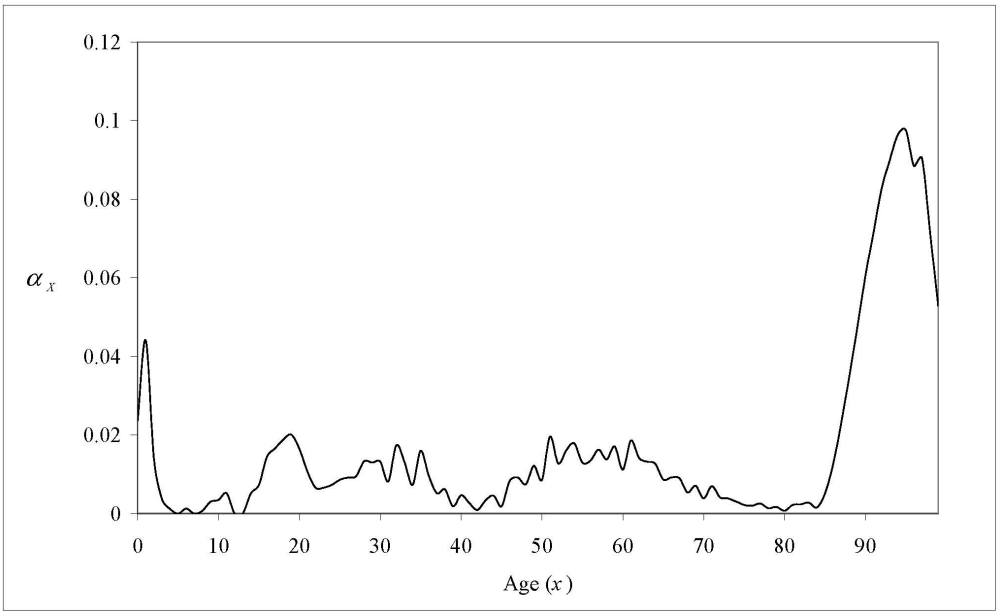


Male

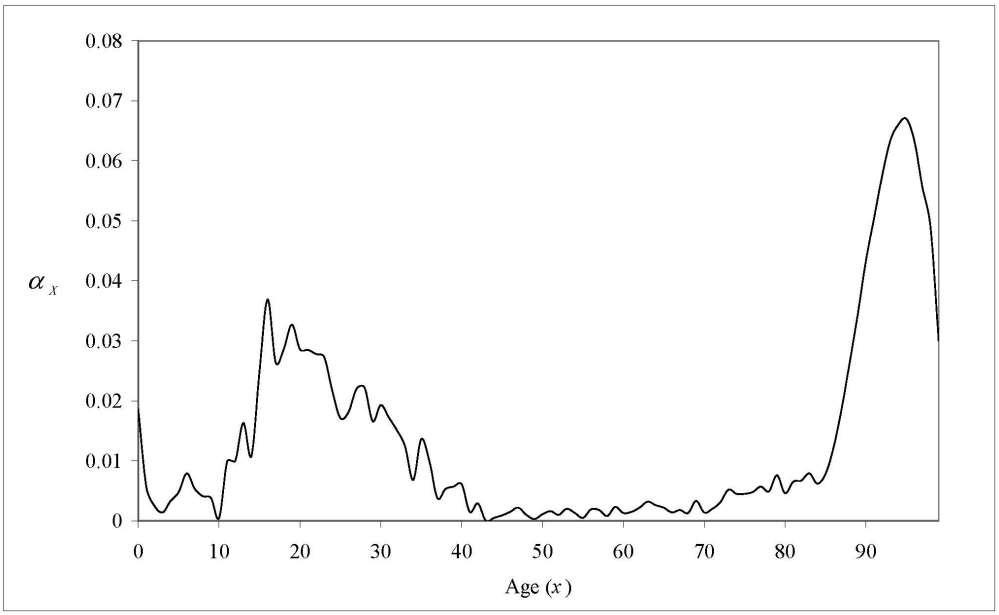


Female

Fig. 2.3. Estimate of parameter k_t.

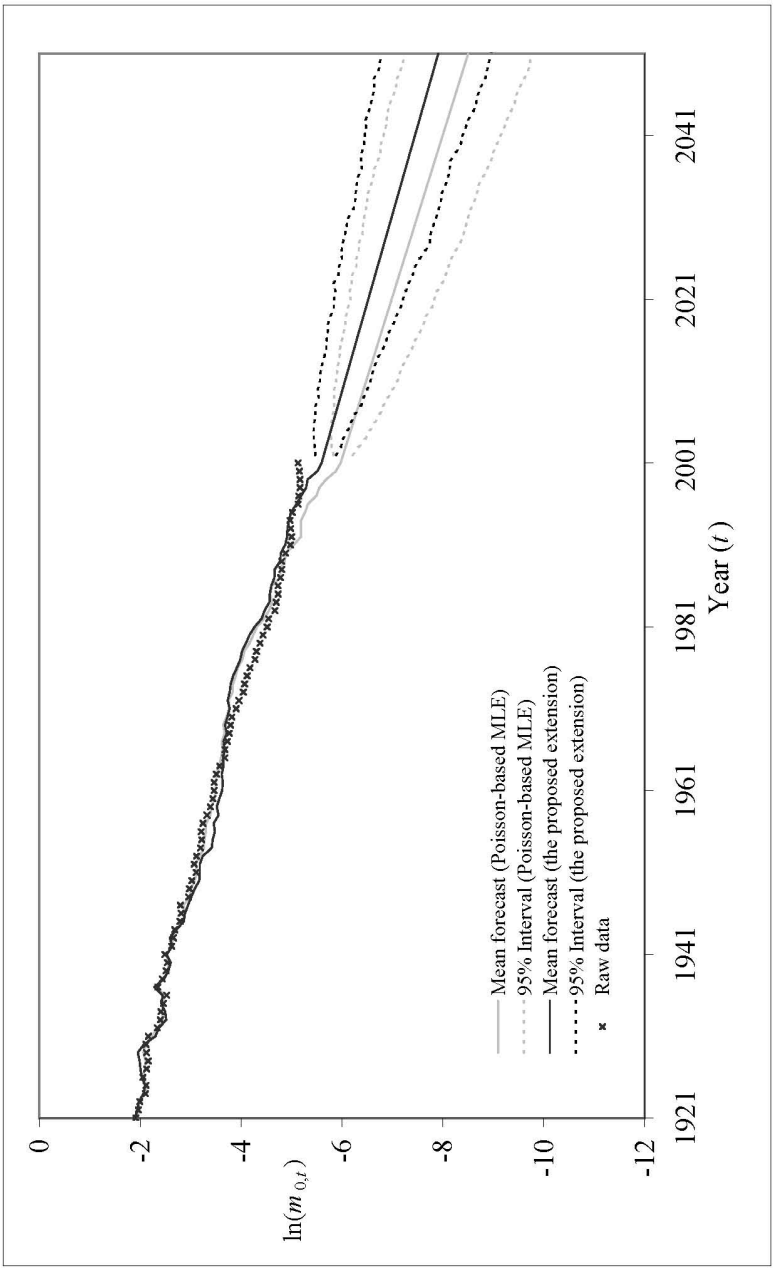


Male



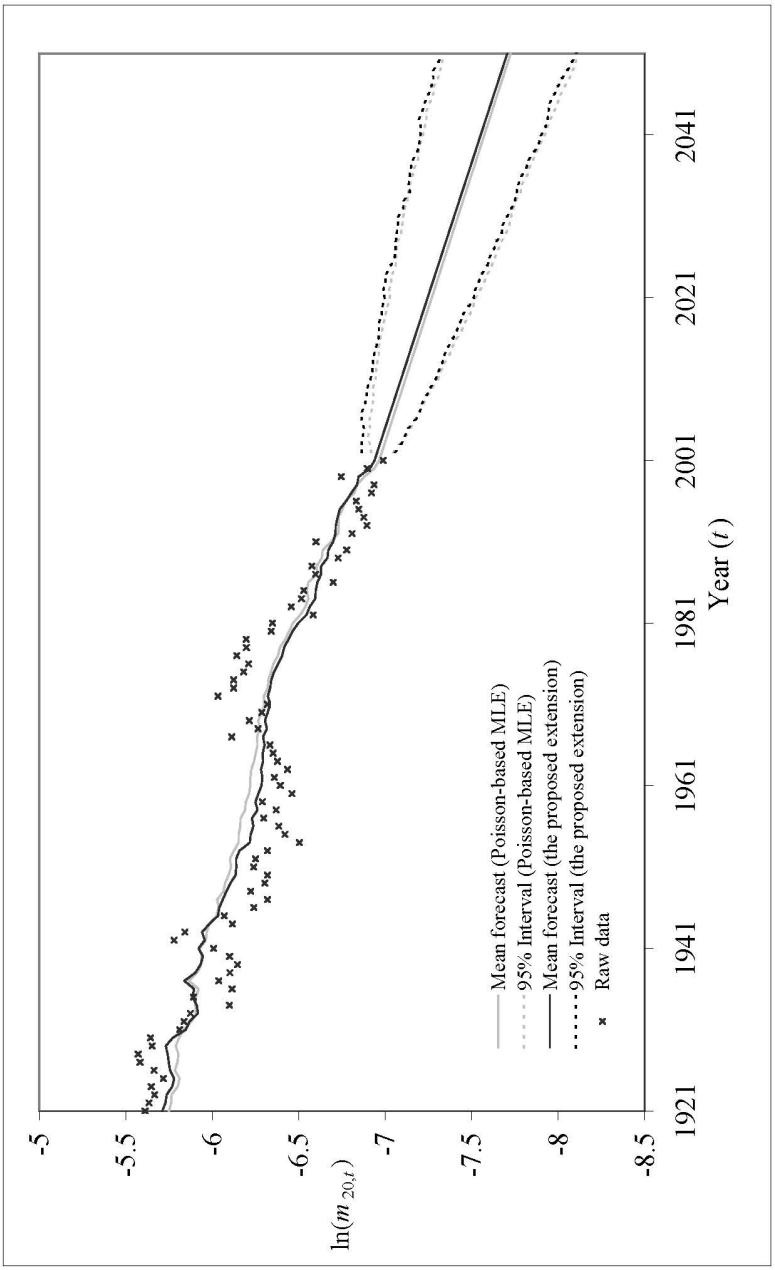
Female

Fig. 2.4. Estimate of parameter α_x.



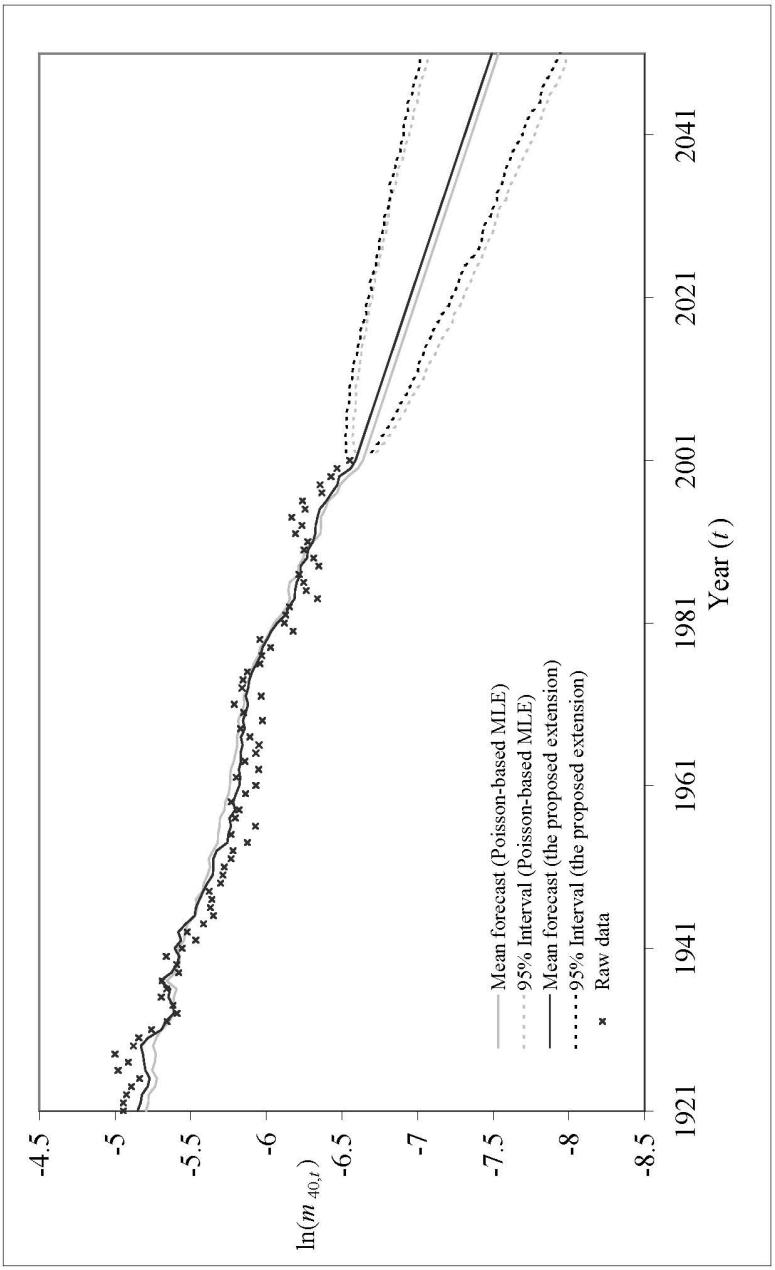
Age 0

Fig. 2.5. The fit (1921-2001) and the projection (2002-2051) of $\ln(m_{x,t})$ at representative ages, Poisson-based MLE and the proposed extension, male.



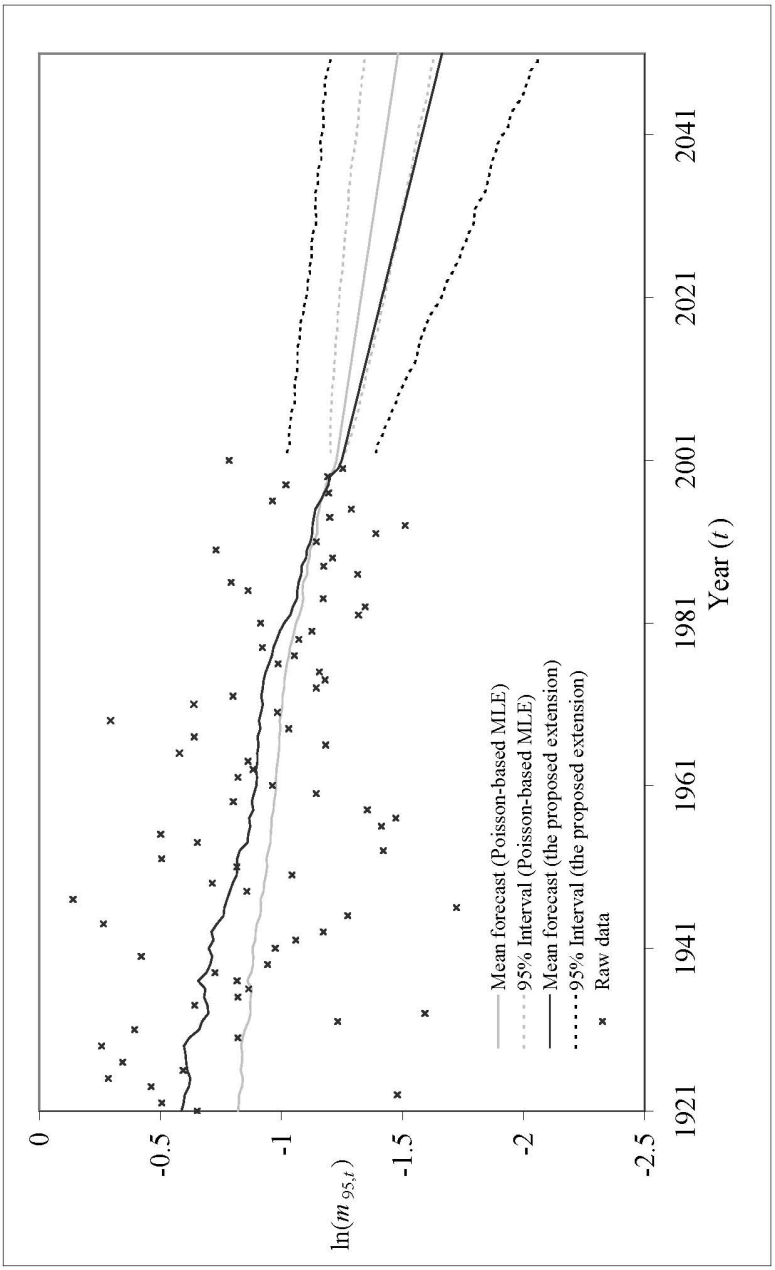
Age 20

Fig. 2.5. (cont'd) The fit (1921-2001) and the projection (2002-2051) of $\ln(m_{x,t})$ at representative ages, Poisson-based MLE and the proposed extension, male.



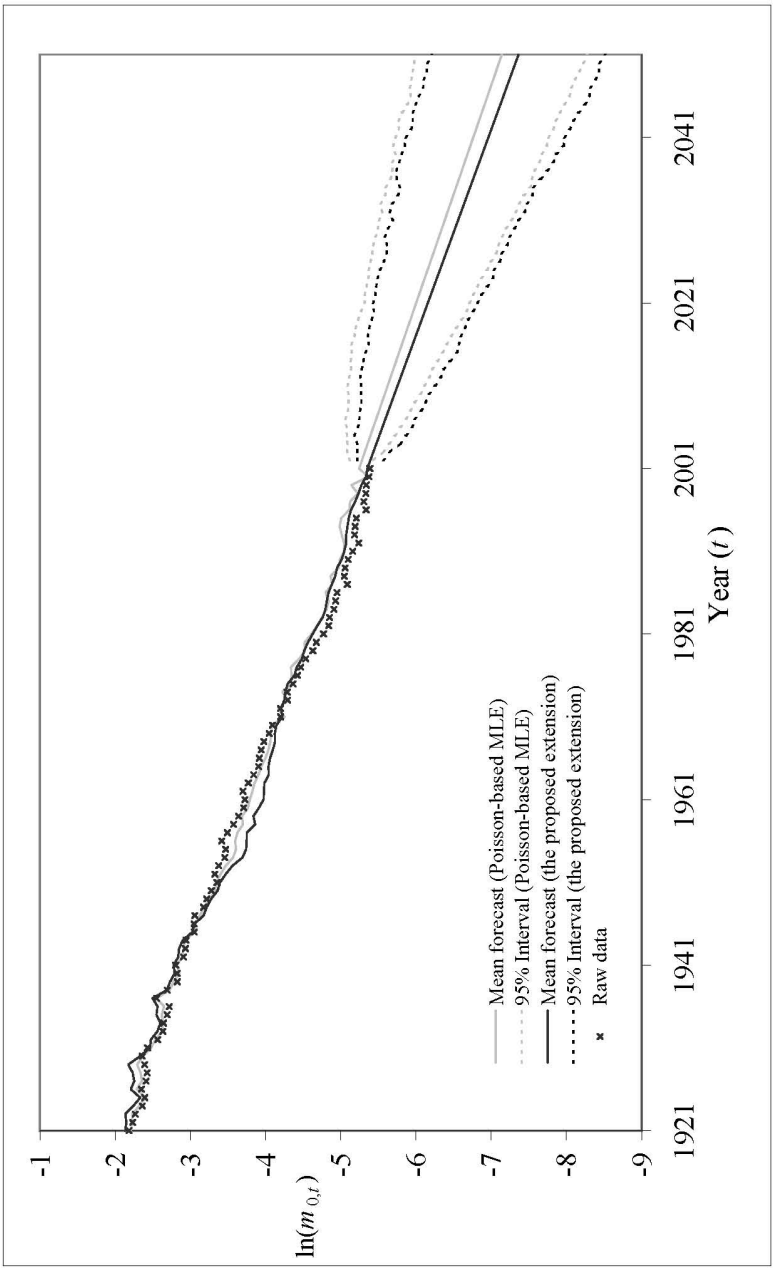
Age 40

Fig. 2.5. (cont'd) The fit (1921-2001) and the projection (2002-2051) of $\ln(m_{x,t})$ at representative ages, Poisson-based MLE and the proposed extension, male.



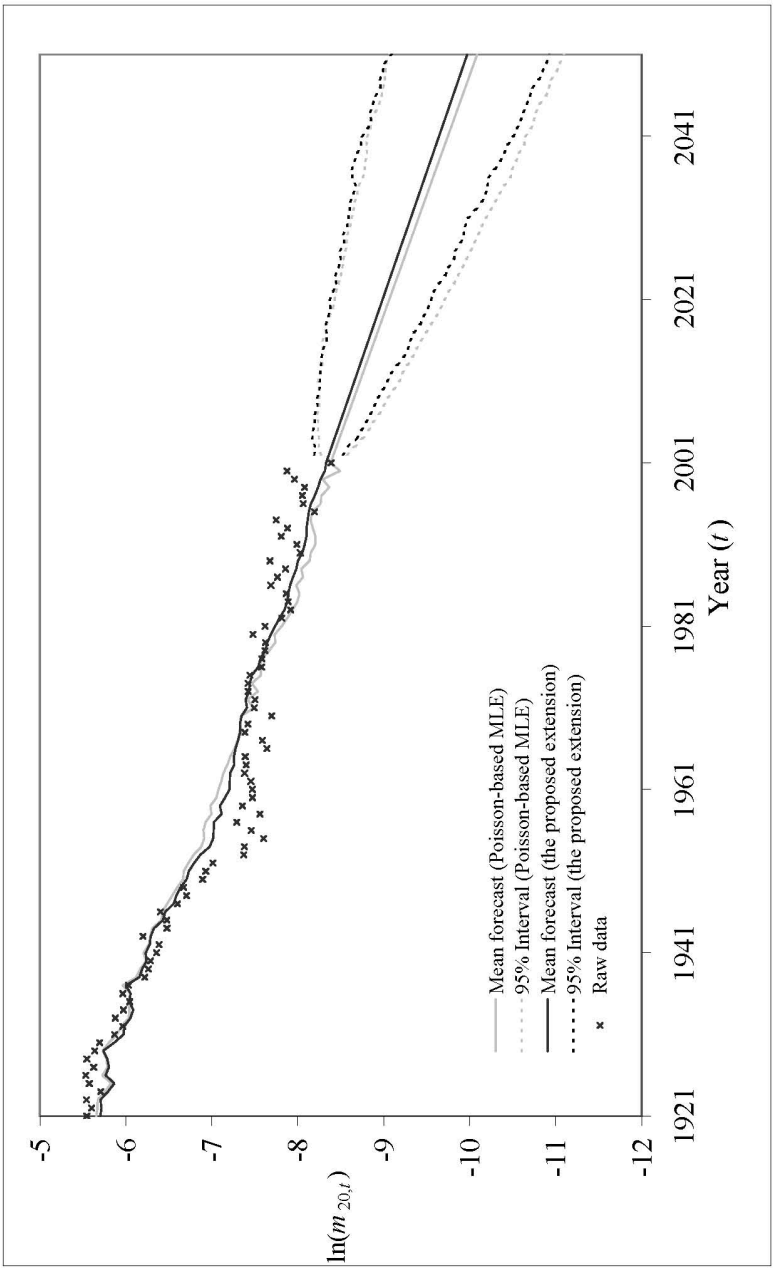
Age 95

Fig. 2.5. (cont'd) The fit (1921-2001) and the projection (2002-2051) of $\ln(m_{x,t})$ at representative ages, Poisson-based MLE and the proposed extension, male.



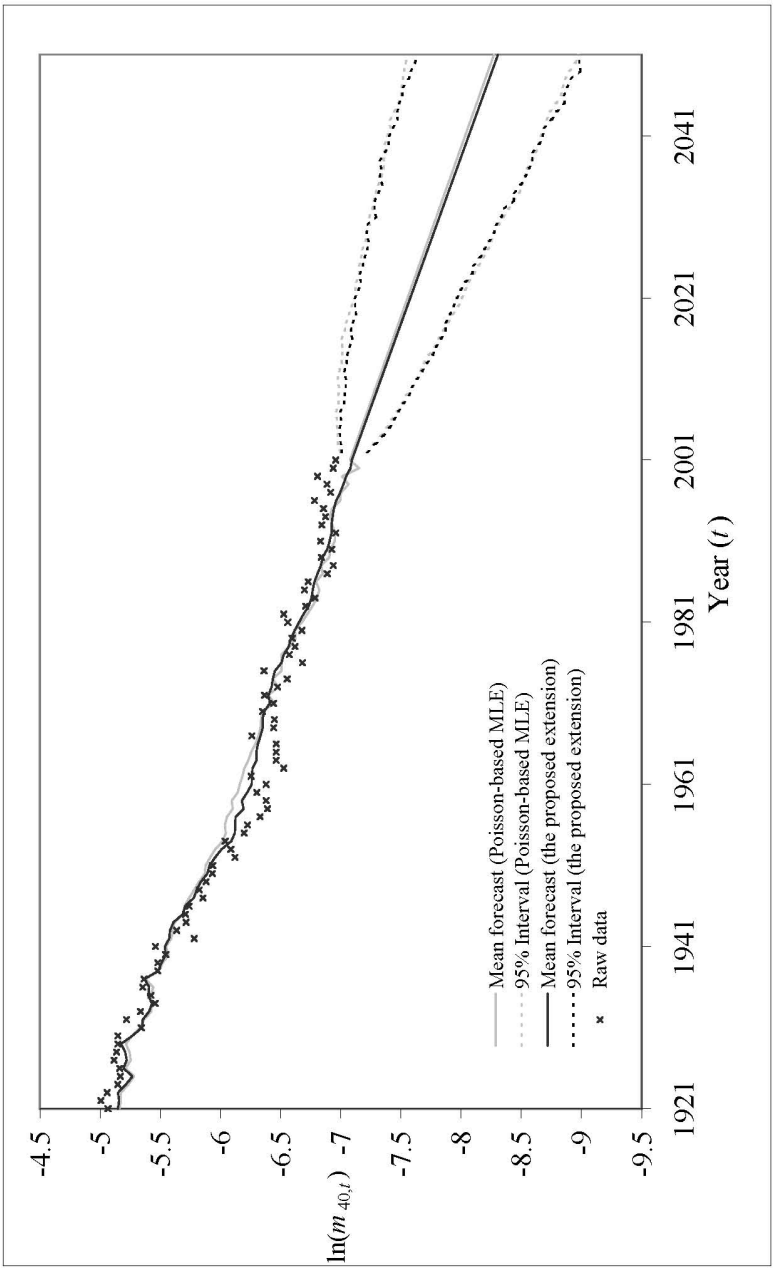
Age 0

Fig. 2.6. The fit (1921-2001) and the projection (2002-2051) of $\ln(m_{x,t})$ at representative ages, Poisson-based MLE and the proposed extension, female.



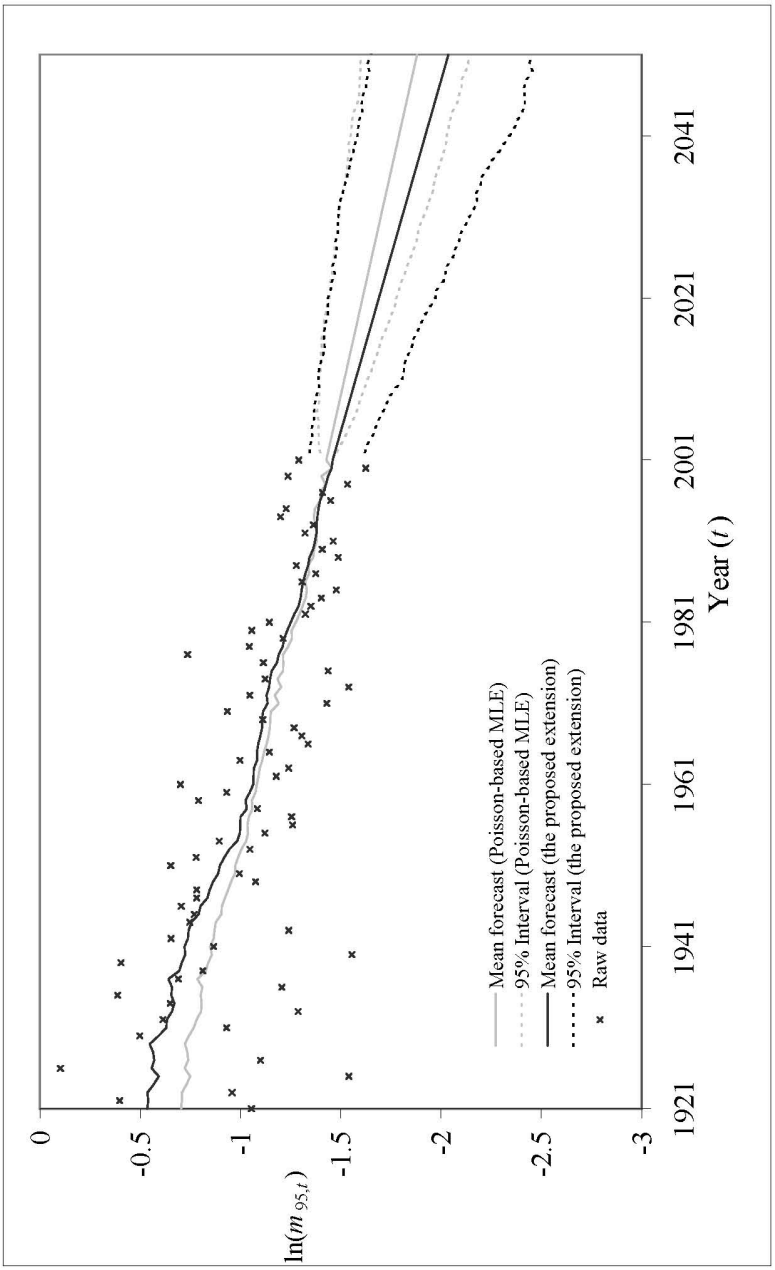
Age 20

Fig. 2.6. (cont'd) The fit (1921-2001) and the projection (2002-2051) of $\ln(m_{x,t})$ at representative ages, Poisson-based MLE and the proposed extension, female.



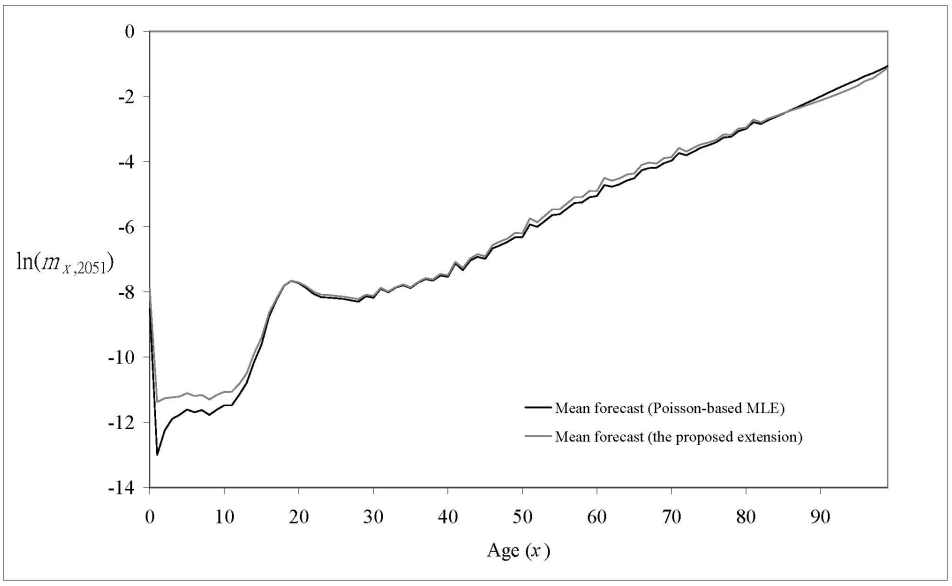
Age 40

Fig. 2.6. (cont'd) The fit (1921-2001) and the projection (2002-2051) of $\ln(m_{x,t})$ at representative ages, Poisson-based MLE and the proposed extension, female.

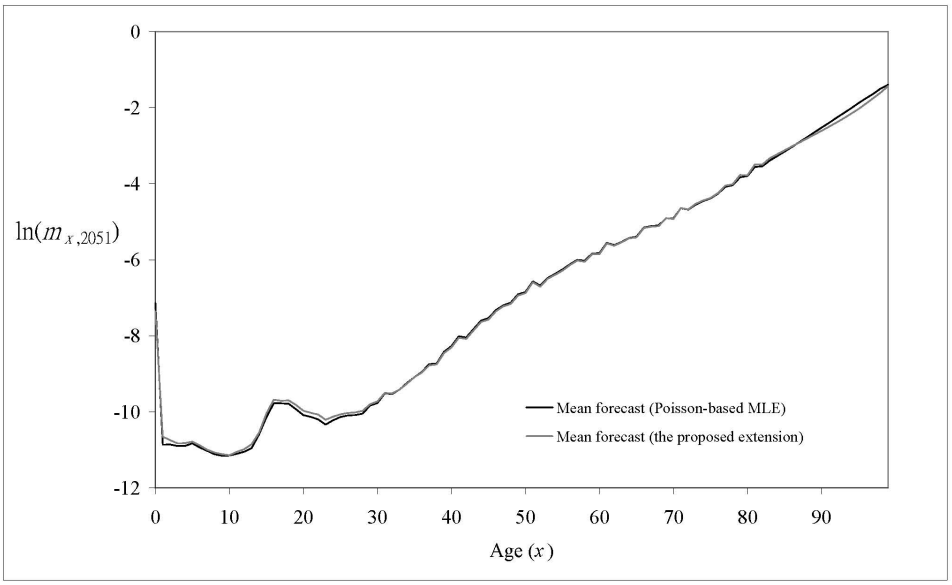


Age 95

Fig. 2.6. (cont'd) The fit (1921-2001) and the projection (2002-2051) of $\ln(m_{x,t})$ at representative ages, Poisson-based MLE and the proposed extension, female.



Male



Female

Fig. 2.7. Projected age pattern of mortality in 2051, Poisson-based MLE and the proposed extension.

	Poisson-based MLE	The Proposed Extension
Number of Parameters	279	379
Log-likelihood	−66,021	−45,517
AIC	−66,300	−43,896
SBC	−67,276	−45,222

Male

	Poisson-based MLE	The Proposed Extension
Number of Parameters	279	379
Log-likelihood	−52,643	−40,748
AIC	−52,922	−41,127
SBC	−53,898	−42,453

Female

Table 2.1. Comparison of model selection information for the Poisson-based MLE and the proposed extension.

2.5 Conclusion

In this chapter, we relaxed the original Lee-Carter model by incorporating heterogeneity in each age-period cell. By assuming that the unobserved heterogeneity follows a gamma distribution, we demonstrated that the distribution of the number of deaths is negative binomial. This generalization implicitly introduced an extra parameter vector (the age-specific dispersion parameters) to the model. As the main specification of the model was unaltered, the extension retained all the appealing features, including the linearity in the time-varying component and the stability of age-patterns, of the original version. Given the more conservative confidence intervals and the improved goodness-of-fit, the proposed generalization seems to be an attractive alternative to the original version.

For mathematical convenience, we assumed a single parameter gamma distribution for the unobserved heterogeneity. Although the gamma distribution may be replaced by any continuous distribution with positive support, the mixture of distributions may not be carried out analytically. In this case, the ML estimates of 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 we mentioned, the modeling proceeds in two steps: we first estimate the model parameters a_x, b_x, k_t and α_x (if applicable), and then the estimate of k_t is further modeled by an ARIMA process for extrapolation. Czado et al. (2005) pointed out that this two-step procedure may give rise to incoherence and they proposed using the Markov Chain Monte Carlo (MCMC) method that could combine the two steps. 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.

Chapter 3

Mortality Improvement Scales for Canadian Insured Lives

3.1 Introduction

In Chapter 1, we reviewed several previous mortality improvement scales. These scales are purely deterministic in the sense that no measure of uncertainty is given. This chapter aims to develop improvement scales for Canadian insured lives using stochastic mortality models. The resultant scales should include a range of possible outcomes, and a probability attached to that range.

Willekens (1990) suggested that stochastic models for forecasting mortality can be roughly divided into two categories, namely, extrapolative and process-oriented. A mutual shortcoming of all extrapolative models is an entire reliance on observed past trends and consequently a lack of information on the forces shaping the changes in mortality. Nevertheless, the implementation of process-oriented methods is always obscured by the statistical difficulties in determining the dependencies between causes of death and the unavailability of the required individual level data − in this study, the available data consist of no information on causes of death and risk factors other than smoker-status. For that reason, we shall focus on various extrapolative models. These models shall give us prediction intervals that could allow for a wide range of possible outcomes, so as to cope with our ignorance of the complex biological mechanisms underlying.

The implementation of extrapolative models is not straightforward, due primarily to the two limitations in the available Canadian insured lives experience. First, the experience is available for only 20 policy years (1982-1983 to 2001-2002), which is probably too short for a direct statistical projection, no matter which model is used. Second, in earlier policy years, the volume of data with known smoker status is scanty, and there is not even any smoker breakdown in the ultimate data before policy year 1992-1993. This inevitably precludes the derivation of improvement scales separately for smokers and non-smokers.

To overcome these problems, we rely additionally on Canadian population mortality data, which covers a far longer period of 81 calendar years (1921 to 2001) and is much richer in the number of exposures-to-risk. Given that the Canadian population is highly insured, the mortality experiences of the general population and insured lives should share common features. Hence, the simpler approach of fitting solely to the insured lives data is, in a sense, inefficient. In our proposed methodology, both experiences are projected simultaneously by means of a joint model, which consists of four fitting stages. First, we project Canadian population mortality. Second, we summarize the projection by some tractable mathematical formulae, which give mortality improvement scales for the general population. Third, we search for a persistent parametric relationship between the experiences of the general population and insured lives. Fourth, based on the parametric relationship, we modify the results in stage two to obtain improvement scales for Canadian insured lives. If necessary, these scales are further adjusted for the effects of selection and/or smoker status.

The flow of this chapter follows the sequence in which the ultimate improvement scales are derived. In Section 2, we state all sources of data. In Section 3, we introduce various stochastic models for forecasting mortality. These models are modified to suit our purposes and then fitted to the mortality experience of the general population. In Section 4, we focus on the insured lives. We firstly perform a two dimensional graduation to obtain smoothed insured lives mortality rates, which are then related to the mortality rates of the general population by an appropriate parametric model. Given this relationship and the results in Section 3, we obtain mortality improvement scales for the insured lives. In Section 5, we provide a brief discussion on the effects of duration and smoker status.

3.2 Sources of Data

Below we list the sources of data to be used in this chapter.

- The general population
  The mortality data for the Canadian population are the same as that used in Chapter 2. According to the HMD documentation, there were a few changes in the coverage of the vital statistics during that period. However, rigorous outlier analyses presented in the later part of this chapter indicate that the effect of these changes is not significant.

- The insured lives
  The insured lives experience was collected from all life insurance companies in Canada by the Institute of Insurance and Pension Research at the University of Waterloo. It covers policy years 1982-1983 to 2001-2002. Numbers of deaths and exposures-to-risk are available for both sexes and all ages up to 99. The data are further segregated by:

  1. sex − male and female;
  2. duration − 1, 2, and so on up to 15, and ultimate (16+);
  3. smoker status − smoker, non-smoker and indeterminate (no smoker breakdown in the ultimate data prior to policy year 1992-1993).

  In each category, data are given in terms of both the number of lives and the amount of insurance. We prefer "numbers" to "amounts" as the trends in the "numbers" data are much more stable.

3.3 Projecting the Mortality Experience of the General Population

In this section, we project the mortality experience of the general population using two genres of stochastic mortality models, namely, the P-splines regression (Currie et al., 2004) and the Lee-Carter model (Lee and Carter, 1992). These models are also considered by the CMIB in projecting the United Kingdom insured lives experience (CMIB, 2005).

3.3.1 The P-splines Regression

Let us begin with the one-dimensional case. Let $D_{x,t}$, $\mu_{x,t}$ and $E_{x,t}$ respectively be the number of deaths, the force of mortality, and the number of exposures-to-risk at age x and time t. For convenience, we write $\mathbf{d}' = \{D_{x,1}, D_{x,2}, ..., D_{x,T}\}$, $\boldsymbol{\mu}' = \{\mu_{x,1}, \mu_{x,2}, ..., \mu_{x,T}\}$, and $\mathbf{e}' = \{E_{x,1}, E_{x,2}, ..., E_{x,T}\}$, where T is number of periods in the time-series mortality data. To model $\mu_{x,t}$, we may regress $\ln(m_{x,t})$ against some basis functions $\mathbf{B}$; that is,

$$\ln(\mu_{x,t}) = \mathbf{Ba} + \epsilon_{x,t}, \tag{3.1}$$

where $\mathbf{a}$ is the matrix of regression coefficients and $\epsilon_{x,t}$ is the error term. In a simple linear regression, $\mathbf{B} = \{1, t\}$, and in a B-splines regression, the basis is a collection of K cubic B-splines, that is, $\mathbf{B} = \{B_1(x), B_2(x), ..., B_K(x)\}$, where $B_i(x)$, $i = 1, 2, ..., K$, is a B-spline formed by joining cubic polynomial pieces smoothly. Therefore, in a B-splines regression, the fitted value of $\ln(\mu_{x,t})$ becomes

$$\hat{\ln}(\mu_{x,t}) = \sum_{i=1}^{K} B_i(x)\hat{a}_i. \tag{3.2}$$

The coefficients can be estimated by maximizing the log-likelihood, $l(\mathbf{a}; \mathbf{d}, \mathbf{e})$, based on the assumed distribution for $D_{x,t}$. We assume that $D_{x,t}$ follows a Poisson distribution with mean $E_{x,t}\mu_{x,t}$.

In fitting a B-splines regression, we usually use a high value of K for a high goodness-of-fit. This, however, may lead to unwanted overfitting. To deal with this problem, we can use a P-splines regression (Eilers and Marx, 1996), which restricts the parameters by a difference penalty on the adjacent coefficients. By incorporating the penalty into the original log-likelihood, l, we have the following penalized log-likelihood function:

$$l_p(\mathbf{a}; \mathbf{d}, \mathbf{e}) = l(\mathbf{a}; \mathbf{d}, \mathbf{e}) - \frac{1}{2}\lambda \mathbf{a}'\boldsymbol{\Delta}'\boldsymbol{\Delta}\mathbf{a}, \tag{3.3}$$

where $\boldsymbol{\Delta}$ is a difference matrix of order n. The maximization of l_p is equivalent to maximizing the goodness-of-fit, measured by l, and simultaneously minimizing the n^{th} order differences between the adjacent coefficients so that the coefficients tend to lie on a degree n adaptive polynomial. The balance between goodness-of-fit and smoothness is controlled

by the smoothing parameter, λ. Let $\mathbf{P} = \lambda \boldsymbol{\Delta}'\boldsymbol{\Delta}$ be the penalty matrix. The maximization of l_p gives the following penalized likelihood equation:

$$\mathbf{B}'(\mathbf{d} - \mathbf{e}\boldsymbol{\mu}) = \mathbf{P}\mathbf{a}, \tag{3.4}$$

which can be solved by the scoring algorithm below:

$$\hat{\mathbf{a}} = (\mathbf{B}'\hat{\mathbf{W}}\mathbf{B} + \mathbf{P})^{-1}(\mathbf{B}'\hat{\mathbf{W}}\mathbf{B}\tilde{\mathbf{a}} + \mathbf{B}'(\mathbf{d} - \mathbf{e}\tilde{\boldsymbol{\mu}})), \tag{3.5}$$

where $\tilde{\mathbf{a}}$, $\tilde{\boldsymbol{\mu}}$ and $\tilde{\mathbf{W}}$ are the current estimates of $\mathbf{a}$, $\boldsymbol{\mu}$; and $\mathbf{W}$, respectively; $\hat{\mathbf{a}}$ is the updated estimate of $\mathbf{a}$; $\mathbf{W} = \text{diag}(\boldsymbol{\mu})$. The approximate variance of the linear predictor is given by

$$\text{Var}(\mathbf{B}\hat{\mathbf{a}}) \approx \mathbf{B}(\mathbf{B}'\hat{\mathbf{W}}\mathbf{B} + \mathbf{P})^{-1}\mathbf{B}'. \tag{3.6}$$

Assuming normality holds, the approximate 95% confidence interval for $\ln(\mu_{x,t})$ can be written as

$$\hat{\ln}(\mu_{x,t}) \pm 1.96\sqrt{\text{Var}(\hat{\ln}(\mu_{x,t}))}, \tag{3.7}$$

where $\text{Var}(\hat{\ln}(\mu_{x,t}))$ is the t^{th} diagonal element in the matrix of $\text{Var}(\mathbf{B}\hat{\mathbf{a}})$.

In forecasting, we treat future values of $D_{x,t}$ and $E_{x,t}$ $(t > T)$ as missing data. Suppose that we forecast future death rates in S years from now. Then we have to extend the original basis matrix $\mathbf{B}$ to $\mathbf{B}^*$ that covers $T + S$ years; that is,

$$\mathbf{B}^* = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix}, \tag{3.8}$$

where $\mathbf{B}_1$ and $\mathbf{B}_2$ contain the B-splines that cover the entire domain, $\{1, 2, ..., T, T + 1, ..., S\}$. Let $\mathbf{V} = \text{blockdiag}(\mathbf{I}; \mathbf{0})$, where $\mathbf{I}$ is an identity matrix of size T and $\mathbf{0}$ is a S-by$S$ square matrix of zeros. Then, the scoring algorithm specified by equation (3.5) can be rewritten as

$$\hat{\mathbf{a}} = (\mathbf{B}'\mathbf{V}\tilde{\mathbf{W}}\mathbf{B} + \mathbf{P})^{-1}(\mathbf{B}'\mathbf{V}\tilde{\mathbf{W}}\mathbf{B}\tilde{\mathbf{a}} + \mathbf{B}'\mathbf{V}(\mathbf{d} - \mathbf{e}\tilde{\boldsymbol{\mu}})), \tag{3.9}$$

which enables us to fit and forecast simultaneously. The mean and interval forecast can be obtained by equations (3.2) and (3.7), respectively.

Note that the above algorithm depends on the choice of K, n and l. Eilers and Marx (1996) used the following rule of thumb to choose K:

$$K = \max\left(\frac{T}{4}, 40\right) + 3. \tag{3.10}$$

The choice of n determines the form of the forecast. Setting $n = 1$ implies no mortality improvement in the future; setting $n = 2$ gives a linear forecast; setting $n = 3$ gives a quadratic forecast. As neither the consequence of $n = 1$ nor that of $n = 3$ seems reasonable, we set $n = 2$. Finally, λ is chosen to minimize the Bayesian Information Criterion (BIC) (Schwarz, 1978), which is defined by

$$\text{BIC} = Dev + \ln(n \times Tr), \tag{3.11}$$

where Dev is the deviance, and Tr is the trace of the hat matrix

$$\mathbf{H} = \mathbf{B}(\mathbf{B}'\mathbf{W}\mathbf{B} + \mathbf{P})^{-1}\mathbf{B}'\mathbf{W}. \tag{3.12}$$

We now move on to the two-dimensional case. Let $D = \{D_{x,t}\}_{X\times T}$, $M = \{m_{x,t}\}_{X\times T}$ and $E = \{E_{x,t}\}_{X\times T}$, where X is the total number of age groups. Let $\mathbf{B}_a$ and $\mathbf{B}_y$ be the basis matrices when the P-splines regression is applied to the columns and rows of $\mathbf{M}$, respectively. It can be shown that the basis matrix for the two-dimensional model is

$$\mathbf{B} = \mathbf{B}_y \otimes \mathbf{B}_a, \tag{3.13}$$

where $\otimes$ is the Kronecker product operator. Similarly, let $\mathbf{D}_a$ and $\mathbf{D}_y$ be the difference matrices when the P-splines regression is applied to the columns and rows of $\mathbf{M}$, respectively. It can be shown that the penalty matrix for the two-dimensional model is

$$\mathbf{P} = \lambda_a \mathbf{I}_{c_y} \otimes \mathbf{D}_a'\mathbf{D}_a + \lambda_y \mathbf{D}_y'\mathbf{D}_y \otimes \mathbf{I}_{c_a}, \tag{3.14}$$

where λ_a and λ_y are the smoothing parameter in the age dimension and the year dimension, respectively; c_a and c_y are the number of columns in $\mathbf{B}_a$ and $\mathbf{B}_y$, respectively. Let vec($\mathbf{Y}$) be the operator that converts $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N]$ into $[\mathbf{y}_1', \mathbf{y}_2', ..., \mathbf{y}_N']$. To obtain the algorithm for smoothing and forecasting in the two-dimensional case, we replace $\mathbf{B}$, $\mathbf{P}$ and $\mathbf{d}$ in the one-dimensional case by equation (3.13), equation (3.14) and vec($\mathbf{D}$), respectively.

Figures 3.1 to 3.3 show the P-splines projection of the general population mortality. The P-splines regression has two attractive features. One is that the interval forecasts are conservative, and the other is an excellent fit to the data.

Nevertheless, the P-splines behave arbitrarily badly outside the region of the data. At infancy and older ages, the projection is nonsensical − the mean forecast of m_{90} exceeds

1 in no more than 20 years. The projected age pattern is inconsistent with the generally accepted ones. For males, it is almost flat from age 30 to 70, and there exists an unexplainable hump at around age 75. Also, for both sexes, there is a anti-intuitive fall at around age 90, leading to mortality crossovers. These unappealing phenomena may be attributed to the fundamental assumptions of the model. First, in contrast to other models, the P-splines regression is two-dimensional in the sense that the entire mortality surface is allowed to change over time. The high degree of freedom allows the P-splines to give a high goodness-of-fit and wide confidence intervals. However, as there are fewer constraints on the behavior of future rates, the age pattern is not stable. Second, as the P-splines method is regression based, death rates are assumed not to be serially correlated. No different from other extrapolative methods, the P-splines forecast is a weighted average of past observations. The inappropriate assumption of independence may lead to an incorrect specification of weights so that too much emphasis is placed on recent observations. In Figures 3.1 and 3.2, we notice that the forecasts are almost totally determined by the slope of the last 5 (perhaps even fewer) observations. In other words, an abnormality near the forecast origin could severely distort the entire forecast.

Overall, the P-splines method does an excellent job in smoothing the data but not in projecting Canadian population mortality. Therefore, we use the P-splines regression for the purpose of data and parameter graduation, and seek alternative approaches for the calibration of improvement scales.

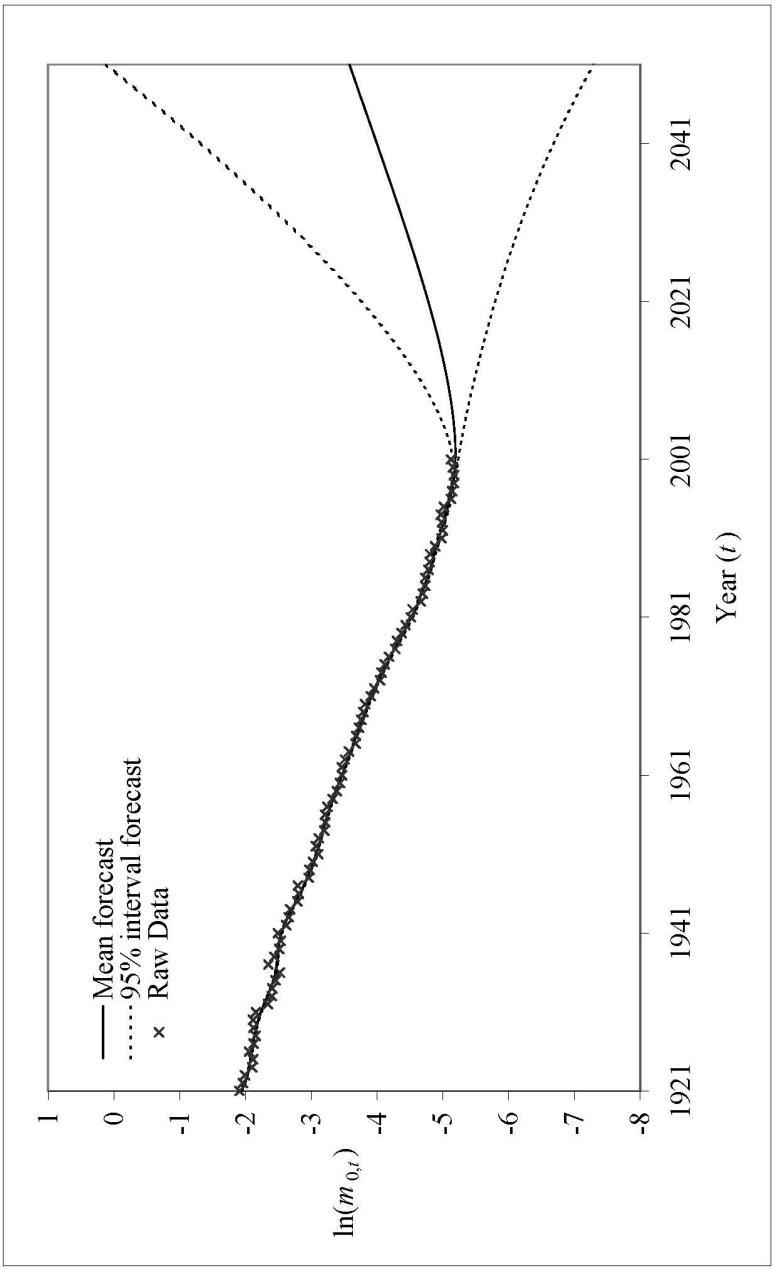


Fig. 3.1. The fitted (1921-2001) and the projected (2002-2051) mortality experience at representative ages, P-splines regression, male.

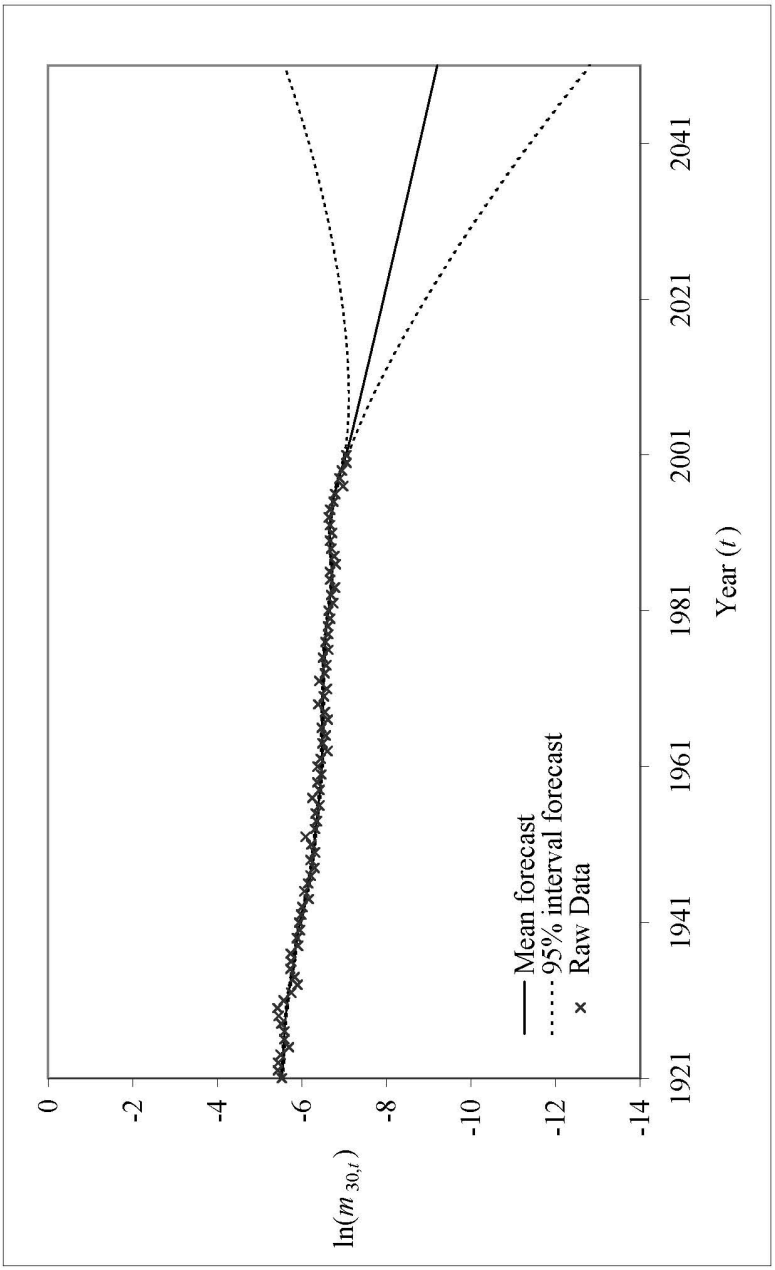
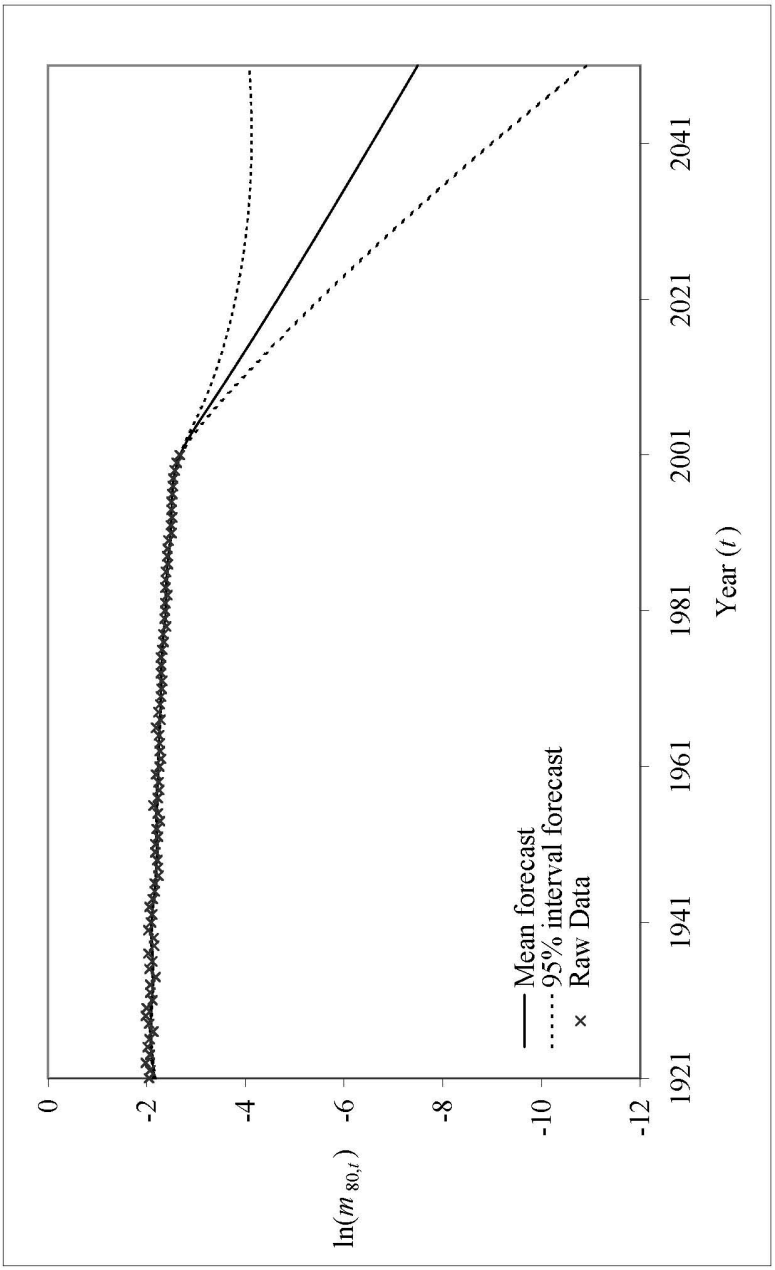
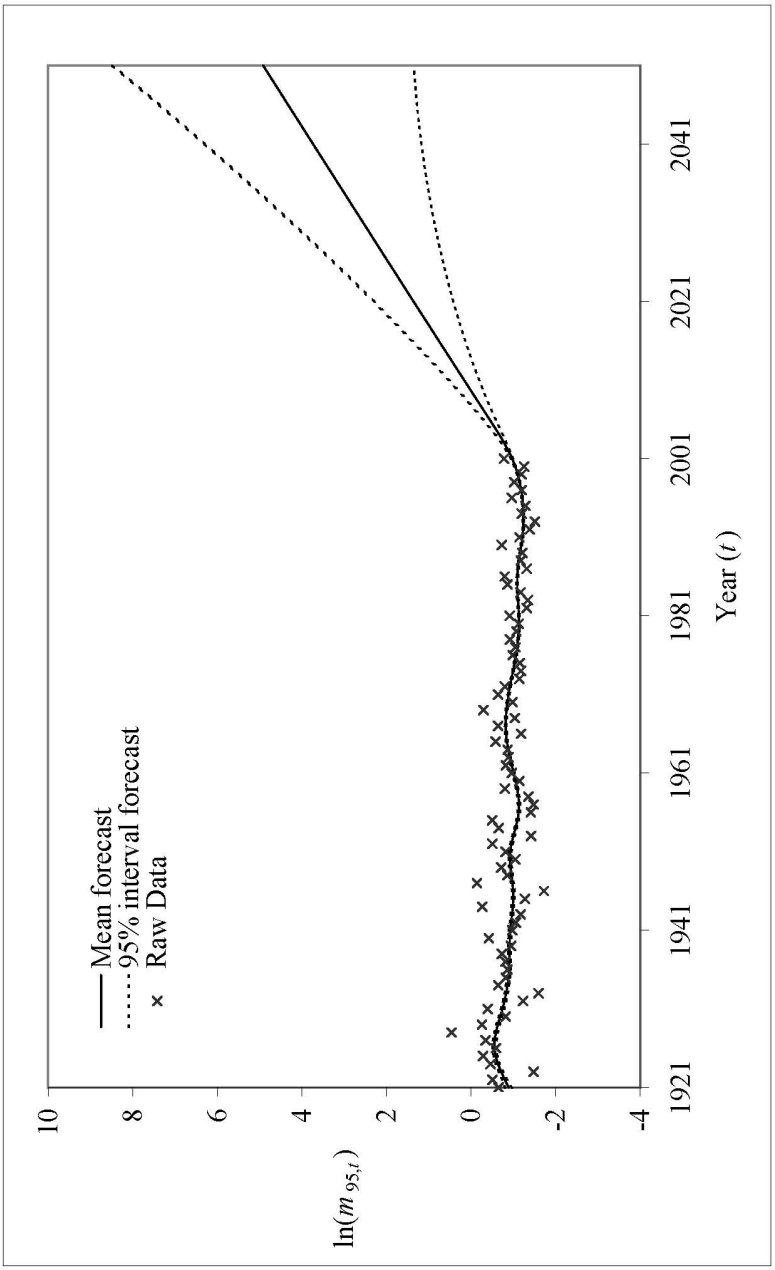


Fig. 3.1 (cont'd). The fitted (1921-2001) and the projected (2002-2051) mortality experience at representative ages, P-splines regression, male.



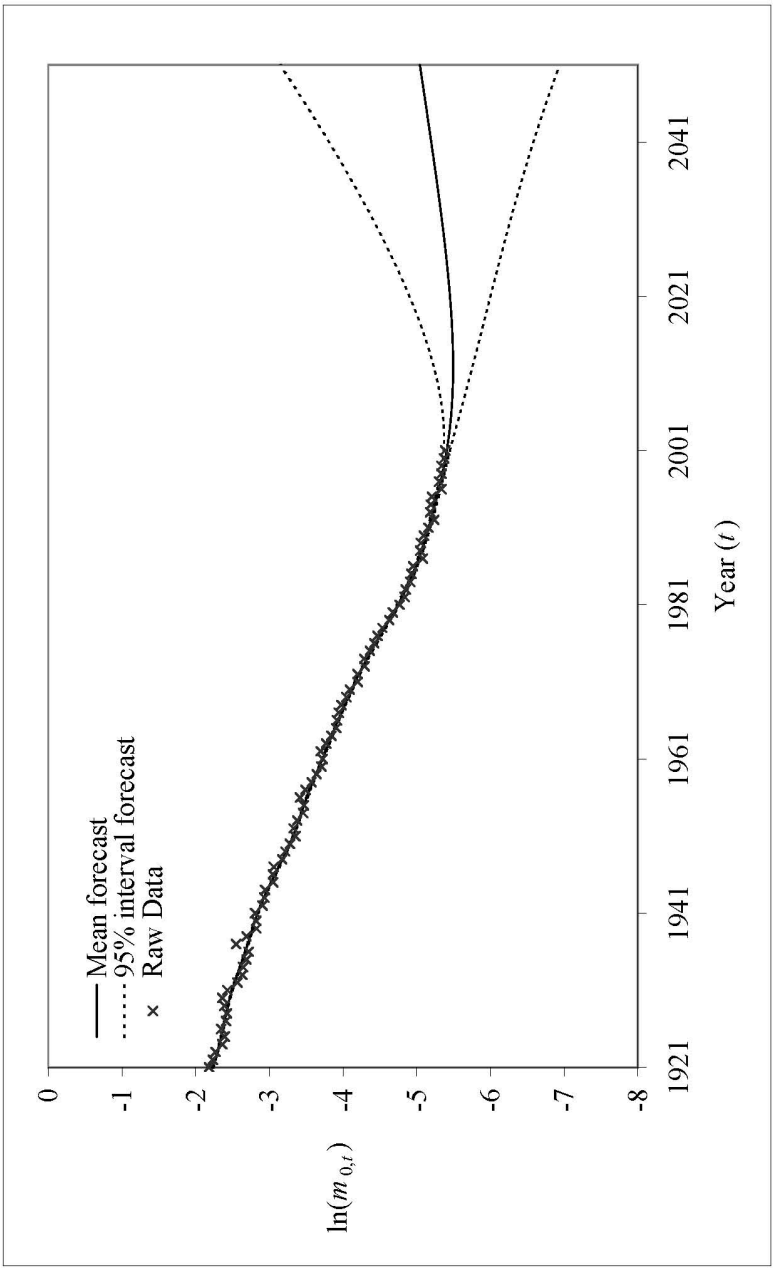
Age 80

Fig. 3.1 (cont'd). The fitted (1921-2001) and the projected (2002-2051) mortality experience at representative ages, P-splines regression, male.



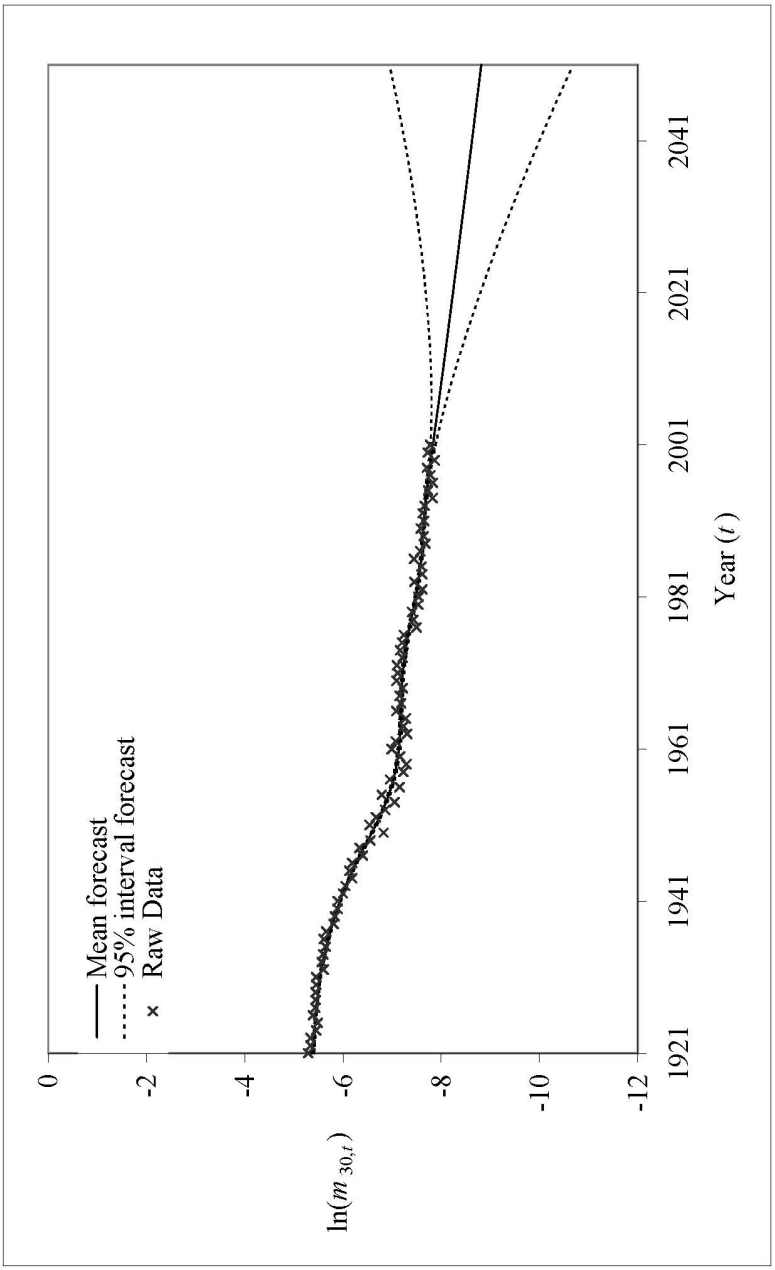
Age 95

Fig. 3.1 (cont'd). The fitted (1921-2001) and the projected (2002-2051) mortality experience at representative ages, P-splines regression, male.



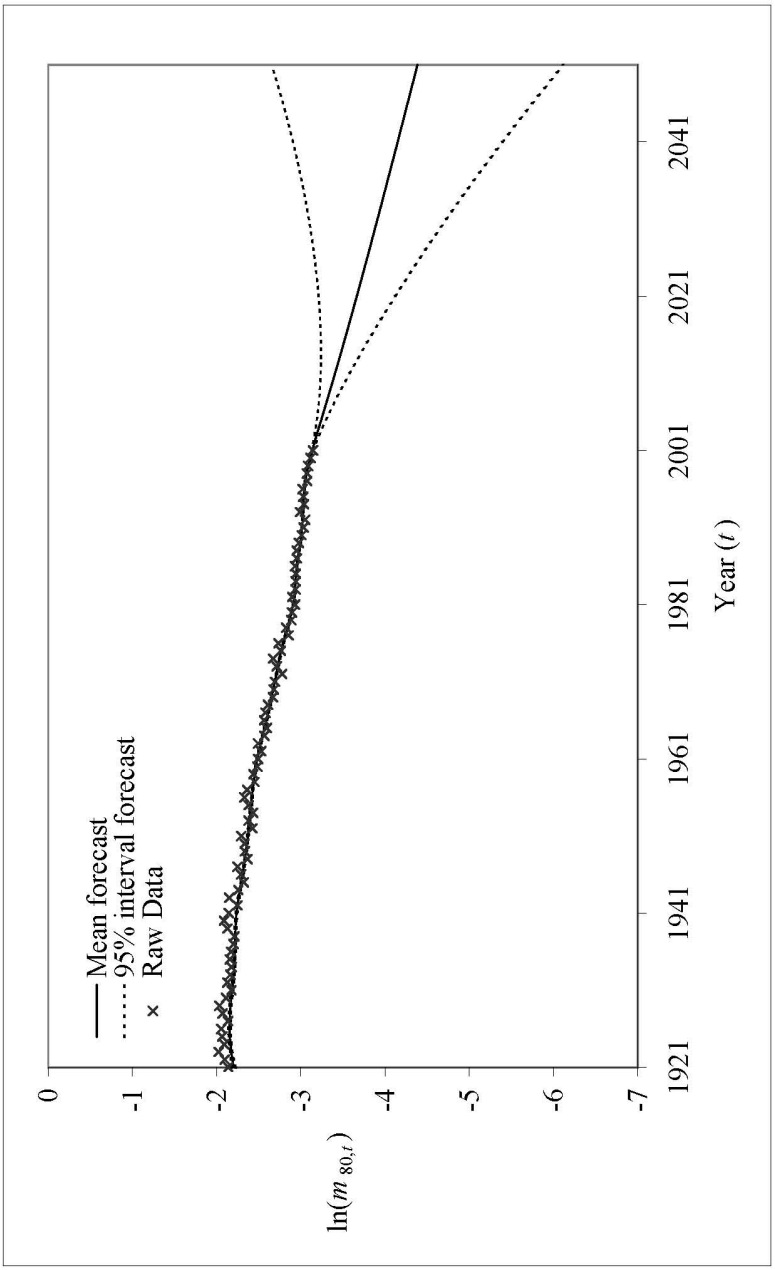
Age 0

Fig. 3.2. The fitted (1921-2001) and the projected (2002-2051) mortality experience at representative ages, P-splines regression, female.



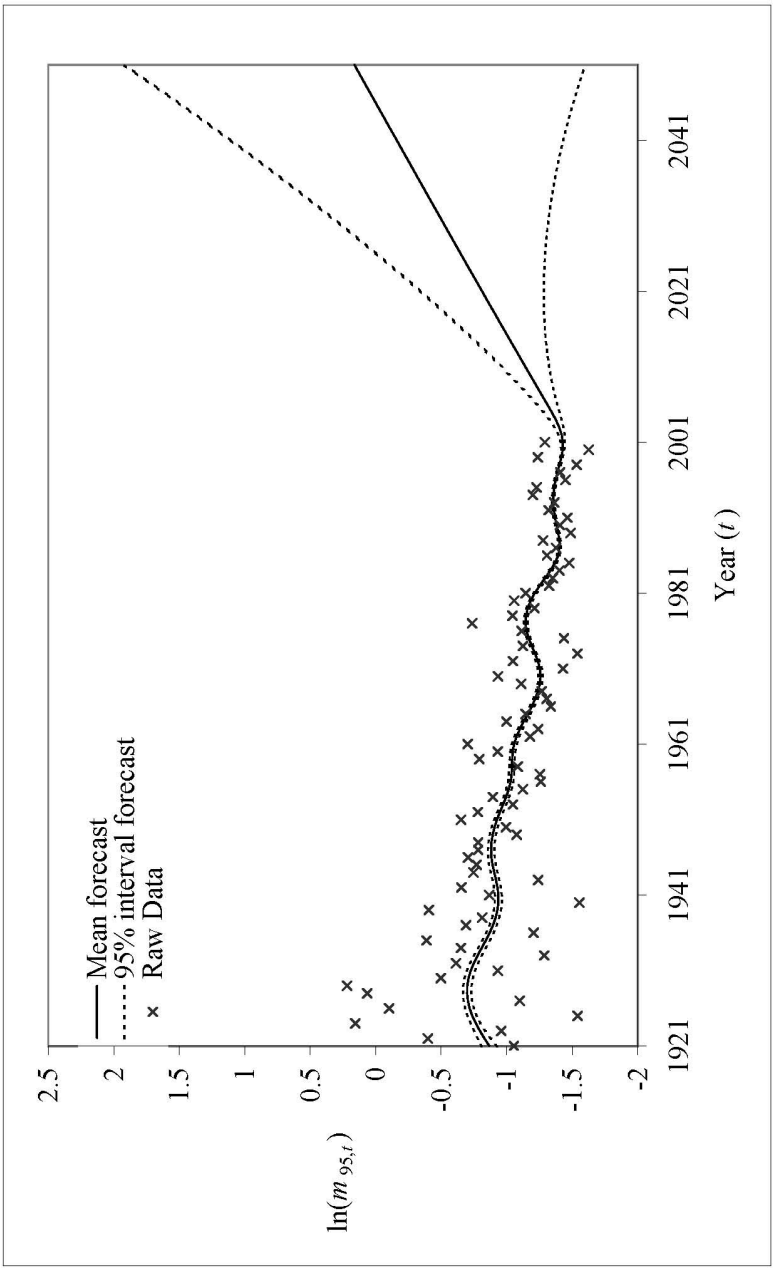
Age 30

Fig. 3.2 (cont'd). The fitted (1921-2001) and the projected (2002-2051) mortality experience at representative ages, P-splines regression, female.



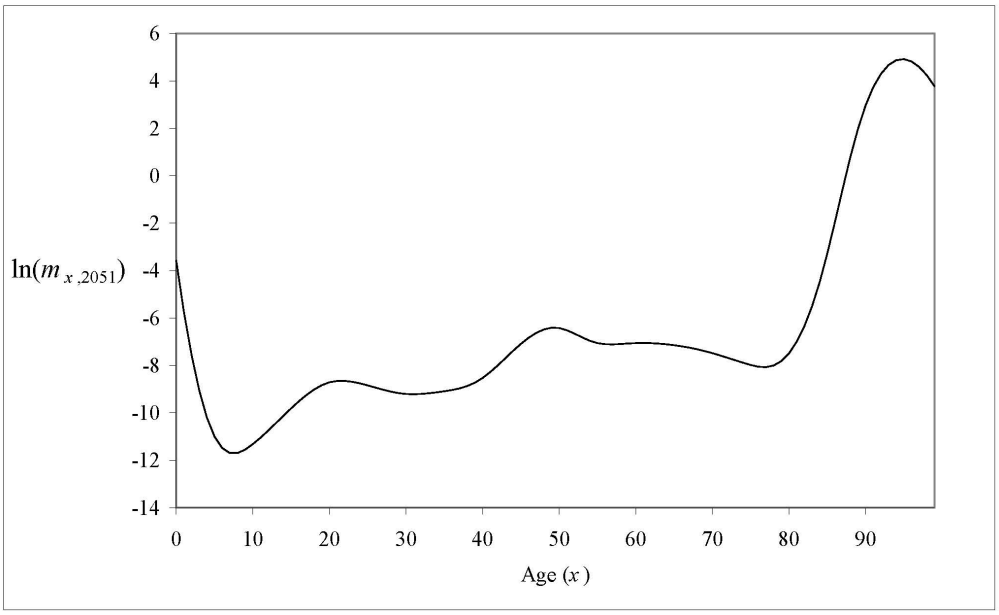
Age 80

Fig. 3.2 (cont'd). The fitted (1921-2001) and the projected (2002-2051) mortality experience at representative ages, P-splines regression, female.

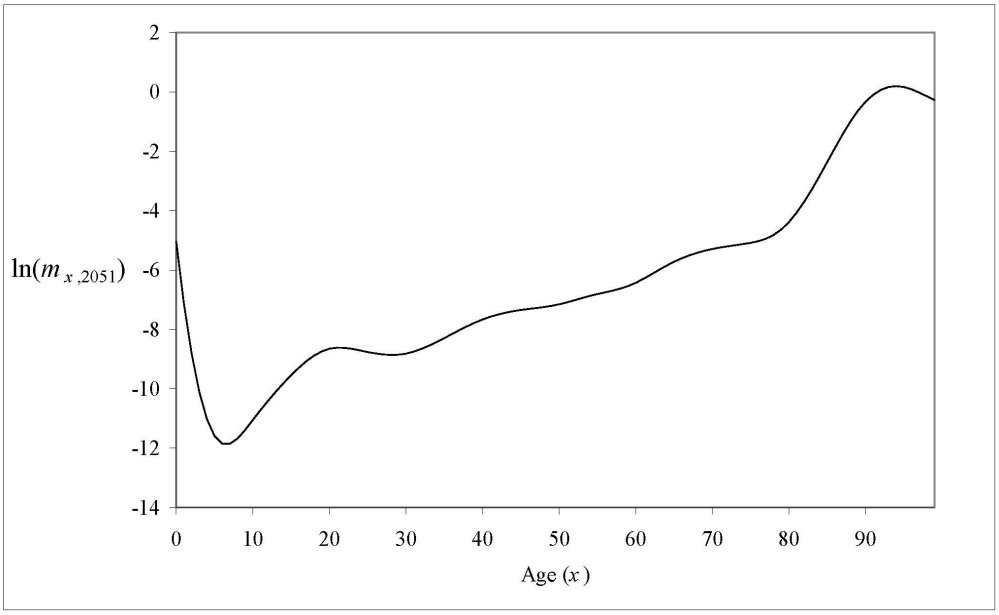


Age 95

Fig. 3.2 (cont'd). The fitted (1921-2001) and the projected (2002-2051) mortality experience at representative ages, P-splines regression, female.



Male



Female

Fig. 3.3. The projected age pattern of mortality, 2051, P-splines regression, male and female.

3.3.2 The Lee-Carter Model

In Chapter 2, we provided an in-depth discussion about the Lee-Carter model. We also pointed out that the traditional model assumption of Poisson death counts imposes the mean-variance equality restriction, requiring the variance to be less than it really is. To overcome this problem, we can replace the Poisson death counts with negative binomial ones. We demonstrated that such a modification can give us more conservative confidence intervals, but it does not lead to the erratic behaviors found in the P-splines forecasts.

Furthermore, in fitting the Lee-Carter model, we might have smoothed out several discrepant observations. These discrepancies may be due to several non-repetitive exogenous interventions, such as pandemics or wars, and thus may contain information that is useful to our projection. To discern how these interventions might have affected the mortality dynamics and to take account of the abnormalities in the interval forecast, we can conduct a statistical outlier analysis. Performing an outlier analysis can also help us avoid erroneous model specifications due to the masking effect of the outliers (see Tsay, 1986), and verify the uniformity of the data by statistically testing for an outlier effect at time points when the coverage of the vital statistics were changed.

The outlier analysis is based on the time-varying parameter k_t, which is modeled by an ARIMA(p,d,q) process:

$$\phi(B)(1-B)^d k_t = \theta(B)a_t, \tag{3.15}$$

where B is the backshift operator such that $B^s k_t = k_{t-s}$, $\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - ... - \theta_p B^p$, and $\{a_t\}$ is a sequence of white noise random variables, iid with mean 0 and constant variance σ^2.

If there exists an outlier, the observed series will be a combination of the outlier-free $\{k_t\}$ and an exogenous intervention effect, denoted by $\Delta_t(T, \omega)$; that is,

$$k_t^* = k_t + \Delta_t(T, \omega), \tag{3.16}$$

where k_t^* is the observed series, T and ω are the location and the magnitude of the outlier, respectively.

We consider four types of outlier, namely, additive outlier (AO), innovational outlier (IO), level shift (LS) and temporary change (TC).

- An AO affects only a single observation; that is, $\Delta_t(T,\omega) = \omega D_t^{(T)}$.
- An IO affects all observations beyond T through the memory of the outlier-free process; that is, $\Delta_t(T,\omega) = \frac{\theta(B)}{\phi(B)(1-B)^d}\omega D_t^{(T)}$.
- A LS has a permanent effect; that is, $\Delta_t(T,\omega) = \frac{\omega}{1-B} D_t^{(T)}$.
- A TC affects a series at a given time, and its effect decays exponentially according to a dampening factor, say δ; that is, $\Delta_t(T,\omega) = \frac{\omega}{1-\delta B} D_t^{(T)}$. We take $\delta = 0.7$ as recommended by Chen and Liu (1993).

In the above expressions, $D_t^{(t)}$ is an indicator variable which equals 1 when $t = T$ and zero otherwise. In general, there may be more than one, say m, outliers. This may be represented by the general outlier model:

$$k_t^* = k_t + \sum_{i=1}^{m} \Delta_t(T_i, \omega_i). \tag{3.17}$$

A complete algorithm for outlier detection and adjustment can be found in Li and Chan (2005). Table 3.1 shows all the detected outliers in the series of k_t for the Canadian population. Note that the positive ones may be regarded as unexpected mortality deteriorations, while the negative ones may be regarded as unexpected improvements. The detected outliers do not match any of the years when the coverage of the vital statistics is altered, justifying the integrity of the data.

Year	Magnitude	t-statistic	Type
	Male		
1937	5.497	3.69	AO
	Female		
1926	6.729	4.05	LS
1929	5.271	3.22	TC
1937	6.420	4.51	AO
1954	−4.940	−3.02	TC

Table 3.1. Outliers detected in k_t, Canadian population, male and female, 1921-2001

Year	Magnitude	t-statistic	Type
	Male		
1918	29.444	11.67	AO
1921	−12.471	−3.88	TC
1927	−7.806	−3.09	AO
1936	9.815	3.05	TC
	Female		
1916	10.117	3.48	LS
1918	30.574	11.87	AO
1921	−11.795	−4.07	TC
1926	7.610	2.94	AO
1927	10.630	3.62	TC
1936	11.531	4.00	TC
1954	−8.059	−2.79	TC
1975	−10.459	−3.61	LS

Table 3.2. Outliers detected in k_t, the US population, male and female, 1921-2001

To learn more about the exogenous interventions, we replicate the exercise using the US population mortality data from 1901 to 2001. The death rates and the mid-year population estimates are provided by the National Center for Health Statistics (2004a, 2004b) and the United States Census Bureau (2004), respectively. Table 3.2 shows all the outliers detected in the series of k_t for the US population. The positive outlier in 1918 is most likely due to the Spanish flu epidemic, and its additive nature suggests that this deadly epidemic had only a temporary effect on the mortality dynamics. We refer interested readers to Li and Chan (2007) for explanations of other detected outliers.

To incorporate the potential recurrence of the detected outliers into the interval forecasts, we add the following step to the bootstrap procedure we discussed in Chapter 2: in each simulation, we sample with replacement from the pool of detected outliers and superimpose the sampled outliers to simulated sample path of future k_t's. The pool comprises all outliers in Table 3.1 and those before 1921 in Table 3.2. By allowing the interruptive events to recur in the future, the interval forecasts can be more conservative and capable of including more stochastic uncertainty. Note that the outlier analysis does not affect the estimates of parameters a_x, b_x and k_t. The outlier adjusted forecast is very similar to that obtained in Chapter 2, except that the confidence intervals are slightly wider.

The derivation of the improvement scales will be based on the Lee-Carter model with the extensions proposed. However, readers should bear in mind that the model structure is still strong so that the confidence intervals might not be conservative enough to reflect the possibility of structural changes in the future.

3.3.3 Summarizing the Projection

Recall that under the Lee-Carter model, we can write the s-step ahead forecast of $m_{x,t}$ as

$$\hat{m}_{x,T+s} = \exp(\hat{a}_x + \hat{b}_x \hat{k}_{T+s}), \tag{3.18}$$

where T is the forecast origin. Substituting $m_{x,T}$ into equation (3.18), we have

$$\begin{aligned} \hat{m}_{x,T+s} &= \hat{m}_{x,T} \exp(\hat{b}_x(\hat{k}_{T+s} - \hat{k}_T)) \\ &\approx m_{x,T} \exp(\hat{b}_x(\hat{k}_{T+s} - \hat{k}_T)). \end{aligned} \tag{3.19}$$

Tuljapurkar et al. (2000) found that $\{k_t\}$ typically follows an ARIMA(0,1,0) process. In this case,

$$\hat{k}_{T+s} - \hat{k}_T = \hat{c}s, \tag{3.20}$$

where c is the drift term in the ARIMA(0,1,0) process. Let us define the improvement scale by the percentage reduction of $m_{x,T}$ in s years; that is,

$$\text{IS}(x,s) = \frac{m_{x,T+s}}{m_{x,T}}. \tag{3.21}$$

Then, combining equations (3.18) to (3.21), we have

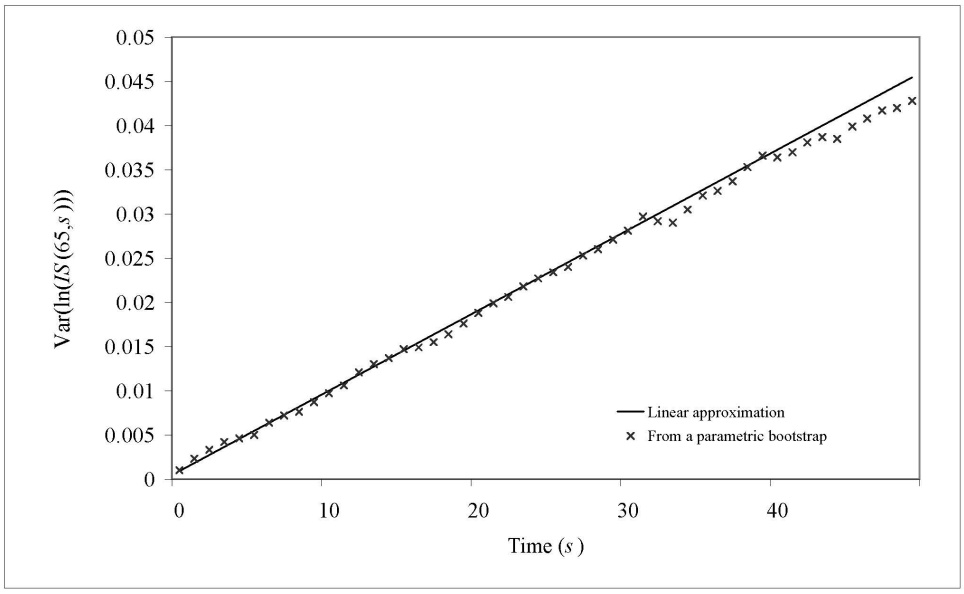
$$\hat{\text{IS}}(x,s) = \exp(\hat{w}_x s), \tag{3.22}$$

where $\hat{w}_x = \hat{b}_x\hat{c}$. This expression means that the projection can be reduced from a two dimensional array of numbers to a vector of parameters ($\hat{w}_x$'s).

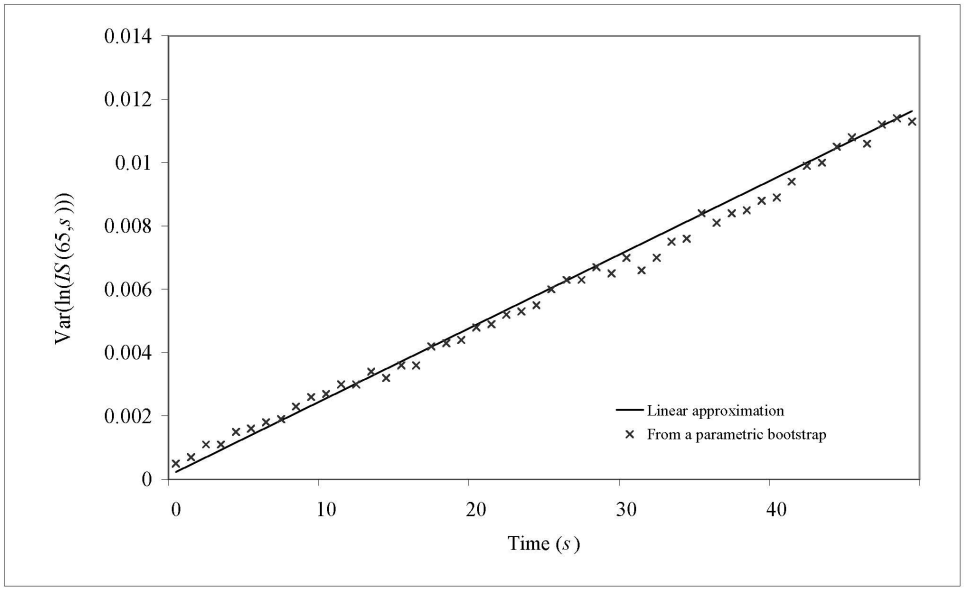
We can use parametric bootstrapping to generate the empirical distribution of the improvement scale. To obtain an algebraic expression of the interval forecast, we have to make an additional assumption. Given that the uncertainty under an ARIMA(0,1,0) process is increasing linearly with time and that the width of the interval forecast should be close to zero at the forecast origin, we approximate the variance of the log of the improvement scale by a straight line without intercept, illustrated in Figure 3.4. We denote the slope of this line by v_x. Then, an 95% point-wise confidence interval for the improvement scale can be expressed as

$$\left[\exp(\hat{w}_x s - 1.96\sqrt{\hat{v}_x s}), \exp(\hat{w}_x s + 1.96\sqrt{\hat{v}_x s})\right]. \tag{3.23}$$

Note that $\hat{v}_x$ has incorporated both parameter and stochastic uncertainty.



Male



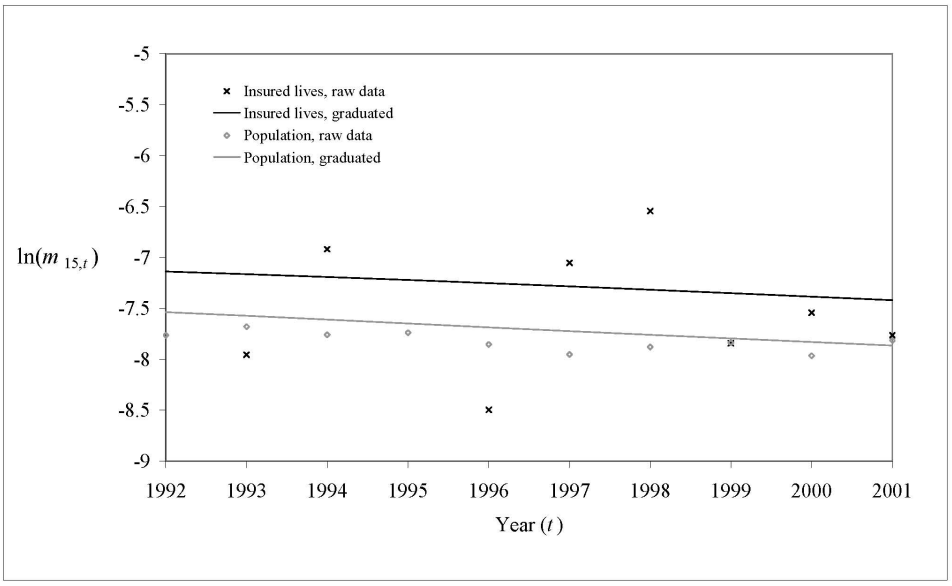
Female

Fig. 3.4. A linear approximation of the variance of the log of the improvement scale obtained from a parametric bootstrap.

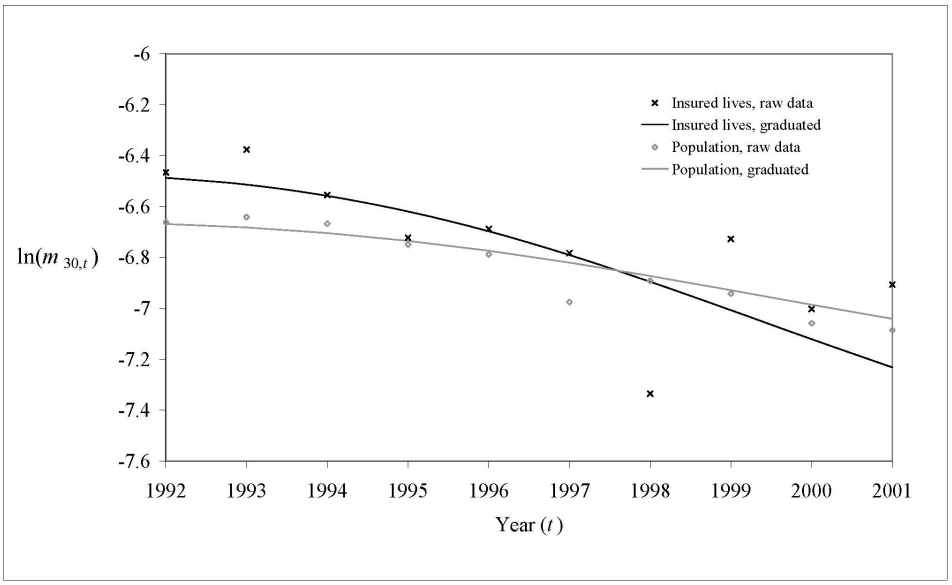
3.4 Improvement Scales for the Insured Lives

Having obtained a mortality projection for the general population, we now proceed to the third fitting stage of the joint model: relating the experiences of the general population and insured lives by a parametric equation. As the relationship has to be reasonably stable over time, it will be based on the ultimate experience (smoker status combined), which has the largest exposure base and therefore is less vulnerable to abnormalities.

We firstly use a two-dimensional P-splines regression to graduate the experiences. This allows us to examine the relationship between the experiences graphically and gives us a smooth base table (given in the Appendix) on which the ultimate improvement scales are to be applied. Figures 3.5 and 3.6 show that the experiences are closely related to each other. It is interesting to note that contrary to common belief, the experience of the insured lives is not necessarily better than that of the population. Also, the relationships, for example, males at age 30, are not necessarily in the form of level shifts, indicating that the experiences might be improving at different rates. This phenomenon should be allowed for in the specification of the relational model.

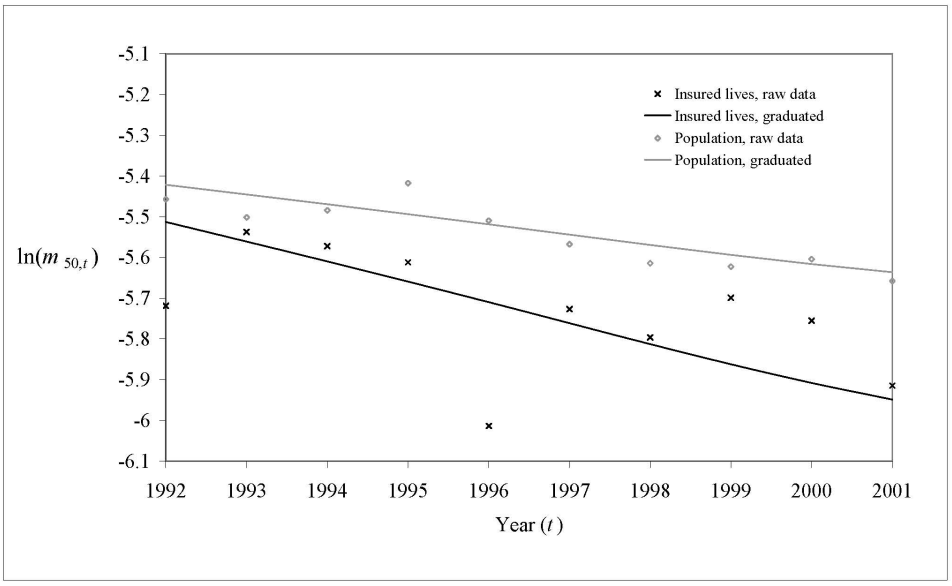


Age 16

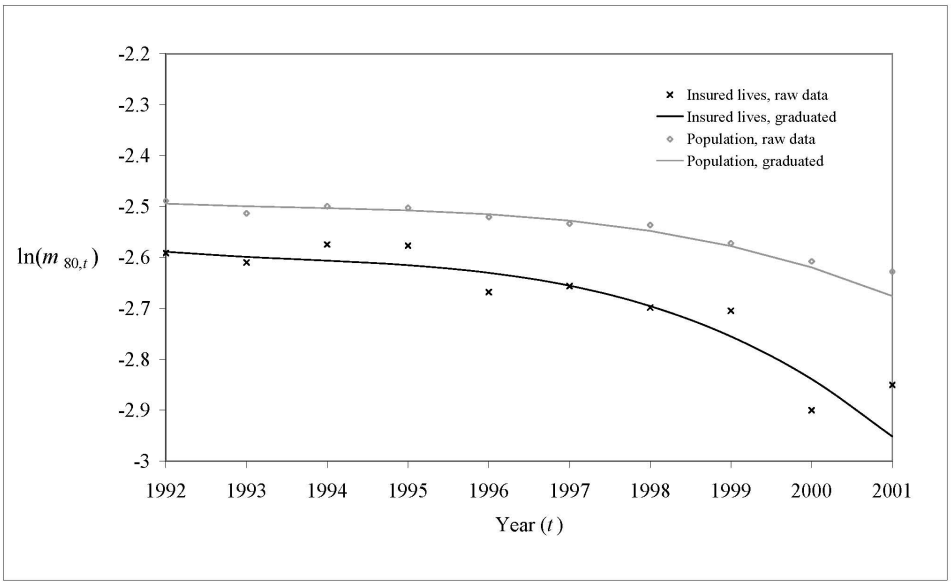


Age 30

Fig. 3.5. Comparison of the insured lives and population experiences at representative ages, male.

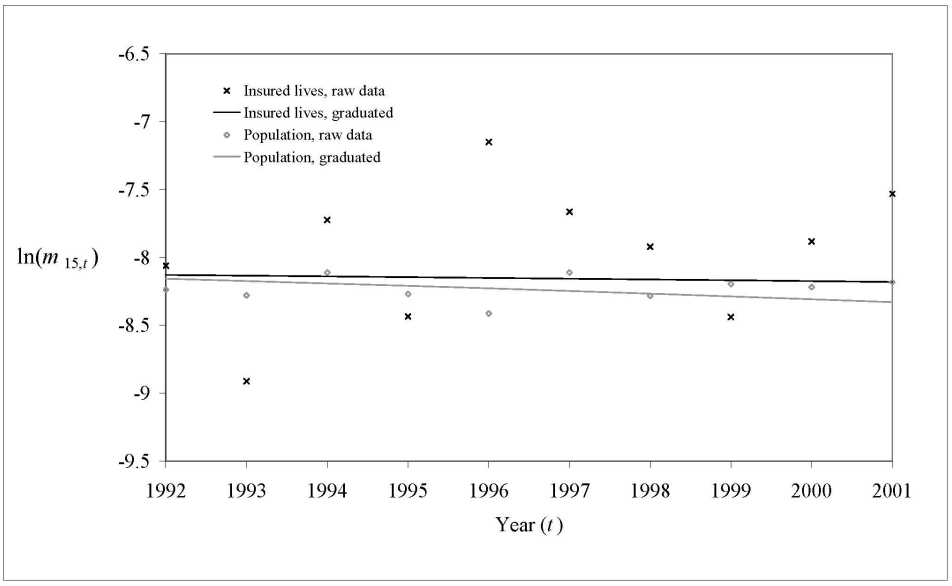


Age 50

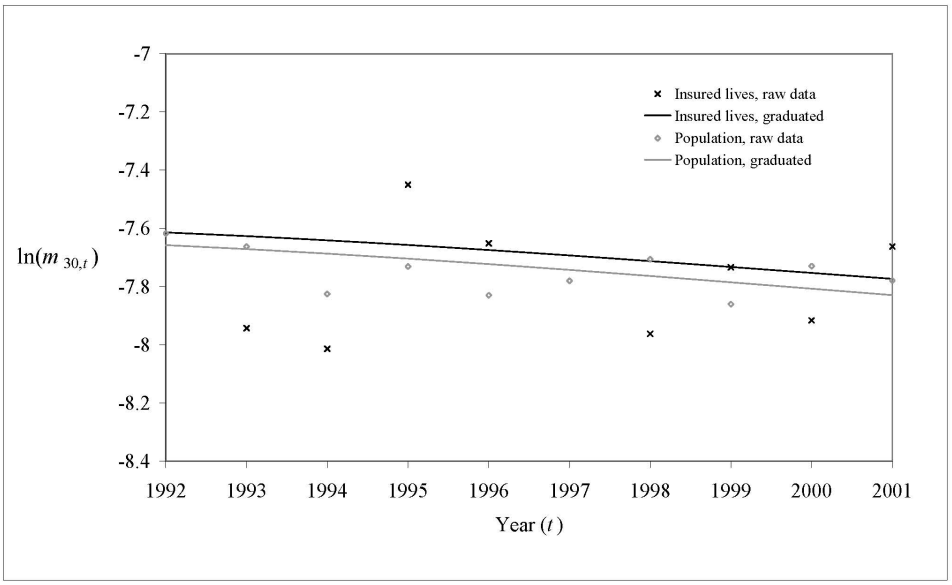


Age 80

Fig. 3.5 (cont'd). Comparison of the insured lives and population experiences at representative ages, male.

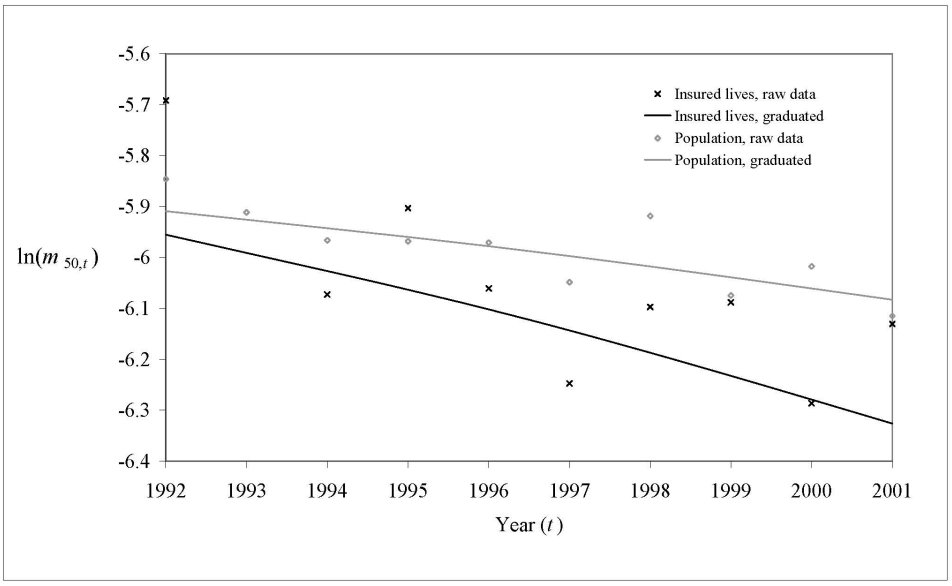


Age 16

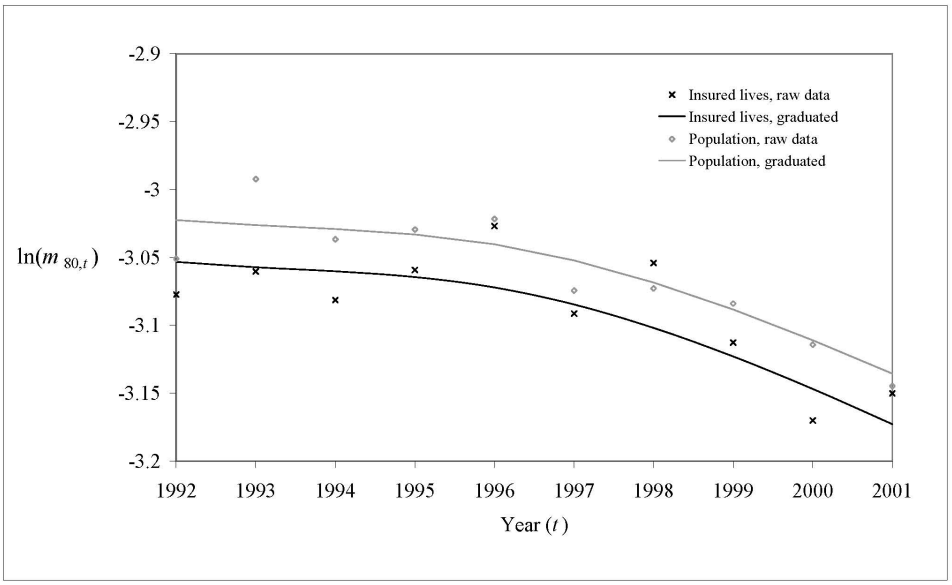


Age 30

Fig. 3.6. Comparison of the insured lives and population experiences at representative ages, female.



Age 50



Age 80

Fig. 3.6 (cont'd). Comparison of the insured lives and population experiences at representative ages, female.

Given the properties of the relationships, the Brass-type model (Brass, 1975) seems to be appropriate for relating the mortality experiences of the general population and insured lives. This model has been widely applied to situations where the required mortality data are limited, deficient or even non-existent. For instance, the United Nations (1997) recommended the use of the Brass-type model in relating the mortality of a specific population to a mortality standard calibrated by the vital statistics of various low-mortality countries. Also, Brouhns et al. (2002) used the Brass-type model to relate the experiences of the Belgian annuitants and the whole Belgian population. The model can be expressed as

$$\ln(m^*_{x,t}) = h_{1,x} + h_{2,x} \ln(m_{x,t}) + \epsilon_{x,t}, \tag{3.24}$$

where $m^*_{x,t}$ is the central death rate of the ultimate insured lives experience, $m_{x,t}$ is the central death rate of the general population (the reference population), $h_{1,x}$ and $h_{2,x}$ are age-specific parameters, which are assumed to be invariant over time, and $\epsilon_{x,t}$ is the error term, which is assumed to be normally distributed with mean zero and constant variance $\sigma^2_{\epsilon,x}$.

In contrast to the practice of fitting parallel mortality surfaces (see, e.g., Currie et al., 2004), the Brass-type model allows $m^*_{x,t}$ and $m_{x,t}$ to vary at different rates and directions, specified by parameter $h_{2,x}$. Assuming that $m_{x,t}$ is decreasing over time, the effect of parameter $h_{2,x}$ can be segregated into the following five cases:

1. $h_{2,x} > 1$: the experience of the insured lives is improving faster than that of the general population;

2. $h_{2,x} = 1$: the experiences of the general population and insured lives are moving at the same speed;

3. $0 < h_{2,x} < 1$: the experience of the insured lives is improving more slowly than that of the population;

4. $h_{2,x} = 0$: the insured lives experience shows no improvement;

5. $h_{2,x} < 0$: the insured lives experience is deteriorating.

The parameters in equation (3.24) can be estimated readily by the principle of least squares. Figure 3.7 shows the estimates of $h_{2,x}$. The variance of $\hat{h}_{2,x}$ is relatively high when x is greater than 90, as both $m^*_{x,t}$ and $m_{x,t}$ are more volatile at the extreme ages. For example, for males, the variance of $\hat{h}_{2,96}$ is 0.1054, giving an approximate 95% confidence interval of $[-0.9175, 0.3811]$ for $\hat{h}_{2,96}$.

We observe that the estimates of $h_{2,x}$ for males are negative when $x \geq 94$, indicating that for males of age 93 and higher, the insured lives and general population mortality experiences tend to change in different directions. Nevertheless, standard t-tests indicate that these estimates are not significantly less than zero. Therefore, we replace them by zero to avoid any anti-intuitive mortality projections.

Fig. 3.7. Estimates of parameter $h_{2,x}$, male and female.

The relationship can be used to estimate $\ln(m^*_{x,T+s})$ and $\ln(m^*_{x,T})$. Furthermore, by differencing the estimates, we obtain

$$\hat{\ln}(m^*_{x,T+s}) - \hat{\ln}(m^*_{x,T}) = \hat{h}_{2,x}(\hat{\ln}(m_{x,T+s}) - \ln(m_{x,T})), \tag{3.25}$$

which gives

$$\hat{\text{IS}}^*(x,s) = \exp(\hat{h}_{2,x})\hat{\text{IS}}(x,s), \tag{3.26}$$

where $\hat{\text{IS}}^*(x,s)$ is the estimate of the improvement scale for the insured lives. Note that parameter $h_{1,x}$, which measures the amount of parallel shifts, is irrelevant to the scales.

Finally, recall that under the Lee-Carter,

$$\hat{\text{IS}}(x,s) = \exp(\hat{w}_x s), \tag{3.27}$$

which can be combined with equation (3.26) to obtain

$$\hat{\text{IS}}^*(x,s) = \exp(\hat{z}_x s), \tag{3.28}$$

where $\hat{z}_x = \hat{h}_{2,x}\hat{w}_x$. In other words, the mortality improvement of the insured lives can be summarized by a single parameter vector ($\hat{z}_x$'s). Having smoothed out the bumps in the crude estimates of z_x by a B-splines regression, we obtain the improvement scales applicable to the insured lives experience. The graduated estimates of z_x are shown in Figure 3.8 and are provided in the Appendix.

We now turn to the measure of uncertainty. Assuming that the model specified by equation (3.24) is correct, the true value of $\text{IS}^*(x,s)$ can be written as

$$\ln(\text{IS}^*(x,s)) = (\hat{h}_{2,x} + \eta_{2,x})(\hat{\ln}(\text{IS}(x,s)) + \iota_{x,s}) + \epsilon_{x,2001+s} - \epsilon_{x,2001}, \tag{3.29}$$

where $\eta_{2,x}$ and $\iota_{x,s}$ denote the error in estimating $h_{2,x}$ and $\ln(\text{IS}(x,s))$, respectively. Hence, the forecast error of the $\ln(\text{IS}^*(x,s))$ can be expressed as

$$E_{x,s} = \eta_{2,x}\hat{\ln}(\text{IS}(x,s)) + \iota_{x,s}\hat{h}_{2,x} + \eta_{2,x}\iota_{x,s} + \epsilon_{x,2001+s} - \epsilon_{x,2001}. \tag{3.30}$$

Assuming independence between $\eta_{2,x}$ and $\iota_{x,s}$, and using the results in Section 3.3.3, the variance of the forecast error can be written as

$$\sigma^2_{E,x,s} \approx \sigma^2_{\eta,2,x}(\hat{w}_x s)^2 + (\hat{h}^2_{2,x} + \sigma^2_{\eta,2,x})\hat{v}_x s + 2\sigma^2_{\epsilon,x}, \tag{3.31}$$

where $\sigma^2_{\eta,2,x}$ is the error variance of $\hat{h}_{2,x}$. We can easily estimate $\sigma^2_{\eta,2,x}$ by the principle of least squares. By letting $u_{1,x} = 2\sigma^2_{\epsilon,x}$, $u_{2,x} = (\hat{h}^2_{2,x} + \sigma^2_{\eta,2,x})\hat{v}_x$, and $u_{3,x} = \sigma^2_{\eta,2,x}\hat{w}^2_x$, equation (3.31) can be simplified further to

$$\sigma^2_{E,x,s} = u_{1,x} + u_{2,x}s + u_{3,x}s^2. \tag{3.32}$$

Finally, on the basis of the above results, we have the following approximate 95% point-wise confidence interval for the insured lives improvement scale:

$$\left[\exp(\hat{z}_x s - 1.96\sqrt{\hat{u}_{1,x} + \hat{u}_{2,x} s + \hat{u}_{3,x} s^2}), \exp(\hat{z}_x s + 1.96\sqrt{\hat{u}_{1,x} + \hat{u}_{2,x} s + \hat{u}_{3,x} s^2})\right]. \quad (3.33)$$

As before, we smooth the crude estimates of parameters $u_{1,x}$, $u_{2,x}$, and $u_{3,x}$ by B-splines regressions. The graduated estimates of these parameters are shown in Figures 3.9 to 3.11, and are provided in the Appendix.

At the younger ages, the value of $u_{3,x}$ is negligibly small. However, towards the advanced ages, parameter $\sigma^2_{\eta,2,x}$, the variance of $\hat{h}_{2,x}$, becomes larger, resulting in parameter $u_{3,x}$ no longer being negligible. Figures 3.12 and 3.13 compare the interval forecast of $m^*_{x,t}$ with and without the third term in equation (3.32). We observe that before age 80, the third term has no material effect on the interval forecasts. Given this observation, parameter $u_{3,x}$ for $x < 80$ may be excluded from the model for a more parsimonious model specification. Nevertheless, beyond age 80, the third term contributes significantly to the interval forecasts, particularly in the distant future. Therefore, for $x \geq 80$, parameter $u_{3,x}$ must be considered to avoid understating the uncertainty associated with the mortality projection.

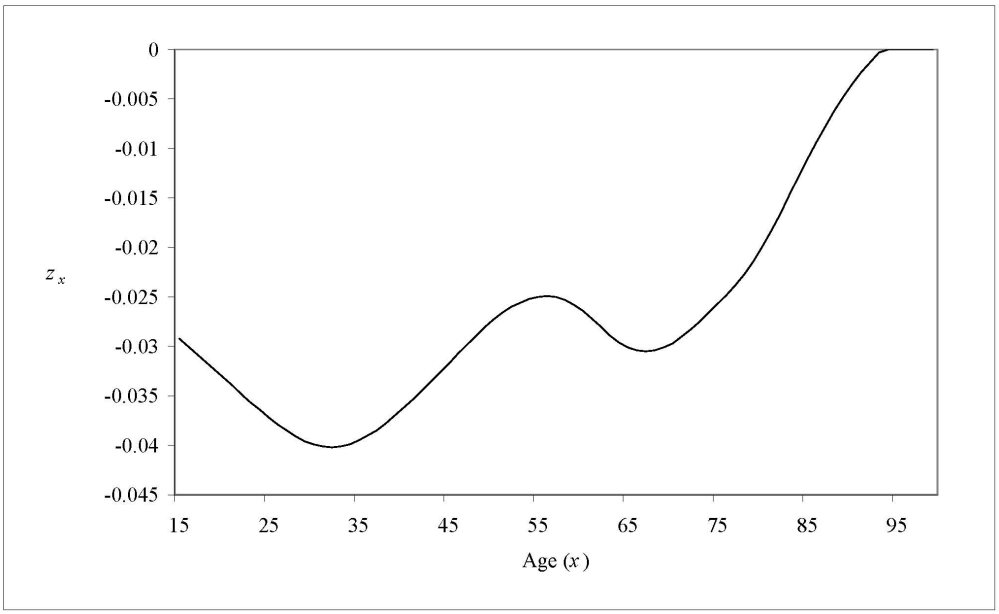
To illustrate the use of the scales, we consider the forecast of m_{80} (male, ultimate, composite smoker/non-smoker) in 2021. From the Appendix, we get $m_{80} = 0.063156$ in 2001, $z_{80} = -0.019931$, $u_{1,80} = 0.000924$, $u_{2,80} = 0.001597$, and $u_{3,80} = 6.23 \times 10^{-6}$. Hence, the best estimate and the confidence interval of m_{80} in 2021 is $0.063156 \times \exp(-0.019931 \times 20) = 0.042393$ and $0.063156 \times [0.464329, 0.970370] = [0.029325, 0.061265]$, respectively. Note that the improvement scales we obtained are different from the older ones in that those scales applied to q_x's instead of m_x's. In case the base table is given in terms of q_x's, we may make use of the assumption of a constant force of mortality in fractional ages to obtain forecasts of q_x's; that is,

$$q_{x,T+s} = 1 - (1 - q_{x,T})^{\hat{\mathrm{IS}}^*(x,s)}. \quad (3.34)$$

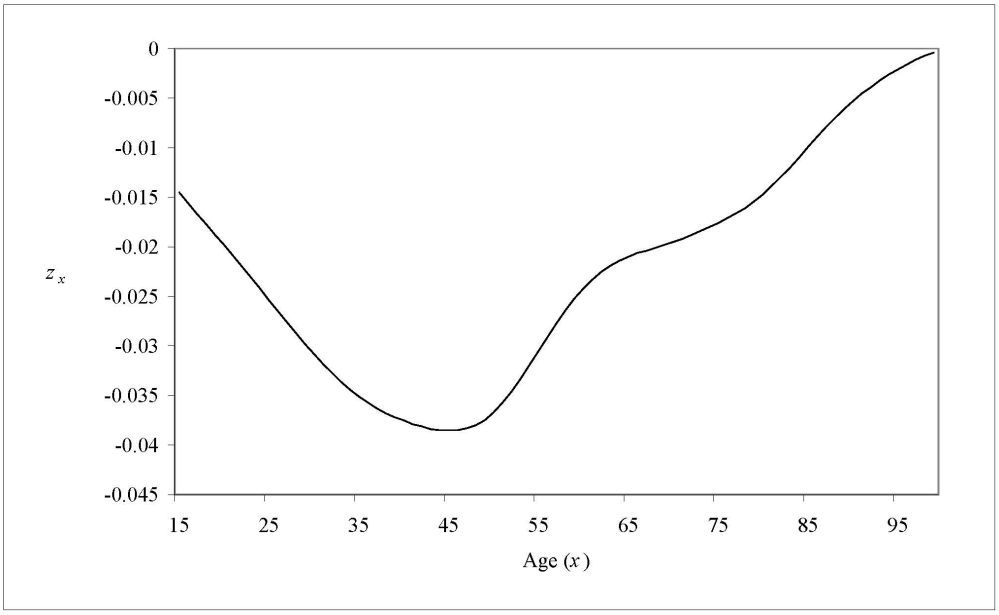
The corresponding interval forecast can be obtained by replacing $\hat{\mathrm{IS}}^*(x, s)$ with the upper and lower limits in expression (3.33).

To access the performance of the improvement scales, we conduct an ex post analysis in which the joint model is fitted to the mortality data prior to year 1993; the forecast based

on the restricted fitting period is compared with the actual experience from 1993 to 2001. In addition, we compare our forecasts with those derived from the improvement scales provided in (1) the Society of Actuaries Group Annuity Mortality Table (GAMT), (2) the Society of Actuaries 2001 Valuation Basic Experience Table (VBT) and (3) the Institute of Actuaries "92" Series Base Table. Figures 3.14 and 3.15 show all the comparisons.

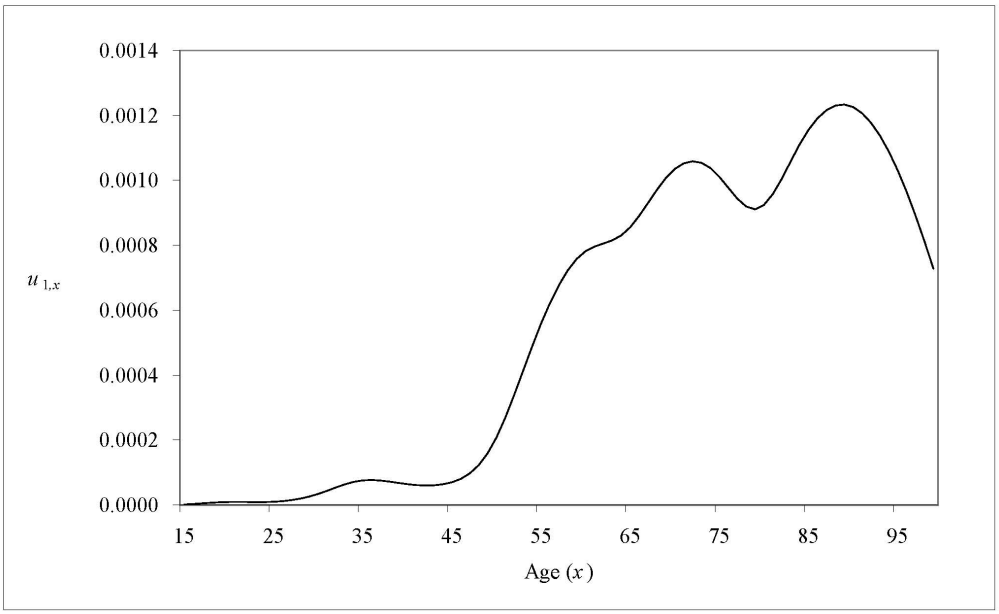


Male

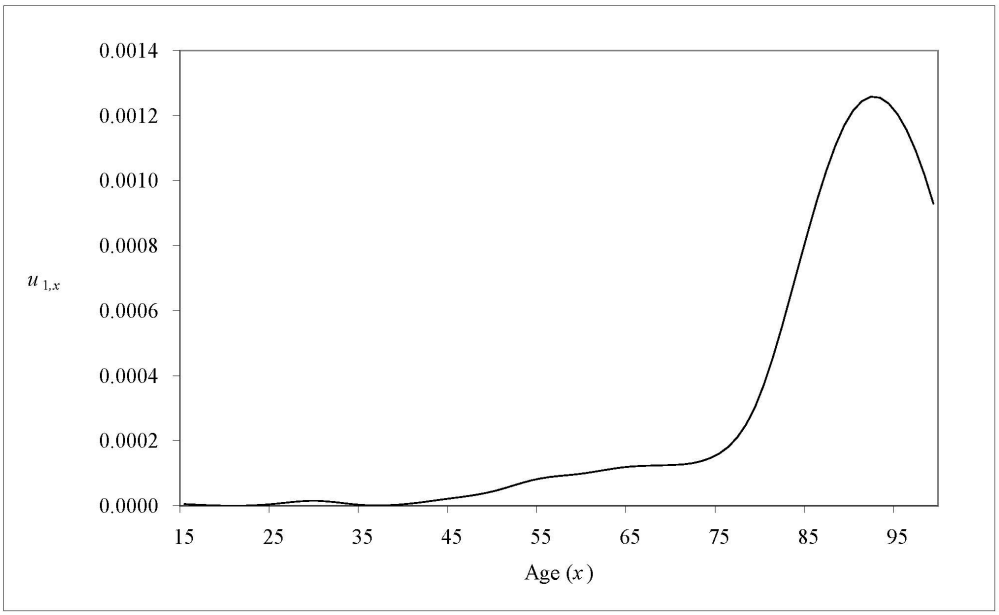


Female

Fig. 3.8. Estimates of parameter z_x, male and female.

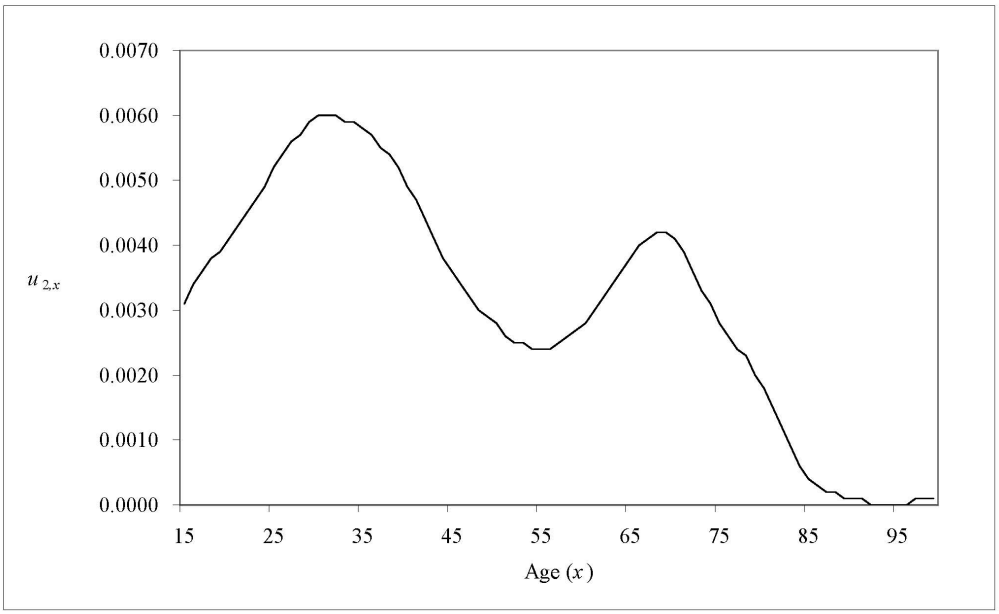


Male

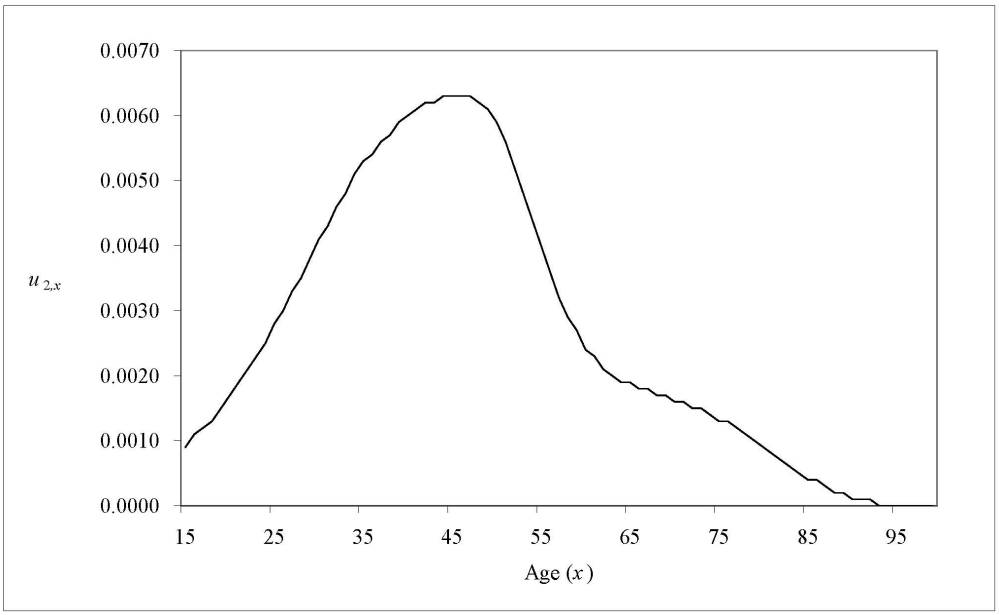


Female

Fig. 3.9. Estimates of parameter $u_{1,x}$, male and female.

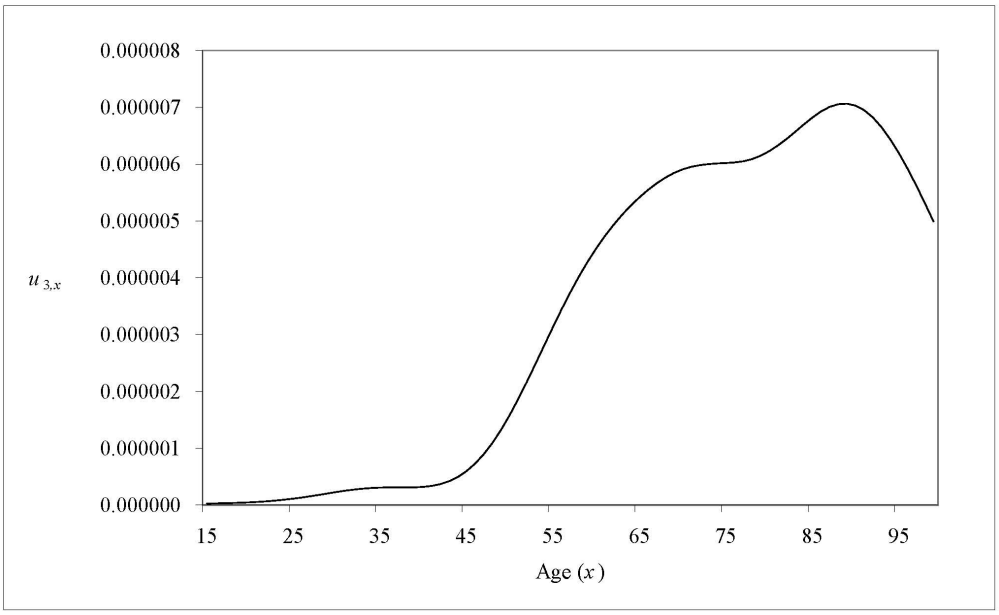


Male

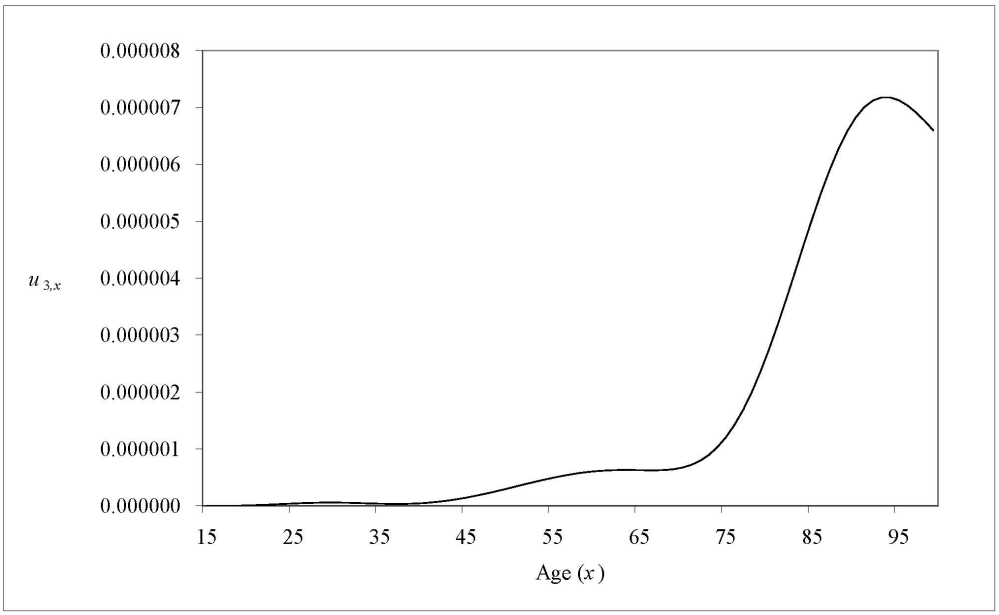


Female

Fig. 3.10. Estimates of parameter $u_{2,x}$, male and female.

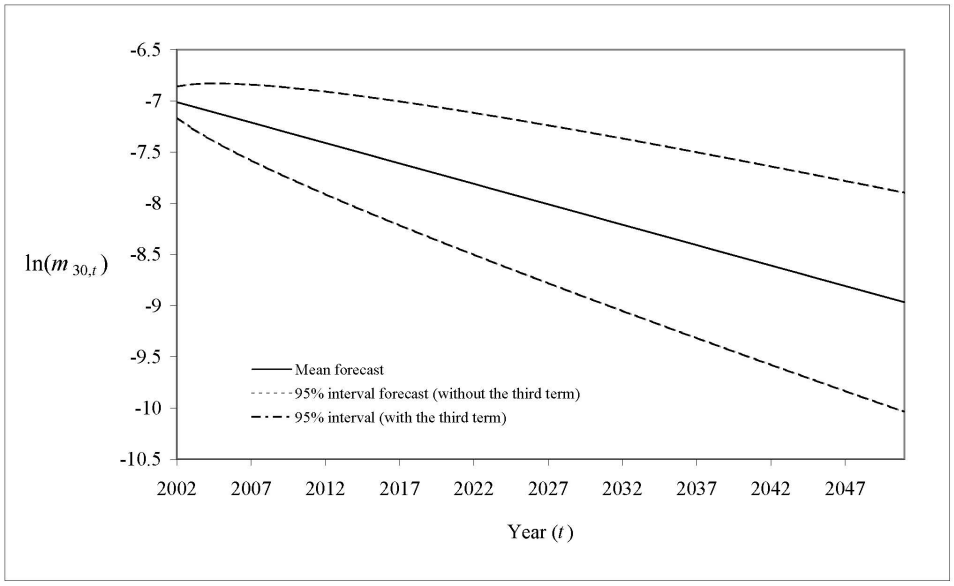


Male

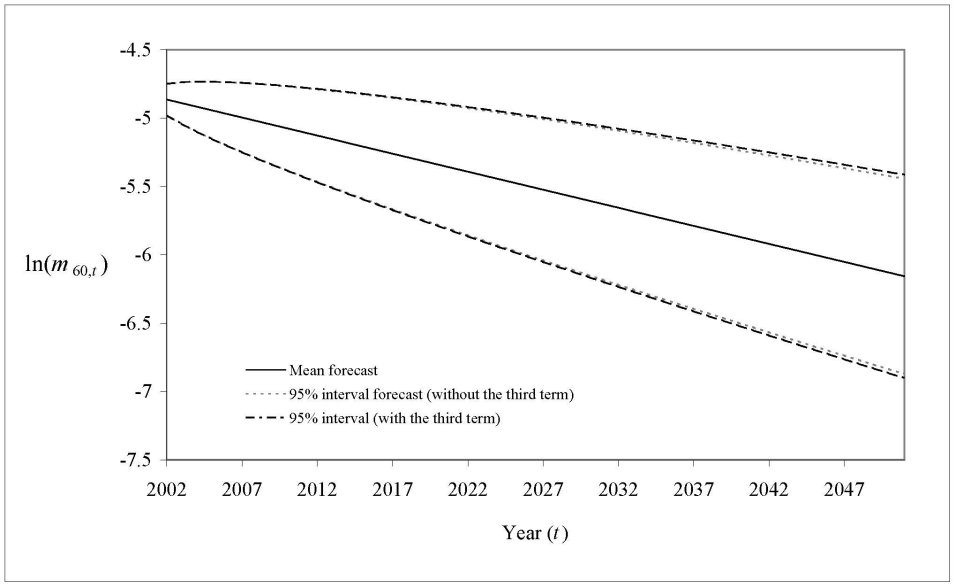


Female

Fig. 3.11. Estimates of parameter $u_{3,x}$, male and female.

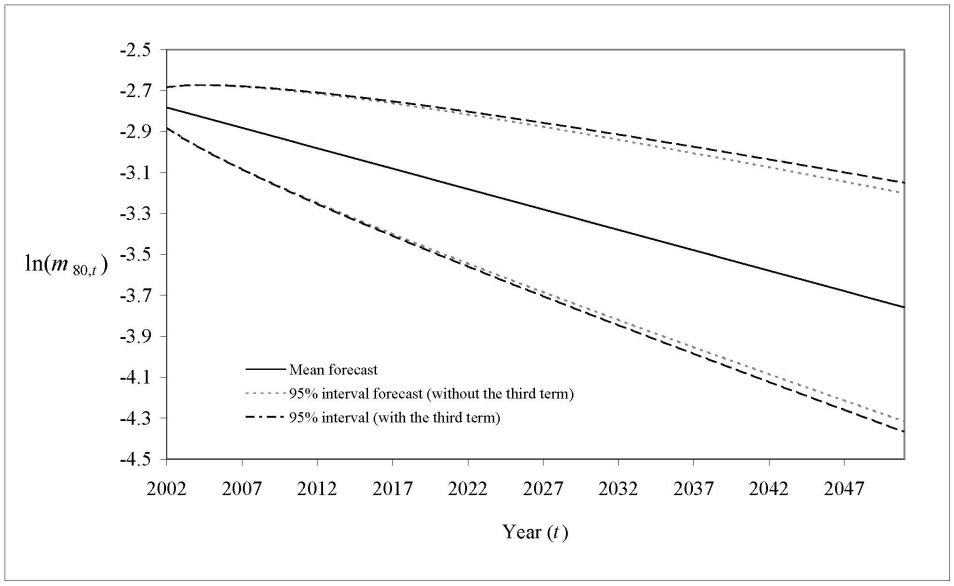


Age 30

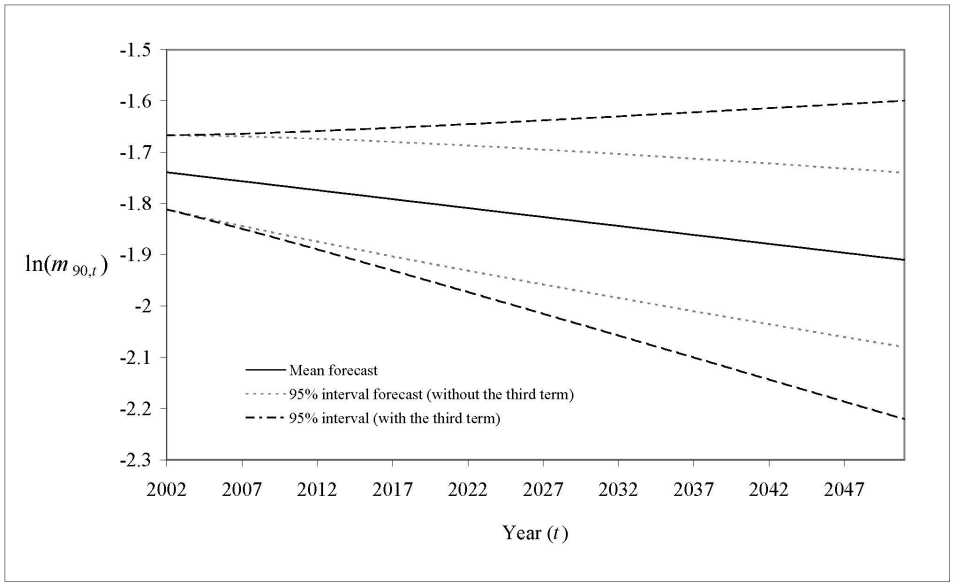


Age 60

Fig. 3.12. Interval forecast of $\ln(m_{x,t})$ with and without the third term in equation (3.32), male.

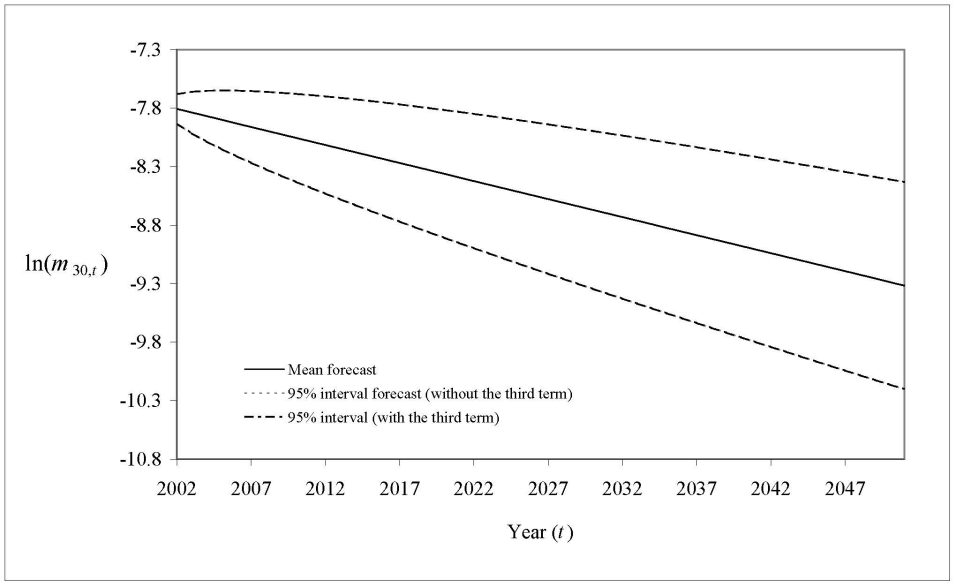


Age 80

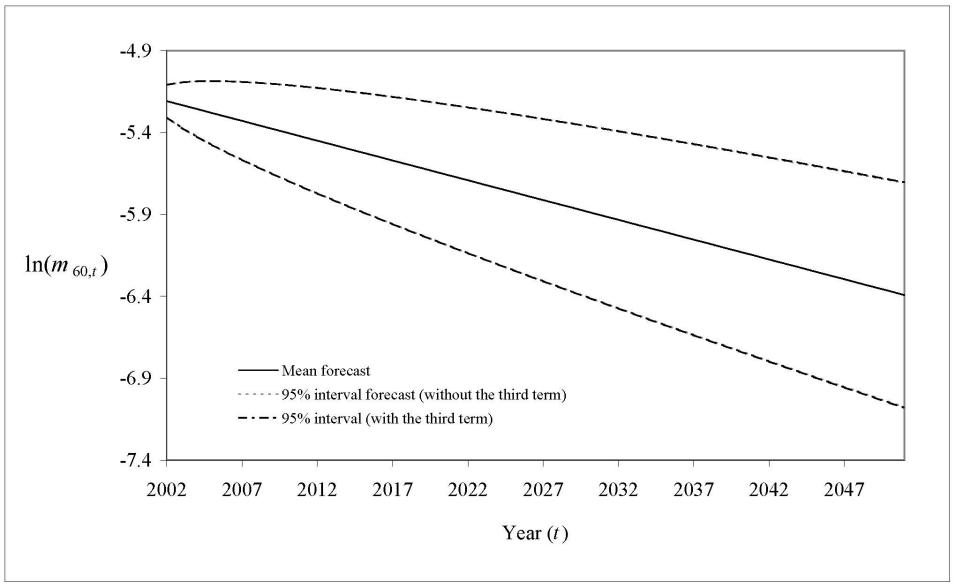


Age 90

Fig. 3.12 (cont'd). Interval forecast of $\ln(m_{x,t})$ with and without the third term in equation (3.32), male.

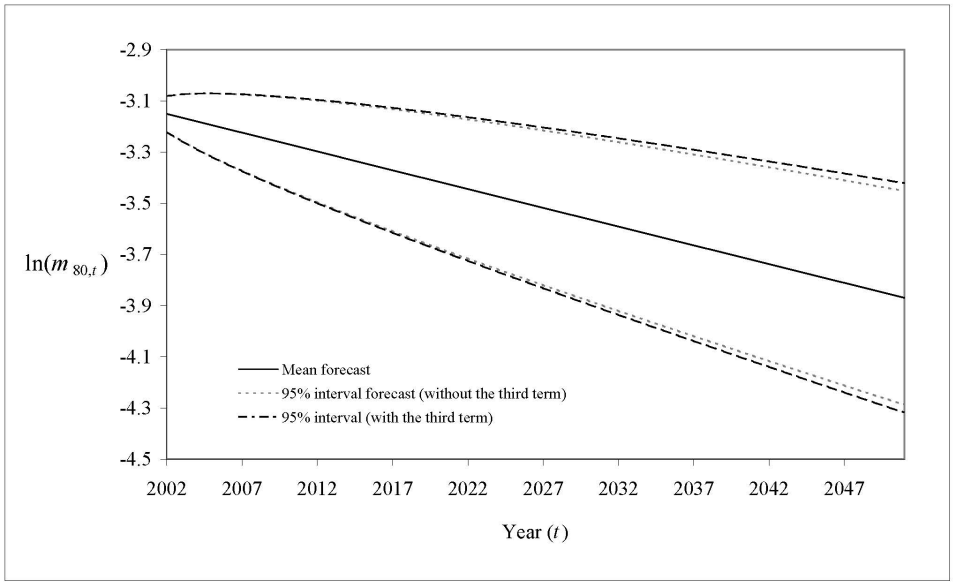


Age 30

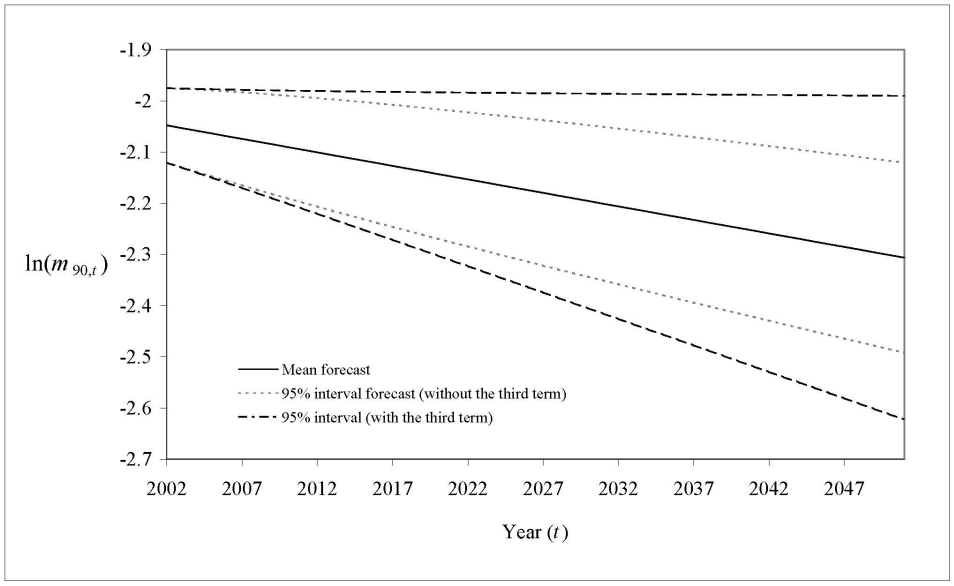


Age 60

Fig. 3.12. Interval forecast of $\ln(m_{x,t})$ with
and without the third term in equation (3.32), female.

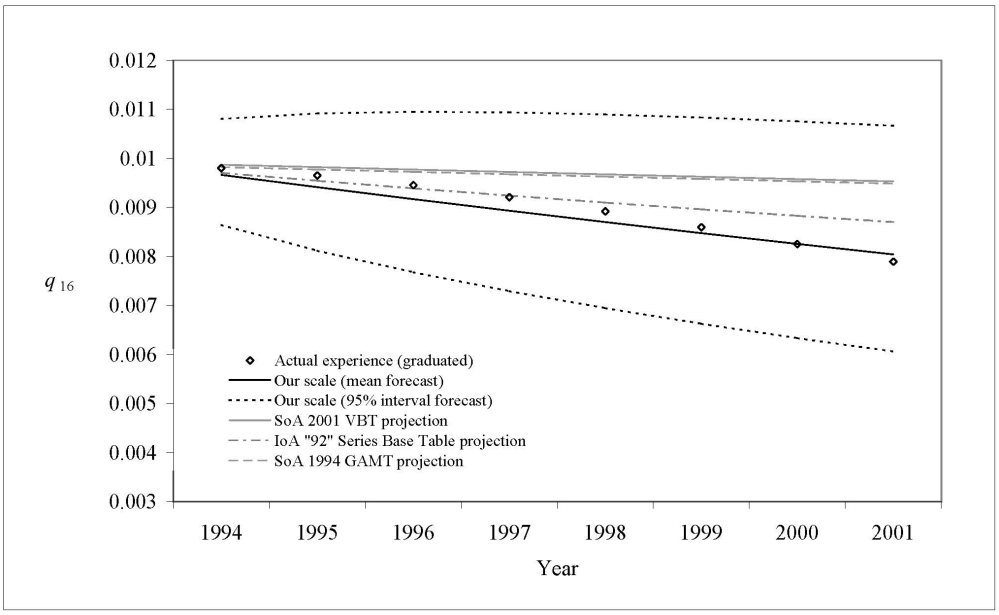


Age 80

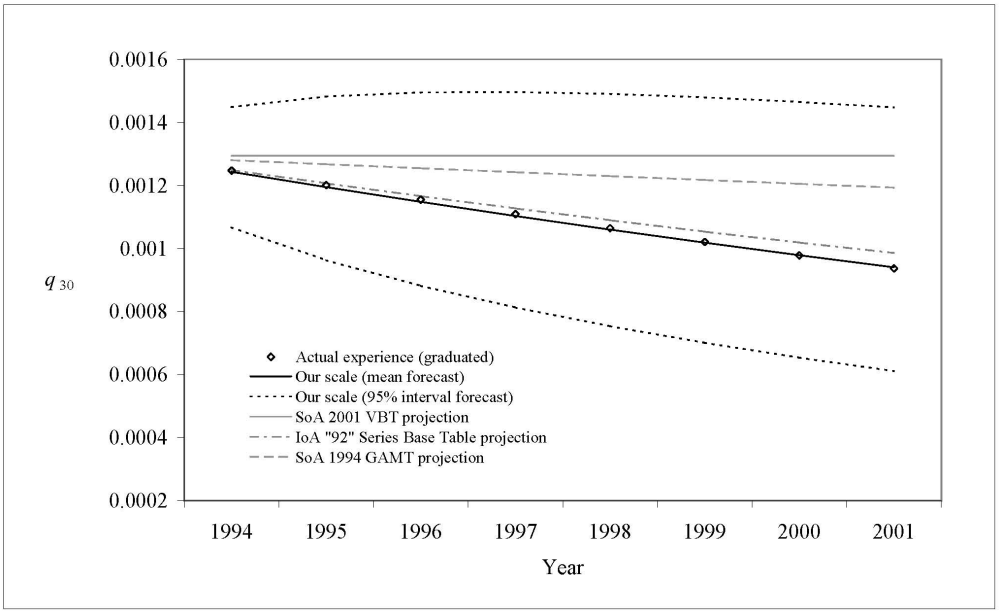


Age 90

Fig. 3.13 (cont'd). Interval forecast of $\ln(m_{x,t})$ with and without the third term in equation (3.32), female.

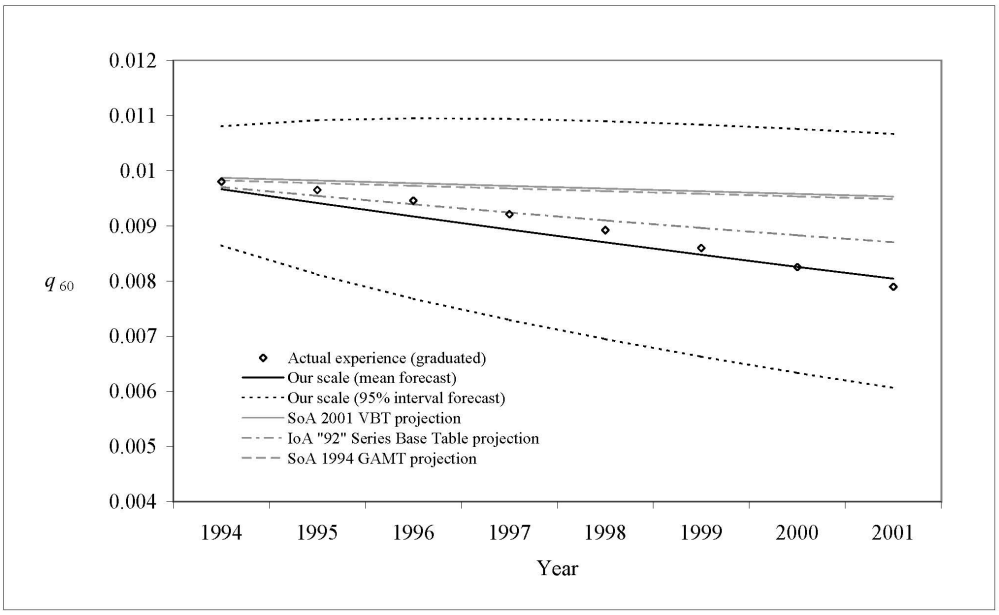


Age 16

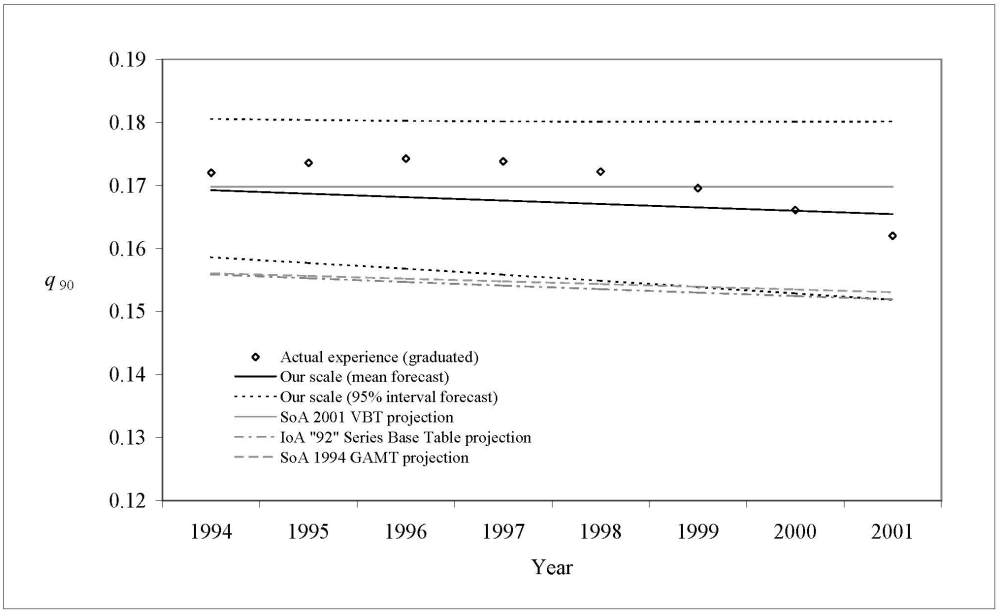


Age 30

Fig. 3.14. Ex post analysis of the improvement scales, male.

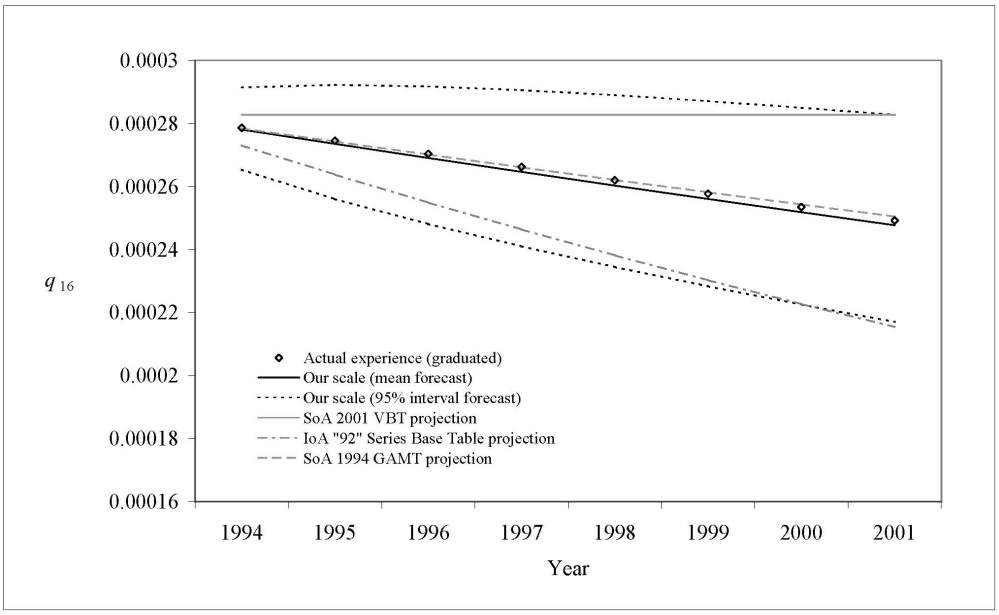


Age 60

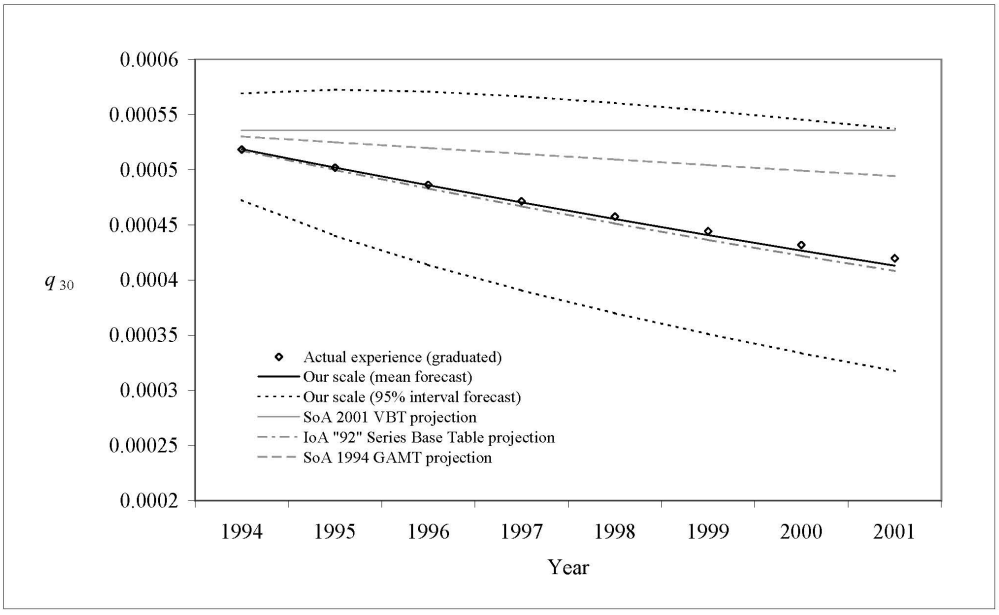


Age 90

Fig. 3.14 (cont'd). Ex post analysis of the improvement scales, male.

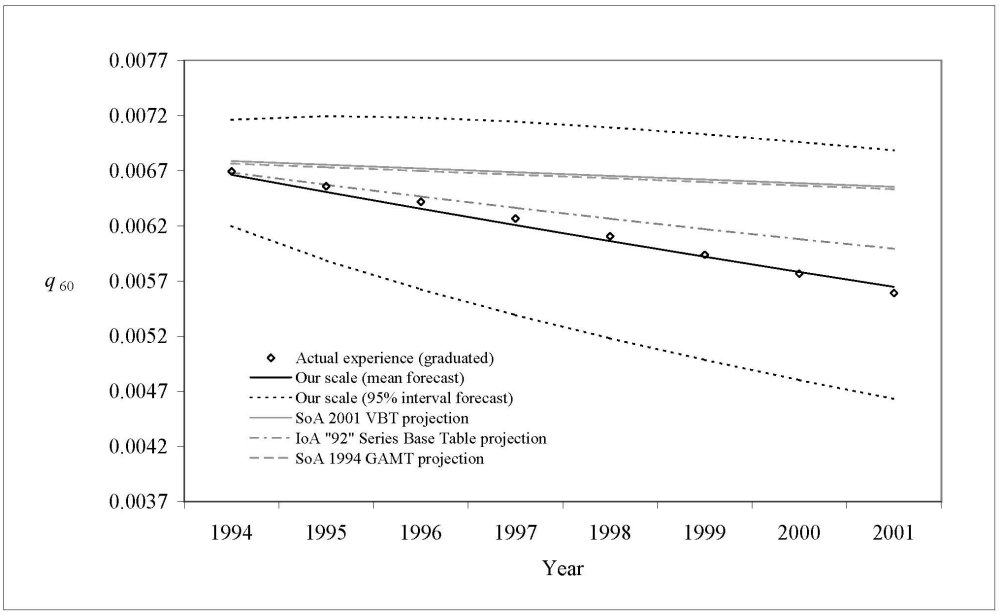


Age 16

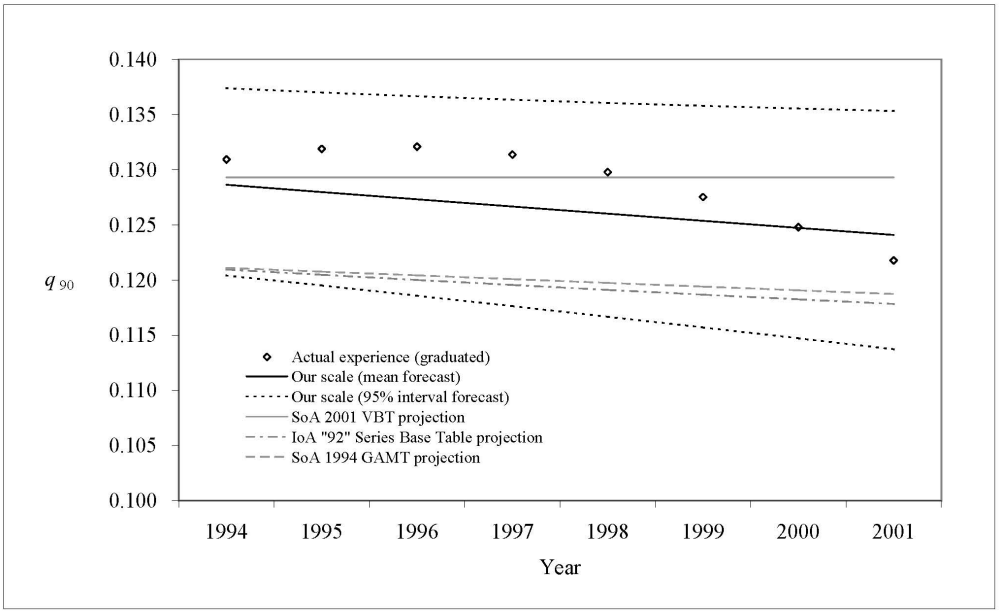


Age 30

Fig. 3.15. Ex post analysis of the improvement scales, female.



Age 60



Age 90

Fig. 3.15 (cont'd). Ex post analysis of the improvement scales, female.

3.5 Further Considerations

3.5.1 The Effect of Duration

Recall that the insured lives experience is segregated by duration 1, 2,..., 15 and ultimate (16+). Denote the central death rate, the number of deaths and the number of exposures-to-risk at duration d by $m^d_{x,t}$, $D^d_{x,t}$ and $E^d_{x,t}$, respectively. In this section, we aim to derive a relationship between $m^{16+}_{x,t}$ and $m^d_{x,t}$, for $d = 1, 2, ..., 15$.

This problem has been considered by various researchers, for example, Currie and Waters (1991), Panjer and Tan (1995), and Renshaw and Haberman (1997). We apply Renshaw and Haberman's methodology.

In contrast to the Lee-Carter assumptions, we assume here that $E_{x,t}$ is random, but $D^d_{x,t}$ is not. Define a random variable $Y^d_{x,t} = \frac{E^d_{x,t}}{D^d_{x,t}}$, and assume that $\mathrm{E}(Y^d_{x,t}) = \frac{1}{m^d_{x,t}}$ and $\mathrm{Var}(Y^d_{x,t}) = \frac{\phi\left(\frac{1}{m^d_{x,t}}\right)^2}{D^d_{x,t}}$. This gives $\left(\frac{1}{m^d_{x,t}}\right)^2$ as the variance function, $D^d_{x,t}$ as the prior weight and ϕ as the scale parameter. Using the variance stabilizing transformation $Q^d_{x,t} = \ln(Y^d_{x,t})$, we have $\mathrm{E}(Q^d_{x,t}) \approx \ln\left(\frac{1}{m^d_{x,t}}\right)$ and $\mathrm{Var}(Q^d_{x,t}) \approx \frac{\phi}{D^d_{x,t}}$, which is free of $\mathrm{E}(Q^d_{x,t})$. This is particularly useful in situations when there is a paucity of data. Denote the graduated ultimate central death rates by $\tilde{m}^{16+}_{x,t}$, and let $Z^d_{x,t} = Q^{16+}_{x,t} - Q^d_{x,t}$. Then,

$$\mathrm{E}(Z^d_{x,t}) \approx \ln(m^d_{x,t}) - \ln(m^{16+}_{x,t}), \tag{3.35}$$

and

$$\mathrm{Var}(Z^d_{x,t}) \approx \phi\left(\frac{D^d_{x,t} + D^{16+}_{x,t}}{D^d_{x,t} D^{16+}_{x,t}}\right). \tag{3.36}$$

Let $v^d_{x,t} = \ln(m^d_{x,t}) - \ln(m^{16+}_{x,t})$. We estimate $v^d_{x,t}$ for each t and d by a linear predictor with $\frac{D^d_{x,t}+D^{16+}_{x,t}}{D^d_{x,t}D^{16+}_{x,t}}$ as the prior weights. If the mortality selection process is effective, we shall anticipate that $v^0_{x,t} \le v^1_{x,t} \le ...v^{15}_{x,t} \le 0$ for fixed x and t.

In Figure 3.16 we observe that $v^d_{x,t}$, the log of the select to ultimate mortality ratio, fluctuates unsystematically for fixed t and d. This motivates us to use a linear predictor in the form of $\hat{v}^d_{x,t} = \hat{\theta}^d_t$.

The estimate of θ^d_t and its standard error can be computed readily by the principle of weighted least squares. Analysis of residuals (not shown) supports the choice of this form

of linear predictor. Finally, we carry out the following procedures to diagnose the behavior of θ_t^d over time.

1. For each d, we sum $E_{x,t}^d$ and $D_{x,t}^d$ over $t = 1982, 1983, ..., 2001$. Using this summation, we arrive at an authoritative value of θ_t^d. This value is denoted by θ^d and is assumed to be invariant over time.

2. For each t, we estimate θ_t^d and its standard error. Then, we use the usual t-test to test the null hypothesis: $\theta_t^d = \theta^d$. Rejection of the null hypothesis implies that θ_t^d may be dependent on time.

The above is illustrated in Figure 3.17. For all d, the null hypothesis is not rejected, suggesting that we may assume that θ_t^d is invariant over time.

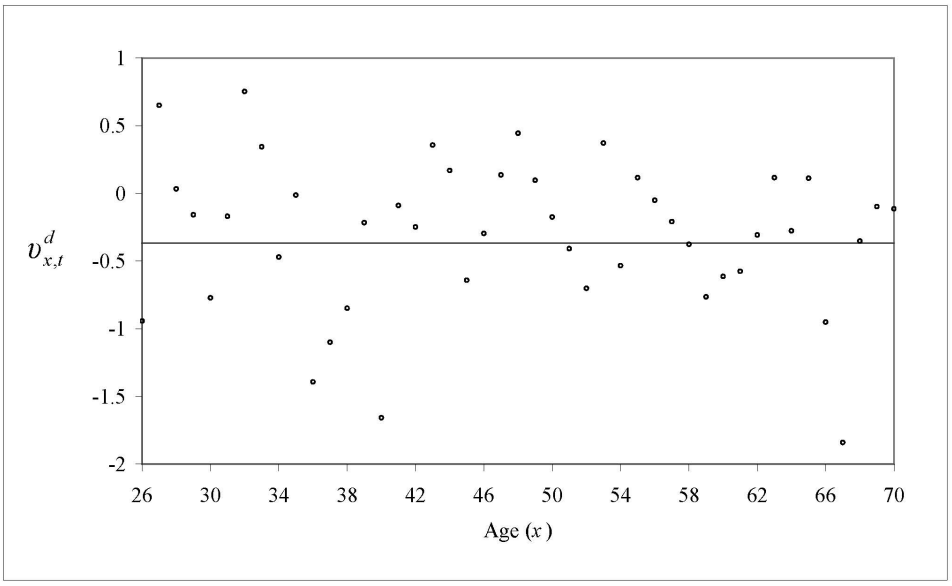
Then, by equation (3.35), we have

$$\hat{m}_{x,t}^d = \tilde{m}_{x,t}^{16+} \exp(\theta^d), \tag{3.37}$$

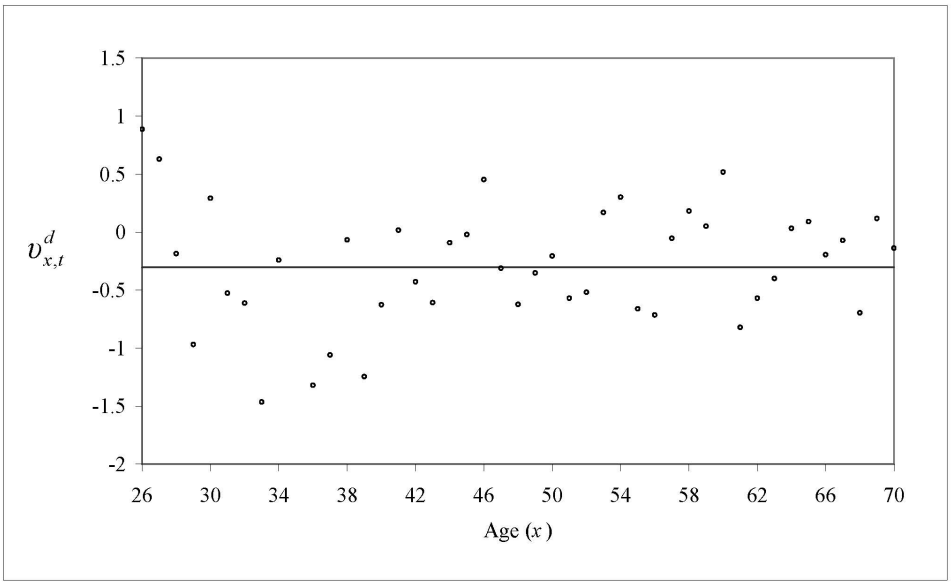
for all x, t, and d. It follows that

$$\mathrm{IS}^d(x, s) = \mathrm{IS}^{16+}(x, s), \tag{3.38}$$

where $\mathrm{IS}^d(x, s)$ and $\mathrm{IS}^{16+}(x, s)$ are the improvement scales for $d = 0, 1, ..., 15$, and the ultimate duration, respectively. Equivalently, there is no strong evidence to support the provision of separate scales for different durations.

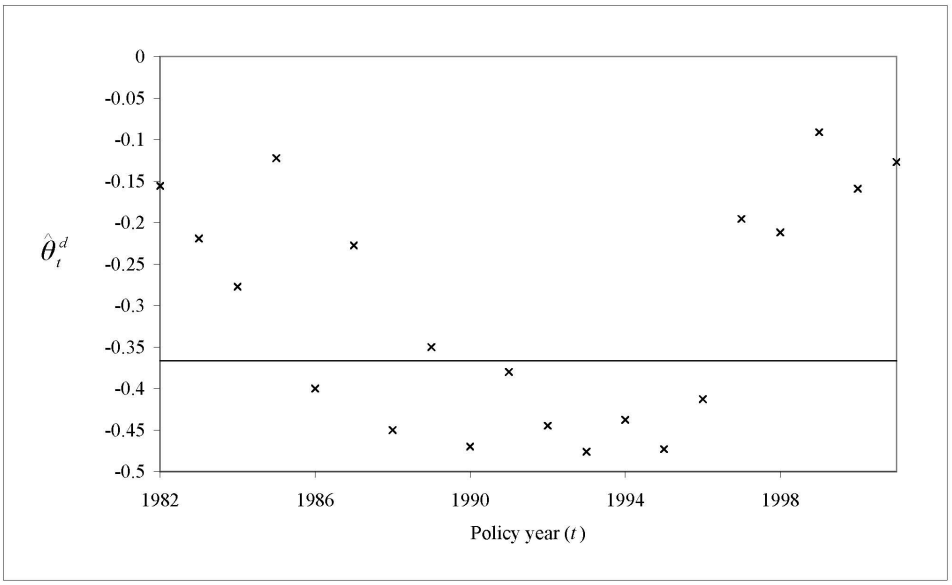


$d = 6$

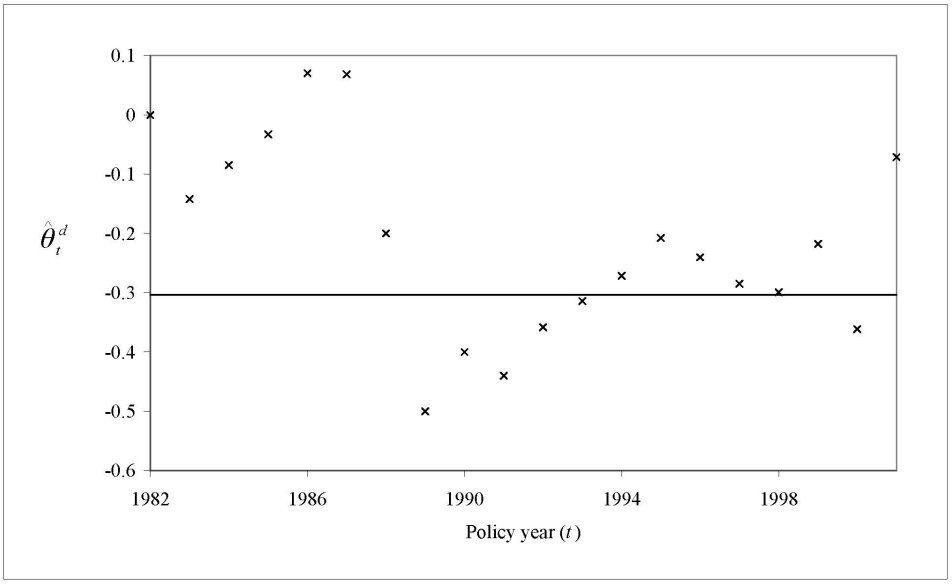


$d = 11$

Fig. 3.16. Trends of $v^d_{x,t}$ over age for selected values of d, male, policy year 2001-2002. The solid line represents the linear predictor $\hat{\theta}^d_t$.



$d = 6$



$d = 11$

Fig. 3.17. Trends of $\hat{\theta}_t^d$ over age for selected values of d, male, policy years 1982-1983 to 2001-2002. The solid line represents the linear predictor $\hat{\theta}^d$.

3.5.2 The Effect of Smoker Status

Parametric models have been used in modeling the relationship between the mortality of different smoker statuses. For instance, in the graduation of Canadian individual insurance mortality experience from 1986 to 1992, Panjer and Tan (1995) modeled the relationship by an age-specific ratio, which is defined as follows:

$$q_x = r(x)q_x^*, \tag{3.39}$$

where

$$r(x) = 1 + l_1(105 - x)^{l_2}x^{l_3}, \tag{3.40}$$

and l_1, l_2 and l_3 are constants that are free of x. In this specification, $r(x)$ tends to 1 at the youngest and highest ages and reaches a maximum at a certain intermediate age. Also, $r(x)$ is assumed to be the same for all durations. Graphical analyses suggest that the model gives a good fit to Canadian individual insurance mortality experience.

If the linkage specified by equation (3.39) is persistent over time, then the same scale can be applied to different smoker statuses, due to the cancellation of $r(x)$[1]. If any of the parameters is time-varying, $r(x)$ may not be canceled and separate scales for different smoker statuses are required. Unfortunately, it is difficult to trend the parameters in equation (3.39) due to the paucity of data: most of the data collected in the beginning of the 1990s are of unknown smoker status; for example, in policy year 1992-1993, the number of exposures-to-risk with identified smoker status accounts for no more than two percent of the aggregate number of exposures-to-risk (see Table 3.3). Consequently, estimates of parameters in the earlier years tend to be highly volatile, and may not be comparable to the more recent ones, which are based on a much richer number of exposures-to-risk. Therefore, the data available at this moment do not allow us to draw a conclusion on the time-varying behavior of the parameters, and we shall revisit this problem when more information is available.

[1]We assume that equation (3.39) is applicable to both q_x and m_x.

	Smoker	Non-smoker	Unclassified
		Male	
1992-1993	12,802	20,300	1,616,410
1996-1997	25,531	23,194	1,538,654
2001-2002	146,829	317,862	1,447,933
		Female	
1992-1993	1,617	10,057	663,043
1996-1997	21,869	8,482	775,397
2001-2002	111,637	220,350	830,030

Table 3.3. Number of exposures-to-risk segregated by smoker status, ultimate, male and female.

3.6 Conclusion

In the first part of this chapter, we examined Canadian population mortality, using 81 years of data up to 2001. We compared the P-splines regression with the Lee-Carter methodology. We concluded that the P-splines regression was highly suitable for graduation, but less useful for projecting future mortality due to an over-emphasis on the most recent mortality trends.

However, the Lee-Carter method in its original form is very restrictive. We have adapted it to allow for over-dispersion by using the negative binomial distribution in place of the Poisson distribution. This was supported by the model diagnostics, and resulted in somewhat wider confidence intervals. Also, we performed an outlier analysis to explore shocks and structural shifts in Canadian population mortality. The evidence for such shocks is not very strong. However, there is a clear highly significant shock in the US data from the 1918 flu pandemic. To allow for the possibility of future epidemics, the detected outliers were included in the bootstrap procedure for estimating the uncertainty in the mortality projection.

On the basis of the Lee-Carter methodology, we proposed a mortality improvement formula that is multiplicative in the central death rates, which is equivalent to saying it is

exponential in the survival probabilities.

Finally, we discussed the selection effect, and concluded that there is no evidence to apply different improvement factors for different durations. We also considered smoker/non-smoker issues, and explained why the data on smoker/non-smoker differential mortality is far too sparse for any conclusions.

Chapter 4

Threshold Life Tables and Their Applications

4.1 Introduction

Since World War II, mortality improvement in developed countries has largely been a product of society's extending 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 (IDL) has been increasing exponentially since the mid-1970s (Robine and Vaupel, 2002). Although this tendency may be subject to an upward bias due to 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 so-called low-frequency-high-severity (LFHS) losses in businesses concerned with life contingencies. Prime examples include life annuity portfolios and defined-benefit (DB) pension plans in which the loss severity is proportional to the lifetime of the annuitant or the pensioner. Reverse mortgages 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 pay-

ment. The loan is paid off by the proceeds of the sale of the property when the homeowner either dies or moves out. In other words, a prolonged life time may result in a severe loss due to 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 the 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 the advanced ages, leading to tremendous 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 age at death (see, e.g., Bernard and Vaupel, 1999; Bourbeau and Desjardins, 2002), and it is often impossible to confirm or disprove the accuracy in the absence of documentary evidence. Without reliable measurements of mortality at the advanced ages, actuaries sometimes close life tables at an arbitrarily chosen age, say 100, by setting the probability of death at that particular age as one. 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 risk measures as the odds of an extreme loss are largely determined by the tail of the assumed 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 the EVT to the parametric modeling of mortality. More specifically, we use the asymptotic distribution of the exceedances over a 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 usual parametric graduation for the earlier ages. This strategy allows us to rationally extrapolate the survival distribution to the extreme ages without the need of mortality data of the supercentenarians, and to statistically determine 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. This gives us a more comprehensive picture of life contingencies: the dynamic version takes account of both mortality risk, which arises from the randomness of the age at death, and longevity risk, which originates

from the uncertainty in the trend of mortality improvement. The dynamic threshold life tables are finally utilized 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, commonly referred to as omega in the actuarial literature.

The rest of the chapter 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 omega. Section 6 concludes.

4.2 Notation and Data

The following notation is used throughout this chapter:

- 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 mid-year 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 in calender year t. For detailed interpretations of the above notation, we refer readers to Chapter 3 of Bowers et al. (1997).

Crude data should be preferred over abridged rates in modeling old-age mortality. Although raw death counts for the Canadians are quite exhaustive, raw census population counts for Canadians are right-censored at age 90; that is, the coverage of raw census population counts by single year of age ceases at age 89. According to the Human Mortality Database documentation (Wilmoth et al., 2005), 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,

$$P(x,t) = \sum_{i=0}^{\omega-x-1} \left[D_U(x+i, t+i) + D_L(x+i+1, t+i) \right], \tag{4.1}$$

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. Figure 4.1 illustrates 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.

While the contaminated data have a relatively smaller impact on the application of the Lee-Carter model, the impact on the analysis of the highest attained age and the pattern of extreme-age mortality 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, instead of the Canadian population, we apply the theoretical results in this chapter to populations where mortality data are more complete. Specifically, we consider Japan and the Scandinavian countries, that is, Denmark, Finland, Norway and Sweden. For these countries, raw death and population counts by single year of age are available up to and including age 99. The required data for this chapter are obtained from the Human Mortality Database (2005).

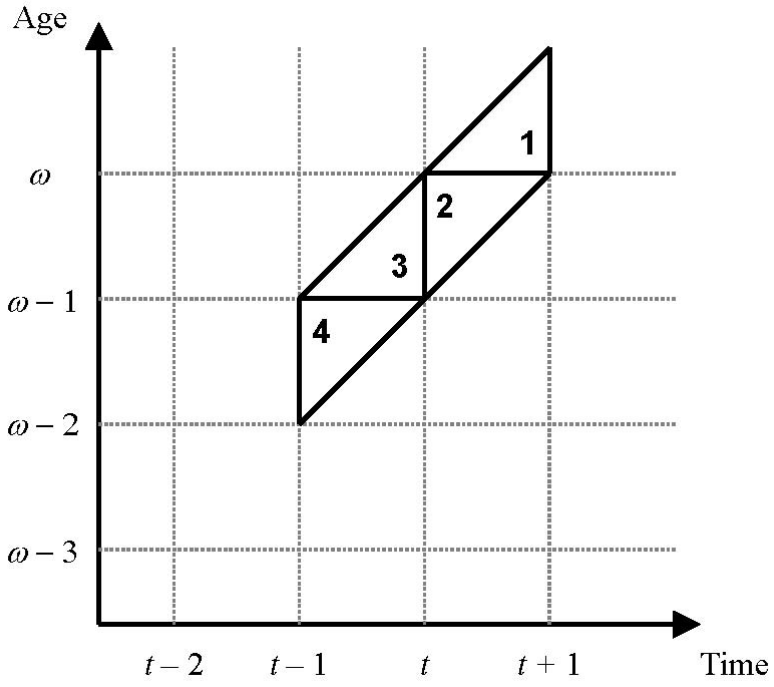


Fig. 4.1. Illustration of the extinct cohorts method via a Lexis diagram.

4.3 Previous Research on Modeling Old-age Mortality

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

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

$$WLS = \sum_{x=z}^{99} E_x(\mu_x^{grad} - a - bx - cd^x)^2, \tag{4.2}$$

subject to the constraint that $\mu_z^{grad} = a - bz - cd^z$, where $\mu_z^{grad} = -\ln(1 - q_x^{grad})$, 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:

$$q_{x+99}^{grad} = ax^3 + bx^2 + cx + d, \tag{4.3}$$

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 $q_{105}^{poly} = 1$.

4.3.2 The Heligman-Pollard Model

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

$$\frac{q_x}{1 - q_x} = A^{(x+B)^C} + D\exp(-E(\ln x - \ln F)^2) + GH^x, \tag{4.4}$$

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.

While the estimation of model parameters can be performed easily by minimizing the weighted sum of squared errors, formulated in a similar manner as equation (4.2), there is strong empirical evidence that old-age mortality is not Gompertzian (see, e.g., Olshansky, 1997). Hence, the Heligman-Pollard model may not be ideal for describing old-age mortality.

4.3.3 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. This 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

$$k(x) = k(x-1) - R, \quad x \geq 84, \tag{4.5}$$

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

$$k(85) + ... + k(110) = 26k(84) - (1 + 2 + ... + 26)R. \tag{4.6}$$

Solving for R, we obtain

$$R = \frac{26k(84) + \ln(m_{84}) - \ln(m_{110})}{351}. \tag{4.7}$$

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 age 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.

4.3.4 The Relational Model of Mortality

The relational model (Himes et al., 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:

$$\text{logit}(m_x^s) = \alpha + \beta x, \tag{4.8}$$

where $\text{logit}(m_x^s) = \ln\left(\frac{m_x^s}{1-m_x^s}\right)$, and m_x^s denotes the "standard" central rate of death at age x. Equation (4.8) 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 a logit regression:

$$\text{logit}(m_x) = \delta + \gamma \text{logit}(m_x^s), \tag{4.9}$$

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

4.4 The Threshold Life Table

We begin with an exploratory graphical analysis of a typical pattern of the central rate of death. Figure 4.2 indicates that the rate of increase after age 85 is much more rapid than before. The modeling of the age pattern of mortality is usually done in a piece-wise approach. First, a parametric graduation is 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 the 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.

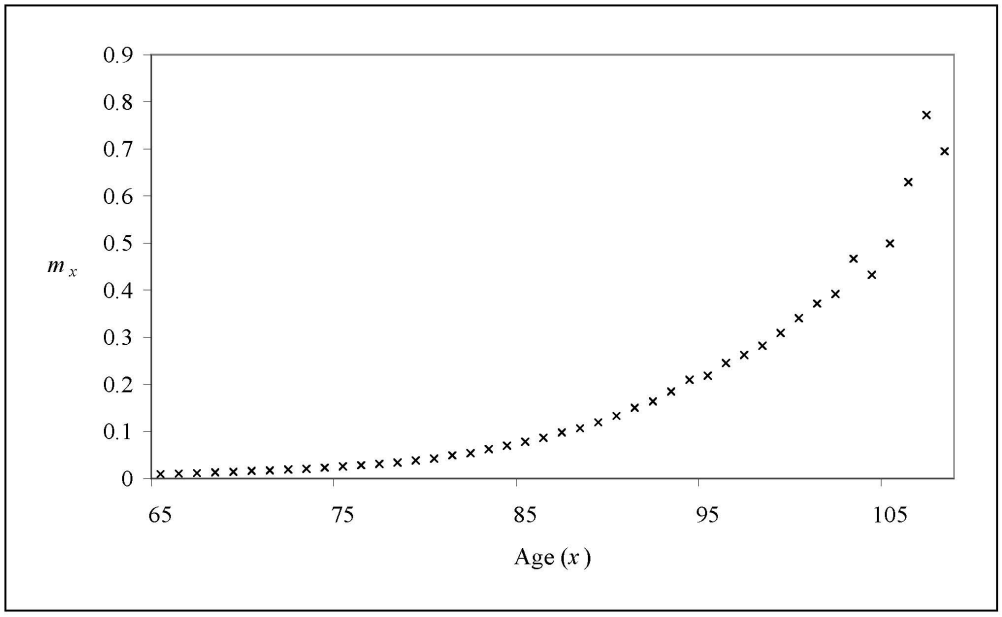


Fig. 4.2. Central rate of death from age 65,
the Japanese population mortality experience, 2003.

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

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 (GP) 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 $x \leq N$,

$$F(x) = 1 - \exp\left(-\frac{B}{\ln C}(C^x - 1)\right); \tag{4.10}$$

for $x > N$,

$$F(x) = 1 - p\left(1 + \gamma\left(\frac{x - N}{\theta}\right)\right)^{-1/\gamma}, \tag{4.11}$$

where $p = F(N)$ 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 $B > 0$, $C > 0$, 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.

The usual Hill's plot (Hill, 1975; Mason, 1982) is not applicable in the selection of the threshold age, due partly to the fact that the mortality data is highly censored, as we know only the number of deaths in different age intervals but not the precise age at death for each individual. Hill's estimator is however not designed for censored data. Moreover, the use of Hill's plot requires eyeballing and this is infeasible in the simulation studies we are going to 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.

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

$$L(B, C, \gamma, \theta; N) = \prod_{x=65}^{99} \left(\frac{s(x) - s(x+1)}{s(65)} \right)^{d_x} \left(\frac{s(100)}{s(65)} \right)^{l_{100}}. \tag{4.12}$$

  For age $x = 65, ..., 99$, 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 (4.12) is divided by $s(65)$ to acknowledge the left-truncation at age 65. We can easily show that the logarithm of equation (4.12) can be decomposed as the sum of two components, $l_1(B, C; N) + l_2(\gamma, \theta; N)$, where

$$l_1(B, C; N) = \sum_{x=65}^{N-1} d_x \ln[s(x) - s(x+1)] + l_N \ln[s(N)] - l_{65} \ln[s(65)], \tag{4.13}$$

  where $s(x) = \exp\left(\frac{-B}{\ln C}(C^x - 1)\right)$, and

$$l_2(\gamma, \theta; N) = \sum_{x=N}^{99} d_x \ln \left[\frac{s(x)}{s(N)} - \frac{s(x+1)}{s(N)} \right] + l_{100} \ln \left[\frac{s(100)}{s(N)} \right], \tag{4.14}$$

  where $\frac{s(x)}{s(N)} = \left(1 + \gamma \left(\frac{x-N}{\theta}\right)\right)^{-1/\gamma}$.
  This observation suggests that for fixed N, parameter estimation for the Gompertz part and the generalized Pareto part can be done separately by maximizing l_1 and l_2, respectively. The choice of N relies on the maximization of the profile likelihood:

$$l_p(N) = l\left(\hat{B}(N), \hat{C}(N), \hat{\gamma}(N), \hat{\theta}(N); N\right), \tag{4.15}$$

where $l = \ln(L)$; $\hat{B}(N)$,$\hat{C}(N)$,$\hat{\gamma}(N)$ and $\hat{\theta}(N)$ are the maximum likelihood estimates 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 likelihood, l_p.
2. Repeat Step (1) for $N = 97, 96, ..., 85$.
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 d_x's and l_x's can be computed by constructing a hypothetical cohort based on the static q_x's. 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 ML 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 m_x's, then the method of weighted least squares can be readily applied. Let us rewrite the threshold life table in terms of the force of mortality:

  for $x \leq N$,

$$\ln(\mu(x)) = \ln(B) + (\ln C)x; \tag{4.16}$$

  for $x > N$,

$$\ln(\mu(x)) = -\ln(\theta) - \ln\left[1 + \gamma\left(\frac{x - N}{\theta}\right)\right]. \tag{4.17}$$

Under the assumption of a 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, defined as follows:

$$SSE = SSE_1 + SSE_2, \tag{4.18}$$

where

$$SSE_1 = \sum_{x=65}^{N-1} E_x \left[\ln(m_x) - \ln(B) - (\ln(C))x\right]^2, \tag{4.19}$$

and

$$SSE_2 = \sum_{x=N}^{99} E_x \left[\ln(m_x) + \ln(\theta) + \ln\left(1 + \gamma\left(\frac{x-N}{\theta}\right)\right)\right]^2. \tag{4.20}$$

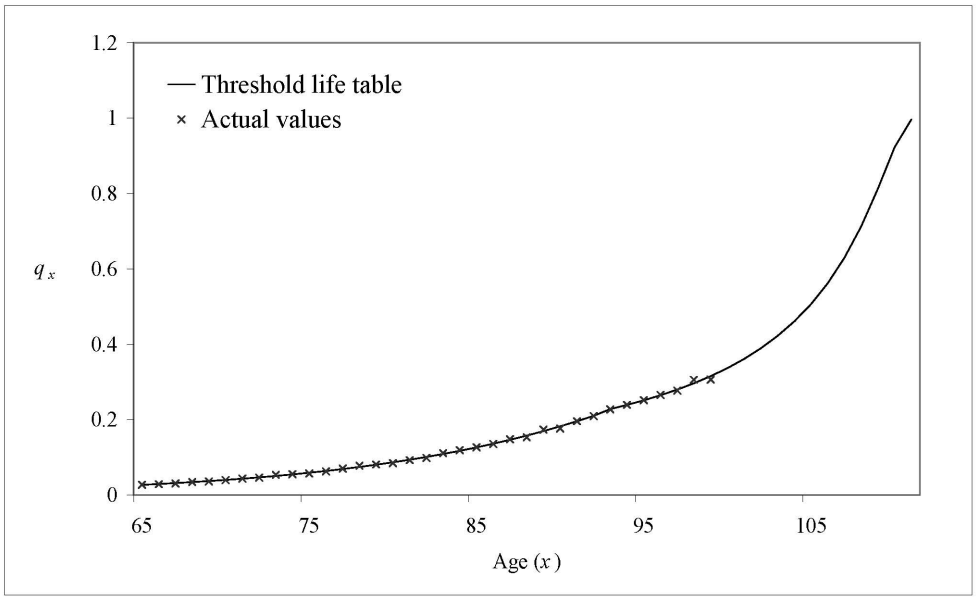
The squared errors are weighted by the number of exposures-to-risk to allow for a better fit for ages with larger risk exposures[1]. Similar to Method 1, the parameters in the Gompertz part and the generalized Pareto part can be estimated separately given a fixed value of N. The estimation process can 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 $N = 97, 96, ..., 85$.
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.

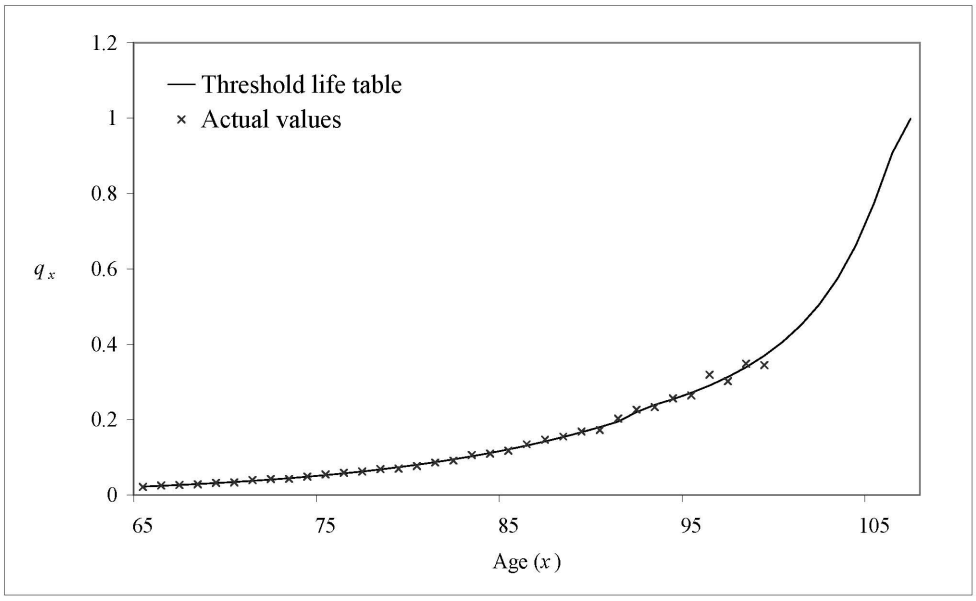
[1] An equal weight of 1 yields a similar result.

Figure 4.3 depicts the fitted threshold life tables for the 1897 birth cohort in Japan and the Scandinavian countries. 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 effect is 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 extends progressively to 1 at the end point of the life table, which is determined statistically by the fitted distribution. Detailed discussion on the end point is deferred to the later section that deals with the prediction of the highest attained age.

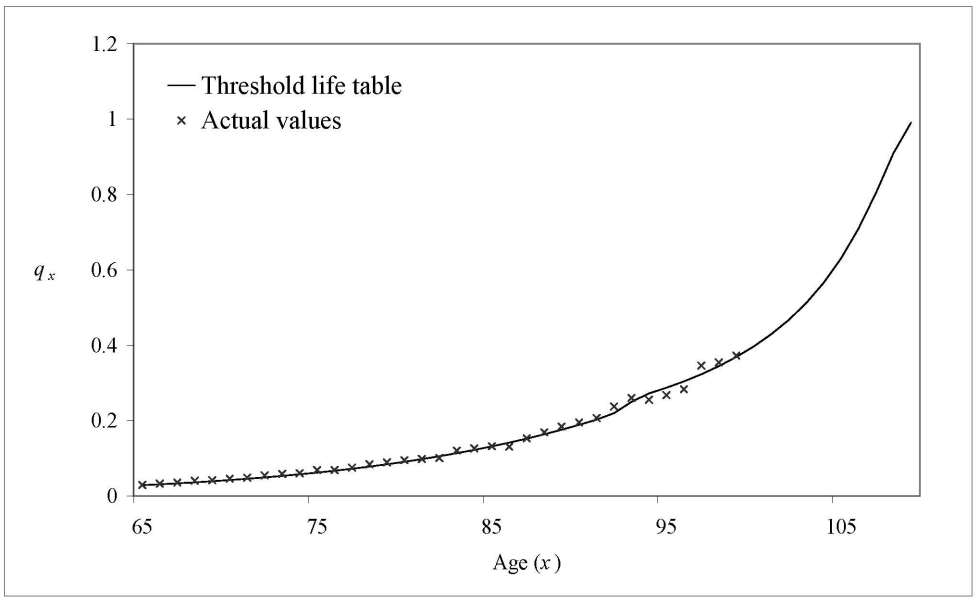


Japan, $N = 93$

Fig. 4.3. Fitted threshold life tables for the 1897 birth cohort in Japan and the Scandinavian countries.

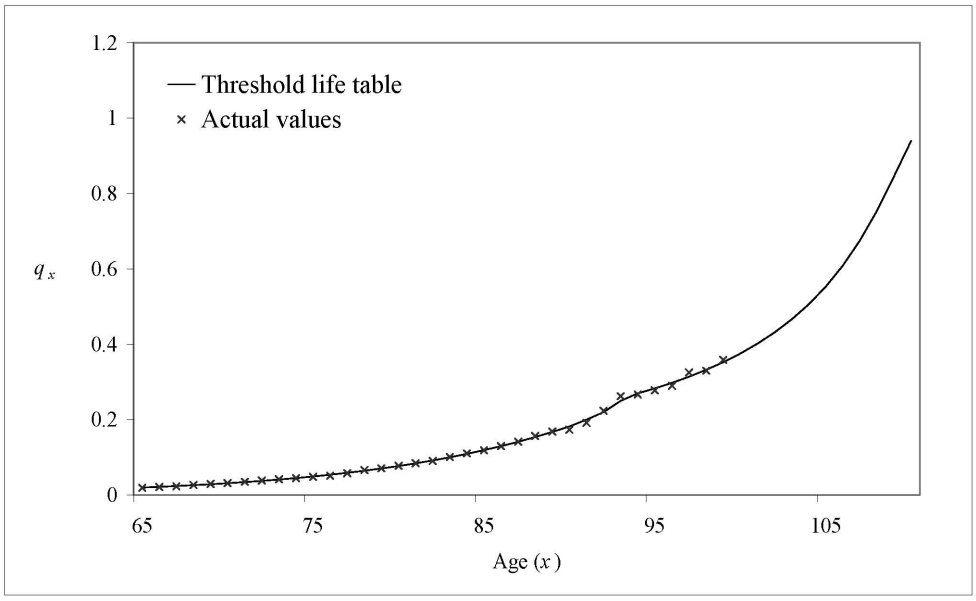


Denmark, $N = 92$

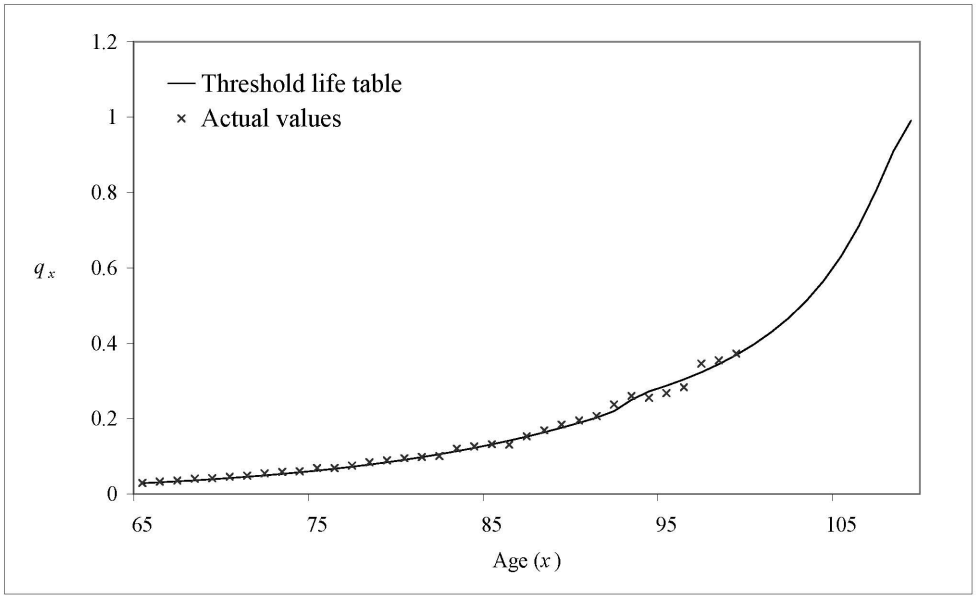


Finland, $N = 92$

Fig. 4.3 (cont'd). Fitted threshold life tables for the 1897 birth cohort in Japan and the Scandinavian countries.



Norway, $N = 93$



Sweden, $N = 95$

Fig. 4.3 (cont'd). Fitted threshold life tables for the 1897 birth cohort in Japan and the Scandinavian countries.

4.5 Applications

4.5.1 Risk Measures

We consider a simple portfolio of a single premium life annuity with unit annual benefit sold to a life-aged-65 in 2003. Valuations are based on 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 expense.

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 year 2003 and close the life table at age 100; that is, we assume that $q_{100} = 1$. To simulate the age at death (curtate future life time, $K(65)$), we generate z from the Uniform (0,1) distribution; then we can obtain the simulated age at death by the following equation:

$$K(65) = \sup\left\{ y : \prod_{x=65}^{y} (1 - q_x) > z \right\}, \tag{4.21}$$

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

$$PV = \frac{1 - 1.05^{-(K(65)+1)}}{0.05/1.05}. \tag{4.22}$$

  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.

- *Method 2*
  We use the static death rates in year 2003, and model these death rates by a threshold life table, which extends the rates to the extreme ages. The evaluation of the loss is similar to that in Method 1 except that the q_x's in equation (4.21) are replaced by those obtained from the fitted threshold life table.

- *Method 3*
  We use the cohort death rates for the cohort born in 1938 (aged 65 in 2003). To take longevity risk into account, we simulate 10,000 Lee-Carter mortality scenarios, based the historical mortality data from 1950 to 2003, by the parametric bootstrapping method mentioned in Chapter 2. For each of these scenarios, we compute the projected cohort death rates: $q_{65,2003}$, ..., $q_{99,2037}$. Here, we do not extrapolate the rates and assume that $q_{100,2038} = 1$. 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 4.4 provides a graphical illustration of the above four methods. In Table 4.1, 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, since Methods 3 and 4 have taken account of the non-diversifiable longevity risk.

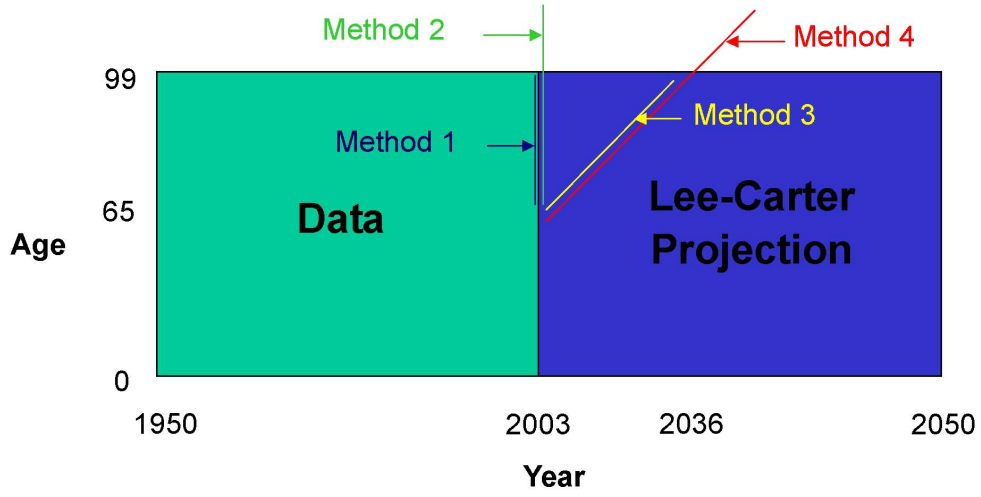


Fig. 4.4. Distinction between the four methods of valuation.

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

Table 4.1. Estimated mean, variance, 95% VaR and 95% CTE of the loss using different methods.

4.5.2 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.

Rigorous statistical modeling of the highest attained age (omega) was pioneered by Thatcher (1999), who predicted omegas for several cohorts using the classical extreme value theory. At the outset, Thatcher assumed that the age-at-death random variable follows a three-parameter logistic distribution:

$$\mu(x) = \frac{\alpha \exp(\beta x)}{1 + \alpha \exp(\beta x)} + \gamma. \tag{4.23}$$

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

$$F_{M_n}(x) = \left\{ 1 - \exp\left(- \int_{x_0}^{x} \mu(t) dt \right) \right\}^n. \tag{4.24}$$

Combining equations (4.23) and (4.24), Thatcher obtained the following closed form expression for the distribution function of M_n:

$$F_{M_n}(x) = \left[1 - \exp\left(\gamma(x_0 - x) + \beta^{-1} \ln\left(\frac{1 + \alpha \exp(\beta x_0)}{1 + \alpha \exp(\beta x)} \right) \right) \right]^n. \tag{4.25}$$

The choice of x_0 is arbitrary, and n can be approximated by the mid-year 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 the historical period death rates to age 150 by the relational model proposed by Himes et al. (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 (4.24). However, the distribution function of M_n must be calculated numerically as the extrapolation of death rates was performed in a non-parametric manner.

Similar to our proposed methodology, 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 United States 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 obtained. Han's method, however,

merits re-consideration for a number of reasons. First, the threshold ages were chosen in a completely arbitrary manner, leading to possible inconsistency with the age-pattern at earlier ages. Second, the application to historical period life tables is inappropriate since the highest attained age makes sense only if it is referenced to a particular birth cohort. Third, 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 et al. (2006) modeled the highest attained age directly by using the Generalized Extreme Value (GEV) distribution:

$$F_{M_n}(x) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{4.26}$$

where μ is the location parameter, σ is the scale parameter, and ξ is the shape parameter. The GEV distribution was fitted to the ten 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:

$$\begin{aligned} L &= \prod_{t=1980}^{2000}\left\{\exp\left\{-\left[1+\xi\left(\frac{z_t^{(10)}-\mu(t)}{\sigma(t)}\right)\right]^{-1/\xi}\right\}\right. \\ &\qquad \left.\prod_{k=1}^{10}\sigma(t)^{-1}\left[1+\xi\left(\frac{z_t^{(k)}-\mu(t)}{\sigma(t)}\right)\right]^{-1/\xi-1}\right\}, \end{aligned} \tag{4.27}$$

where $z_t^{(k)}$ is the k^{th} 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 were obtained by substituting μ and σ in equation (4.26) by $\mu(s)$ and $\sigma(s)$ respectively. Problems associated with this method include: (1) the assumption of linearity in the location parameter implicitly preset a linear trajectory of omega; (2) this direct approach requires precise data about the highest age at death, which is heavily subject to the problem of misreporting − it was shown that the right-hand-end support of the distribution of M_n would change from infinite to finite upon removal of an outlier.

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:

$$F_Y(y) = 1 - \left(1 + \gamma \frac{y}{\theta}\right)^{-1/\gamma}, \tag{4.28}$$

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 is experienced when threshold life tables are fitted to the mortality rates for various birth cohorts), 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, $i = 1, 2, ..., n$, be a sequence of exceedances over the threshold age and assume that Y_i follows the generalized Pareto distribution specified by equation (4.28) independently. Also, we define $M_n = \max\{Y_i, i = 1, 2, ..., n\}$. If $\gamma < 0$, then the distribution function of M_n can be expressed as

$$F_{M_n}(y) = \begin{cases} 0, & y < 0 \\ [F_Y(y)]^n, & 0 \le y < -\theta/\gamma. \\ 1, & y \ge -\theta/\gamma \end{cases} \tag{4.29}$$

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

$$\lim_{n\to\infty} F_{M_n}(y) = \begin{cases} 0, & y < -\theta/\gamma \\ 1, & y \ge -\theta/\gamma. \end{cases} \tag{4.30}$$

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 $\omega = N - \frac{\theta}{\gamma}$, 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:

$$\mathrm{Var}(\hat{\omega}) = \begin{bmatrix} \frac{\partial \omega}{\partial \gamma} & \frac{\partial \omega}{\partial \theta} \end{bmatrix} [I(\gamma, \theta)]^{-1} \begin{bmatrix} \frac{\partial \omega}{\partial \gamma} \\ \frac{\partial \omega}{\partial \theta} \end{bmatrix}, \tag{4.31}$$

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

$$\left[\hat{\omega} - 1.96\sqrt{\text{Var}(\hat{\omega})}, \hat{\omega} + 1.96\sqrt{\text{Var}(\hat{\omega})}\right]. \tag{4.32}$$

Table 4.2 shows the predicted omegas for the 1897 birth cohort in Japan and the Scandinavian countries.

To predict the omegas 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 $q_{65,1996}$, $q_{66,1997}$, ..., $q_{99,2020}$. The values of $q_{65,1996}$,..., $q_{72,2003}$ can be obtained from the data provided by the Human Mortality Database, while those of $q_{73,2004}$, ..., $q_{99,2020}$ may be derived from a Lee-Carter mortality projection, based on the historical death rates. An accurate projection of senescent mortality requires a sufficiently lengthy time-series of exhaustive and uncontaminated mortality data. However, the data for the Scandinavian countries do not satisfy these criteria; for instance, the raw population counts for Denmark are right-censored at age 90 prior to year 1975. We shall therefore examine the impact of the year of birth on the highest attained age for Japanese population only.

On the basis of the historical mortality experience of the Japanese population from 1950 to 2003, we obtain a Lee-Carter 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 for the omegas using equations (4.31) and (4.32). Figure 4.5 graphically illustrates our methodology.

Table 4.3 shows the predicted omegas for selected birth cohorts. Moving down the column, we observe that the improvement of mortality 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 years per annum. Also, we observe that the width of the confidence interval increases with the year of birth. This trend is because the continual mortality improvement allows more people to survive to the open age group, leading to more uncertainty about the tail of the survival distribution.

We can further incorporate the uncertainty in mortality improvement into the probability distribution of omega by the following algorithm:

1. Generate N mortality scenarios by a parametric bootstrap (see Chapter 2).
2. For each scenario:
   (a) derive the cohort death rates for the birth cohort in which we are interested;
   (b) estimate the omega of the cohort by fitting a threshold life table to the cohort death rates.
3. Obtain an empirical distribution of omega for the birth cohort.

Table 4.4 compares the predicted omega for the cohort born in 1941 with and without considering the uncertainty in mortality improvement. While the best estimates are similar, the standard error is significantly larger if the uncertainty about future mortality is taken into account. Note that the confidence intervals are obtained by different methods: the one without considering the uncertainty in mortality improvement is based on the asymptotic normality of the maximum likelihood estimates, while the other one is obtained from the percentiles of the empirical distribution, which is permitted to be skewed.

Country	Predicted omega
Japan	112.4605 (108.7544, 116.1580)
Denmark	108.2390 (106.1907, 110.2871)
Finland	109.0127 (106.4956, 111.5297)
Norway	112.5596 (109.1448, 115.9747)
Sweden	111.9840 (106.9487, 117.0193)

Table 4.2. Predicted omegas for the 1897 birth cohort in Japan and the Scandinavian countries.

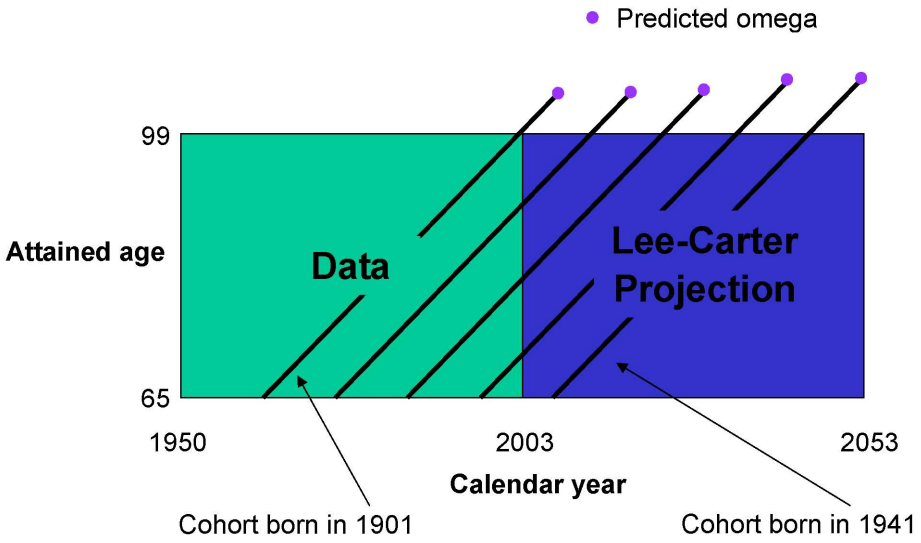


Fig. 4.5. Graphical illustration of the methodology used in predicting omegas for various cohorts of the Japanese population.

Year of birth	Attained age in year 2006	Predicted omega
1901	105	113.4477 (109.4321, 115.3740)
1911	95	114.6435 (109.0398, 120.1731)
1921	85	115.0247 (107.8065, 122.2430)
1931	75	119.9964 (110.0398, 132.1508)
1941	65	121.0953 (107.0769, 132.9158)

Table 4.3. Predicted omegas for various cohorts of the Japanese population.

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)

Table 4.4. Predicted omega for the cohort born in year 1941, Japanese population, with and without considering the uncertainty in mortality improvement.

4.6 Conclusion

The tail behavior of the survival function has long been a controversial issue. There are three schools of thought on the subject. The first one argues that q_x may reach one at a finite age (e.g., Vincent, 1951). The second belief 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 which 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 the crude data remains open. Unverified ages at death obtained from government agencies are often exaggerated, leading to a positive bias and possibly an unwanted kurtosis in the predicted distribution of the highest attained age. Watts et al. (2006) provided an excellent illustration. On the other hand, the use of verified mortality data is not an alternative, as the number of validated supercentenarians in Japan is too scanty for statistical inferences of any kind[2]. Until sufficient validated data among the oldest-old are available, a minimal extrapolation, either mathematical or statistical, is apparently the only feasible solution.

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

The accuracy of the predicted omegas is highly dependent on the validity of the assumptions used. The predicted omegas in this study would tend to 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 tend to be accurate if the relational model (see Section 4.3.4) they used is appropriate. These conditions, unfortunately, can never be verified.

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 would likely affect the forecast of the highest attained age.

[3]A distribution function F belongs to the peaks-over-threshold domain of attraction if and only if $n\bar{F}(a_n x + b_n) \to -\ln(W(x))$ as $n \to \infty$, for all x, where W is the distribution function of a generalized Pareto random variable, for some sequence 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).

Chapter 5

On Pricing and Hedging the No-Negative-Equity-Guarantee in Equity Release Mechanisms

5.1 Introduction

In the World Bank Report on Averting the Old Age Crisis (The World Bank, 1994), it is suggested that financial security for the old should be based on three pillars. These are (1) a publicly managed system with mandatory participation and the limited goal of reducing poverty among the old, (2) a privately managed mandatory savings system, and (3) a voluntary savings system. Working together, these three pillars would ensure against the many risks of old age.

In the United Kingdom, issues with an increasing dependency ratio have resulted in government policies being formulated to reduce the proportion of pension provided by the public sector. The state pension reforms proposed in the White Paper "Personal accounts: a new way to save" will promote a reduced reliance on pension credit and ensure that the pension credit is targeted towards those who need it (Department for Work and Pension, 2006). On the other hand, defined benefit schemes are in decline, with many being closed to new entrants and even new contributions. Also, levels of contributions from employers tend to fall when they move to other schemes. The weakening of the first two pillars

means that the general population will have to make material increases to their private savings if they are to achieve the same level of retirement income. Nevertheless, there is a strong inertia towards saving for retirement. Independent research commissioned by the Association of British Insurers showed that Britain has an annual aggregate savings gap of approximately £27 billion. This represents the difference between what people should be saving to fund a reasonable retirement and what they are actually saving.

For many people, the savings gap could be bridged by participating in home equity release mechanisms (ERMs), which enables a householder to draw down part of the equity in the house. Such mechanisms are particularly helpful to the so-called "asset-rich, cash-poor" pensioners who receive low incomes but own homes of fairly substantial value. Sodha (2005) estimated that in Britain, of those who were retired, 10.2 percent had an income below the "Modest but Adequate" standard (£157 per week before housing costs), but owned more than £100,000 of housing wealth. The UK equity release market has been expanding rapidly since the early 1990s. It is now at the level of £1.1 billion new loans per annum, which is five times what it was five years ago. The potential for further growth is justified by the massive property wealth owned by pensioners. The aggregate unmortgaged property wealth owned by the 65+ age group is around £1,100 billion (Institute of Actuaries, 2005a). An extensive study conducted by the Institute of Actuaries (ibid.) concluded that the UK equity release market could grow several times its current size over the next few years if the large mortgage providers enter the market. We expect equity release to become a core financial product, and an integral part of retirement provision for many.

Among various types of equity release products, fixed interest lifetime mortgages, or roll-up mortgages, are the most popular in the market today. Under a roll-up mortgage, the borrower is advanced a lump sum of money by the provider, and interest on the amount advanced is compounded at a fixed rate. The capital and interest are repaid from the property sale proceeds when the borrower dies, sells the house, or moves out of the house permanently. Of course, given that the value of the property when the loan is repaid is uncertain, shortfall in the home sales proceeds relative to the mortgage repayable is possible. However, most roll-up mortgages are equipped with a no-negative-equity-guarantee (NNEG) which protects the borrower by capping the redemption amount of the mortgage at the lesser of the face amount of the loan and the sale proceeds of the home.

The NNEG in roll-up mortgages is analogous to the guaranteed minimum death benefit (GMDB) offered in unit-linked insurance. In both guarantees, there is only limited diversification amongst each cohort of contracts, and this means that traditional deterministic valuation with some margins in the assumptions will be inadequate (see Hardy, 2003). Nevertheless, stochastic and financial economics methods of valuation are hampered by the complexity in modeling house price dynamics. Strong autocorrelation and varying volatility effects in house price returns have rendered the classical Black-Scholes formula, which assumes the returns on the underlying asset follow a geometric Brownian motion, inappropriate. In this study, we first search for a stochastic volatility process that can reasonably model house price returns. Under the identified stochastic volatility process, however, there exists more than one equivalent risk-neutral probability measure, and therefore the price of the NNEG is not unique. This phenomenon is often known as market incompleteness. The core part of this study is an investigation into the pricing formula, and the hedging and capital reserving strategies for the NNEG in such an incomplete market.

The remainder of this chapter is organized as follows. Section 2 provides a brief description of various types of equity release products available in the market. In Section 3, we discuss the risks associated with equity release and how they may be inter-related. We begin Section 4 with an in-depth analysis of house price dynamics. Then we value the NNEG and develop appropriate hedging and capital reserving strategies. Section 5 concludes this chapter.

5.2 Different Types of Equity Release Products

There are a wide variety of products that are currently available, or used to be available, in the UK equity release market. Below we provide a synopsis of the key features of each of these products.

- *Roll-up mortgages*
  In a roll-up mortgage, a lump sum is loaned to the homeowner. Interest on the loan rolls up and is added to the outstanding debt, which is cleared by the sale of the house when the homeowner dies, sells the house or moves out of the house permanently.

The roll-up rate is often fixed, but some providers offer variable interest schemes in which the roll-up rate is linked to a standard variable rate, usually with a cap.

- *Fixed-repayment life time mortgages*
  Fixed-repayment life time mortgages are similar to roll-up mortgages. However, the face amount of the loan is fixed regardless of the length of the loan, and is based on the life expectancy of the borrower when the loan is taken out.

- *Home Reversion Schemes*
  In a home reversion scheme, the homeowner sells his/her house to the provider in return for a lump sum, which depends on the value of the house and the life expectancy of the homeowner when the loan is taken out. The (original) homeowner is entitled to live in the house for life, but he/she is responsible for keeping the house well maintained and repaired.

- *Shared Appreciation Mortgages*
  In a shared appreciation mortgage (SAM), the bank advances a percentage of the value of the property. There is no interest payment required. The loan repayment is the original loan plus a multiple of the capital appreciation in the property value on the percentage advanced on the loan. SAMs used to be very successful, but they have been withdrawn due to the providers' inability to raise funds from long-term institutional investors (see Institute of Actuaries, 2001).

- *Home Income Plans*
  In a home income plan, the homeowner is offered a regular lifetime income instead of a lump sum. The income could be arranged on an increasing basis − typically 3% to 5% increase per annum, or linked to the retail price inflation index.

- *Drawdown mortgages*
  Drawdown mortgages are a variant of roll-up mortgages. Instead of taking out a lump sum, the homeowner had approval to draw down payments up to a predefined borrowing limit. The timing and amount of cash advances is at the homeowner's discretion.

The equity release market has seen a significant shift from reversion products to mortgage products over the past decade. In 2005, roll-up mortgages accounted for over 95% of the total equity release sales (Safe Home Income Plans, 2006a). In the following sections, we shall restrict ourselves to considering the issues regarding roll-up mortgages.

5.3 Risks Associated with Equity Release Mechanisms

ERMs expose product providers to a combination of demographic and financial risks. In the following we describe briefly the nature of these risks and discuss how they may be inter-related.

- *Mortality/longevity risk*
  Mortality losses may be incurred if the homeowners' mortality is either lighter or heavier than expected. Heavier mortality will shorten the duration of the roll-up, reducing the period for which a profit margin is earned. On the other hand, excess longevity will effectively increase the value of the NNEG through the extension of the "time to maturity" and the increase of the "exercise price" of the guarantee.

- *Morbidity risk*
  The nature of the providers' exposure to morbidity risk is similar to that of mortality risk.

- *Early-redemption risk*
  Early-redemptions may occur due to a change in personal circumstances, such as marriage and divorce, or due to a voluntary decision, such as remortgaging. Remortgaging is most likely to occur following a decline in interest rates when cheaper new mortgages are available in the market. This implies that in addition to the adverse impact on profitability, the providers are likely to suffer an interest rate loss if they have raised funds at a fixed rate of interest, or if they have a variable cost of capital and have swapped out their interest rate exposure. The providers may be able to mitigate remortgaging risk through the use of early-redemption penalties.

- *House price inflation risk*
  The NNEG can go "deep in the money" in a prolonged period of stagnant or very low house price growth.

- *Interest rate risk*
  The fixed roll-up rate may fall behind the providers' (floating rate) cost of funding. Exposures to such risk may be managed by entering into an interest rate swap.

- *Moral hazard risk*
  Given that the risk of house price inflation is borne by the provider, the borrower may be reluctant to maintain his/her home, particularly when the NNEG becomes "in the money". This is likely to result in a depreciation of the property value, reducing the profit of the provider.

- *Other risks*
  The providers are inevitably exposed to other risks, such as operational risk and reputational risk, that cannot be modeled easily.

Note that it is the NNEG that has exposed the product providers to the risks of (low) house price inflation and increased longevity. A better understanding of the NNEG is therefore of crucial importance in the risk management for equity release products. In particular, an implementation of appropriate hedging and capital reserving strategies for the NNEG may allow the providers to mitigate a significant portion of the demographic and financial risks associated with ERMs.

5.4 The No-Negative-Equity-Guarantee

5.4.1 Overview

Roll-up mortgages generally come with a NNEG. This means that, regardless of what happens to house prices or interest rates, the borrower will never be forced out of his/her home, but will be allowed to live there until he/she dies or goes into long-term care. Jon King, the chairman of Safe Home Income Plans (SHIP), offered an succinct explanation of the importance of NNEG to the equity release market:

> *"The SHIP no negative equity guarantee is a vital cornerstone of equity release products, providing customers with absolute peace of mind that whatever happens to property prices, their home is protected. There will be no cost to them or their estate if their property experiences negative equity, nor will they lose their right to live there. The guarantee, alongside all of the other SHIP product guidelines, is designed to protect a vulnerable category of consumer making a very important decision about a long term product, and to ensure that their experience of equity release is a happy one."* Jon King, in SHIP (2005).

The SHIP code of practice mandates members of SHIP to provide the NNEG feature in their equity release products. Also, since it is hard, if not impossible, to establish whether the borrower will be in a financial position that allows him/her to top up the potential shortfall, omission of the NNEG is practically unfeasible.

From the providers' point of view, provision of a NNEG is equivalent to writing the borrower a series of European put options on the mortgaged property. To illustrate, let us suppose that the amount of cash advanced is X_0 and that the borrower will have to repay the loan at an unknown future time T. The amount repayable (X_T) will be the sum of the principal and the interest accrued at a fixed roll-up rate, u, compounded continuously; that is, $X_T = X_0 \exp(uT)$. If house price inflation is strong enough so that at time T, the value of the mortgaged property (S_T) exceeds the amount repayable, then the provider will receive X_T without further financial obligations. However, in the unfortunate case that the property sale proceeds cannot fully cover the amount repayable, the provider will obtain only S_T, due to the NNEG. Equivalently speaking, the provider will receive a amount X_T plus the payoff from the NNEG:

$$-\max[X_T - S_T, 0], \tag{5.1}$$

which is precisely the payoff (from the writer's perspective) of an European put option with strike price X_T. Note that T is random when the mortgage is written. Therefore, in effect, by offering a NNEG in a roll-up mortgage, the provider has written a series of put options with different times to maturity. It is noteworthy that mortality improvement will result in a more valuable NNEG due to the increase in both the time to maturity[1] and the

[1]The value of an European put option does not necessarily increase with the time to maturity if the strike price is fixed. However, for a straight Black-Scholes framework, the value of a put option will be

strike price of the option.

The nature of the NNEG is identical to that of the guaranteed maturity death benefit (GMDB) feature, which is commonly offered in unit-linked insurance in Britain, segregated fund contracts in Canada, and variable annuities in the United States. The GMDB feature guarantees the policyholder a specific sum of money on death, regardless of the performance of the underlying index or fund. The guaranteed death benefit may be fixed, or may increase at a fixed rate, which is analogous to the roll-up rate in a NNEG. In both features, the financial risks involved cannot be diversified, since any change in the stock index or property market would affect the entire cohort of policies at the same time. The risk management for these features should be done by a stochastic or financial economics approach, rather than traditional deterministic methods that rely on the law of large numbers.

5.4.2 The Dynamics of House Price Returns

We mentioned that the NNEG can be regarded as a series of European put options with the mortgaged property being the underlying asset. However, unlike put options on listed stocks, the underlying asset of the NNEG is unique and is infrequently traded, leaving us no historical prices to base our analysis on. To deal with this problem, we resort to a consideration of house price indices that reflect the changes in residential property prices within a particular geographical region. Examples include the Halifax Bank House Price Index, the Nationwide House Price Index, the HM Land Registry data and the Investment Property Databank Property Index. We refer interested readers to Booth and Marcato (2004) for a comprehensive review of the pros and cons of these indices.

In this study, we use the Nationwide House-Price Index (Nationwide Building Society, 2006), which is available on a quarterly basis from 1952 Q4 to 2006 Q1 (213 observations), and is not seasonally adjusted. The index is derived from the average purchase price of the properties in Nationwide's own mortgage portfolio, which is composed of residential properties all over Britain, with a bias towards the south of England. To prevent the geographical bias from affecting the index, regional weightings in the index are based on

an increasing function of the time to maturity if the the strike price increases at a rate greater than the risk-free interest rate.

figures from the Department of the Environment, Transport and the Region, instead of the actual composition of Nationwide's portfolio. In contrast to valuation-based indices, the Nationwide House-Price index is solely based on purchase prices, and is independent of subjective valuations. In other words, we should expect the index to be free from the problem of valuation smoothing, and we should not be required to perform any de-smoothing procedure prior to modeling.

Nevertheless, readers should keep in mind that changes in the index and changes in the price of the actual underlying property are not likely to be identical, no matter how reliable the index is. This means that the providers are exposed to basis risk, which will be discussed in the section about hedging and capital reserving strategies.

The Institute of Actuaries (2005b) pointed out two important characteristics of the Nationwide House-Price Index:

1. there exists a (very) strong positive autocorrelation effect;

2. the volatility varies with time.

The first characteristic can be observed directly from the plot of the logarithmically transformed return series. In Figure 5.1, we observe that in the log return series, large changes tend to follow large changes, and small changes tend to follow small changes. This phenomenon is commonly known as volatility clustering. In addition, the log returns seem to be correlated with the changes in the volatility. Specifically, the volatility tends to rise in response to higher than expected returns and fall in response to lower than expected returns. This relationship is often referred to as the leverage effect. The presence of the leverage effect suggests that we should consider stochastic processes that permit the volatility to respond asymmetrically to positive and negative innovations.

The second characteristic can be confirmed by checking qualitatively for correlation in the log return series. Figure 5.2 displays the sample autocorrelation function (ACF) of the log returns, along with the upper and lower confidence bounds, based on the assumption that all autocorrelations are zero beyond lag zero. We observe that the correlation is strong in magnitude, and that the correlation effect is persistent: the sample ACF does not cut off until the eighth lag.

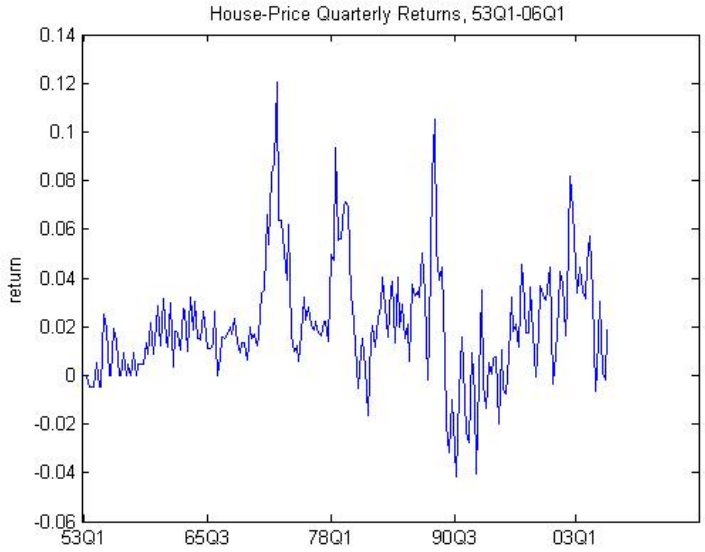


Fig. 5.1. House Price Quarterly Returns, 1953 Q1 - 2006 Q1.

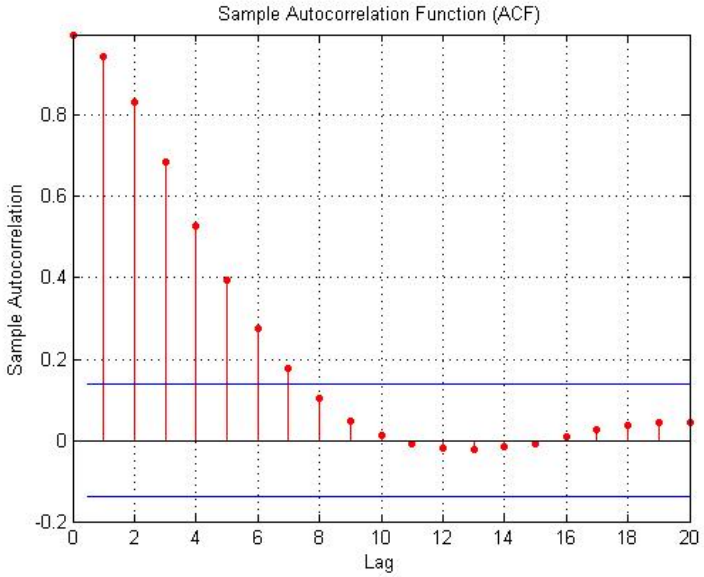


Fig. 5.2. Sample ACF of the House Price Returns.

Unfortunately, there is little discussion on the subject of house price index modeling in the actuarial literature. In Britain, a few stochastic real estate investment models have been proposed for modeling commercial real estate performance. These include the Wilkie (1995) model that describes the logarithm of the property yield, $P(t)$, as a function of a mean value, PMU, a first-order auto-regressive parameter, PA, and a random term, $PSD.\mathrm{N}(0,1)$:

$$\ln P(t) = \ln PMU + PA.(\ln P(t-1) - \ln PMU) + PSD.\mathrm{N}(0,1). \tag{5.2}$$

Booth and Marcato (2004) extended the Wilkie model by introducing several parameters that relate the property yield to equity and bond yields. In the US, Calhoun (1996) expressed the logarithm of the price of an individual house by the sum of (1) a market price index, (2) a random term that describes the relationship between the individual house price and the market price index and (3) another random term that represents the idiosyncratic features of individual houses.

Unfortunately, the existing models do not allow for the varying volatility and leverage effects that are seen in the Nationwide House-Price Index. An alternative model specification is therefore necessary.

We first use the Box-Jenkins (1976) approach to identify an appropriate linear time-series process for the log return series. An ARMA(1,3) model gives an excellent fit. Figure 5.3 plots the sample ACF of the residuals. We observe that the residuals are uncorrelated, indicating that the fitted ARMA(1,3) model is able to capture most of the autocorrelation effect in the log return series. Figure 5.4 plots the sample ACF of the squared residuals. We observe that the squared residuals are significantly correlated. This observation suggests that, even though the residuals are largely uncorrelated, the variance process exhibits some correlations. This phenomenon is known as conditional heteroscedasticity in the time-series literature. We can further quantify the preceding qualitative checks by using formal hypothesis tests to test for the null hypothesis that the squared innovations are not serially correlated. The Ljung-Box Portmanteau test statistic[2] (Ljung and Box, 1978) gives a p-

[2]The Ljung-Box Portmanteau test statistic tests for the null hypothesis $\mathrm{H}_0 : \rho_1 = \rho_2 = ... = \rho_m = 0$ against the alternative hypothesis $\mathrm{H}_1 : \rho_i \neq 0$ for some $i \in \{1, 2, ..., m\}$, where ρ_i is lag-i autocorrelation in the raw return series. The selection of m may affect the performance of the test statistic. Tsay (2002)

value of 0.0015. This p-value supports the conclusion that conditional heteroscedasticity exists.

recommended the use of $m \approx \ln(T)$, where T is the total number of observations. Following Tsay's recommendation, we use $m = 5 \approx \ln(213)$.

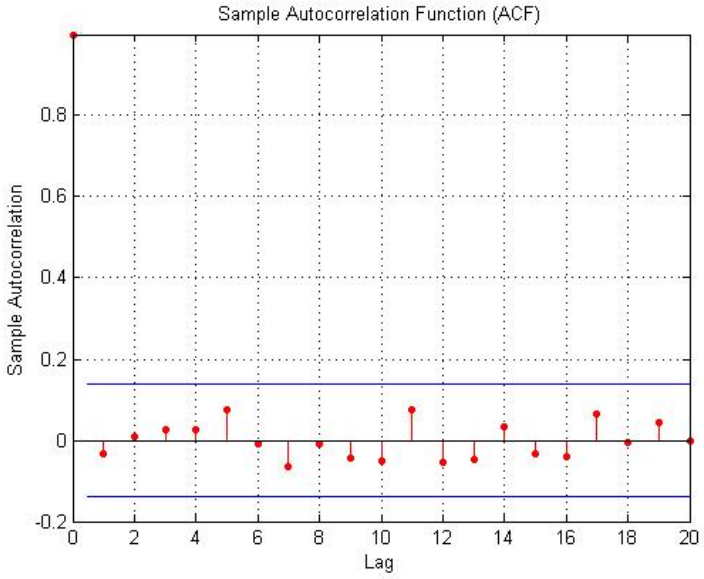


Fig. 5.3. Sample ACF of the residuals of the fitted ARMA(1,3) model.

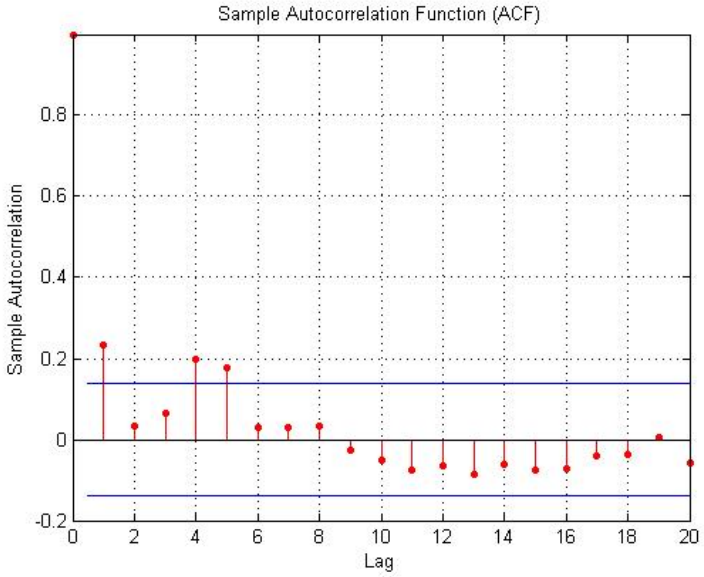


Fig. 5.4. Sample ACF of the squared residuals of the fitted ARMA(1,3) model.

The above analysis indicates that a conditional heteroscedastic time-series model is appropriate for modeling the house-price index. A conditional heteroscedastic model consists of one sub-model for the conditional mean process and another for the conditional variance process. The ARMA(R,M) conditional mean model is applicable to all conditional heteroscedastic models. Let $Y_t = \ln(S_t/S_{t-1})$, where S_t is the value of the house price index at time t. The ARMA(R,M) conditional mean model for the log returns can be expressed as

$$Y_t = c + \sum_{i=1}^{R} \phi_i Y_{t-i} - \sum_{j=1}^{M} \theta_j a_{t-j} + a_t,$$

where c is the drift term. The coefficients ϕ_i's and θ_j's govern the relative emphasis on the previous Y_t's and a_t's, respectively. We require the polynomials $1-\phi_1 B-\phi_2 B^2-...-\phi_R B^R$ and $1-\theta_1 B-\theta_2 B^2-...-\theta_M B^M$ have no common factors, and have zeros lying outside the unit circle to ensure stationarity. We assume that the innovations, a_t's, are normally distributed with mean zero and variance h_t. Given the information up to and including time $t-1$, the variance of Y_t is simply h_t. The serial dependence of h_t is then captured by a conditional variance model. Typical conditional variance models include:

- The ARCH(Q) model
  The ARCH(Q) model proposed by Engle (1982) is the simplest conditional variance model. It assumes that

  $$h_t = k + \sum_{j=1}^{Q} \beta_j a_{t-j}^2. \tag{5.3}$$

  The basic idea of ARCH models is that the serial dependence of h_t can be described by a simple quadratic function of the previous innovations. The weights on the previous innovations are determined by the coefficients β_j's. We require that the drift term, k, is strictly positive and that $\beta_j \geq 0$ for $j = 1, 2, ..., Q$ to ensure that the conditional variance is always positive. Also, β_j, $j = 1, 2, ..., Q$, must satisfy some regularity conditions to ensure that the unconditional variance of a_t is finite.

- The GARCH(P,Q) model
  Although ARCH models are simple, they often require many parameters to describe

the evolution of h_t adequately. To solve this problem, Bollerslev (1986) extended the ARCH(Q) to the GARCH(P,Q) model, which is defined as follows:

$$h_t = k + \sum_{i=1}^{P} \alpha_i h_{t-i} + \sum_{j=1}^{Q} \beta_j a_{t-j}. \tag{5.4}$$

The GARCH(P,Q) model allows the conditional variance to be dependent on both the previous innovations and the previous conditional variances. This flexibility allows us to use fewer parameters for adequate modeling. For instance, a volatility process that requires an ARCH(9) model may be well described by a GARCH(1,2) model. The dependency on the previous conditional variances is governed by the coefficient α_i's. As in the ARCH model, we require $k > 0$, $\alpha_i \geq 0$ for $i = 1, 2, ..., P$ and $\beta_i \geq 0$ for $j = 1, 2, ..., Q$ to ensure that $h_t > 0$ for all t. We also require $\sum_{i=1}^{\max\{P,Q\}} (\alpha_i + \beta_i) < 1$ so that the unconditional variance of a_t is finite.

- The EGARCH(P,Q) model
  To overcome some weaknesses of the GARCH model in handling financial time-series, Nelson (1991) proposed the exponential GARCH (EGARCH) model, which is based on the weighted innovation $g(\epsilon_t) = \kappa\epsilon_t + \lambda[|\epsilon_t| - \mathrm{E}(|\epsilon_t|)]$, where $\epsilon_t = a_t/\sqrt{h_t}$, and κ and λ are real constants. Since both ϵ_t and $|\epsilon_t| - \mathrm{E}(|\epsilon_t|)$ are zero-mean sequences with continuous distributions, $\mathrm{E}[g(\epsilon_t)] = 0$. Also, it can be shown that $\mathrm{E}(|\epsilon_t|) = \sqrt{2/\pi}$. Having defined the weighted innovation, we can express the EGARCH(P,Q) model as

$$\ln(h_t) = k + \sum_{i=1}^{P} \alpha_i \ln(h_{t-i}) + \sum_{j=1}^{Q} \beta_j g(\epsilon_t). \tag{5.5}$$

  The model uses the usual ARMA parameterization to describe the evolution of $\ln(h_t)$. Analogous to the ARMA model, we require the polynomials $1 - \alpha_1 B - \alpha_2 B^2 - ... - \alpha_P B^P$ and $1 + \beta_1 B + \beta_2 B^2 + ... + \beta_Q B^Q$ have no common factors and have zeros lying outside the unit circle. The EGARCH model differs from the GARCH in a number of ways. First, it models the logarithm of conditional variance to relax the constraints that are required in GARCH to ensure the positiveness of h_t. Second, it allows for the leverage effect. This property can be visualized immediately from the following

reparameterization:

$$\log(h_t) = k + \sum_{i=1}^{P} \alpha_i \log(h_{t-i}) + \sum_{j=1}^{Q} \beta_j^* \left[|\epsilon_{t-j}| - \mathrm{E}(|\epsilon_{t-j}|)\right] + \sum_{j=1}^{Q} \gamma_j \epsilon_{t-j}, \qquad (5.6)$$

where $\beta_j^* = \lambda\beta_j$ and $\gamma_j = \kappa\beta_j$. The leverage parameters γ_j's explicitly enable the conditional variance to respond asymmetrically to positive and negative innovations.

A detailed description of these conditional heteroscedastic models can be found in Tsay (2002). We choose the EGARCH model to take advantage of its ability to handle the leverage effect found in the dynamics of the house price index. Specifying the order of an EGARCH model is not easy. However, lower order EGARCH models are sufficient in most applications, and we use the simplest form of EGARCH(1,1) for parsimony. Very often, the use of a volatility model can reduce the ARMA order, since the conditional variance model may explain away part of the autocorrelation effect in the historical returns. Nevertheless, we found that an ARMA(1,3) representation is still necessary for an adequate fit.

Table 5.1 shows that all parameters of the jointly estimated ARMA(1,3)-EGARCH(1,1) model are highly significant. One can check the adequacy of a fitted EGARCH model by examining the series of standardized innovations, $\tilde{a}_t = a_t/\sqrt{h_t}$. The sample ACF of $\{\tilde{a}_t\}$ shown in Figure 5.5 indicates that the ARMA-EGARCH model does an excellent job in modeling the autocorrelation effect in the historical returns. Also, the comparison of the sample ACF of $\{\tilde{a}_t^2\}$ shown in Figure 5.6 to that of the squared innovations prior to fitting the conditional variance model suggests that the EGARCH representation sufficiently explains the heteroscedasticity in the raw returns. Again, we can quantify the preceding qualitative checks by using the Ljung-Box Portmanteau test. The Ljung-Box test statistic for the standardized shocks $\{\tilde{a}_t\}$ gives a p-value of 0.3877 and that of $\{\tilde{a}_t^2\}$ gives a p-value of 0.8786. These p-values further confirm the explanatory power of the fitted ARMA-EGARCH model.

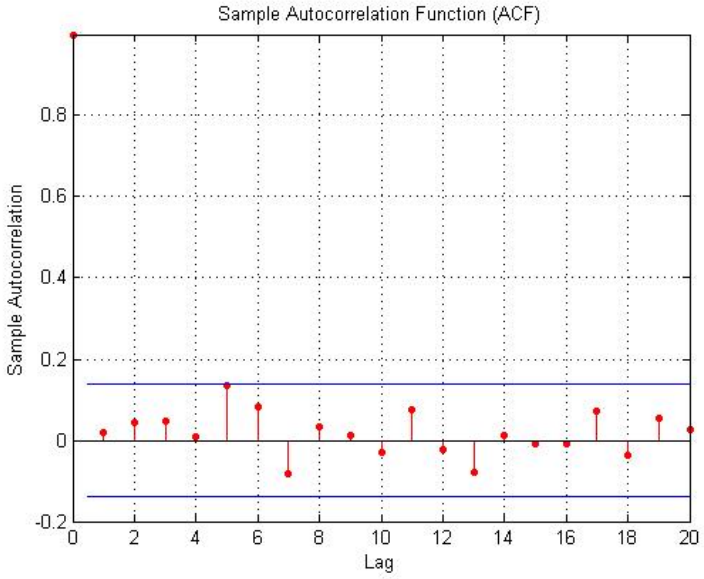


Fig. 5.5. Sample ACF of the standardized innovations of the fitted ARMA(1,3)-EGARCH(1,1) model.

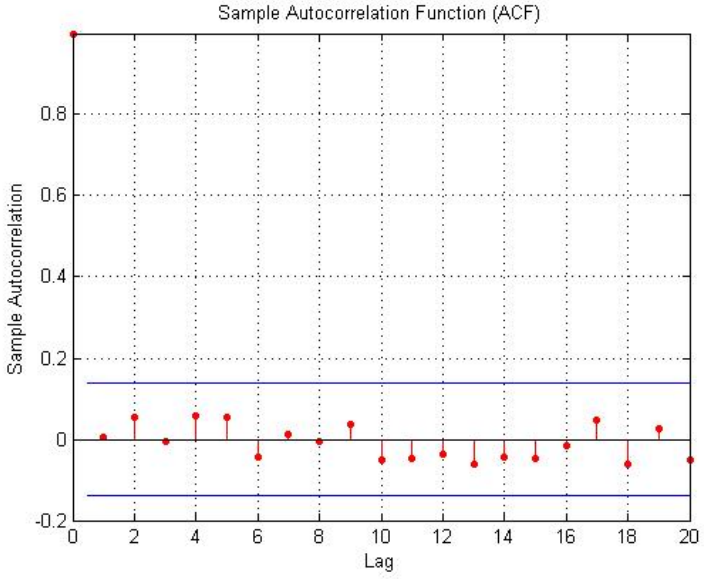


Fig. 5.6. Sample ACF of the squared standardized innovations of the fitted ARMA(1,3)-EGARCH(1,1) model.

Parameter	Value	Standard Error	t-statistic
c	0.02278	0.00566	4.02550
ϕ_1	0.66736	0.05766	11.5732
θ_1	0.82166	0.06170	13.3169
θ_2	0.79242	0.06732	11.7710
θ_3	0.68304	0.04813	14.1912
k	-1.20180	0.39607	-3.03440
α_1	0.84939	0.05004	16.9748
β_1^*	0.36924	0.09052	4.07920
γ_1	0.17941	0.06222	2.88340

Table 5.1. Parameter estimates of the fitted ARMA(1,3)-EGARCH(1,1) model.

5.4.3 Identification of an Equivalent Martingale Measure

In Section 5.4.1, we demonstrated the equivalence between the payoff function of a NNEG and that of an European put option. Therefore, a natural starting point for placing a value on the NNEG is to price the NNEG using the classical Black-Scholes option pricing methodology. A key assumption in the Black-Scholes formula is that the returns on the underlying asset follow a geometric Brownian motion (GBM). Under a GBM, the returns have a constant volatility over time, and are Markovian; that is, future returns depend only on the current return, but not on previous returns. These properties imply that the GBM assumption is not applicable to the house price returns in which the effects of autocorrelation and volatility clustering are very strong. While the ARMA-EGARCH model is more appropriate for modeling the dynamics of the house price index, the use of the ARMA-EGARCH model would lead to market incompleteness.

Market completeness refers to the situation that the payoff of an option can be obtained as the terminal value of a self-financing portfolio. Hence, by the principle of no arbitrage, the price of the option at any time before the maturity date must be the value of the replicating portfolio at that time. Equivalently speaking, under a complete market, there

exists one and only one equivalent martingale (risk-neutral) probability measure, and the unique price of any option is given by its expected discounted payoff at the maturity date under the martingale measure. On the other hand, if a market is incomplete, there exists more than one equivalent martingale measure, and this implies that there are a range of no-arbitrage prices for an option. Therefore, one critical issue in this study is to identify an equivalent martingale that gives an economically consistent and justifiable price for the NNEG.

The identification of a unique equivalent martingale measure under an incomplete market can be performed in various ways. Rubinstein (1976) adopted the equilibrium or utility maximization-based approach. Föllmer and Sodermann (1986), Föllmer and Scheweizer (1991) and Schweizer (1996) identified a unique equivalent measure by minimizing the variance of the hedging loss. This method was later applied to the valuation of equity indexed annuities by Lin and Tan (2003). Gerber and Shiu (1994) suggested a feasible way to choose an equivalent martingale measure using an Esscher transform, introduced by Esscher (1932). The option price chosen by the Esscher transform can be justified by maximizing the expected power utility of an economic agent.

In recent years, researchers have begun to study conditional heteroskedastic models for option pricing due to their superior performance in describing various types of asset returns. Duan (1995) was the first to provide a solid theoretical foundation for option valuation in the GARCH framework. Duan's results are primarily based on the utility maximization approach, and the concept of a locally risk neutral relationship (LRNVR) in which the conditional variances under the physical and risk-neutral measures are equal. Siu et al. (2004) utilized the concept of a conditional Esscher transform introduced by Bühlmann et al. (1996) to identify an equivalent martingale measure when asset returns follow a GARCH model. The advantage of using a conditional Esscher transform is that different distributions, such as a translated gamma, can be assumed for the GARCH innovations. Under the conditional normality assumption for the GARCH innovations, their pricing result is consistent with that of Duan. Given that the statistical distribution of the cumulative return is unknown for all GARCH specifications, analytical formulae for pricing European options are not available. Researchers have tried to speed up the valuation of European options under conditional heteroskedastic models by developing analytical ap-

proximations. Heston and Nandi (2000) developed a numerical technique to obtain prices of European options when the dynamics of the conditional variance are driven by a GARCH process. Duan et al. (1999, 2006) developed analytical European option pricing formulae when the dynamics of the conditional variance are governed by a GARCH, EGARCH or GJR-GARCH (Glosten et al., 1993) process.

Nevertheless, in previous studies of option pricing under GARCH, the conditional mean was assumed to follow a GARCH-in-mean specification, instead of an ARMA process that we found to be highly suitable for describing the autocorrelation effects in the house price returns. To overcome this problem, we extend the work of Siu et al. (2004) slightly to obtain an equivalent martingale measure when the returns in the physical measure are governed by an ARMA-GARCH process.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, where $\mathbb{P}$ is the data-generating probability measure. We assume that all economic activities take place at time t, where $t \in \mathcal{T}$ and $\mathcal{T}$ is the time index set $\{0, 1, ..., T\}$. Also, let $\Phi = \{\Phi_t\}_{t\in\mathcal{T}}$ be the natural filtration such that, for each $t \in \mathcal{T}$, Φ_t contains all market information up to and including time t, and that $\Phi_T = \mathcal{F}$.

Let us assume that, under $\mathbb{P}$, the log house price returns, $\{Y_t\}_{t\in\mathcal{T}}$, follow a ARMA(R,M)-EGARCH(P,Q) process with Gaussian innovations. It follows that, under $\mathbb{P}$, $Y_t|\Phi_{t-1} \sim \mathrm{N}(\mu_t, h_t)$, where $\mu_t = c + \sum_{i=1}^{R} \phi_i Y_{t-i} - \sum_{j=1}^{M} \theta_j a_{t-j} + a_t$, and h_t is specified by equation (5.6). Note that μ_t and h_t are non-random given Φ_{t-1}.

Following Bühlmann et al. (1996), we define a sequence $\{Z_t\}_{t\in\mathcal{T}}$ with $Z_0 = 1$, and for $t \geq 1$,

$$Z_t = \prod_{k=1}^{t} \frac{\exp(\lambda_k Y_k)}{\mathrm{E}(\exp(\lambda_k Y_k)|\Phi_{k-1})}, \tag{5.7}$$

for some constants $\lambda_1, \lambda_2, ..., \lambda_t$. It can be shown easily that $\{Z_t\}_{t\in\mathcal{T}}$ is a martingale.

Let $\mathbb{P}_t$ be the restriction of the measure $\mathbb{P}$ on the information Φ_t, for each $t \in \mathcal{T}\backslash\{0\}$, where $\mathbb{P}_T = \mathbb{P}$. The fact that $\{Z_t\}_{t\in\mathcal{T}}$ is a martingale allows us to construct a family of measures $\{\tilde{\mathbb{P}}_t\}_{t\in\mathcal{T}}$ such that $d\tilde{\mathbb{P}}_t = Z_t d\mathbb{P}_t$ and $\tilde{\mathbb{P}}_t = \tilde{\mathbb{P}}_{t+1}|\Phi_t$, and a probability measure $\tilde{\mathbb{P}} = \tilde{\mathbb{P}}_T$ on the sample space $(\Omega, \mathcal{F})$. Let A be an open interval on the real line, and $I_A(Y_t)$

be the indicator function for the event $\{Y_t \in A\}$. The conditional distribution

$$\tilde{\mathbb{P}}_t(Y_t \in A|\Phi_{t-1}) = \mathrm{E}_{\mathbb{P}_t}\left[I_A(Y_t)\frac{\exp(\lambda_t Y_t)}{\mathrm{E}_{\mathbb{P}_t}(\exp(\lambda_t Y_t))}|\Phi_{t-1}\right] \tag{5.8}$$

is called the conditional Esscher transform. Substituting $(-\infty, y]$, where y is a real number, for A in equation (5.8), we obtain the following distribution function of Y_t given Φ_{t-1} under the measure $\tilde{\mathbb{P}}_t$:

$$F_{\tilde{\mathbb{P}}_t}(y; \lambda_t|\Phi_{t-1}) = \frac{\int_{-\infty}^{y}\exp(\lambda_t x)dF_{\mathbb{P}_t}(x|\Phi_{t-1})}{\mathrm{E}_{\mathbb{P}_t}(\exp(\lambda Y_t)|\Phi_{t-1})}. \tag{5.9}$$

It immediately follows that the moment generating function of Y_t given Φ_{t-1} under the measure $\tilde{\mathbb{P}}_t$ is given by

$$\mathrm{E}_{\tilde{\mathbb{P}}_t}[\exp(Y_t z); \lambda_t|\Phi_{t-1}] = \frac{\mathrm{E}_{\mathbb{P}_t}[\exp(Y_t(z+\lambda_t))|\Phi_{t-1}]}{\mathrm{E}_{\mathbb{P}_t}[\exp(Y_t(\lambda_t))|\Phi_{t-1}]}. \tag{5.10}$$

Using equation (5.10) and the fact that $Y_t|\Phi_{t-1} \sim \mathrm{N}(\mu_t, h_t)$, we have

$$\mathrm{E}_{\tilde{\mathbb{P}}_t}[\exp(Y_t z); \lambda_t|\Phi_{t-1}] = \exp\left(\left(\mu_t + h_t\lambda_t\right)z + \frac{1}{2}h_t z^2\right). \tag{5.11}$$

To construct a risk-neutral probability measure $\mathbb{Q}$ which is equivalent to the data-generating measure $\mathbb{P}$ on the sample space $(\Omega, \mathcal{F})$, we choose a sequence of conditional Esscher parameters $\{\lambda_t^q\}_{t\in\mathcal{T}\setminus\{0\}}$ such that the following set of equations are satisfied:

$$\mathrm{E}_{\tilde{\mathbb{P}}_t}[\exp(Y_t); \lambda_t^q|\Phi_{t-1}] = \exp(r), \quad t \in \mathcal{T}\setminus\{0\}, \tag{5.12}$$

where r is the risk-free interest rate. We refer interested readers to Bühlmann et al. (1996) for a detailed proof. Substituting λ_t^q into equation (5.11), we obtain

$$\mathrm{E}_{\mathbb{Q}_t}[\exp(Y_t z)|\Phi_{t-1}] = \exp\left(\left(r - \frac{1}{2}h_t\right)z + \frac{1}{2}h_t^2\right), \tag{5.13}$$

where $\mathbb{Q}_t$ is the restriction of $\mathbb{Q}$ on the information Φ_t, and $\mathbb{Q}_T = \mathbb{Q}$. Hence, under $\mathbb{Q}$, $Y_t|\Phi_{t-1} \sim \mathrm{N}(r - 1/2h_t, h_t)$, or equivalently, $Y_t = \mu_t + a_t$ and $a_t \sim \mathrm{N}(-\mu_t + r - 1/2h_t, h_t)$. In other words, the dynamics of Y_t under the equivalent risk-neutral measure are the same as that under the physical measure, except that the distribution of a_t is translated by an amount of $-\mu_t + r - 1/2h_t$.

Finally, under the equivalent martingale measure $\mathbb{Q}$, the time t, $t \in \mathcal{T}$, the price of an option that gives a payoff V_T at maturity is given by

$$V_t = \mathrm{E}_{\mathbb{Q}}[\exp(-r(T-t))V_T|\Phi_t]. \tag{5.14}$$

5.4.4 Pricing the NNEG

The underlying property in an NNEG is analogous to a dividend paying stock. This is because the total return from a property can be viewed as the sum of its capital growth and its rental income net of insurance and maintenance costs. Such rental income is sometimes referred to as imputed rent by economists. To simplify the analysis, we assume that the rental yield of the underlying property is constant over time.

Recall that in offering a NNEG, the provider has written a series of European put options with different times to maturity. We let $P(t, S_0, X_0, u, r, g)$ be the value of each of these options at time zero, where t is the time to maturity, S_0 is the value of the mortgaged property at time zero, X_0 is the amount of loan advanced, u is the roll-up rate, r is the risk-free interest rate, and g is the rental yield. Then, using discrete annual time steps, the value of a NNEG in a portfolio of roll-up mortgages sold to persons who are aged x at inception can be expressed as

$$V_{NNEG} = \sum_{k=0}^{\omega-x-1} {}_kp_x q_{x+k} P\left(k + \frac{1}{2} + \delta, S_0, X_0, u, r, g\right), \tag{5.15}$$

where δ is the average delay in time from the point of home exit until the actual sale of the property, ω is the highest attained age, ${}_kp_x$ is the probability that a borrower who is aged x at inception survives to age $x+k$, and q_{x+k} is the probability that a borrower who is aged x at inception dies in the time interval k to $k+1$ given that he/she has survived to age $x+k$. We assume that all home exits occur at mid-year. Therefore, the time to maturity is $k + 1/2 + \delta$.

Furthermore, if we assume that the house price returns follow a ARMA-GARCH process, then $P\left(k + \frac{1}{2} + \delta, S_0, X_0, u, g\right)$ is given by

$$\mathrm{e}^{-r\left(k+\frac{1}{2}+\delta\right)}\mathrm{E}_{\mathbb{Q}}\left[\left(X_0\mathrm{e}^{u\left(k+\frac{1}{2}+\delta\right)} - S_0\mathrm{e}^{\left(-g\left(k+\frac{1}{2}+\delta\right)+\prod_{t=0}^{4\left(k+\frac{1}{2}+\delta\right)} Y_t\right)}, 0\right)^+\right], \tag{5.16}$$

where $\mathbb{Q}$ is the equivalent martingale measure we identified in Section 5.4.3. The expectation in expression (5.16) can be evaluated numerically using Monte Carlo simulations.

The key assumptions that we have made for the purpose of illustration are as follows:

- The average delay in time from the point of home exit until the actual sale of the property is six months.
- The risk-free interest rate is 4.75% per annum, compounded continuously.
- The roll-up rate is 7.50% per annum, compounded continuously.
- The rental yield is 2% per annum, compounded continuously.
- The amount of cash advanced is £30,000.
- The following two sets of life tables are used:

  1. Period life tables derived from English and Welsh population mortality in 2003.
  2. Cohort life tables constructed from the Lee-Carter projection of English and Welsh population mortality.

  For both life tables, mortality rates for ages 100 and above are estimated by the threshold life table technique suggested in Chapter 4. The required data are provided by the Human Mortality Database (2005).

- Death is the only mode of decrement. The impact of long-term care incidence on the NNEG values are discussed later in Section 5.4.5.
- The initial house values are equal to the minimum permitted values shown in Table 5.2. These values are broadly in line with the products that are currently available in the UK equity release market.

Age at inception	60	70	80	90
Initial house value	£176,500	£111,000	£81,000	£60,000

Table 5.2. Minimum initial house values.

The resultant NNEG values, expressed as a percentage of the cash advanced, are displayed in Tables 5.3 to 5.6. For readers' information, we compute also the NNEG values using the classical Black-Scholes methodology by replacing $P\left(k+\frac{1}{2}+\delta, S_0, X_0, u, g\right)$ in equation (5.15) with

$$X_0 \mathrm{e}^{\left((u-r)\left(k+\frac{1}{2}+\delta\right)\right)} \mathrm{N}(-d_2) - S_0 \mathrm{e}^{\left(-g\left(k+\frac{1}{2}+\delta\right)\right)} \mathrm{N}(-d_1), \tag{5.17}$$

where $d_1 = \frac{\ln\left(\frac{S_0}{X_0}\right)+\left(r-g-u+\frac{\sigma^2}{2}\right)\left(k+\frac{1}{2}+\delta\right)}{\sigma\sqrt{k+\frac{1}{2}+\delta}}$, and $d_2 = d_1 - \sigma\sqrt{k+\frac{1}{2}+\delta}$.

Model \ Age at inception	60	70	80	90
EGARCH	13.27%	7.98%	3.02%	1.25%
GBM	9.83%	6.13%	2.51%	1.17%

Table 5.3. NNEG costs as a percentage of the cash advanced, period life table (2003), male.

Model \ Age at inception	60	70	80	90
EGARCH	19.09%	10.22%	3.63%	1.43%
GBM	13.54%	7.57%	2.96%	1.28%

Table 5.4. NNEG costs as a percentage of the cash advanced, projected cohort life table, male.

Model \ Age at inception	60	70	80	90
EGARCH	23.97%	14.75%	5.71%	2.27%
GBM	17.03%	10.73%	4.44%	1.87%

Table 5.5. NNEG costs as a percentage of the cash advanced, period life table (2003), female.

Model \ Age at inception	60	70	80	90
EGARCH	33.72%	19.27%	7.02%	2.58%
GBM	23.92%	13.94%	5.37%	2.09%

Table 5.6. NNEG costs as a percentage of the cash advanced, projected cohort life table, female.

Note that each term on the right hand side of equation (5.15) reflects the relative importance of the age at death $(x + k)$ to the aggregate cost of the NNEG. Figures 5.7 and 5.8 illustrate this relationship for a NNEG written to borrowers who were aged 60 and 80 at inception, respectively. Even at old ages, where the mortality rate is very low, the contribution is significant as the option price for the longer lived is so high. Therefore, the tail of the assumed survival distribution is highly crucial to the valuation of a NNEG. In particular, if the assumed life table is closed too early, say, at age 100, the NNEG could be seriously underpriced.

Alternatively, we can express the NNEG values as a yield cost per annum. Let y be the annual yield cost, compounded continuously. If the option expires at a future time t, then the present value of the NNEG margin is given by

$$M(t, X_0, u, r, y) = e^{-rt} X_0 \left(e^{ut} - e^{(u-y)t} \right). \tag{5.18}$$

The annual yield cost y is derived so as to produce the expected present value of the NNEG margin equal to the monetary cost of the NNEG; that is, we solve the following equation for y:

$$\sum_{k=0}^{\omega-x-1} {}_k p_x q_{x+k} \left[P\left(k + \frac{1}{2} + \delta, S_0, X_0, u, r, g \right) - M\left(k + \frac{1}{2} + \delta, X_0, u, r, y \right) \right] = 0. \tag{5.19}$$

Tables 5.7 to 5.10 show the NNEG values, expressed as an annual yield cost, for different genders and ages at inception.

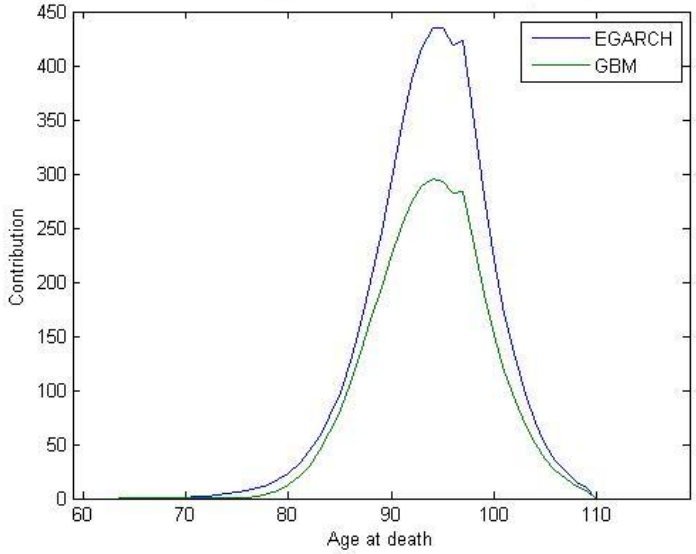


Fig. 5.7. Contribution to the aggregate NNEG cost at different ages at death, male, aged 60 at inception.

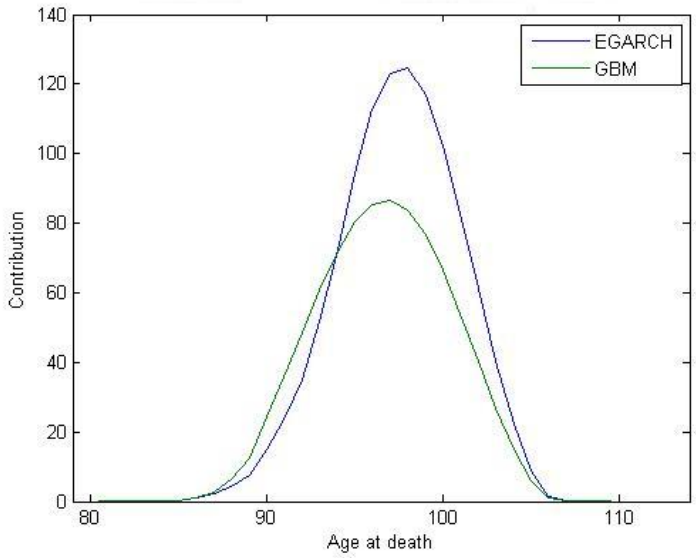


Fig. 5.8. Contribution to the aggregate NNEG cost at different ages at death, male, aged 80 at inception.

Model/Age	60	70	80	90
EGARCH	0.33%	0.37%	0.28%	0.25%
GBM	0.24%	0.28%	0.24%	0.23%

Table 5.7. NNEG costs as an annual yield cost, period life table (2003), male.

Model/Age	60	70	80	90
EGARCH	0.43%	0.46%	0.34%	0.28%
GBM	0.30%	0.33%	0.27%	0.25%

Table 5.8. NNEG costs as an annual yield cost, projected cohort life table, male.

Model/Age	60	70	80	90
EGARCH	0.48%	0.55%	0.43%	0.38%
GBM	0.33%	0.40%	0.33%	0.31%

Table 5.9. NNEG costs as an annual yield cost, period life table (2003), female.

Model/Age	60	70	80	90
EGARCH	0.62%	0.68%	0.51%	0.41%
GBM	0.42%	0.47%	0.38%	0.33%

Table 5.10. NNEG costs as an annual yield cost, projected cohort life table, female.

5.4.5 Modeling Long-Term Care Assumptions

In Section 5.4.4, all numerical illustrations are based on the assumption that death is the only mode of decrement. This section examines the impact of another mode of decrement − long-term care − on the NNEG prices. Long-term care incidence and mortality have strong interactions; for instance, people moving into a residential long-term care facility are generally less healthy that those remaining in their own homes. Hence, the incorporation of long-term care incidence into the pricing framework requires an additional model that quantifies such interactions. This objective may be fulfilled by using a multiple state model.

Our multiple state model can be represented conveniently by the diagram in Figure 5.9. The boxes represent the three possible states that a borrower can be in at any time and the arrows indicate the possible transitions between these states. We interpret being in state 1 as living in the mortgaged property, being in state 2 as living in a residential long-term care facility, and being in state 3 as dead. Transfer to either one of the other two states is possible for states 1 and 2, but not for state 3; that is, state 3 is absorbing. Furthermore, we assume that transitions can occur at most once in a year.

More precisely, we let $\{S(x); x = 60, ..., \omega - 1\}$[3] be a discrete time, time inhomogeneous Markov chain with a finite state space $\{1, 2, 3\}$. We interpret the statement "$S(x) = 1$", for example, to mean "the borrower is in state 1 at age x". We let

$$p_x^{gh} = P[S(x+1) = h | S(x) = g] \qquad g, h = 1, 2, 3; \quad x = 60, ..., \omega - 1. \tag{5.20}$$

The process is time (age) inhomogeneous since we assume p_x^{gh} depends on x, the age of the borrower. Also, the process is Markov since we assume the value of p_x^{gh} is unchanged if we are given any information about the behavior of the process before age x. By definition, $p_x^{31} = p_x^{32} = 0$ for all x; and also,

$$\sum_{h=1}^{3} p_x^{gh} = 1 \qquad g = 1, 2, 3; \quad x = 60, ..., \omega - 1. \tag{5.21}$$

As a result, for each age, the effective number of parameters is four.

In the ideal case, the free parameters should be estimated by using longitudinal data from extensive household surveys that follow the same cohort of lives and record information on disabilities. Unfortunately, such longitudinal data are not available in Britain.

[3] In modeling the NNEG, we require transition probabilities for age 60 and above only.

Previous mathematical analyses of the long-term care demand in the UK are based on either some small scale household surveys (e.g., Scott et al., 2001) or some extensive long-term care studies conducted in other countries (e.g., Rickazen and Walsh. 2002). The former approach may lead to high sampling error in estimating the transition rates, while the latter approach may not truly reflect the situation in Britain. To avoid these problems, we adopt an alternative method that makes use of the following information:

1. period life tables for the English and Welsh population in 2001;
2. population counts for the English and Welsh population on 1 January 2001;
3. number of residents (non-staff) in communal establishments in England and Wales in 2001; these numbers are based on the 2001 census in England and Wales.

Items (1) and (2) are provided by the Human Mortality Database (2005), while item (3) is provided by the Office of National Statistics (2007). Figures in item (3) are given by age groups. To compute the number of residents in communal establishments by single year of age, we disaggregate the raw numbers by the method proposed by Boot et al. (1967). Then we estimate the age-specific long-term care prevalence rates by dividing, for each age, the number of residents in communal establishments by the population count.

Let us define the following notation:

- l_x: the total number of people aged x last birthday on 1 January 2001;
- h_x: the number of people aged x last birthday on 1 January 2001 not living in long-term care facilities;
- c_x: the number of people aged x last birthday on 1 January 2001 living in long-term care facilities;
- q_x: the probability of death between ages x and $x+1$ in 2001, conditional on survival to age x;
- γ_x: the long-term care prevalence rate at age x in 2001.

Assuming that the population is closed in 2001 and that the long-term care prevalence rates remain unchanged through 2002, we have the following system of linear equations for $x = 60, ..., \omega - 1$:

$$h_x p_x^{13} + c_x p_x^{23} = l_x q_x; \tag{5.22}$$

$$h_x p_x^{12} + c_x (1 - p_x^{21} - p_x^{23}) = l_x (1 - q_x)\gamma_{x+1}; \tag{5.23}$$

$$h_x (1 - p_x^{12} - p_x^{13}) + c_x p_x^{21} = l_x (1 - q_x)(1 - \gamma_{x+1}), \tag{5.24}$$

where $\gamma_\omega = 0$. It is easy to see that one of above equations is redundant. Therefore, we require two more equations concerning the p_x^{gh}'s to obtain a unique solution. First, we employ the Rickayzen and Walsh (2002) model for extra mortality for different levels of OPCS (Office of Population, Census and Surveys) disability categories. The model expresses the death probability for someone aged x, $x \geq 50$, in disability category n, $n = 1, 2, ..., 10$, as

$$q_x + \frac{0.20}{1 + 1.1^{50-x}} \frac{\text{Max}(n - 5, 0)}{5}, \tag{5.25}$$

where q_x is the probability of death between ages x and $x + 1$ for the overall population. Nuttall et al. (1994) classify people with disability categories 9 and 10 as having an ongoing long-term care need. Therefore, we assume that people in long-term care facilities have the same amount of extra mortality as people with a disability catagory 9. As a result, we have the following equation for computing p_x^{23}'s:

$$p_x^{23} = q_x + \frac{0.16}{1 + 1.1^{50-x}} \qquad x = 60, ...\omega - 1. \tag{5.26}$$

Second, we compute p_x^{21}'s by using the parameterized transition intensities between disability states calculated from the United States National Long-Term Care Study (NLTCS) in 1982 and 1984 (Pritchard, 2006)[4]. We borrow p_x^{21}'s but not other p_x^{gh}'s from the transition probability matrices based on the US NLTCS data for the reason that among all p_x^{gh}'s, p_x^{21}'s should be the least likely to be affected by the mortality differentials between Britain and the US.

[4]Pritchard's multiple state model consists of seven states; they are (1) healthy, (2) 1 ADL only (inability to perform one activity of daily living), (3) 1-2 ADLs, (4) 3-4 ADLs, (5) 5-6 ADLs, (6) institutionalized, and (7) dead. We let p_x^{21} be the probability that an individual aged x moves from state 6 to either one of the first five states in one year.

Given p_x^{21}'s and equations (5.22) to (5.26), we can estimate the remaining transition probabilities for year 2001. Figures 5.10 and 5.11 show the estimated transition probabilities.

The next step is to project the transition rates for year 2001 to the future. Given that residents in long-term care facilities generally have one or more limiting long-term illness, huge effort is required to raise the likelihood of moving from state 2 to state 1. For this reason, we assume that p_x^{21}'s remain unchanged over time. However, the likelihood of moving to state 3 is, of course, dependent on the mortality of the general population. Therefore, we assume that for all age, both p_x^{13} and p_x^{23} reduce at the same rate as that specified in the Lee-Carter projection of future q_x.

The likelihood of moving to state 1 from state 2 is unlikely to remain stationary. Future trends in long-term care incidence will depend on a number of different factors; for example,

- health status − the population of Britain has been living longer, but the extra years have not necessarily been lived in good health; over the past 20 years, the life expectancy has been increasing at a faster rate than the healthy life expectancy (expected years of life in good or fairly good health) (Office of National Statistics, 2004),
- government policy − if government policy emphasizes home care, then the tendency to move to a long-term care facility may be reduced,
- supply of long-term care places − long-term care incidence rates may rise if there are more long-term care places available; incidence rates may also increase if long-term care insurance becomes more popular.

Given the uncertainty about the factors affecting long-term care incidence and the absence of time-series data regarding p_x^{12}'s, it is hard to legitimately predict the magnitude, and even the direction, of future changes in p_x^{12}'s. In this study, we consider the following three scenarios:

1. p_x^{12}'s are reduced by 30%;
2. p_x^{12}'s remain unchanged;
3. p_x^{12}'s are increased by 30%.

Finally, the projected values of p_x^{11}'s and p_x^{22}'s can be calculated by equation (5.21).

To compute the NNEG values based on the multiple state model, we replace ${}_kp_xq_{x+k}$ in equation (5.15) by

$$\left(\prod_{j=0}^{k-1} \tilde{p}_{x+j}^{11}\right)(1-\tilde{p}_{x+k}^{11}), \tag{5.27}$$

where $\tilde{p}_{x+k}^{11}$ is the projected value of p_{x+k}^{11} for persons who were aged x at inception.

Tables 5.10 and 5.11 present the NNEG values under different assumptions about the future trend in p_x^{12}'s. These values are based on the ARMA-EGARCH pricing framework. All model assumptions except decrement are the same as those in Section 5.4.4.

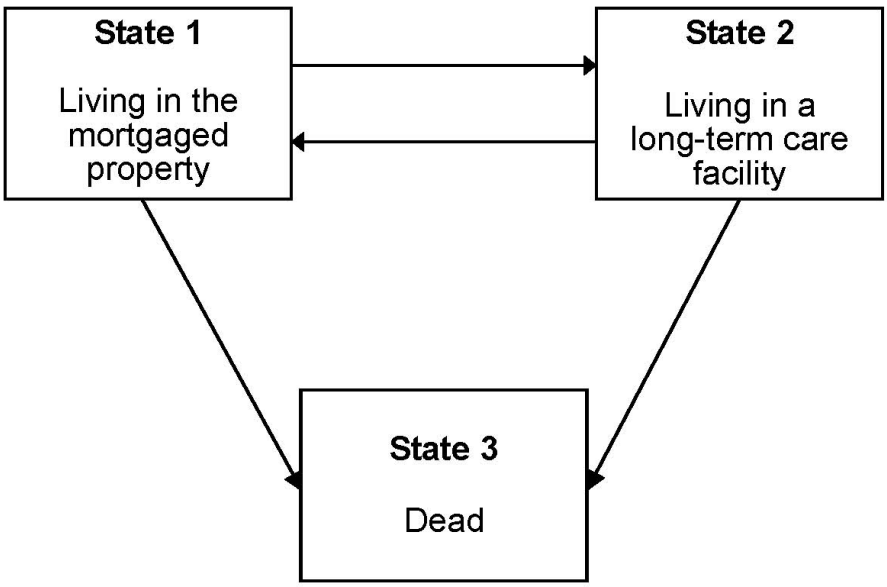


Fig. 5.9. Multiple state model for modeling long-term care incidence and mortality.

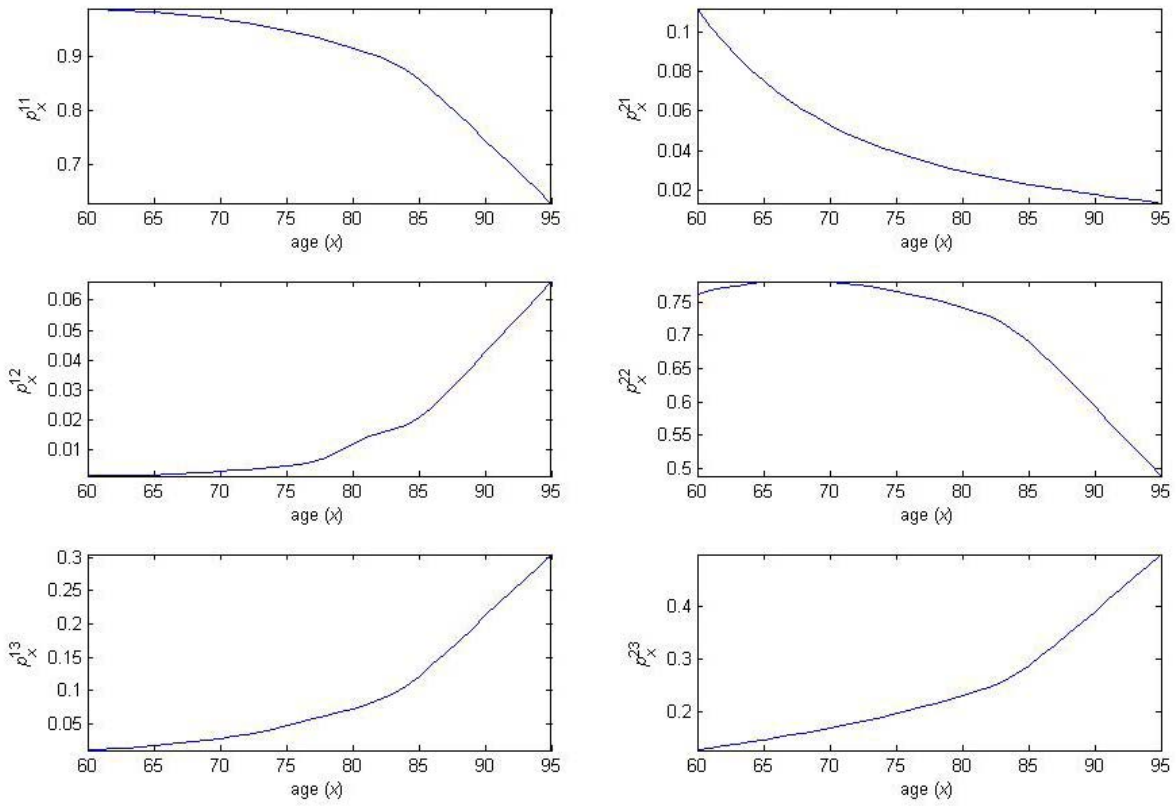


Fig. 5.10. Estimates of transition probabilities, male.

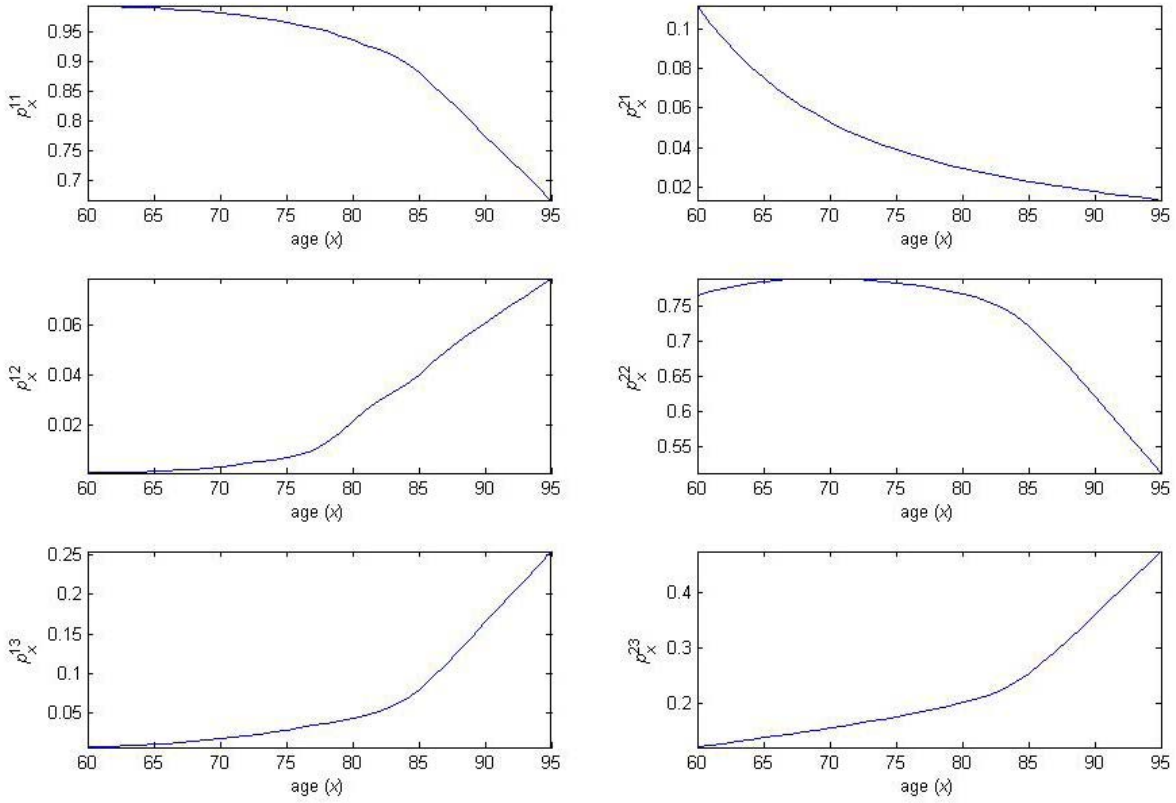


Fig. 5.11. Estimates of transition probabilities, female.

Age at inception	Without LTC[5]	$p_x^{12} \downarrow 30\%$	p_x^{12} unchanged	$p_x^{12} \uparrow 30\%$
60	19.09%	17.13%	15.77%	13.88%
70	10.22%	9.21%	8.10%	7.17%
80	3.63%	3.28%	2.71%	2.26%
90	1.43%	1.40%	1.08%	0.83%

Table 5.10 NNEG costs as a percentage of cash advanced under different decrement assumptions, male.

Age at inception	Without LTC	$p_x^{12} \downarrow 30\%$	p_x^{12} unchanged	$p_x^{12} \uparrow 30\%$
60	33.72%	28.26%	23.42%	19.64%
70	19.27%	16.24%	13.05%	10.59%
80	7.02%	6.23%	4.67%	3.52%
90	2.58%	2.37%	2.19%	1.61%

Table 5.11 NNEG costs as a percentage of cash advanced under different decrement assumptions, female.

5.4.6 Hedging and Capital Reserving Strategies for the NNEG

Recall that the actual underlying asset (the mortgaged property) of the NNEG is unique and infrequently traded. In practice, however, we can apply dynamic hedging to the NNEG by forming a hedge portfolio which has a price sensitivity profile similar to that of the NNEG liability. More specifically, the hedge portfolio and the NNEG liability should have similar deltas (and possibly gammas). Possible assets for forming a hedge portfolio include residential property index funds and real estate income trusts (REITs). These assets are traded, and the dynamics of their returns should be close to that of the Nationwide house price index − the proxy for the true underlying asset.

[5] "Without LTC" refers to the simplest case that death is the only mode of decrement.

Unfortunately, there is not much choice of residential property index funds, since most property index funds listed on the London Stock Exchange are tied to commercial rather than residential properties. In addition, the UK REIT market is still in its infancy: the first UK residential REIT, the HBOS residential REIT, was just launched in early 2007, giving us little historical data for the required stochastic modeling.

Even if we are able to find suitable assets for constructing a hedge portfolio, dynamic hedging will not be perfect due to (1) discrete hedging error, (2) transaction cost, (3) longevity risk and (4) basis risk. These factors may give rise to additional costs on the top of the hedge portfolio for dynamic hedging. Such additional costs, which we call unhedged liability, can be handled by an actuarial approach − using stochastic simulations, we generate an empirical distribution of the unhedged liability; then the upper 95^{th} percentile of the empirical distribution, for example, gives the amount of risk capital required for cushioning against the unhedged liability. In the following we discuss and quantify each of the factors affecting the unhedged liability by using a hypothetical hedge portfolio that consists of a risk-free bond and a risky asset that is perfectly correlated with the Nationwide house price index. To keep the illustration simple, we compute hedge costs using the classical Black-Scholes formula and perform stochastic simulations on the basis of the log-normal model in the physical measure.

- *Discrete hedging error*
  The Black-Scholes-Merton approach assumes continuous trading. However, continuous trading is impossible in practice. When we relax the assumption of continuous trading, discrete hedging error is introduced. With discrete time steps, between which the hedge portfolio is not adjusted, the hedge portfolio may not be self-financing. As a result, additional funds may be required when the hedge portfolio is rebalanced.
  We let $H^c(t)$ be the value of the hedge at time t, given that the borrower survives at time t; then according to the pricing framework in Section 5.15, we have

$$H^c(t) = \sum_{k=0}^{\omega-x-t-1} {}_kp_{x+t}\, q_{x+t+k} P\left(k + \frac{1}{2} + \delta, S_t, X_0 e^{ut}, u, g\right), \tag{5.28}$$

  where x is the age at issue. Note that $H^c(0)$ equals the NNEG value at inception.

According to the Black-Scholes-Merton framework, the hedge portfolio, given that the borrower survives at time t, consists of

1. Δ_t^c units of the risky asset, and
2. $H^c(t) - \Delta_t^c S_t$ dollar amount of risk-free bonds, where $\Delta_t^c = \frac{\partial H^c(t)}{\partial S_t}$.

Let us assume that the hedge portfolio is rebalanced yearly. Immediately before rebalancing at time $t+1$, the hedge portfolio from time t will have accumulated to

$$H^c(t+1^-) = \Delta_t^c S_{t+1} + (H^c(t) - \Delta_t^c S_t)e^r + \Delta_t^c S_{t+1} g. \tag{5.29}$$

Recall that the NNEG may be interpreted as a portfolio of put options with different times to maturity. At time 1 from inception, q_x units of the put options will expire, while the remaining units will have a total value of $p_x H^c(1)$. Therefore, the funds required at time 1 are

$$q_x(X_0 e^u - S_1)^+ + p_x H^c(1), \tag{5.30}$$

which means the hedging error at time 1 can be written as

$$HE_1 = q_x(X_0 e^u - S_1)^+ + p_x H^c(1) - H^c(1^-). \tag{5.31}$$

After rebalancing at time 1, we hold a hedge portfolio of $p_x H^c(1)$. Inductively, the hedging error at time 2 is given by

$$HE_2 = p_x \left[(q_{x+1} X_0 e^{2u} - S_2)^+ + p_{x+1} H^c(2) - H^c(2^-)\right], \tag{5.32}$$

and in general, the hedging error at time t, $t = 2, 3, ..., \omega - x$, can be expressed as

$$HE_t = \left(\prod_{k=0}^{t-2} p_{x+k}\right) \left[(q_{x+t-1} X_0 e^{ut} - S_t)^+ + p_{x+t-1} H^c(t) - H^c(t^-)\right]. \tag{5.33}$$

We are interested in the empirical distribution of the present value of the hedging errors. Such an empirical distribution can be generated using the following procedure:

1. simulate a large number of sample paths using the stochastic model for $\{S_t\}$ under the physical measure;

2. for each sample path, calculate the hedging error at time points when the hedge portfolio is rebalanced, and compute the present value of the hedging errors.

Note that equation (5.28) assumes that all home exits occur in mid-year. We can relax this assumption by allowing home exits to occur in the middle of any fraction, say $1/m$, of a year; this generalizes equation (5.28) to

$$H^c(t) = \sum_{k=0}^{m(\omega-x-t)-1} {}_{\frac{k}{m}}p_{x+t} \quad q_{x+t+\frac{k}{m}} P\left(\frac{k+\frac{1}{2}}{m} + \delta, S_t, Ke^{ut}, u, g\right). \tag{5.34}$$

By adjusting equations (5.29) to (5.33) accordingly, we can obtain hedging errors based on different rebalancing frequencies.

Table 5.12 shows the percentiles of the empirical distributions for different ages at inception. The numbers, expressed as a ratio of the present value of hedging errors to the NNEG value, are based on the assumption that the hedge portfolio is rebalanced quarterly.

- *Transaction cost*

  A transaction cost is incurred whenever the hedge portfolio is rebalanced. We assume transaction costs of τ times the value of the risky assets that have to be traded when the hedge portfolio is rebalanced. Let us assume again that the hedge portfolio is rebalanced yearly. At time 1, we require a short position of q_x units of the risky asset if the options expiring at time 1 are in the money; we also require a short position of $-p_x\Delta_1^c$ units of the risky asset for hedging the options not maturing at time 1. Therefore, the transaction cost at time 1 is given by

$$TC_1 = \tau S_1 \left| p_x\Delta_1^c + q_x I_{\{X_0e^u - S_1 > 0\}}(-1) - \Delta_0^c \right|.$$

  Inductively, we have the following expression for the transaction cost at time t, $t = 2, 3, ..., \omega - x$:

$$TC_t = \left(\prod_{k=0}^{t-2} p_{x+k}\right) \tau S_t \left| p_{x+t-1}\Delta_t^c + q_{x+t-1} I_{\{X_0e^{ut} - S_t > 0\}}(-1) - \Delta_{t-1}^c \right|.$$

The above expressions can be generalized easily to a version applicable to hedging with more frequent rebalancing. To incorporate transaction costs into the empirical distribution of the unhedged liability, we can simply add the transaction costs to their corresponding hedging errors in Step 2 of the procedure for generating the unhedged liability distribution.

Table 5.13 shows for different ages at inception the percentiles of the empirical distributions, with consideration of transaction costs. The numbers are based on a quarterly rebalanced hedge portfolio, and on the assumption that $\tau = 0.2\%$ per annum. Again, the numbers are expressed as a ratio of the unhedged liability to the NNEG value.

- *Longevity risk*

  In Section 5.4.4, we have demonstrated that mortality assumptions play a crucial role in the valuation of the NNEG. The small sample risk which arises from random departures from the assumed survival distribution can be mitigated effectively by diversification. However, the providers are also exposed to longevity risk, that is, the uncertainty about the future mortality improvement. Such risk is non-diversifiable. In the ideal case, the providers could lay off some longevity risk exposures at a reasonably low cost by using mortality-linked securities, exemplified by the European Investment Bank (EIB) longevity bond in which coupon payments are linked to a survivor index based on the realized mortality experience. However, the longevity bond market is still at its very early stage so that the required hedging instruments may not be available. Therefore, for the time-being, we have to resort to actuarial approaches.

  To illustrate the impact of longevity risk, we consider a simple case in which the hedge portfolio is rebalanced yearly. Let $\tilde{q}_x$ ($\tilde{p}_x$) be the realized conditional probability of death (survival). The actual number of put options expiring at time 1 is $\tilde{q}_x$, which is not necessarily equal to q_x, the assumed conditional probability of death, due to the uncertainty of future mortality. Therefore, if stochastic mortality is considered, the hedging error at time 1 becomes

$$HE_1 = \tilde{q}_x(X_0 e^u - S_1)^+ + \tilde{p}_x H^c(1) - H^c(1^-), \tag{5.35}$$

and in general,

$$HE_t = \left(\prod_{k=0}^{t-1} \tilde{p}_{x+k}\right) \left(\tilde{q}_{x+t-1}(S_0 e^{ut} - S_t)^+ + \tilde{p}_{x+t-1} H^c(t) - H^c(t^-)\right), \tag{5.36}$$

for $t = 2, 3, ..., \omega - 1$.

We can incorporate longevity risk into the unhedged liability distribution by the following procedure:

1. attach, to each of the house price sample paths, a mortality scenario generated by the parametric bootstrapping method discussed in Chapter 2;
2. treat the death probabilities in each mortality scenario as the realized death probabilities;
3. compute the hedging errors (and the transaction costs) on the basis of the realized death probabilities obtained in Step 2.

Table 5.14 shows for different ages at inception the percentiles of the empirical distributions of the unhedged liability, with consideration of stochastic mortality. As before, we assume that $\tau = 0.2\%$ per annum and that the hedge portfolio is rebalanced quarterly. The numbers are expressed as a ratio of the unhedged liability to the NNEG value.

- *Basis risk*

  In the above illustrations, it is assumed that returns on the true underlying asset, the mortgaged property, are exactly the same as that on its proxy, the Nationwide house price index. This is unlikely to be true in reality, so the providers are exposed to basis risk. To incorporate such basis risk into the unhedged liability distribution, we may base the empirical distribution on sample paths that are simulated via a stochastic model with a larger volatility parameter, on the grounds that property prices tend to exhibit greater variations when we drill down to individual housing districts.

  Furthermore, as the Nationwide house price index is not a traded security, we are required to use alternative assets in forming the hedge portfolio. This, obviously, will

incur additional basis risk, since the returns on the house price index and that on the assets in the hedge portfolio may not be identical.

Finally, we examine the impact of rebalancing frequency on the unhedged liability distributions. Table 5.15 depicts the 95^{th} percentiles of the unhedged liability distributions based on different rebalancing frequencies. We observe that the benefit of moving from quarterly to monthly is only marginal, particularly for the case in which stochastic mortality is considered. This is because, as we observe in Tables 5.12 to 5.14, the effect of stochastic mortality outweighs that of discrete hedging errors and that of transaction costs.

Note that the above illustrations can be performed using the ARMA-EGARCH framework instead of the Black-Scholes-Merton. We expect that the unhedged liability based on the ARMA-EGARCH framework will be even higher, since under the conditional Esscher transformation, the replicating portfolio is no longer self-financing. When the ARMA-EGARCH model is assumed, the Greek letters have to be computed by simulations. Such simulations should be performed using the methods proposed by Broadie and Glasserman (1996) rather than crude estimation techniques which may produce unreliable estimates.

Age at inception	70^{th}	90^{th}	95^{th}
60	0.0745	0.1444	0.1947
70	0.1154	0.2185	0.2796
80	0.1680	0.3463	0.4628
90	0.2347	0.5512	0.7445

Table 5.12 Selected percentiles of the empirical distributions of the present value of hedging errors.

Age at inception	70^{th}	90^{th}	95^{th}
60	0.1312	0.1999	0.2572
70	0.1874	0.2926	0.3527
80	0.2681	0.4514	0.5717
90	0.3456	0.7114	0.9037

Table 5.13. Selected percentiles of the empirical distributions of the present value of hedging errors and transaction costs.

Age at inception	70^{th}	90^{th}	95^{th}
60	0.4774	0.9083	1.1421
70	0.4687	0.8468	1.1922
80	0.5046	0.9639	1.3393
90	0.4813	1.0647	1.4786

Table 5.14. Selected percentiles of the empirical distributions of the present value of hedging errors and transaction costs, with consideration of stochastic mortality.

Rebalancing Frequency	HE only	HE + HC	HE + HC, with SM
Yearly	1.8208	1.9022	2.0604
Quarterly	0.7445	0.9037	1.4786
Monthly	0.4003	0.6708	1.2240

Table 5.15. 95^{th} percentiles of the empirical distributions of the present value of (1) hedging errors (HE), (2) hedging errors and hedging costs (HC), and (3) hedging errors and hedging costs with consideration of stochastic mortality (SM).

5.5 Conclusion

Pricing the NNEG is not easy since the underlying asset, the mortgaged property, is unique and infrequently traded. Although the Nationwide house price index may be used as

a proxy for the true underlying asset, the returns on the index demonstrate effects of varying volatility and autocorrelation, rendering the classical Black-Scholes approach for option pricing inappropriate. In this study, we have found that the house price returns can be well described by an ARMA-EGARCH model. On the basis of the fitted ARMA-EGARCH model, we identified an equivalent martingale measure using the conditional Esscher transformation and derived a pricing formula for the NNEG. We have found that the NNEG can be a heavy burden to equity release product providers. Depending on the borrower's age at inception and the loan to value ratio, the NNEG can cost in excess of 30% of the cash advanced. The estimated NNEG costs, however, have not taken account of the basis risk that arises from the discrepancies between the returns on the true underlying asset and that on its proxy. This risk is likely to further increase the NNEG liability, primarily because individual house prices tend to be more volatile than the aggregate price index.

On the other hand, morbidity risk and early redemption risk may lower the NNEG values, since they reduce the likelihood of survival to the very advanced ages when negative equity claims are the most likely to occur. In this study, we have developed a multiple state model to find out how long-term care incidence may impact the NNEG costs. We have found that the option prices are lowered slightly when long-term care incidence is considered. To keep the analysis simple, we have not taken account of other modes of decrement, such as remortgaging. Remortgaging and its interaction with interest rate assumptions may be considered in future research on equity release mechanisms.

With equity release being no exception, product features drive policyholder behaviors in many financial products; for example, given that roll-up mortgage purchasers are making a positive statement about their desire to remain in their own home as long as possible, they are likely to experience lower long-term care incidence rates than ordinary homeowners. Likewise, since, in most cases, the mortgaged property has to be liquidated when the borrower moves into a long-term care facility, we expect those who purchased a roll-up mortgage to experience a lower probability of moving from the state of living in a long-term care facility to the state of living at home. In other words, the central decrement assumptions used in this study may not precisely reflect the actual situation experienced by purchasers of equity release products. In practice, decrement assumptions should be

adjusted according to relevant experience studies.

In Section 5.15, we have pointed that the tail of the assumed survival distribution has a significant impact on the NNEG values. This phenomenon has two important implications. First, the reduction of mortality rates, particularly at the advanced ages, will considerably increase the NNEG cost. To cope with future reduction of mortality, equity release product providers may have to consider lowering the maximum loan to value ratio in order to maintain the NNEG liability at a manageable level, although this may make the product less competitive. Second, the providers should take account of their longevity risk exposures in pricing equity release products and appraise regularly the appropriate level of risk capital set aside as the portfolio evolves.

We have proposed a combination of dynamic hedging and capital reserving strategies for managing the risks associated with the NNEG. Dynamic hedging can be practically performed by forming a hedge portfolio with traded assets that are highly correlated to the Nationwide house price index. However, no matter how high the correlation is, dynamic hedging will not be perfect due to discrete hedging errors, transaction costs, longevity risk and basis risk. The aggregate additional costs on the top of the hedge portfolio for a practical dynamic hedging strategy are known as unhedged liability. Unhedged liability can be handled by an actuarial approach that determines the amount of risk capital required for cushioning against the unhedged liability. In this study, we have illustrated the generation of the unhedged liability distribution on the basis of the Black-Scholes-Merton framework. An immediate extension of this study would be an investigation of the unhedged liability distribution based on the ARMA-EGARCH framework, although such an investigation may demand a huge computational effort.

In the United States, contracts similar to roll-up mortgages are called reverse mortgages. The topic of reverse mortgages has been considered by researchers in various fields; for example, Davidoff and Welke (2004) investigated the risks of selection and moral hazard in reverse mortgages; Artle and Varaiya (1978) studied the appeal of transferring housing wealth from the period after death to the period while still alive. The majority of reverse mortgages in the US are sold via the Home Equity Conversion Mortgage (HECM) program introduced by the United States Department of Housing and Urban Development in the late 1980s. Borrowers of reverse mortgages sold via the HECM program have to pay a

guarantee fee of 2% of the initial property value to the Federal Housing Administration (FHA), who will top up the shortfall of funds if the property sale proceeds is less than the outstanding loan balance when the loan is terminated. Therefore, the guarantee fee may be regarded as the premium for the NNEG. To date, high rates of house price inflation have left the FHA with few claims to pay. However, it is entirely possible that house price inflation will slow down. It would be warranted to examine whether or not the fee of 2% of the initial property value would be sufficient to fund the guarantee in the long run.

Chapter 6

Concluding Remarks and Further Research

It is hard to conceive that mortality projections without measures of uncertainty would be acceptable for use in actuarial risk management in the future. While uncertainty can be effectively forecasted by means of stochastic methodologies, the probability statements about mortality projections may, to a certain extent, be based on untestable assumptions and model specification.

In this thesis, we relaxed the structure of the Lee-Carter model by introducing a more realistic assumption that the death counts follow a Negative Binomial distribution. We demonstrated that such an assumption is equivalent to presuming the heterogeneity within each age-period cell follows a gamma distribution. The revised version of the Lee-Carter model gives a better fit to historical data and captures more uncertainty about future mortality dynamics.

The model was then utilized in the derivation of mortality improvement scales for Canadian insured lives. To solve the problem of data insufficiency, we developed a joint model that could borrow information from the mortality data of the general population. In addition to the generalization proposed in the previous chapter, we also took account of shocks and structural shifts in the measures of uncertainty by introducing outliers detected in the historical data to the bootstrapping procedure.

Due to the lack of old-age mortality data, the improvement scales are available only up

to age 99. To estimate survival probabilities beyond this age, we introduced a method for actuaries to extrapolate mortality rates and to determine the age at which the life table should be closed. We also applied this method to the prediction of the highest attained age for various birth cohorts.

Finally, we studied the valuation of the NNEG offered in most roll-up mortgages. Such a valuation requires a simultaneous consideration of demographic and financial risks. We handled mortality risk by the stochastic mortality models we considered, and house-price inflation risk by an ARMA-EGARCH model. Under the ARMA-EGARCH assumption, the Black-Scholes formula is no longer applicable, and therefore we resort to the risk-neutral valuation based on an equivalent martingale measure identified by the conditional Esscher transformation. The pricing formula is accompanied with some hedging and capital reserving strategies that are yet to be implemented on the basis of the ARMA-EGARCH framework.

In addition to the implementation of hedging and capital reserving strategies under the ARMA-EGARCH assumption, we may also consider the following suggestions for further research.

- *Generalizing the threshold life table to a version for multiple-life.*

  The threshold life table in Chapter 4 may be generalized to a version that is capable of describing the statistical dependence between the mortality of two or more lives. Such a generalization may be performed by considering copulas of different types. The generalized threshold life table can then be applied to the valuation of the NNEG so that NNEGs sold to couples can be priced.

- *Investigation into the moral hazard risk in equity release mechanisms.*

  When the NNEG becomes "in the money", the borrower would have little incentive to maintain his/her home. The lack of maintenance may lead to a depreciation of the property value and consequently a reduction of the provider's profit. The risk of such problems may be reduced if the contracts are redesigned. For example, Shiller and Weiss (2000) argued that the moral hazard risk in reverse mortgages might be reduced if the borrower is penalized for deviations of home selling price from the value predicted by a house price index and the value of the mortgaged property at

inception of the contract. It would be interesting to search for alternative contract designs that can reduce the risk of moral hazard and keep the product attractive to potential buyers.

- *Refitting the stochastic mortality models using alternative data sets.*

  The joint model for estimating mortality improvement scales may be refitted to Canadian annuitants mortality experience. The aims of refitting are threefold: (1) to obtain improvement scales for the annuitants, (2) to test the robustness of our model, and (3) to obtain a relationship between the mortality of the annuitants and the general population. Similarly, the model may also be refitted to the mortality data from the Canada/Quebec Pension Plan.

- *Investigation into other equity release products.*

  Although the UK equity release market is currently dominated by roll-up mortgage products, the development of other equity release businesses has been fairly strong in recent years. In FY 2005, home reversions accounted for £54.6 million worth of new business, up 35% year on year from FY 2004 (£40.5 million) (SHIP, 2006a). Drawdown mortgage business has also exhibited an impressive growth. Indeed, a recent SHIP Members Survey (SHIP, 2006b) showed drawdown mortgages as one of the most popular features for consumers in 2005.

  The risk management for these products is rather different from that for roll-up mortgages. For instance, home reversion providers are exposed to house price inflation risk directly rather than via the NNEG, and drawdown mortgage providers are exposed to the uncertainty about policyholder behaviors, since policyholders are allowed to draw down either regular or ad hoc amounts. A separate study is needed to analyze the financial and demographic risks associated with these equity release products.

- *Estimating the impact of cohort effect on the mortality improvement scales.*

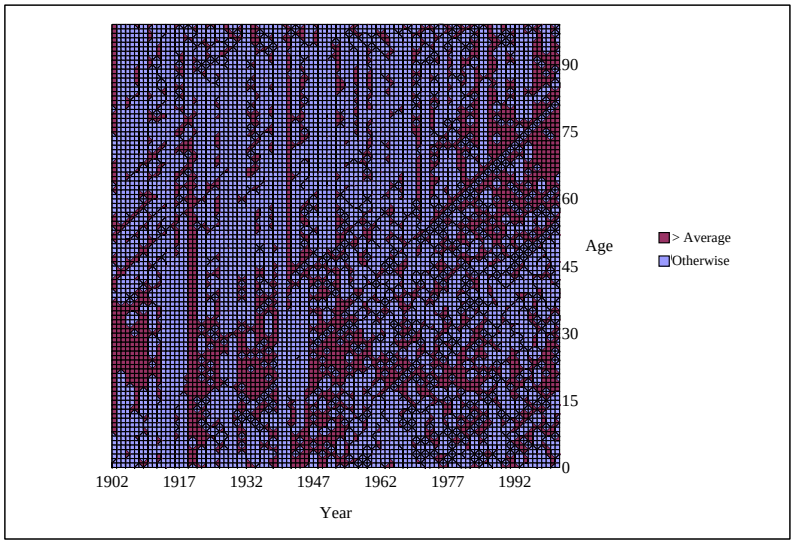
  We define a mortality improvement factor as $1 - \frac{m_{x,t}}{m_{x,t-1}}$. The contour maps in Figure 6.1 show the mortality improvement factors for the English and Welsh population (for ease of exposition, we use two contour levels only). Diagonals in the contour maps, from bottom left to top right, show cells for succeeding ages for the same year

of birth. The strong diagonal features in the contour maps indicate that the mortality improvement for some cohorts is exceptionally large (or small). This phenomenon is known as the cohort effect.

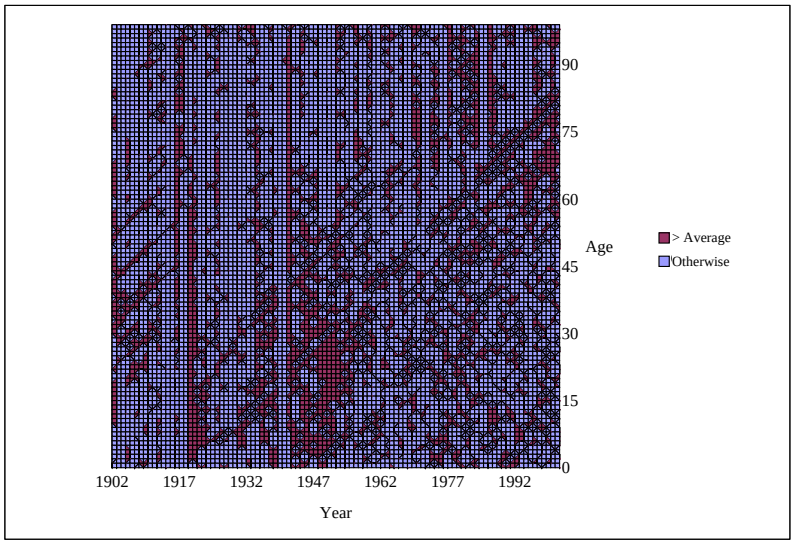
To model the cohort effect, Richards et al. (2005) applied the P-splines regression to the age-cohort dimension, while Renshaw and Haberman (2006) used the Lee-Carter model with an additional bilinear term; that is,

$$\ln(m_{x,t}) = a_x + b_x k_t + c_x \iota_{t-x} + \epsilon_{x,t}, \tag{6.1}$$

where a_x, b_x and k_t are the original Lee-Carter parameters, ι_{t-x} is an additional driving force of mortality improvement due to the cohort effect, c_x determines the sensitivity to ι_{t-x} at different ages, and $\epsilon_{x,t}$ is the error term. In the derivation of the mortality improvement scales in Chapter 3, we have not taken any cohort effect into account. It would be interesting to explore how the mortality improvement scales in Chapter 3 may change when a cohort effect is considered.



Male



Female

Fig. 6.1. Improvement factors based on English and Welsh population mortality, male and female.[1]

[1]Source of data: Human Mortality Database (2005).

References

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

American Academy of Actuaries (2002). Final Report of the American Academy of Actuaries' Commissioners Standard Ordinary Task Force. Available at www.actuaries.org.

Artle, R. and Varaiya, P. (1978). Life Cycle Consumption and Homeownership, *Journal of Economic Theory*, **18**, 38-58.

Balkema, A. A. and de Haan, L. (1974). Residual Life Times at Great Age. *Annals of Probability*, **2**, 792-804.

Bernard, J. and Vaupel, J.W. (1999). *Validation of Exceptional Longevity*. Odense, Demark: Odense University Press.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **31**, 307-327.

Boot, J.C.G. and Lisman, J.H.C. (1967). Further Methods of Derivation of Quarterly Figures from Annual Data, *Applied Statistics*, **16**, 65-75.

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

Booth, P.M. and Marcato, G. (2004). The Measurement and Modelling of Commercial Real Estate Performance. *British Actuarial Journal*, **10**, 5-73.

Booth, H. and Tickle, L. (2003). The future aged: new projections of Australia's elderly population. Department of Actuarial Studies Research Paper Series. Macquarie University, Sydney, Australia.

Bourbeau, R. and Desjardins, B. (2002). Dealing with Problems in Data Quality for the Measurement of Mortality at Advanced Ages in Canada, *North American Actuarial Journal*, **1**, 1-25.

Bowers, Jr., N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., Nesbitt, C.J. (1997). *Actuarial Mathematics*. Schaumburg, IL: Society of Actuaries.

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

Brass, W. (1975). *Methods for Estimating Fertility and Mortality From Limited and Defective Data*. Chapel Hill, North Carolina: Carolina Population Center, Laboratories for Population Statistics.

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*, **3**, 212-224.

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

Buettner, T. (2002). Approaches and Experiences in Projecting Mortality Patterns for the Oldest-old, *North American Actuarial Journal*, **6**, 14–29.

Bühlmann, H., Delbaen, F., Embrechts, P. and Shiryaev, A.N. (1996). No-Arbitrage, Change of Measure and Conditional Esscher Transforms, *CWI Quarterly*, **9**, 291-317.

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

Calhoun, C.A. (1996). *OFHEO House Price Indexes: HPI Technical Description.* Washington D.C.: Office of Federal Housing Enterprise Oversight.

Chen, C. and Liu, L.M. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series, *Journal of American Statistical Association*, **88**, 284-297.

Chia, N.C. and Tsui, A.K.C. (2003). Life annuities and compulsory savings and income adequacy of the elderly in Singapore, *Journal of Pension Economics and Finance*, **2**, 41-65.

Coale, A. and Guo, G. (1989). Revised Regional Model Life Tables at Very Low Mortality, *Population Index*, **55**, 614-643.

Coale, A. and Kisker, E. (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.

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

Continuous Mortality Investigation Bureau (1990). Standard Tables of Mortality Based on the 1979-82 Experiences. CMI Report no. 10. London: Institute of Actuaries and Faculty of Actuaries.

Continuous Mortality Investigation Bureau (1999). Standard Tables of Mortality Based on the 1991-94 Experiences. CMI Report no. 17. London: Institute of Actuaries and Faculty of Actuaries.

Continuous Mortality Investigation Bureau (2004). Projecting Future Mortality: A Discussion Paper. CMI Working Paper no. 3. London: Institute of Actuaries and Faculty of Actuaries.

Continuous Mortality Investigation Bureau (2005). Projecting Future Mortality: Towards a proposal for a stochastic methodology. CMI Working Paper no. 15. London: Institute of Actuaries and Faculty of Actuaries.

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.

Currie, I.D. and Waters, H.R. (1991). On Modelling Select Mortality, *Journal of the Institute of Actuaries*, **118**, 453-481.

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

Davidoff, T. and Welke, D. (2004). Selection and Moral Hazard in the Reverse Mortgage Market. Available at Social Science Reserach Network: http://ssrn.com/abstract=608666.

Department for Work and Pension. (2006). *Personal accounts: a new way to save.* London: The Stationary Office.

Duan, J.C. (1995). The GARCH Option Pricing Model, *Mathematical Finance*, **5**, 13-32.

Duan, J.C., Gauthier, G. and Simonato, J.G. (1999). An Analytical Approximation for the GARCH Option Pricing Model, *Journal of Computational Finance*, **2**, 75-116.

Duan, J.C., Gauthier, G. and Simonato, J.G. and Sasseville, C. (2006). Approximating the GJR-GARCH and EGARCH option pricing models analytically. *Journal of Computational Finance*, to appear.

Eilers, P.H.C. and Marx, B.D. (1996). Flexible Smoothing with B-splines and Penalties, *Statistical Science*, **18**, 251-262.

Embrechts, P., Klüppelberg, C and Mikosch, T. (1997). *Modelling extremal events for insurance and finance.* New York : Springer.

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflations, *Econometrica*, **50**, 987-1007.

Esscher, F. (1932). On the Probability Function in the Collective Theory of Risk, *Skandinavisk Aktuarietidskrift*, **15**, 175-195.

Föllmer, H. and Sodermann, D. (1986). Hedging of Contingent Claims under Incomplete Information. In Hindenbrand, W. and Mas-Colell, A. (eds.) *Contributions to Mathematical Economics*, 205-223. Amsterdam: North Holland.

Föllmer, H. and Schweizer, M. (1986). Hedging of Contingent Claims under Incomplete Information. In Davis, M.H.A. and Eilliot, R.J. (eds.) *Applied Stochastic Analysis*, 289-414. London: Gordon and Breach.

Gerber, H.U. and Shiu, E.S.W. (1994). Option Pricing by Esscher Transforms, *Transactions of the Society of Actuaries*, **46**, 99-191.

Glosten, L., Jagannathan, R. and Runkle, D. (1993). Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance*, **48**, 1779-1801.

Gompertz, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality and on a New Mode of Determining Life Contingencies, *The Royal Society of London, Philosophical Transactions. Series A*, **115**, 513–585.

Han, Z. (2005). Living to 100 and Beyond: An Extreme Value Study. Living to 100 and Beyond Symposium Monograph. Available at: http://www.soa.org/ccm/content/research-publications/library-publications/monographs/life-monographs/living-to-100-and-beyond-monograph/

Hardy, M.R. (2003). *Investment Guarantees: Modeling and Risk Management for Equity-linked Life Insurance.* Hoboken, New Jersey: John Wiley & Sons.

Hörmann, W. (1993). The Transformed Rejection Method for Generating Poisson Random Variables, *Insurance: Mathematics and Economics*, **12**, 39-45.

Heligman, L and Pollard, J.H. (1980). The Age Pattern of Mortality, *Journal of the Institute of Actuaries*, **107**, 437–455.

Heston, S.L. and Nandi, S. (2000). A Closed-Form GARCH Option Valuation Model, *The Review of Financial Studies*, **13**, 585-625.

Hill, B.M. (1975). A Simple Approach to Inference about the Tail of a Distribution, *Annals of Statistics*, **3**, 1163–1174.

Himes, C.L., Preston, S.H., and Condran, G.A. (1994). A Relational Model of Mortality at Older Ages in Low Mortality Countries, *Population Studies*, **48**, 269–291.

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.humanmortality.de (data downloaded on 25 May 2005).

Institute of Actuaries. (2001). Report on Equity Release Mechanisms. Available at http://www.actuaries.org.uk/files/pdf/equity_release/rpt.pdf.

Institute of Actuaries. (2005a). Equity Release Report 2005. Volume I: Main Report. Available at http://www.actuaries.org.uk/files/pdf/equity_release/equityreleaserepjan05V1.pdf.

Institute of Actuaries. (2005b). Equity Release Report 2005. Volume II: Technical Supplement: Pricing Considerations. Available at http://www.actuaries.org.uk/files/pdf/equity_release/equityreleaserepjan05V2.pdf.

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

Koissi, M.C., Shapiro, A.F. and Hognas, G. (2005). Evaluating and Extending the Lee-Carter Model for Mortality Forecasting: Bootstrap Confidence Interval. *Insurance: Mathematics and Economics*, **38**, 1-20.

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 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*, **3**, 187-211.

Li, S.H. and Chan, W.S. (2007). The Lee-Carter Model for Forecasting Mortality, Revisited, *North American Actuarial Journal*, **11**, 68-89.

Lin, X.S. and Tan, K.S. (2003). Valuation of Equity-Indexed Annuities under Stochastic Interest Rates, *North American Actuarial Journal*, **7**, 72-91.

Ljung, G. and Box, G.E.P. (1978). On a Measure of Lack of Fit in Time Series Models, *Biometrika*, **66**, 67-72.

London, D. (1985). *Graduation: The Revision of Estimates.* Winsted, Connecticut: ACTEX Publications.

Mason, D.M. (1982). Laws of Large Numbers for Sums of Extreme Values, *Annals of Probability*, **10**, 754-764.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models, 2nd edition.* London: Chapman and Hall.

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

National Center for Health Statistics (2004a). HIST290 Death Rates for Selected Causes by 10-Year Age Groups, Race, and Sex: Death Registration States, 1900-32, and

United States, 1933-98. Available at http://www.cdc.gov/nchs/datawh/statab/ unpubd/mortabs/ hist290.htm. Accessed January 25, 2004.

National Center for Health Statistics (2004b). GMWK23R Death Rates by 10-Year Age Groups: United States and Each State, 1999-2001. Available at http://www.cdc.gov/nchs/datawh/statab/unpubd/mortabs.htm. Accessed January 25, 2004.

Nationwide Building Society (2006). UK House Prices Since 1952. Available at http://www.nationwide.co.uk/hpi/historical.htm. Data downloaded on 25 June 2006.

Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, **59**, 347-370.

Nuttall, S.R., Blackwood, R.J.L., Bussell, M.H., Cliff, J.P., Cornall, M.J., Cowley, A., Gatenby, P.L. and Webber, J.M. (1994). Financing Long-Term Care in Great Britian, *Journal of the Institute of Actuaries*, **121**, 1-68.

Office of National Statistics (2004). *Health Expectancy: Living Longer, More Years in Poor Health*. Available at: http://www.statistics.gov.uk/CCI/nugget.asp?ID=918&Pos=6&ColRank=2&Rank=1000.

Olshansky, S.J. and Carnes, B.A. (1997). Ever Since Gompertz, *Demography*, **34**, 1-15.

Panjer, H.H. and Russo, G. (1992). Parametric Graduation of Canadian Individual Insurance Mortality Experience: 1982–1988. *Proceedings of The Canadian Institute of Actuaries*, **23**, 378–449.

Panjer, H.H. and Tan, K.S. (1995). Graduation of Canadian Individual Insurance Mortality Experience: 1986 - 1992. Canadian Institute of Actuaries.

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

Renshaw, A.E. and Haberman, S. (1997). Dual Modelling and Select Mortality, *Insurance: Mathematics and Economics*, **19**, 105-126.

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.

Richards, S.J., Kirkby, J.G. and Currie, I.D. (2005). The Importance of Year of Birth in Two-Dimensional Mortality Data. Paper Presented to the Institute of Actuaries. Available at http://www.actuaries.org.uk/files/pdf/sessional/sm20051024.pdf.

Rickayzen, B.D. and Walsh D.E.P. (2002). A Multiple-State Model of Disability for the United Kingdom: Implications for Future Need for Long-Term Care for the Elderly, *British Actuarial Journal*, **8**, 341-393.

Robine, J.M. and Vaupel, J.W. (2001). Supercentenarians: Slower Ageing Individuals or Senile Elderly? *Experimental Gerontology*, **36**, 915-930.

Robine, J.M. and Vaupel, J.W. (2002). Emergence of supercentenarians in low-mortality countries, *North American Actuarial Journal*, **6**, 54-63.

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

Rubinstein, M. (1976). The Valuation of Uncertain Income Streams and the Pricing of Options, *Bell Journal of Economic Review*, **35**, 16-27.

Safe Home Income Plans. (2005). No Negative Equity - Guaranteed: SHIP Code Brings Peace of Mind in an Uncertain Future. Available at http://www.ship-ltd.org/bm∼doc/10-september-2005.pdf.

Safe Home Income Plans. (2006a). SHIP Full Year Results 2005: Reversions Revival Continues and the Equity Release Sector Witnesses the Birth of the Drawdown Mortgage. Available at: http://www.ship-ltd.org/bm∼doc/30-january-2006.pdf.

Safe Home Income Plans. (2006b). SHIP Member Survey: Equity Release Industry Predicts Consumers Will be the Big Winners in 2006. Available at: http://www.ship-ltd.org/bm∼doc/28-january-2006.pdf.

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

Schweizer, M. (1996). Approximation Pricing and Variance-Optimal Martingale Measure, *Annals of Probability*, **24**, 206-236.

Scott, A., Evandrou, M., Falkingham, J. and Rake, K. (2001). *Going into Residential Care: Evidence from BPHS 1991-1998: SAGE Discussion Paper No. 5*, London: London School of Economics.

Shiller, R.J. and Weiss, A.N. (2000). Moral Hazard In Home Equity Conversion, *Real Estate Economics*, **28**, 1-31.

Siu, T.K., Tong, H. and Yang, H. (2004). On Pricing Derivatives under GARCH Models: A Dynamic Gerber-Shiu Approach, *North American Actuarial Journal*, **8**, 17-31.

Society of Actuaries Group Annuity Valuation Table Task Force. (1995). 1994 Group Annuity Mortality Table and 1994 Group Annuity Reserving Table, *Transactions of the Society of Actuaries*, **XLVII**, 865-915.

Sodha, S. (2005). Housing-Risk, Income-Poor: The Potential of Housing Wealth in Old Age. A paper for Housing Across the Lifecycle.
http://ippr.nvisage.uk.com/ecomm/files/housing_rich.pdf.

Tabeau, E. (2001). A Review of Demographic Forecasting Models for Mortality. In Tabeau, E., Jeths, A.V.D.B. and Heathcote, C. (eds.), *Forecasting Mortality in Developed Countries - Insights from a Statistical, Demographic and Epidemiological Perspective.* Dordrecht/Boston/London: Kluwer Academic Publishers.

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

The World Bank. (1994). *Averting the Old Age Crisis: Policies to Protect the Old and Promote Growth*. Oxford, England: Oxford University Press.

Tsay, R.S. (1986). Time Series Model Specification in the Presence of Outliers, *Journal of American Statistical Association*, **81**, 132-141.

Tsay, R.S. (2002). *Analysis of Financial Time Series.* New York: John Wiley and Sons.

Tuljapurkar, S. (1997). Taking the Measure of Uncertainty, *Nature*, **387**, 760-761.

Tuljapurkar, S. (2005). Future Mortality: A Bumpy Road to Shangri-La? *Science of Aging Knowledge Environment*, **14**, pe9.

Tuljapurkar, S., Li, N. and Boe, C. (2000). A Universal Pattern of Mortality Decline in the G7 Countries, *Nature*, **405**, 789-792.

United Nations. (1997). Report of the Working Group on Projecting Old-Age Mortality and Its Consequences. New York: United Nations.

U.S. Census Bureau (2004). Monthly Population Estimates, 1990-2000. Available at www.census.gov/popest/archives/1990s/nat_detail.html. Accessed January 25, 2004.

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.

Vincent, P. (1951) La Mortalité des Vieillards. *Population*, **6**, 181–204.

Watts, K.A, Dupuis, D.J. and Jones, B.L. (2006). An Extreme Value Analysis of Advanced Age Mortality Data, *North American Actuarial Journal*, **10**, 162-178.

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

Wilkie, A.D. (1995). More on a Stochastic Asset Model for Actuarial Use, *British Actuarial Journal*, **1**, 777-964.

Willekens, F.J. (1990). Demographic Forecasting; State-of-art and Research Needs. In: C.A. Hazeu and G.A.B. Frinkin (eds.): *Emerging Issues in Demographic Research.* Elsevier Science Publishers B.V., pp. 9-75.

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.

Wilmoth, J.R., Andreev, K., Jdanov, D. and Glei, D.A. (2005). Methods Protocol for the Human Mortality Database. Available at: www.mortality.org.

Zelterman, D.(1993). A Semiparametric Bootstrap Technique for Simulating Extreme Order Statistics, *Journal of the American Statistical Association*, **88**, 477-485.

Appendix

Base mortality tables ($m_{x,2001}$) and mortality improvement scales (z_x, $u_{1,x}$, $u_{2,x}$, and $u_{3,x}$) for the Canadian insured lives, ultimate, composite smoker/non-smoker.

	Male				Female			
Age (x)	$m_{x,2001}$	z_x	$u_{1,x}$	$u_{2,x}$	$m_{x,2001}$	z_x	$u_{1,x}$	$u_{2,x}$
15	0.000593	-0.029211	0.000005	0.003057	0.000242	-0.014534	0.000006	0.000938
16	0.000613	-0.030018	0.000004	0.003364	0.000249	-0.015623	0.000004	0.001063
17	0.000633	-0.030819	0.000005	0.003598	0.000257	-0.016680	0.000003	0.001198
18	0.000654	-0.031619	0.000007	0.003785	0.000265	-0.017719	0.000002	0.001345
19	0.000675	-0.032421	0.000008	0.003949	0.000273	-0.018757	0.000001	0.001506
20	0.000697	-0.033229	0.000009	0.004115	0.000282	-0.019809	0.000000	0.001682
21	0.000718	-0.034042	0.000010	0.004300	0.000292	-0.020885	0.000000	0.001874
22	0.000739	-0.034853	0.000010	0.004502	0.000302	-0.021982	0.000001	0.002081
23	0.000761	-0.035653	0.000009	0.004714	0.000312	-0.023095	0.000002	0.002300
24	0.000783	-0.036433	0.000009	0.004933	0.000324	-0.024220	0.000004	0.002532
25	0.000806	-0.037184	0.000010	0.005151	0.000336	-0.025351	0.000006	0.002775
26	0.000830	-0.037894	0.000012	0.005362	0.000350	-0.026482	0.000009	0.003026
27	0.000855	-0.038542	0.000015	0.005557	0.000365	-0.027603	0.000012	0.003285
28	0.000881	-0.039107	0.000020	0.005727	0.000381	-0.028706	0.000014	0.003548
29	0.000908	-0.039569	0.000026	0.005861	0.000399	-0.029780	0.000015	0.003813
30	0.000937	-0.039905	0.000034	0.005950	0.000420	-0.030817	0.000015	0.004078
31	0.000968	-0.040103	0.000044	0.005990	0.000443	-0.031806	0.000014	0.004340

Base mortality tables ($m_{x,2001}$) and mortality improvement scales (z_x, $u_{1,x}$, $u_{2,x}$, and $u_{3,x}$) for the Canadian insured lives, ultimate, composite smoker/non-smoker (cont'd).

	Male				Female			
Age (x)	$m_{x,2001}$	z_x	$u_{1,x}$	$u_{2,x}$	$m_{x,2001}$	z_x	$u_{1,x}$	$u_{2,x}$
32	0.001001	-0.040161	0.000054	0.005986	0.000468	-0.032738	0.000011	0.004594
33	0.001035	-0.040081	0.000064	0.005944	0.000497	-0.033606	0.000008	0.004836
34	0.001073	-0.039866	0.000071	0.005872	0.000529	-0.034399	0.000005	0.005061
35	0.001113	-0.039516	0.000076	0.005775	0.000565	-0.035108	0.000003	0.005264
36	0.001156	-0.039039	0.000077	0.005657	0.000606	-0.035732	0.000001	0.005443
37	0.001203	-0.038454	0.000075	0.005515	0.000651	-0.036279	0.000001	0.005600
38	0.001255	-0.037780	0.000072	0.005347	0.000702	-0.036758	0.000002	0.005738
39	0.001314	-0.037036	0.000068	0.005149	0.000759	-0.037178	0.000004	0.005861
40	0.001381	-0.036241	0.000064	0.004918	0.000822	-0.037550	0.000006	0.005970
41	0.001457	-0.035409	0.000061	0.004658	0.000892	-0.037875	0.000009	0.006067
42	0.001543	-0.034544	0.000060	0.004382	0.000970	-0.038145	0.000012	0.006152
43	0.001643	-0.033649	0.000061	0.004101	0.001057	-0.038353	0.000016	0.006222
44	0.001757	-0.032729	0.000064	0.003831	0.001154	-0.038489	0.000020	0.006278
45	0.001890	-0.031785	0.000070	0.003584	0.001263	-0.038545	0.000024	0.006318
46	0.002042	-0.030830	0.000081	0.003367	0.001384	-0.038502	0.000027	0.006334
47	0.002218	-0.029884	0.000097	0.003176	0.001520	-0.038328	0.000031	0.006312
48	0.002418	-0.028970	0.000123	0.003009	0.001672	-0.037993	0.000035	0.006236
49	0.002646	-0.028108	0.000159	0.002861	0.001842	-0.037463	0.000041	0.006090
50	0.002905	-0.027321	0.000208	0.002730	0.002033	-0.036708	0.000047	0.005859
51	0.003197	-0.026625	0.000270	0.002614	0.002247	-0.035731	0.000055	0.005546
52	0.003526	-0.026033	0.000340	0.002516	0.002486	-0.034575	0.000063	0.005174
53	0.003896	-0.025557	0.000414	0.002438	0.002753	-0.033287	0.000072	0.004766
54	0.004308	-0.025209	0.000489	0.002381	0.003049	-0.031912	0.000079	0.004349
55	0.004768	-0.025000	0.000559	0.002350	0.003376	-0.030497	0.000085	0.003946
56	0.005279	-0.024941	0.000622	0.002345	0.003739	-0.029084	0.000089	0.003571

Base mortality tables ($m_{x,2001}$) and mortality improvement scales (z_x, $u_{1,x}$, $u_{2,x}$, and $u_{3,x}$) for the Canadian insured lives, ultimate, composite smoker/non-smoker (cont'd).

	Male				Female			
Age (x)	$m_{x,2001}$	z_x	$u_{1,x}$	$u_{2,x}$	$m_{x,2001}$	z_x	$u_{1,x}$	$u_{2,x}$
57	0.005845	-0.025042	0.000677	0.002369	0.004139	-0.027707	0.000092	0.003230
58	0.006472	-0.025312	0.000723	0.002425	0.004580	-0.026404	0.000094	0.002925
59	0.007164	-0.025761	0.000758	0.002515	0.005068	-0.025211	0.000097	0.002660
60	0.007926	-0.026397	0.000782	0.002641	0.005607	-0.024163	0.000101	0.002439
61	0.008767	-0.027179	0.000796	0.002797	0.006200	-0.023270	0.000105	0.002260
62	0.009696	-0.028025	0.000806	0.002971	0.006855	-0.022519	0.000109	0.002118
63	0.010727	-0.028851	0.000815	0.003148	0.007576	-0.021896	0.000114	0.002007
64	0.011874	-0.029571	0.000831	0.003317	0.008370	-0.021386	0.000118	0.001920
65	0.013155	-0.030105	0.000856	0.003463	0.009243	-0.020977	0.000121	0.001851
66	0.014587	-0.030420	0.000891	0.003575	0.010203	-0.020644	0.000122	0.001795
67	0.016184	-0.030524	0.000931	0.003647	0.011261	-0.020362	0.000124	0.001749
68	0.017962	-0.030427	0.000971	0.003668	0.012429	-0.020102	0.000124	0.001708
69	0.019935	-0.030140	0.001007	0.003630	0.013720	-0.019835	0.000125	0.001667
70	0.022119	-0.029671	0.001035	0.003527	0.015152	-0.019535	0.000126	0.001624
71	0.024536	-0.029047	0.001052	0.003366	0.016742	-0.019194	0.000128	0.001575
72	0.027215	-0.028303	0.001059	0.003167	0.018515	-0.018820	0.000131	0.001522
73	0.030190	-0.027473	0.001054	0.002952	0.020497	-0.018420	0.000138	0.001463
74	0.033499	-0.026594	0.001038	0.002739	0.022719	-0.018000	0.000147	0.001398
75	0.037189	-0.025696	0.001011	0.002548	0.025220	-0.017567	0.000161	0.001327
76	0.041305	-0.024768	0.000977	0.002378	0.028040	-0.017111	0.000181	0.001250
77	0.045900	-0.023768	0.000944	0.002212	0.031219	-0.016612	0.000210	0.001169
78	0.051030	-0.022655	0.000920	0.002037	0.034808	-0.016052	0.000250	0.001083
79	0.056759	-0.021388	0.000911	0.001836	0.038860	-0.015410	0.000304	0.000993

Base mortality tables ($m_{x,2001}$) and mortality improvement scales (z_x, $u_{1,x}$, $u_{2,x}$, and $u_{3,x}$) for the Canadian insured lives, ultimate, composite smoker/non-smoker (cont'd).

	Male					Female				
Age (x)	$m_{x,2001}$	z_x	$u_{1,x}$	$u_{2,x}$	$u_{3,x}$	$m_{x,2001}$	z_x	$u_{1,x}$	$u_{2,x}$	$u_{3,x}$
80	0.063156	-0.019931	0.000924	0.001597	6.23×10^{-6}	0.043435	-0.014670	0.000373	0.000901	2.76×10^{-6}
81	0.070280	-0.018304	0.000958	0.001330	6.33×10^{-6}	0.048592	-0.013836	0.000456	0.000807	3.19×10^{-6}
82	0.078176	-0.016558	0.001005	0.001056	6.44×10^{-6}	0.054382	-0.012928	0.000550	0.000713	3.65×10^{-6}
83	0.086884	-0.014746	0.001058	0.000797	6.57×10^{-6}	0.060856	-0.011965	0.000649	0.000621	4.12×10^{-6}
84	0.096432	-0.012921	0.001111	0.000577	6.70×10^{-6}	0.068060	-0.010966	0.000750	0.000534	4.59×10^{-6}
85	0.106916	-0.011131	0.001156	0.000415	6.82×10^{-6}	0.076021	-0.009949	0.000850	0.000451	5.06×10^{-6}
86	0.118414	-0.009406	0.001191	0.000307	6.93×10^{-6}	0.084806	-0.008934	0.000944	0.000375	5.50×10^{-6}
87	0.131010	-0.007762	0.001216	0.000239	7.01×10^{-6}	0.094486	-0.007941	0.001031	0.000307	5.91×10^{-6}
88	0.144792	-0.006216	0.001230	0.000195	7.05×10^{-6}	0.105137	-0.006988	0.001107	0.000247	6.27×10^{-6}
89	0.159854	-0.004786	0.001234	0.000162	7.06×10^{-6}	0.116841	-0.006094	0.001170	0.000196	6.57×10^{-6}
90	0.176297	-0.003486	0.001226	0.000126	7.03×10^{-6}	0.129683	-0.005276	0.001216	0.000155	6.82×10^{-6}
91	0.194224	-0.002313	0.001207	0.000087	6.95×10^{-6}	0.143755	-0.004532	0.001245	0.000123	7.01×10^{-6}
92	0.213749	-0.001258	0.001177	0.000053	6.82×10^{-6}	0.159152	-0.003852	0.001258	0.000098	7.12×10^{-6}
93	0.234986	-0.000308	0.001138	0.000028	6.65×10^{-6}	0.175974	-0.003223	0.001255	0.000080	7.18×10^{-6}
94	0.258060	0.000000	0.001089	0.000020	6.44×10^{-6}	0.194329	-0.002636	0.001237	0.000065	7.18×10^{-6}
95	0.283100	0.000000	0.001031	0.000033	6.19×10^{-6}	0.214326	-0.002082	0.001203	0.000054	7.13×10^{-6}
96	0.310241	0.000000	0.000965	0.000060	5.92×10^{-6}	0.236081	-0.001567	0.001155	0.000045	7.03×10^{-6}
97	0.339623	0.000000	0.000891	0.000090	5.62×10^{-6}	0.259715	-0.001106	0.001093	0.000038	6.91×10^{-6}
98	0.371395	0.000000	0.000812	0.000113	5.31×10^{-6}	0.285354	-0.000710	0.001018	0.000032	6.76×10^{-6}
99	0.405708	0.000000	0.000728	0.000118	4.99×10^{-6}	0.313126	-0.000393	0.000929	0.000026	6.59×10^{-6}