On Simulation-Based Approaches to Risk Measurement in Mortality with Specific Reference to Binomial Lee-Carter Modelling

Steve Haberman* and Arthur Renshaw

Presented at the Living to 100 and Beyond Symposium

Orlando, Fla.

January 7-9, 2008

Copyright 2008 by the Society of Actuaries.
All rights reserved by the Society of Actuaries. Permission is granted to make brief excerpts for a published review. Permission is also granted to make limited numbers of copies of items in this monograph for personal, internal, classroom or other instructional use, on condition that the foregoing copyright notice is used so as to give reasonable notice of the Society's copyright. This consent for free limited copying without prior consent of the Society does not extend to making copies for general distribution, for advertising or promotional purposes, for inclusion in new collective works or for resale.

* Cass Business School, City University, London, EC1Y 8TZ, UK

Abstract

This paper develops the binomial version of the Lee-Carter model and provides a comparative study of simulation strategies for assessing risk in mortality rate predictions and associated estimates of life expectancy and annuity values in both period and cohort frameworks.

Keywords: Binomial modelling; Over-dispersion; Joint-modelling; Mortality projections; Mortality statistics; Simulated risk

1. Introduction

The Lee-Carter model for modelling mortality rates, introduced by Lee and Carter (1992), has become one of the most widely used reference techniques, being adopted by various demographic agencies, and the leading statistical mortality model in the demographic literature: Deaton and Paxson (2001).

As noted by Lee and Carter (1992) and later researchers, it is important to be able to quantify the uncertainty in projections through the computation of prediction intervals. This is particularly relevant for insurance companies and pensions institutions which are transacting annuity type products and which need to understand and quantify the risks inherent in any portfolio and demonstrate appropriate risk management strategies to regulators, customers, pension plan members, shareholders and other stakeholders.

Analytical derivations are rarely possible for the LC family of models. This is because two different sources of uncertainty need to be combined: errors in the estimation of the parameters of the LC model, and forecast errors in the projected ARIMA parameters. Also the indices of interest (e.g., hazard rates, life expectancies, expected annuity values) are complex, non-linear functions of the Lee-Carter parameters $\alpha_x, \beta_x, \kappa_t$ and the ARIMA parameters used for modelling the time dependence of κ_t.

In order to address this problem, the suggestion in the literature is to use simulation techniques (Brouhns et al. (2002), (2005); Koissi et al. (2006)) as a means of measuring risk when modelling dynamic mortality rates and their impact on future predictions of life expectancy and annuity values. Renshaw and Haberman (2008) have conducted an extensive comparative study of three such simulation strategies (denoted A, B, C), in the context of Poisson bilinear (Lee-Carter) modelling and linear modelling, both of which involve extrapolation. In this paper, we extend the study to binomial bilinear and linear modelling with extrapolation. However, as noted by Renshaw and Haberman (2008), simulation Strategy B (parametric Monte-Carlo) should not be used for risk assessment purposes since different choices of the constraints which are needed to fit the Lee-Carter model result in widely differing confidence and prediction intervals.

Hence, in this paper, we do not present Strategy B but we continue to refer to the other two strategies as Strategy A (semi parametric bootstrap) and Strategy C (residual bootstrap), for continuity of description.

The binomial model specifications, involving three different link functions in combination with bilinear and linear parametric structures, are described in Section 2, together with modelling, fitting, model extrapolation and the two simulation strategies. In Section 3, a comparative study of simulated life expectancy and fixed rate annuity, confidence and prediction intervals, for the U.K. male pensioner 1983-2003 mortality experience, is presented as an illustration. In addition to comparing the range of binomial based intervals, the corresponding Poisson based intervals are included for comparison. In Section 4, we describe how to conduct binomial modelling with extra provision for variable dispersion and investigate the effect which this has on Strategy C, in terms of simulated confidence and prediction intervals, for the U.K. male pensioner 1983-2003 mortality experience, in Section 5. In Section 6, we offer some concluding comments.

2. Methodology

2.1 Model Specification

Let the random variable D_{xt} denote the number of deaths in a population at age *x* in period *t*. It is envisaged that a rectangular data array $\left(d_{xt}, e_{xt}^{i}\right)$ is available for analysis, comprising the respective numbers of deaths and matching initial exposures to the risk of death. Cross-classification is by individual year of age $x = x_1, x_2, ..., x_k$ and by individual calendar year $t = t_1, t_2, ..., t_n$. Empty data cells (zero exposure) are identified by the allocation of zero-one indicator weights, denoted $\omega_{xt} = 1$ if $e_{xt}^{i} > 0$ and $\omega_{xt} = 0$ if $e_{xt}^{i} = 0$.

We target the probability of death q_{xt} by modelling the numbers of deaths as independent binomial variables $D_{xt} \sim bin\left(e_{xt}^{i}, q_{xt}\right)$, and write

$$\mathrm{E}\left(D_{xt}\right)=e_{xt}^{i}q_{xt},\ \mathrm{Var}\left(D_{xt}\right)=\varphi\frac{\mathrm{V}\left(\mathrm{E}\left(D_{xt}\right)\right)}{\omega_{xt}};\ \mathrm{V}\left(u\right)=u\left(1-\frac{u}{e_{xt}^{i}}\right)$$

with a scale parameter $\varphi=1$ and characteristic variance function V. We are interested in the following two parametric predictor structures:

$$\text{LC: } \eta_{xt}=\alpha_x+\beta_x\kappa_t;\ \sum_x\beta_x=1,\ \kappa_{t_n}=0 \tag{1}$$

$$\text{LP: } \eta_{xt}=\alpha_x+\beta_x\left(t_n-t\right)+\gamma_t\left(t_n-t\right);\ \sum_x\beta_x=1,\ \gamma_{t_1}=\gamma_{t_n}=0 \tag{2}$$

The non-linear (LC) structure is that generally attributed to Lee and Carter (1992), subject to a change in the pair of parameter constraints that are normally chosen (see later). This binomial version of the Lee-Carter model has been explored also by Cossette et al. (2007). The linear predictor (LP) structure has been suggested and explored previously in a similar context by Renshaw and Haberman (2003). The purpose of the term involving γ_t under LP modelling is to ensure that the annual actual and expected total deaths are the same, (in addition to improving the quality of the fit); these terms play no role when the model is used for extrapolation (see later). The structures in (1) and (2) are linked to $\mathrm{E}\left(D_{xt}\right)$, by one of the following 1-1 functions:

I: complementary log-log link-

$$\eta_{xt}=\log\left\{-\log\left(1-q_{xt}\right)\right\}\Leftrightarrow q_{xt}=1-\exp\left(-\exp\eta_{xt}\right) \tag{3}$$

II: log-odds link- $\eta_{xt}=\log\left(\frac{q_{xt}}{1-q_{xt}}\right)\Leftrightarrow q_{xt}=\frac{\exp\eta_{xt}}{1+\exp\eta_{xt}}$ (4)

III: probit link- $\eta_{xt}=\Phi^{-1}\left(q_{xt}\right)\Leftrightarrow q_{xt}=\Phi\left(\eta_{xt}\right)$. (5)

2.2 Force of Mortality

When mapping q_{xt} to the associated force of mortality μ_{xt}, we use the approximate relationship

$$\mu_{xt} \approx -\log\left(1-q_{xt}\right) \Leftrightarrow q_{xt} \approx 1-\exp\left(-\mu_{xt}\right) \tag{6}$$

throughout, irrespective of the choice of both the parameterised predictor ((1) or (2)) and the link function ((3) –(5)). Thus, as a special case, under the non-linear LC predictor (1) in combination with the complementary log-log link (3), but not otherwise, it follows that

$$\mu_{xt} = \exp\left(\alpha_x + \beta_x \kappa_t\right)$$

which is the structure of the standard Lee-Carter Poisson model, incorporating the log link, described in Lee and Carter (1992), Renshaw and Haberman (2008) and elsewhere in the literature.

2.3 Model Fitting

Model fitting is by maximising the binomial log likelihood, or, equivalently, minimising the binomial deviance

$$D\left(d_{xt}, e_{xt}^i q_{xt}\right) = 2\sum_{x,t} \omega_{xt} \left\{ d_{xt} \log\left(\frac{d_{xt}}{e_{xt}^i q_{xt}}\right) + \left(e_{xt}^i - d_{xt}\right) \log\left(\frac{e_{xt}^i - d_{xt}}{e_{xt}^i - e_{xt}^i q_t}\right) \right\}. \tag{7}$$

For LC modelling, in combination with any of the three links (3)-(5), because of the non-linear nature of the parametric predictor (1), we use the following iterative fitting procedure:

set starting values $\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t$; compute $\hat{d}_{xt} = e^i_{xt}\hat{q}_{xt}$

$\downarrow$

update $\hat{\alpha}_x$, compute $\hat{d}_{xt}$

update $\hat{\kappa}_t$, adjust s.t. $\sum_t \hat{\kappa}_t = 0$, compute $\hat{d}_{xt}$

update $\hat{\beta}_x$, compute $\hat{d}_{xt}$

compute $D\left(d_{xt}, \hat{d}_{xt}\right)$

$\downarrow$

repeat the above cycle, stop when $D\left(d_{xt}, \hat{d}_{xt}\right)$ converges

The adjustment $\sum_t \hat{\kappa}_t = 0$ in the core of the procedure, complies with the more usual LC parametric identification constraints $\sum_x \beta_x = 1,\ \sum_t \kappa_t = 0$, which are finally adjusted to comply with (1) using the appropriate η_{xt} invariance transformations, once convergence has been achieved. Typical starting values are:

$$\hat{\alpha}_x = \frac{1}{n}\sum_t \omega_{xt} \log\left(\frac{d_{xt}}{e_{xt}}\right),\ \hat{\beta}_x = \frac{1}{k},\ \hat{\kappa}_t = 0\,.$$

Details of the updating relationships, determined according to

$$u(\theta) = \theta - \frac{\partial D}{\partial \theta} \Big/ \frac{\partial^2 D}{\partial \theta^2}$$

are given in Table 1. Here, $u(\theta)$ denotes the updated version of the typical parameter θ and this requires the first and second order partial derivatives of the model deviance *D*.

TABLE 1

LC Parameter Updating Relationships

Comple-mentary Log-log link	$u(\hat{\alpha}_x) = \hat{\alpha}_x - \dfrac{\sum_t \omega_{xt}\left(d_{xt} - \hat{d}_{xt}\right)\log \hat{p}_{xt} / \hat{q}_{xt}}{\sum_t \omega_{xt}\left\{d_{xt}\left(\hat{q}_{xt} + \hat{p}_{xt}\log \hat{p}_{xt}\right) - \hat{d}_{xt}\hat{q}_{xt}\right\}\log \hat{p}_{xt} / \hat{q}_{xt}^2}$
	$u(\hat{\beta}_x) = \hat{\beta}_x - \dfrac{\sum_t \omega_{xt}\left(d_{xt} - \hat{d}_{xt}\right)\hat{\kappa}_t \log \hat{p}_{xt} / \hat{q}_{xt}}{\sum_t \omega_{xt}\left\{d_{xt}\left(\hat{q}_{xt} + \hat{p}_{xt}\log \hat{p}_{xt}\right) - \hat{d}_{xt}\hat{q}_{xt}\right\}\hat{\kappa}_t^2 \log \hat{p}_{xt} / \hat{q}_{xt}^2}$
	$u(\hat{\kappa}_t) = \hat{\kappa}_t - \dfrac{\sum_x \omega_{xt}\left(d_{xt} - \hat{d}_{xt}\right)\hat{\beta}_x \log \hat{p}_{xt} / \hat{q}_{xt}}{\sum_x \omega_{xt}\left\{d_{xt}\left(\hat{q}_{xt} + \hat{p}_{xt}\log \hat{p}_{xt}\right) - \hat{d}_{xt}\hat{q}_{xt}\right\}\hat{\beta}_x^2 \log \hat{p}_{xt} / \hat{q}_{xt}^2}$
Log-odds link	$u(\hat{\alpha}_x) = \hat{\alpha}_x + \dfrac{\sum_t \omega_{xt}\left(d_{xt} - \hat{d}_{xt}\right)}{\sum_t \omega_{xt}\hat{d}_{xt}\hat{p}_{xt}}$
	$u(\hat{\beta}_x) = \hat{\beta}_x + \dfrac{\sum_t \omega_{xt}\left(d_{xt} - \hat{d}_{xt}\right)\hat{\kappa}_t}{\sum_t \omega_{xt}\hat{d}_{xt}\hat{p}_{xt}\hat{\kappa}_t^2}$
	$u(\hat{\kappa}_t) = \hat{\kappa}_t + \dfrac{\sum_x \omega_{xt}\left(d_{xt} - \hat{d}_{xt}\right)\hat{\beta}_x}{\sum_x \omega_{xt}\hat{d}_{xt}\hat{p}_{xt}\hat{\beta}_x^2}$
Probit link	$u(\hat{\alpha}_x) = \hat{\alpha}_x + \dfrac{\sum_t \omega_{xt}\dfrac{d_{xt} - \hat{d}_{xt}}{\hat{q}_{xt}\hat{p}_{xt}}\phi(\hat{\eta}_{xt})}{\sum_t \omega_{xt}\left\{\dfrac{d_{xt}\left(\hat{p}_{xt} - \hat{q}_{xt}\right) + \hat{d}_{xt}\hat{q}_{xt}}{\hat{p}_{xt}^2\hat{q}_{xt}^2}\phi(\hat{\eta}_{xt}) + \dfrac{d_{xt} - \hat{d}_{xt}}{\hat{p}_{xt}\hat{q}_{xt}}\hat{\eta}_{xt}\right\}\phi(\hat{\eta}_{xt})}$
	$u(\hat{\beta}_x) = \hat{\beta}_x + \dfrac{\sum_t \omega_{xt}\dfrac{d_{xt} - \hat{d}_{xt}}{\hat{q}_{xt}\hat{p}_{xt}}\phi(\hat{\eta}_{xt})\hat{\kappa}_t}{\sum_t \omega_{xt}\left\{\dfrac{d_{xt}\left(\hat{p}_{xt} - \hat{q}_{xt}\right) + \hat{d}_{xt}\hat{q}_{xt}}{\hat{p}_{xt}^2\hat{q}_{xt}^2}\phi(\hat{\eta}_{xt}) + \dfrac{d_{xt} - \hat{d}_{xt}}{\hat{p}_{xt}\hat{q}_{xt}}\hat{\eta}_{xt}\right\}\phi(\hat{\eta}_{xt})\hat{\kappa}_t^2}$

	$u(\hat{\kappa}_t) = \hat{\kappa}_t + \dfrac{\sum_x \omega_{xt} \dfrac{d_{xt} - \hat{d}_{xt}}{\hat{q}_{xt}\hat{p}_{xt}} \phi(\hat{\eta}_{xt}) \hat{\beta}_x}{\sum_x \omega_{xt} \left\{ \dfrac{d_{xt}(\hat{p}_{xt} - \hat{q}_{xt}) + \hat{d}_{xt}\hat{q}_{xt}}{\hat{p}_{xt}^2 \hat{q}_{xt}^2} \phi(\hat{\eta}_{xt}) + \dfrac{d_{xt} - \hat{d}_{xt}}{\hat{p}_{xt}\hat{q}_{xt}} \hat{\eta}_{xt} \right\} \phi(\hat{\eta}_{xt}) \hat{\beta}_x^2}$

Note: $\hat{d}_{xt} = e_{xt}^i \hat{q}_{xt}$, $\hat{p}_{xt} = 1 - \hat{q}_{xt}$, $\phi(\hat{\eta}_{xt}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\hat{\eta}_{xt}^2\right)$.

Under LP modelling, in combination with any of the three links, we note that the predictor (2) is linear in the parameters, and so the model may be fitted using any statistical package with the facility for fitting GLMs, such as GLIM (Francis et al. (1993)). While the constraint $\sum_x \beta_x = 1$ does not feature as part of the fitting process, it is imposed in (2), (subject to the invariance of η_{xt}), in order to facilitate parameter comparison with (1).

2.4 Model Extrapolation

Denoting the link functions (3)-(5) collectively by *g*, so that

$$\eta_{xt} = g(q_{xt}) \Leftrightarrow q_{xt} = g^{-1}(\eta_{xt}),$$

we conduct the extrapolation according to

$$\text{LC}: \dot{q}_{x,t_n+s} = g^{-1}\left(\hat{\alpha}_x + \hat{\beta}_x \dot{\kappa}_{t_n+s}\right),\ s > 0$$

$$\text{LP}: \dot{q}_{x,t_n+s} = g^{-1}\left(\hat{\alpha}_x - \hat{\beta}_x s\right),\ s > 0. \tag{8}$$

Under LC modelling, $\dot{\kappa}_{t_n+s}$, $s > 0$ denotes the forecast value generated in the analysis of the time series $\{\hat{\kappa}_t : t = t_1, t_2, ..., t_n\}$, typically obtained using an ARIMA model. Under LP modelling, based on expression (2), we extrapolate the main period effects term

$\hat{\beta}_x(t_n - t)$, while setting the extrapolation of the period fitting adjustment term $\hat{\gamma}_t(t_n - t)$ equal to zero. Under LC modelling, for the special case, when the time series analysis leads to the adoption of a random walk with drift parameter θ

$$\text{LC}: \dot{q}_{x,t_n+s} = g^{-1}\left(\hat{\alpha}_x - \hat{\theta}\hat{\beta}_x s\right),\ s > 0;\ \theta = \hat{\kappa}_{t_1} / (t_n - t_1),$$

which is directly comparable with extrapolation under the LP model, represented by (8).

2.5 Simulation Strategies

We are primarily concerned with the simulation of confidence and prediction intervals for the age-period statistics of interest: life expectancy $e_x(t)$ and fixed rate annuity $a_x(t)$ with a discount factor *v*, under both the cohort method of computing, so that

$$\hat{e}_x(t) = \frac{\sum_{i\geq 0} l_{x+i}(t+i)\left\{1 - \tfrac{1}{2}\hat{q}_{x+i}(t+i)\right\}}{l_x(t)},\ \hat{a}_x(t) = \frac{\sum_{i\geq 1} l_{x+i}(t+i)v^i}{l_x(t)}$$

where

$$l_{x+1}(t+1) = \left\{1 - \hat{q}_x(t)\right\} l_x(t),$$

and the period method of computing, in which the variation in *t* in these expressions is suppressed.

Following Renshaw and Haberman (2008), we consider two of the three possible simulation strategies that have been discussed in the literature. These may be re-expressed for binomial Lee-Carter modelling purposes as follows: at each simulation *j* (= 1, 2, …, *N*)

A: simulate responses $d_{xt}^{(j)}$ by sampling $bin\left(e_{xt}^{i}, \hat{q}_{xt}\right)$,

then compute $q_{xt}^{(j)}$ by fitting $D_{xt}^{(j)} \sim bin\left(e_{xt}^{i}, q_{xt}^{(j)}\right)$,

all subject to the preservation of empty data cells,

before computing the statistics of interest.

C: simulate residuals $r_{xt}^{(j)}$ by sampling $\{r_{xt}\}$ with replacement,

map the simulated residuals to simulated responses $r_{xt}^{(j)} \mapsto d_{xt}^{(j)}$,

then compute $q_{xt}^{(j)}$ by fitting $D_{xt}^{(j)} \sim bin\left(e_{xt}^{i}, q_{xt}^{(j)}\right)$,

all subject to the preservation of empty data cells,

before computing the statistics of interest.

Under Strategy C, the bootstrap deviance residuals $r_{xt}^{(j)}$ are mapped by solving the relationship

$$r_{xt}^{(j)} = sign\left(d_{xt} - \hat{d}_{xt}\right)\sqrt{2}\sqrt{d_{xt}\log\left(\frac{d_{xt}}{\hat{d}_{xt}}\right) + \left(e_{xt}^{i} - d_{xt}\right)\log\left(\frac{e_{xt}^{i} - d_{xt}}{e_{xt}^{i} - \hat{d}_{xt}}\right)}$$

for d_{xt} given $\hat{d}_{xt}, e_{xt}^{i}$. Hence, for clarity of notation, on suppressing the suffices *x* and *t* and using the prefix * instead of (*j*) to denote bootstrap values, this implies we require the appropriate root d^{*} of

$$g\left(d\right) = d\log d + \left(e^{i} - d\right)\log\left(e^{i} - d\right) - \hat{b}d - \hat{c}^{*}$$

where

$$\hat{b} = \log\hat{d} - \log\left(e^{i} - \hat{d}\right),\ \hat{c}^{*} = \frac{r^{*2}}{2} + e^{i}\log\left(e^{i} - \hat{d}\right)$$

when mapping $r^* \mapsto d^*$. The derivatives of g

$$g'(d) = \log d - \log\left(e^i - d\right) - \hat{b}, \quad g''(d) = \frac{e^i}{d\left(e^i - d\right)} > 0 \ \forall \ 0 < d < e^i$$

imply that $d = \hat{d}$, where $g\left(\hat{d}\right) < 0$, is a minimum and the graph of $g(d)$ *vs* d $\left(0 < d < e^i\right)$ is concave. There are either one or two roots in this range. The required root is determined by the sign of the residual r^*, satisfying $d > \hat{d}$ when $r^* > 0$ and $d < \hat{d}$ when $r^* < 0$. The root is readily determined by the Newton-Raphson iterative process using starting values for d which match the above constraints, within the limits of the domain $\left\{d : 0 < d < e^i\right\}$.

3. Application: UK Male Pensioner 1983-2003 Mortality Experience

This application uses data from the U.K. male pensioner 1983-2003 mortality experience, with ages ranging from 51 to 104 and with roughly 95 percent of the total exposure to the risk of death in the age range 62 to 89. Approximately 5 percent of the 21×54 data cells are empty. Insight into the data is provided by Figure 1, in which log crude mortality rates are plotted against period for select ages (continuous profiles). We have decided not to update the data, so that in addition to comparing the outcomes from binomial modelling under both predictor structures (1)-(2) in combination with each of the three link functions (3)-(5), we can additionally compare the outcomes under Poisson modelling in combination with both predictors and the log link, as reported in Renshaw and Haberman (2008).

Model fitting proceeds as described in Section 2.3 and the resulting parameter estimates depicted in Figures 2a, 2b and 2c. Here, in each figure, we devote a separate row of picture frames for each link function (in the prescribed order I to III) while superimposing parameter estimates for the two predictors structures within the left hand frames, and depicting their differences in the matching right hand frames. Thus, the

parameter estimates for α_x, β_x and κ_t (where, in this context, we refer loosely to $\kappa_t = t_n - t$ under LP modelling as a parameter), are shown respectively in Figures 2a, 2b and 2c. We note the near identical patterns in superimposed parameters throughout. However, the differences in the respective magnitudes of $\hat{\alpha}_x$ and $\hat{\kappa}_t$ for the probit link, compared with the other two links, are noteworthy. Obviously, such comparisons are only possible because of the equivalent parameter constraints placed on the two predictor structures (1) and (2).

The scaled versions, $\hat{\gamma}_t(t_n - t)$, of the remaining γ_t parameter estimates under LP modelling, are depicted in the right hand frames of Figure 3. In the left hand frames of Figure 3, we display the annual differences between the actual and expected total deaths, with plots superimposed by predictor type. Here, both the magnitude of the differences and the lack of any systematic patterns over period indicate a good model fit. On comparing like with like, the relative smaller differences under LP compared with LC modelling, for links I and III are noteworthy, together with the zero differences generated under LP modelling in the context of the log-odds link II. It is well known that this last feature is directly attributable to the inclusion of the parameters γ_t in the linear predictor (2) under Poisson log link modelling, known as the 'Poisson trick,' as reported for this data set in Renshaw and Haberman (2008). However, we are unclear as to whether this is a general property of binomial LP modelling under the canonical log-odds link.

Deviance residual plots under both predictor types (1)-(2) in combination with all three links (3)-(5) generate similar patterns when plotted. We illustrate these patterns, which are indicative of a good fit, for LP modelling under the complementary log-log link in Figure 4. Full details for the other cases are available from the authors. In addition, we have inserted the respective fitted log mortality rates under LC and LP binomial modelling, in combination with the complementary log-log link, against the background of the crude log mortality rates in Figure 1.

Mortality rates are extrapolated as described in Section 2.4, using the random walk with drift in the case of LC modelling and using the linear extrapolation of $\hat{\beta}_x(t_n - t)$ only in the case of LP modelling, while the right hand frames in Figure 3 are consistent with setting the extrapolation of $\hat{\gamma}_t(t_n - t)$ to zero (for LP modelling). There is an argument for using the ARIMA (1,1,0) model for males but on grounds of simplicity and convenience we have adopted the same ARIMA (0,1,0) model for both genders.

We compare the simulated point and prediction intervals for the U.K. male pensioner's mortality experience for life expectancy and a fixed life annuity value, based on a 4 percent interest rate, for a life age 65, in 2003, the latest period for which data are reported. The results are summarised in Figure 5, in which we present the 2.5, 50 and 97.5 percentiles for the statistics of interest, based on 5,000 simulations under the following cross-classification:

- Simulation Strategy A or C (within each frame)
- All three link functions (within each frame)
- LC or LP predictors (respective left and matching right frames)
- Computation of the statistics of interest by either the period or cohort methods (alternative rows of frames).

Additionally, we include within each frame a fourth set of results, denoted as link 4—these are the equivalent results for the log link, Poisson LC and LP model which has been reported in Renshaw and Haberman (2008), together with the relevant model based maximum likelihood estimates (m.l.e.).

For computations by period, the m.l.e. 95 percent CIs displayed, are based on the approximate expressions for the respective variances:

$$Var\left\{\hat{e}(x)\right\} \approx \sum_{j\geq 0}\left(\frac{\hat{q}_{x+j}}{\hat{p}_{x+j}e^{i}_{x+j}}\right)\left(\frac{l_{x+j}\hat{e}(x+j)}{l_x}\right)^2$$

$$Var\left\{\hat{a}(x)\right\} \approx \sum_{j\geq 0}\left(\frac{\hat{q}_{x+j}}{\hat{p}_{x+j}e^{i}_{x+j}}\right)\left(\frac{l_{x+j}v^{j}\hat{a}(x+j)}{l_x}\right)^2$$

(e.g., Benjamin and Pollard (1980), Chapter 17). The scales in adjacent frames are set consistently in order to facilitate comparison, with LC modelling on the left and LP modelling on the right. Computation of the statistics of interest by period and then by cohort, are represented in alternative rows of frames.

We note the following features:

1. The symmetry of the simulated prediction and confidence intervals, reflected in the associated simulated histograms (not shown).
2. Within each frame, the close agreement of the interval widths across all of the links (and with the width of the m.l.e. intervals, where applicable).
3. Within each frame, the close vertical alignment of the simulated binomial complementary log-log link (link 1), the Poisson log link (link 4) and the estimated (m.l.e.) central or first moment measures. The theoretical explanation for this alignment, is to be found in the close association between the approximate relationship (6), used when mapping μ_{xt} to q_{xt}, and the complementary log-log link, referred to in Section 2.2.
4. Within each frame, the degree of lateral displacement in the central measures under binomial probit link modelling (link 3), compared with complementary log-log link and Poisson log link modelling (links 1 & 4), which is appreciable for computations by cohort. This feature is directly attributable to the relative curvature of the respective link functions at the extremity of their domain. Historically, binomial complementary log-log modelling for q_{xt} and Poisson log link modelling for μ_{xt}, have featured strongly in the construction of static actuarial life tables (for fixed *t*) and have included the well-known Gompertz law of mortality. Further, we note that the binomial probit link approach to modelling

has only been given marginal attention in the actuarial literature (e.g., Forfar et al. (1988), Renshaw (1991)).

5. Within each frame, the degree of agreement in the central measures under binomial log-odds link modelling (link 2), compared with links 1 & 4, for computations by period, and the modest lateral displacement for computations by cohort.
6. Comparing like with like within horizontally adjacent frames, the close degree of agreement between LC and LP modelling with respect to central measures, especially for computations by period, and with respect to interval widths in general.
7. A comparison of like with like in vertically adjacent frames, indicates the degree of lateral displacement and measure of increased risk (interval widths), for computations by cohort, compared with computations by period.

4. Age-Specific Scale Parameters and Joint Modelling

The binomial distribution $D \sim bin\left(q, e^i\right)$ (with fixed initial exposure e^i), is characterised by the single first moment parameter q and has no inbuilt provision for the separate targeting of second moment properties (which would be useful in risk measurement); while model fitting requires the probability function or likelihood. We follow Renshaw (1992) and Renshaw and Haberman (2008) and resolve this issue through the introduction of variable dispersion parameters in order to capture second moment properties. At the same time, we retain the essential first order moment properties of the binomial distribution and apply (distribution free) two stage joint model fitting, as follows:

Stage 1: Model D_{xt} as independent binomial responses

$$E\left(D_{xt}\right) = e_{xt}^i q_{xt},\ Var\left(D_{xt}\right) = \varphi_{xt} \frac{V\left\{E\left(D_{xt}\right)\right\}}{\omega_{xt}};\ V\left(u\right) = u\left(1 - \frac{u}{e_{xt}^i}\right)$$

with variable dispersion parameters φ_{xt}, the link function I, II or III and the LC or LP parametric predictor structures.

Then, define $R_{xt} = \omega_{xt} \dfrac{\left(D_{xt} - \hat{d}_{xt}\right)^2}{\hat{d}_{xt}\hat{p}_t}$, $\hat{d}_{xt} = e^i_{xt}\hat{q}_{xt}$, $\hat{p}_{xt} = 1 - \hat{q}_{xt}$.

Stage 2: Model R_{xt} as independent gamma responses

$$E\left(R_{xt}\right) = \varphi_{xt},\ Var\left(R_{xt}\right) = \tau \frac{V\left\{E\left(R_{xt}\right)\right\}}{\omega_{xt}};\ V\left(u\right) = u^2$$

with scale parameter τ, the log link

LG: $\log \varphi_{xt} = \varsigma_x$

and linear predictor parametric structure in age effects, denoted LG.

The joint modelling process is implemented by repeatedly fitting each stage alternatively, terminating with the convergence of both stage specific deviances: $D\left(d_{xt}, \hat{d}_{xt}\right)$ as defined by (7) for Stage 1 and

$$D\left(r_{xt}, \hat{\varphi}_{xt}\right) = 2\sum_{x,t} \omega_{xt} \left\{ -\log\left(\frac{r_{xt}}{\hat{\varphi}_{xt}}\right) + \frac{r_{xt} - \hat{\varphi}_{xt}}{\hat{\varphi}_{xt}} \right\}$$

for Stage 2. Stage 1 residuals form the Stage 2 gamma responses while the Stage 2 fitted values form the Stage 1 weights $\omega_{xt}/\hat{\varphi}_{xt}$ where ω_{xt}, are the zero-one empty cell indicators. We set $\hat{\varphi}_{xt} = 1$ as effective Stage 1 starting values.

The theoretical justification for the approach to the joint modelling of first and second moment parameterised structures, is based on the optimisation of the pseudo (log) likelihood *P* where

$$-2P = \sum_{x,t} \frac{r_{xt}}{\varphi_{xt}} + \sum_{x,t} \log\left\{2\pi\varphi_{xt} V\left(E\left(D_{xt}\right)\right)\right\}.$$

The properties and applications of *P* are discussed in Carroll and Ruppert (1982), and in Davidian and Carroll (1987), (1988).

5. Joint Modelling of the U.K. Male Pensioner Mortality Experience

Model fitting is as described in Section 4 for each Stage 1 predictor structure (1)-(2) in combination with each link function (3)-(5). Typically, convergence of the Stage 1 and Stage 2 deviances is rapid, (to a pre-specified level of accuracy), and occurs within 10 or so iterations. The resulting Stage 1 parameter estimates are as depicted in Figures 2a, b and c, and the right hand frames of Figure 3. The respective Stage 1 deviance residual plots are essentially identical to Figure 4 and are hence not duplicated. The patterns in the left hand frames in Figure 3, depicting the annual differences between the actual and expected total deaths under single stage modelling, differ from their counter-parts under joint modelling and these are reproduced in the left hand frames of Figure 6. The feature that these differences are no longer zero under binomial LP log odds joint modelling, compared with single modelling, has its equivalence under Poisson LP log link modelling, which is illustrated in Renshaw and Haberman (2008).

The Stage 2 parameter estimates $\hat{\varphi}_{xt} = \hat{\varphi}_x = \log \hat{\varsigma}_x$, encapsulating the second moment structure of the joint modelling process, are depicted in the right hand frames of Figure 6. For each link, within each frame, we have superimposed the results obtained under LC/LG and LP/LG joint modelling for comparison. We note the convex shape of the near identical matching profiles within each frame, a feature which is similar to the profiles obtained under Poisson log link joint modelling (Renshaw and Haberman (2007)). We also observe that under dispersion ($\varphi_x < 1$) occurs largely at the age extremities, a feature that coincides with the paucity of exposure to mortality risk at these ages.

Again focusing on lives age 65, period 2003, we conduct simulation Strategy C for each joint modelling predictor structure LC/LG and LP/LG, in combination with each Stage 1 link function (3)-(5), displaying the 2.5, 50, 97.5 (N = 5,000) percentiles for life expectancy and a fixed life annuity value (calculated using a 4 percent rate of interest)

predictions, computed both by the period and cohort methods (Figure 7). Here, the layout of the individual frames is the same as Figure 5, and we reproduce the respective Strategy C single modelling results within each frame, for comparison. We have not attempted to implement the binomial joint modelling version of simulation Strategy A because, unlike Strategy C, which is distribution free, a two parameter probability distribution function representing the binomial distribution with dispersion is required in order to map the fitted responses on to the simulated responses. This issue is discussed further in the context of Poisson joint modelling by Renshaw and Haberman (2008).

Referring to Figure 7, we note the following:

1. Comparing like with like, the increase in the simulated confidence and prediction intervals under joint modelling, compared to single modelling, throughout.
2. Comparing like with like, within each frame, the consistent small lateral displacements of the central measures under joint modelling. These are due to the switch from single to joint modelling and occur also when joint modelling is conducted with a constant dispersion parameter *viz.* $\varsigma_x = \varsigma$.
3. However, still comparing like with like within each frame, when joint modelling is conducted under constant dispersion, the widths of the simulated intervals are comparable with those under single modelling, so that the increased width under joint modelling is directly attributable to the age variation in the dispersion parameters.
4. A number of the features identified under single modelling are preserved under joint modelling: the preservation of the alignment of simulated central measures under link 1 and link 4 and the relative displacements of the central measures under link 2 and link 3.
5. The element of asymmetry, present in the simulated intervals under Poisson log link joint modelling (link 4), would appear to be limited to this choice of models. We are not sure of the explanation for this feature.

6. Concluding Comments

We conclude with some comments about the effectiveness of the simulation strategies.

We note that simulation strategy C requires a good fit at Stage 1, and this gives rise to a set of pattern-free residuals, from which we may sample (with replacement) at each simulation. The strategy is distribution free, which facilitates its use in the formulation and implementation of the joint modelling process.

In contrast, simulation strategy A is not distribution free, and requires a distribution function to generate new (Stage 1) responses at each simulation; otherwise, it cannot be applied.

We also highlight the following two points:

- Joint modelling incorporates an extra provision for targeting second moment properties, compared to single stage modelling. This is reflected in wider simulated confidence and prediction intervals under Strategy A (both approaches) and Strategy C (joint modelling only), than would otherwise be the case.

- LC modelling allows for greater variability in the period component than LP modelling. We have compared confidence and prediction intervals under LC modelling and under extrapolation by random walk with drift, with matching intervals under LP modelling and linear extrapolation. For the two simulation strategies investigated (A and C), in combination with both modelling approaches (single and joint), none of the matching simulated prediction intervals are materially wider in absolute terms under LC modelling compared with LP modelling. This may suggest that LC modelling fails to capture the full magnitude of the time series forecast error.

Overall, we believe that further work is required to investigate the basis for using the bootstrap framework for simulating prediction intervals, in the presence of model extrapolation.

References

Benjamin, B., and Pollard, J.H. 1980. *The Analysis of Mortality and Other Actuarial Statistics*. London: Heineman.

Brouhns, N., Denuit, M., and Vermunt, J.K. 2002. "Measuring the Longevity Risk in Mortality Projections." Bulletin of the Swiss Association of Actuaries, 105-130.

Brouhns, N., Denuit, M., and van Keilegom, 2005. "Bootstrapping the Poisson Log-Linear Model for Mortality Forecasting." Scandinavian Actuarial Journal 3, 212-224.

Cossette, H., Delwarde, A., Denuit, M., Guillot, F., and Marceau, E. 2007. "Pension Plan Valuation and Mortality Projection; A Case Study with Mortality Data." North American Actuarial Journal 11 (2), 1-34.

Carroll, R.J., and Ruppert, D. 1982. "Robust Estimation of Heteroscedastic Linear Models." Annals of Statistics 10, 429-441.

Davidian, M., and Carroll, R.J. 1987. "Variance Function Estimation." Journal of the American Statistical Association 82, 1079-1091.

Davidian, M., and Carroll, R.J. 1988. "A Note on Extended Quasi-Likelihood." Journal of the Royal Statistical Society, Series B 50, 74-82.

Deaton, A., and Paxson, C. 2001. Mortality Income and Income Inequality Over Time in Britain and the United States. NBER Working Paper 8534, NBER.

Forfar, D.O., McCutcheon, J.J., and Wilkie, A.D. 1988. "On Graduation by Mathematical Formula." Journal of the Institute of Actuaries 115, 1-135.

Francis, B., Green, M., and Payne, C. 1993. *The Glim System: Release 4 Manual*. Oxford: Clarendon.

Koissi, M-C., Shapiro, A.F., and Hognas, G. 2006. "Evaluating and Extending the Lee-Carter Model for Mortality Forecasting: Bootstrap Confidence Intervals." Insurance: Mathematics and Economics 38, 1-20.

Lee, R., and Carter, L. 1992. "Modelling and Forecasting U.S. Mortality." Journal of the American Statistical Association 87, 659-671.

Renshaw, A.E. 1991. "Actuarial Graduation Practice and Generalised Linear and Non-linear Models." Journal of the Institute of Actuaries 118, 295-312.

Renshaw, A.E. 1992. "Joint Modelling for Actuarial Graduation and Duplicate Policies." Journal of the Institute of Actuaries 119, 69-85.

Renshaw, A.E., and Haberman, S. 2003. “Lee Carter Mortality Forecasting: A Parallel Generalized Linear Modelling Approach for England and Wales Mortality Projections.” Applied Statistics 52, 199-137.

Renshaw, A.E., and Haberman, S. 2008. “On Simulation-Based Approaches to Risk Measurement in Mortality with Specific to Poisson Lee-Carter Modelling.” Insurance: Mathematics and Economics. To appear.