

A quantitative comparison of simulation strategies for mortality projection

Jackie Li*

Curtin University, Australia

Abstract

We compare quantitatively six simulation strategies for mortality projection with the Poisson Lee–Carter model. We test these strategies on New Zealand mortality data and discuss the simulated results of the mortality index, death rates, and life expectancy.

Keywords

simulation methods; mortality projection; Poisson Lee-Carter model

1. Introduction

Governments providing pensions and insurers selling annuities encounter the so-called longevity risk, which is the risk of paying more than anticipated due to unexpected mortality improvement. To assess the probable extent of future mortality decline, an extensive amount of research has been conducted in the past 20 years in the area of mortality projection. In the literature, most of the work is related to the Lee & Carter (1992) method and its various extensions (e.g. Lee, 2000). Because of analytical intractability of these models, simulation methods represent a practically viable means to measure the variability of mortality changes and so the longevity risk, and have been adopted in many applications (e.g. Kogure & Kurachi, 2010).

Initially, Lee & Carter (1992) focused only on the error term of the random walk with drift for the mortality index (though they also provided suggestions on how to include the errors of parameter estimation). Li *et al.* (2004) further incorporated the estimation error of the drift term. Brouhns *et al.* (2002b) took a different approach and generated samples of the parameters from the multivariate normal distribution. On the other hand, Brouhns *et al.* (2005) considered bootstrapping the number of deaths from the Poisson distribution. Koissi *et al.* (2006) also examined bootstrapping the residuals to produce pseudo data samples. Renshaw & Haberman (2008) then provided a comparative study of the last three simulation methods, which tested different model structures, parameter constraints, scale parameters, and distribution assumptions. They suggested that using the multivariate normal distribution to generate samples of the parameters would lead to widely differing prediction intervals under different constraints. They also found that using variable scale parameters and the negative binomial distribution would result in wider prediction intervals when bootstrapping the number of deaths. Moreover, Czado *et al.* (2005) applied Bayesian modelling and Markov chain Monte Carlo (MCMC) simulation to mortality projection.

*Correspondence to: Jackie Li, Department of Mathematics & Statistics, GPO Box U1987 Perth, 6845 Western Australia. Tel: +618 92667171; Fax: +618 92663197. E-mail: jackie@actuarialworkshop.com

Under the current regulatory environment that has become more risk based, it is of practical interest to find a probability distribution to depict how actual outcomes may differ from expected. This distribution can be used to compute such risk metrics as standard deviation, variance, and value-at-risk in pricing, hedging, reserving, and capital management. For example, QIS5 Technical Specifications for Solvency II states that “the nature of the risks underlying the insurance contracts could be described by the probability distribution of the future cash flows arising from the contracts”, and that all of process error, parameter error, and model error should be considered. Prudential Standard GPS 320 by Australian Prudential Regulation Authority (2013) also stipulates that the minimum value of insurance liabilities must be the greater of the “75% level of sufficiency” and the “central estimate plus one half of a standard deviation above the mean”. As such, it is important for practitioners to choose suitable simulation methods to generate random samples of future outcomes, from which different probability inferences can be drawn. In particular, uncertainty in parameter estimation needs to be sufficiently allowed for when the portfolio size is small.

In this paper, we carry out a quantitative comparison of all these simulation methods, using population data from New Zealand. Particularly, we apply the Poisson Lee–Carter (LC) model proposed by Brouhns *et al.* (2002a). We construct prediction intervals for projected death rates and life expectancy, and pay particular attention to selecting an appropriate data fitting period.

The paper is organised as follows. Section 2 provides an overview of the six simulation strategies. Section 3 presents the results of applying these simulation techniques to the mortality data of New Zealand. Section 4 sets forth our concluding remarks.

2. Simulation Strategies

The Poisson LC model proposed by Brouhns *et al.* (2002a) is defined as follows

$$D_{x,t} \sim \text{Poisson}(e_{x,t} m_{x,t}), \quad \ln m_{x,t} = a_x + b_x k_t \quad (1)$$

in which $D_{x,t}$ is the number of deaths at age x in year t , $e_{x,t}$ the corresponding (known) exposure, $m_{x,t}$ the central death rate, a_x the general mortality schedule, and b_x measures the sensitivity of the log death rate to changes in the mortality index k_t . The mortality index is usually modelled as a random walk with drift, i.e., $k_t = \mu + k_{t-1} + \varepsilon_t$ where μ is the drift term and ε_t the error term with mean zero and variance σ^2 . This model offers a natural way to model the number of deaths and avoids the assumption of homoscedastic errors in the original LC method. The maximum likelihood parameter estimates are calculated from an iterative updating scheme (based on Newton’s method), subject to two constraints $\sum b_x = 1$ and $\sum k_t = 0$. In the next section, we will check the resulting residuals for the goodness-of-fit and take the most recently observed data as the starting point of projection.

After equation (1) is fitted to the data, the following six simulation strategies are tested:

- (I) Lee & Carter (1992) focused on ε_t and used $\pm 2\sigma$ to approximate the 95% prediction intervals. Here, assuming $\varepsilon_t \sim N(0, \sigma^2)$, we simulate n samples of future k_t via $k_t = \hat{\mu} + k_{t-1} + \varepsilon_t$ and then obtain values of future death rates using the estimated parameters.
- (II) Li *et al.* (2004) assumed $\varepsilon_t \sim N(0, \sigma^2)$ and simulated n samples of future k_t with $k_t = \hat{\mu} + \text{se}(\hat{\mu})\eta + k_{t-1} + \varepsilon_t$, where $\text{se}(\hat{\mu})$ was the standard error in estimating μ and $\eta \sim N(0, 1)$. Note that ε_t and η were treated as independent. They then obtained values of future death rates using the estimated parameters.

- (III) Brouhns *et al.* (2002b) adopted a parametric Monte Carlo simulation approach. They first generated a sample of the model parameters ($a_x^{(j)}$, $b_x^{(j)}$, $k_t^{(j)}$) from the asymptotic multivariate normal distribution, in which the mean was approximately taken as the maximum likelihood parameter estimates and the covariance matrix was given by the inverse of the Fisher information matrix. The Cholesky decomposition process was used to produce multivariate random samples. Then they fitted a new time series model to the sampled mortality index ($k_t^{(j)}$) and projected it to the future, allowing for the error term of the time series model. Finally, they produced values of future death rates ($m_{x,t}^{(j)}$) using the sampled parameters and the projected mortality index. The whole process was repeated to generate n scenarios, i.e., $j = 1, 2, \dots, n$. Renshaw & Haberman (2008) also considered incorporating the scale parameter ϕ into the covariance matrix.
- (IV) Brouhns *et al.* (2005) implemented a semi-parametric bootstrapping approach. They first bootstrapped a pseudo sample of the number of deaths ($d_{x,t}^{(j)}$) from the Poisson distribution with mean $\hat{d}_{x,t}$, the observed number of deaths. They applied equation (1) to this pseudo sample and computed the model parameters ($a_x^{(j)}$, $b_x^{(j)}$, $k_t^{(j)}$). Then they fitted a new time series model to the mortality index ($k_t^{(j)}$) of the pseudo sample and projected it to the future, taking into account the error term of the time series model. Correspondingly, they generated values of future death rates ($m_{x,t}^{(j)}$) with the computed parameters and the projected mortality index of the pseudo sample. The entire process was repeated many times to produce n scenarios, i.e., $j = 1, 2, \dots, n$. Wang & Lu (2005) also took a similar approach (with the binomial distribution) to calculate the standard error of the parameter estimates. On the other hand, Renshaw & Haberman (2008) sampled from the fitted Poisson distribution (with mean $\hat{d}_{x,t}$) instead.
- (V) Koissi *et al.* (2006) suggested the use of residual bootstrapping. First, they resampled, with replacement, the deviance residuals ($r_{x,t}^{(j)}$) based on the fitted Poisson distribution. Then they used the inverse formula to turn these resampled residuals into a pseudo sample of the number of deaths ($d_{x,t}^{(j)}$), and applied equation (1) to this pseudo sample and computed the model parameters ($a_x^{(j)}$, $b_x^{(j)}$, $k_t^{(j)}$). They fitted a new random walk with drift to the mortality index ($k_t^{(j)}$) of the pseudo sample and projected it to the future, allowing for the error term ε_t . Finally, they generated values of future death rates ($m_{x,t}^{(j)}$) with the computed parameters and the projected mortality index of the pseudo sample. The whole process was repeated to give n scenarios, i.e., $j = 1, 2, \dots, n$. Here, we follow the inverse formula used by Renshaw & Haberman (2008), as it complies with the principle in Efron & Tibshirani (1993), such that it gives a pseudo sample of the number of deaths (instead of the fitted number of deaths).
- (VI) Czado *et al.* (2005) applied Bayesian modelling and MCMC simulation (the Gibbs sampler and Metropolis–Hastings algorithm) to project mortality. The major step of Bayesian analysis is to find the posterior distribution of the unknown parameters and quantities given the data. In principle, the posterior density function can be derived from $f(\boldsymbol{\theta}|\mathbf{D}) \propto f(\mathbf{D}|\boldsymbol{\theta})f(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ denotes the unknown parameters and quantities and \mathbf{D} means the data. As this problem is mathematically intractable, one can perform MCMC simulation to generate samples from the posterior distribution. Similar to Czado *et al.* (2005) and also Kogure *et al.* (2009), we choose the prior distributions below:

$$\begin{aligned}
 a_x &\sim N(0, \sigma_a^2) \\
 b_x &\sim N\left(\frac{1}{\text{number of age groups}}, \sigma_b^2\right) \\
 \mu &\sim N(\mu_0, \sigma_\mu^2) \\
 \varepsilon_t &\sim N(0, \sigma^2) \\
 \sigma^{-2} &\sim \text{Gamma}(\alpha, \beta)
 \end{aligned}$$

In effect, (I) considers only the random error of projecting the mortality index, while (II) also allows for the fact that different data samples would give rise to different estimates of the drift term. On the other hand, (III) to (V) incorporate both the error of estimating the model parameters due to sampling fluctuation and the random error of projecting the mortality index. For (VI), Bayesian analysis automatically includes all these errors via the prior distributions. Roughly speaking, (I) caters for process error only, while (II) to (VI) take both process error and parameter error (or parameter uncertainty) into account. In terms of programming and computation, (I) and (II) are fairly straightforward and do not require much computation time, whereas (III) to (VI) are generally much more demanding and time-consuming. In our analysis below, the former takes 15 minutes or so to run, but the latter ones can take quite a few hours to even a couple of days.

We run at least 5,000 iterations (i.e. $n = 5,000$) for each strategy, from which we can obtain the sample means, sample standard deviations, and 95% prediction intervals for projected death rates and life expectancy. To perform simulations, we utilise the software R for (I) to (V) and WinBUGS (Spiegelhalter *et al.*, 2003) for (VI). Note that some practitioners may find the MCMC simulation process in (VI) too tedious to implement, as it involves complex scientific programming. The specialised software WinBUGS provides an alternative, in which the programming language is relatively easy to use (e.g. Li, 2014). For demonstration purposes, we adopt WinBUGS to carry out MCMC simulation. The simulation process, however, can still take up to a couple of days to run. If greater flexibility is required, one may develop MCMC algorithms from scratch, but this calls for significant expertise in building the models and writing the codes. Note also that vague priors are further tested, such as multiplying σ_a^2 and σ_b^2 by 10. The sample means are basically not affected, and the resulting changes in the sample variances are immaterial.

3. Simulation Results

The number of deaths and exposed to risk data of New Zealand for 1948–2008, by sex and single year of age, are collected from the Human Mortality Database (2012). As noted in Li (2010, 2013), the linearity of the mortality index in the LC structure is an important property for projecting the overall mortality decline and makes the projection a sensible and straightforward exercise. Moreover, the fundamental drivers of mortality improvement have changed significantly for the recent decades, and it is often preferable to include only the most recent and relevant data periods. Figure 1 shows the mortality indices computed from applying equation (1) to the whole set of data, in which there is a prominent change in the slope in around 1985 for both sexes. This change is found to be broadly in line with the emergence of old-age improvement during that period of time. As such, following Li (2010), we select the starting year 1985 for all the projections below. In that paper, the selection was based on the change in the slope, the stability of b_x 's, and the R^2 values of testing the linearity of the k_t series. In addition, as the death rates from age 90 onwards are rather volatile due to small exposures, we exclude these data from our analysis and adopt the Coale & Guo (1989) method to “close out” the life table when computing life expectancy.

There are two main reasons behind the use of New Zealand data for testing the simulation methods. First, its population size is much smaller than those of the United Kingdom and the United States, the data of which are often used in mortality projection studies. The smaller sample size means higher uncertainty in parameter estimation, which makes it important to incorporate parameter error into the simulation process. Second, data of insurance portfolios and pension plans are generally much less than population data, and the experience of a relatively small country is more comparable to those of the largest portfolios and plans.

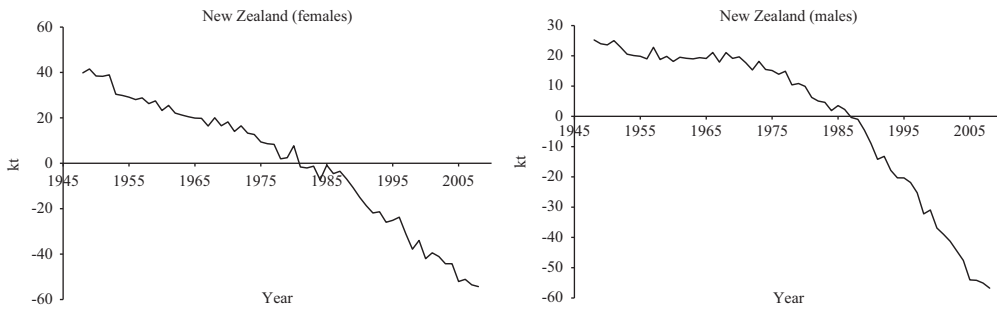


Figure 1. Mortality indices for the data period 1948–2008.

The estimated parameters based on the fitting period 1985–2008 are plotted in Figure 2. The a_x estimates display the typical shape of the mortality schedule. For the b_x estimates, there are three peaks at early childhood, around age 20, and age 60, which refer to the highest sensitivity to the general trend of mortality decline. Moreover, as expected, the k_t series are reasonably linear, and this feature helps justify the use of the random walk with drift. The corresponding standardised deviance residuals are then shown in Figure 3. Overall, these residuals are randomly distributed against age, calendar year, and cohort year. The absence of significant unusual patterns in the residuals indicates that the Poisson LC model (equation (1)) can be deemed as being able to capture the main characteristics of the data. The randomness of the residuals also provides a strong basis for using the residual bootstrapping method (V).

Provided with the parameter estimates, we apply the six simulation methods (I) to (VI) to generate samples of future death rates and life expectancy and construct 95% prediction intervals. As mentioned earlier, we further divide each of (III) and (IV) into two cases: (IIIa) without the scale parameter versus (IIIb) with the scale parameter; (IVa) using the observed number of deaths as the Poisson parameter versus (IVb) using the fitted Poisson distribution. The scale parameter is a measure of over-dispersion and its incorporation would lead to higher variability if it is larger than 1. Moreover, for (III) to (V), we do not fit a new time series model (e.g. ARIMA models) but simply a new random walk with drift in each iteration, since the mortality index samples are largely linear. It should also be noted that all the parameters and future values are sampled coherently under (VI).

Tables 1 and 2 list the projected values (central estimates calculated from the maximum likelihood estimates of the parameters and the projected mortality index), sample means, sample standard deviations, and 95% prediction intervals for future mortality index, (period) life expectancy at birth, and death rates at different ages in 2030 and 2050. Figure 4 plots the sample means and 95% prediction intervals for future values in 2050. We have made a number of observations as below:

- (1) The sample means agree closely with the projected values based on maximum likelihood.
- (2) The widths of the 95% prediction intervals are on average around 3.9 times the standard deviations. This ratio is the same as that of a normal distribution.
- (3) Compared with all the other methods, (I) clearly leads to smaller variances and narrower prediction intervals. This result is a direct consequence of allowing for only the random error term of the random walk with drift. For life expectancy, the magnitude of the differences appears to increase over time in the projections.

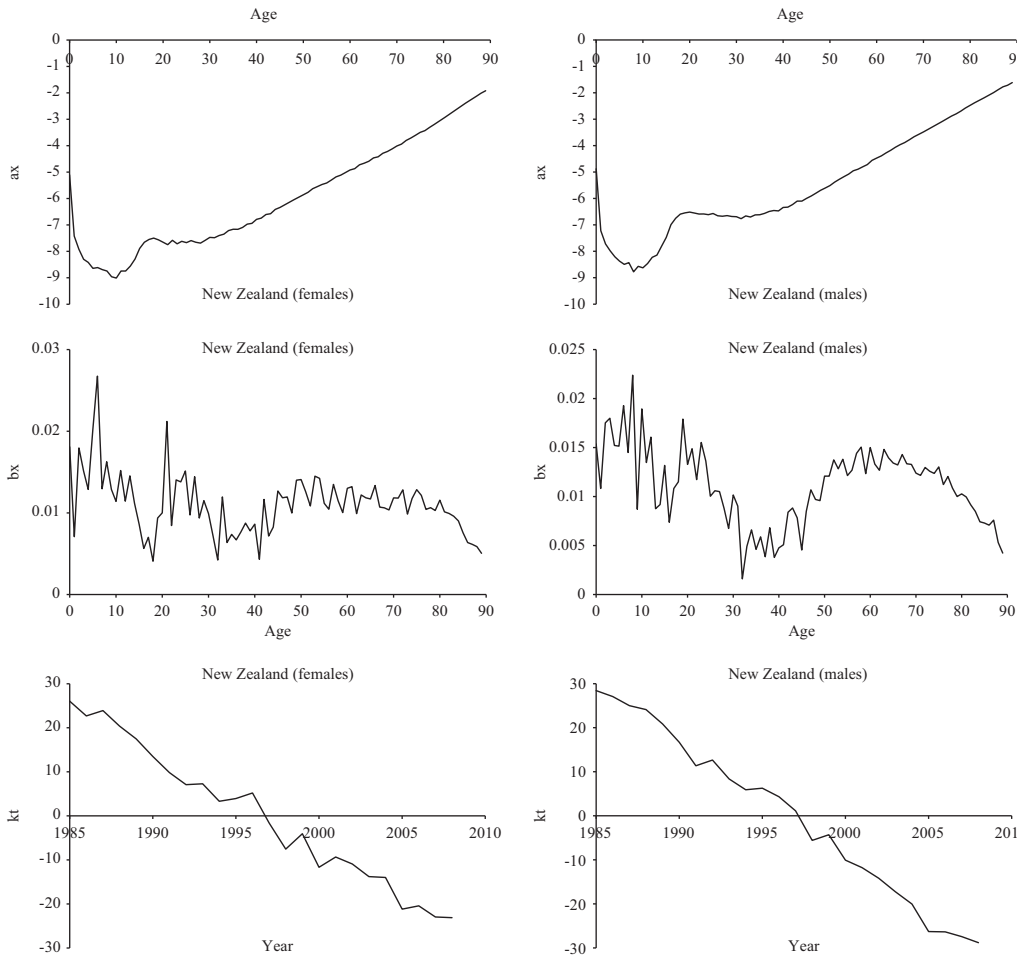


Figure 2. Parameter estimates for the data period 1985–2008.

- (4) For the mortality index and life expectancy, it can be seen that (II) gives rise to the largest variances and widest prediction intervals. For the death rates, (II) produces larger variances and wider prediction intervals at older ages but the results are opposite at younger ages, when compared with the other methods (except (I)).
- (5) The results of (IIIa) and (IIIb) are very close. Because the scale parameters for females and males are only 1.09 and 1.11, there is effectively not much difference between the two variations.
- (6) There does not appear to be any material difference between using the observed number of deaths and the fitted number of deaths as the Poisson parameter. The figures produced by (IVa) and (IVb) are very similar.
- (7) Overall, the results of (III) to (V) are highly comparable. This agreement seems to further suggest that the Poisson LC model (equation (1)) provides a satisfactory fit to the data. Brouhns *et al.* (2005) made a similar comment when they compared (IIIa) and (IVa).

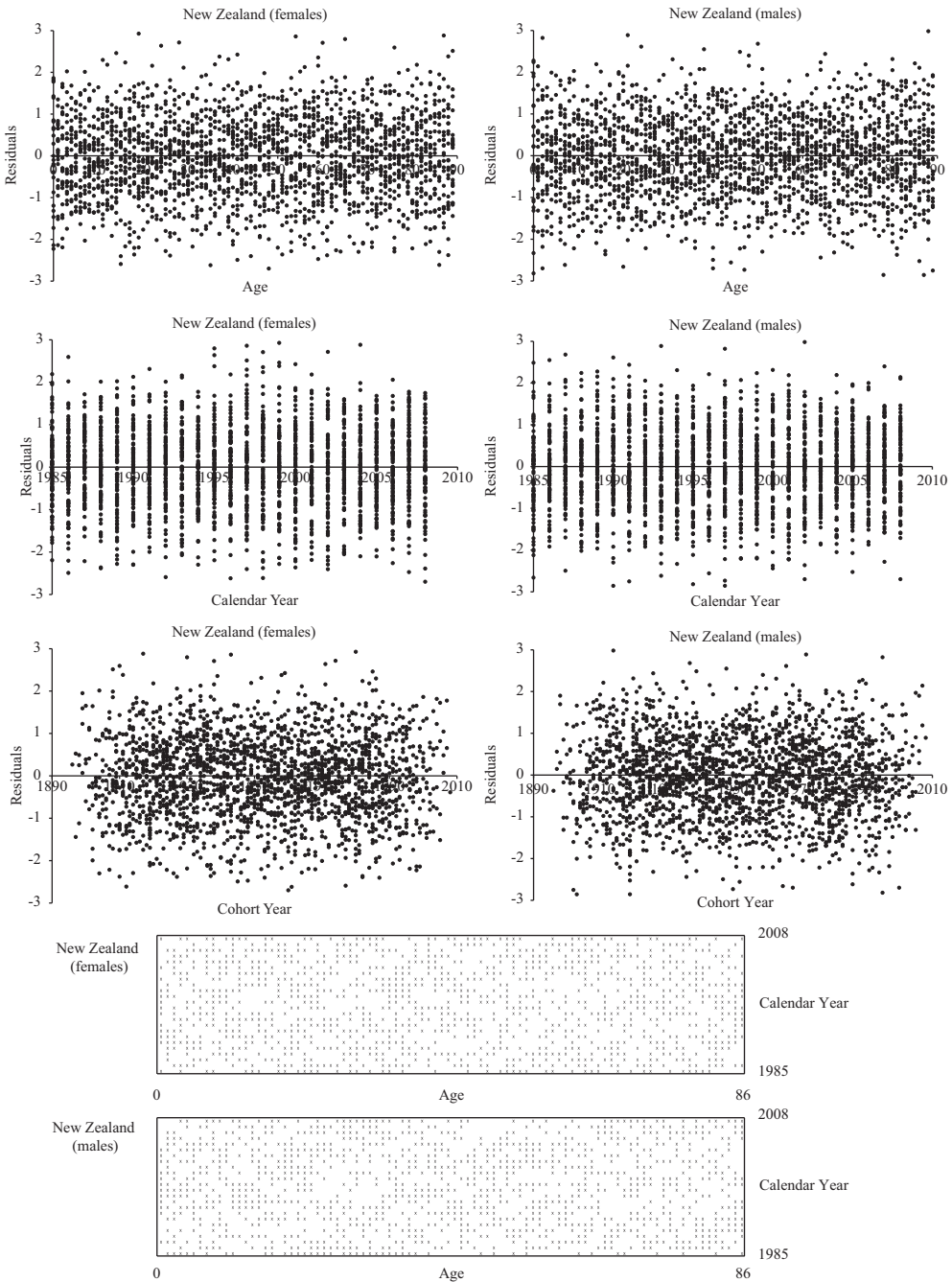


Figure 3. Standardised deviance residuals for the data period 1985–2008. Note: The empty cells refer to positive residuals and the marked ones refer to negative residuals.

Table 1. Projected values, sample means, sample standard deviations, and 95% prediction intervals for future mortality index, life expectancy at birth, and death rates in 2030 and 2050 (females).

Method	$k_{2030} = -70.15$			$k_{2050} = -112.91$		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	-70.33	14.62	(-98.69, -42.05)	-113.17	20.25	(-153.60, -73.72)
(II)	-70.02	20.40	(-109.82, -29.14)	-112.67	33.87	(-176.18, -46.01)
(IIIa)	-70.44	15.80	(-101.54, -39.47)	-113.32	21.73	(-155.35, -70.15)
(IIIb)	-70.27	15.77	(-101.81, -40.45)	-112.84	21.92	(-156.69, -71.14)
(IVa)	-69.95	15.81	(-101.17, -38.30)	-112.43	22.07	(-155.38, -69.22)
(IVb)	-69.58	15.91	(-100.61, -38.13)	-112.72	22.16	(-156.63, -70.18)
(V)	-69.69	15.68	(-100.29, -38.77)	-112.12	21.63	(-153.91, -69.80)
(VI)	-69.43	14.42	(-97.70, -41.20)	-111.70	23.30	(-158.80, -64.68)

Method	Life expectancy in 2030 = 86.76			Life expectancy in 2050 = 89.59		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	86.72	1.11	(84.44, 88.73)	89.53	1.18	(87.02, 91.65)
(II)	86.65	1.56	(83.25, 89.41)	89.36	2.03	(84.79, 92.58)
(IIIa)	86.73	1.29	(84.14, 89.25)	89.60	1.62	(86.62, 93.16)
(IIIb)	86.72	1.30	(84.26, 89.34)	89.56	1.64	(86.65, 93.20)
(IVa)	86.71	1.28	(84.12, 89.21)	89.56	1.61	(86.54, 92.98)
(IVb)	86.68	1.31	(84.09, 89.30)	89.56	1.63	(86.65, 93.14)
(V)	86.72	1.29	(84.16, 89.25)	89.58	1.61	(86.63, 93.00)
(VI)	86.92	1.21	(84.42, 89.49)	89.88	1.74	(86.26, 93.58)

Method	$m_{20,2030} = 0.00037$			$m_{20,2050} = 0.00024$		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	0.00038	0.00006	(0.00028, 0.00049)	0.00025	0.00005	(0.00016, 0.00036)
(II)	0.00038	0.00008	(0.00025, 0.00056)	0.00026	0.00009	(0.00013, 0.00047)
(IIIa)	0.00039	0.00012	(0.00019, 0.00064)	0.00027	0.00012	(0.00009, 0.00056)
(IIIb)	0.00039	0.00012	(0.00019, 0.00066)	0.00027	0.00013	(0.00009, 0.00058)
(IVa)	0.00039	0.00012	(0.00018, 0.00064)	0.00027	0.00013	(0.00008, 0.00057)
(IVb)	0.00039	0.00012	(0.00019, 0.00063)	0.00027	0.00012	(0.00009, 0.00055)
(V)	0.00039	0.00012	(0.00019, 0.00064)	0.00027	0.00012	(0.00009, 0.00056)
(VI)	0.00037	0.00009	(0.00022, 0.00056)	0.00025	0.00009	(0.00011, 0.00047)

Method	$m_{40,2030} = 0.00068$			$m_{40,2050} = 0.00047$		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	0.00068	0.00009	(0.00053, 0.00086)	0.00047	0.00008	(0.00033, 0.00066)
(II)	0.00069	0.00012	(0.00048, 0.00096)	0.00049	0.00015	(0.00027, 0.00083)
(IIIa)	0.00069	0.00015	(0.00041, 0.00099)	0.00049	0.00016	(0.00023, 0.00083)
(IIIb)	0.00069	0.00015	(0.00040, 0.00100)	0.00049	0.00016	(0.00022, 0.00085)
(IVa)	0.00069	0.00015	(0.00041, 0.00100)	0.00050	0.00017	(0.00022, 0.00085)
(IVb)	0.00069	0.00015	(0.00042, 0.00099)	0.00049	0.00016	(0.00023, 0.00084)
(V)	0.00069	0.00015	(0.00042, 0.00098)	0.00049	0.00016	(0.00023, 0.00083)
(VI)	0.00065	0.00013	(0.00042, 0.00091)	0.00045	0.00014	(0.00021, 0.00074)

Table 1. Continued

Method	$m_{60,2030} = 0.00266$			$m_{60,2050} = 0.00153$		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	0.00270	0.00052	(0.00184, 0.00384)	0.00157	0.00042	(0.00090, 0.00254)
(II)	0.00276	0.00075	(0.00159, 0.00454)	0.00169	0.00078	(0.00067, 0.00364)
(IIIa)	0.00272	0.00060	(0.00169, 0.00404)	0.00159	0.00050	(0.00081, 0.00273)
(IIIb)	0.00273	0.00060	(0.00167, 0.00398)	0.00161	0.00051	(0.00080, 0.00276)
(IVa)	0.00272	0.00060	(0.00167, 0.00406)	0.00160	0.00051	(0.00080, 0.00275)
(IVb)	0.00274	0.00061	(0.00170, 0.00406)	0.00160	0.00050	(0.00081, 0.00272)
(V)	0.00273	0.00060	(0.00168, 0.00406)	0.00160	0.00050	(0.00081, 0.00275)
(VI)	0.00270	0.00056	(0.00172, 0.00393)	0.00160	0.00054	(0.00076, 0.00288)

Method	$m_{80,2030} = 0.02310$			$m_{80,2050} = 0.01409$		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	0.02338	0.00398	(0.01661, 0.03196)	0.01444	0.00342	(0.00880, 0.02216)
(II)	0.02378	0.00569	(0.01460, 0.03710)	0.01526	0.00622	(0.00678, 0.03053)
(IIIa)	0.02346	0.00443	(0.01586, 0.03314)	0.01454	0.00382	(0.00846, 0.02339)
(IIIb)	0.02344	0.00441	(0.01589, 0.03279)	0.01456	0.00386	(0.00823, 0.02293)
(IVa)	0.02346	0.00443	(0.01576, 0.03326)	0.01457	0.00389	(0.00817, 0.02341)
(IVb)	0.02360	0.00449	(0.01584, 0.03367)	0.01456	0.00389	(0.00824, 0.02322)
(V)	0.02342	0.00441	(0.01577, 0.03319)	0.01451	0.00387	(0.00836, 0.02330)
(VI)	0.02296	0.00406	(0.01603, 0.03209)	0.01420	0.00413	(0.00738, 0.02383)

- (8) The figures computed from (VI) are quite close to those from (III) to (V). For the mortality index, death rates at older ages, and life expectancy, (VI) tends to give larger variances and wider prediction intervals than (III) to (V) later in 2050, but the situation is reversed earlier in 2030. On the other hand, for the death rates at younger ages, (VI) tends to produce smaller variances and narrower prediction intervals in both periods.
- (9) Figure 5 displays the sampled distributions of life expectancy at birth in 2050. It can be seen that the shapes of the distributions are very similar for (III) to (V), which agree with the observations above.
- (10) It is interesting to observe in Figure 5 that there is a small extent of skewness in the sampled distributions of life expectancy. While the distributions generated by (I) and (II) are negatively skewed, those by (III) to (VI) are slightly positively skewed. We realise that the former case actually becomes more negatively skewed over time during the projection period, whereas the latter changes from being negatively skewed to positively skewed gradually. Li *et al.* (2004) and Li (2013) demonstrated the same effects as the former case when they adopted (II). This negative skewness in life expectancy appears to be driven by the fact that changes in the mortality index are the only source of uncertainty allowed for and that these changes follow the normal distribution, which means that the projected death rates are then lognormally distributed and have positive skewness. On the other hand, for (III) to (VI), there is also uncertainty of the other parameters, and the aggregate effects on skewness are mixed and less clear-cut.

Table 2. Projected values, sample means, sample standard deviations, and 95% prediction intervals for future mortality index, life expectancy at birth, and death rates in 2030 and 2050 (males).

Method	$k_{2030} = -83.58$			$k_{2050} = -133.36$		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	-83.45	10.60	(-103.99, -62.68)	-133.29	14.50	(-161.69, -104.07)
(II)	-83.70	14.58	(-111.93, -55.25)	-133.55	24.29	(-182.11, -84.96)
(IIIa)	-83.36	11.45	(-106.38, -61.11)	-133.21	15.77	(-167.79, -102.68)
(IIIb)	-83.37	11.49	(-106.36, -61.20)	-133.02	15.90	(-164.52, -102.14)
(IVa)	-83.12	11.86	(-106.12, -59.66)	-133.00	16.59	(-164.86, -100.49)
(IVb)	-83.51	11.78	(-107.15, -61.11)	-133.39	16.46	(-166.18, -101.02)
(V)	-83.16	11.81	(-106.35, -59.57)	-132.76	16.33	(-164.44, -100.56)
(VI)	-83.26	10.65	(-103.40, -60.89)	-132.90	16.90	(-166.20, -100.50)

Method	Life expectancy in 2030 = 83.57			Life expectancy in 2050 = 86.98		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	83.53	0.86	(81.75, 85.12)	86.94	0.83	(85.12, 88.44)
(II)	83.52	1.19	(81.04, 85.66)	86.87	1.42	(83.68, 89.31)
(IIIa)	83.53	0.97	(81.61, 85.36)	87.00	1.13	(84.92, 89.38)
(IIIb)	83.53	0.97	(81.62, 85.42)	87.00	1.13	(84.88, 89.42)
(IVa)	83.52	1.01	(81.48, 85.47)	87.00	1.17	(84.79, 89.43)
(IVb)	83.54	0.99	(81.58, 85.45)	87.01	1.14	(84.84, 89.41)
(V)	83.53	1.00	(81.48, 85.43)	87.00	1.16	(84.88, 89.47)
(VI)	83.72	0.95	(81.74, 85.47)	87.28	1.25	(84.99, 90.02)

Method	$m_{20,2030} = 0.00052$			$m_{20,2050} = 0.00027$		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	0.00053	0.00008	(0.00040, 0.00069)	0.00028	0.00005	(0.00019, 0.00040)
(II)	0.00053	0.00010	(0.00036, 0.00076)	0.00028	0.00009	(0.00014, 0.00051)
(IIIa)	0.00054	0.00011	(0.00033, 0.00078)	0.00028	0.00009	(0.00014, 0.00049)
(IIIb)	0.00054	0.00012	(0.00033, 0.00079)	0.00029	0.00009	(0.00014, 0.00049)
(IVa)	0.00054	0.00011	(0.00033, 0.00078)	0.00028	0.00009	(0.00014, 0.00048)
(IVb)	0.00054	0.00011	(0.00033, 0.00077)	0.00028	0.00009	(0.00014, 0.00048)
(V)	0.00054	0.00012	(0.00034, 0.00078)	0.00029	0.00009	(0.00014, 0.00049)
(VI)	0.00055	0.00011	(0.00035, 0.00078)	0.00030	0.00009	(0.00015, 0.00052)

Method	$m_{40,2030} = 0.00072$			$m_{40,2050} = 0.00057$		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	0.00072	0.00004	(0.00066, 0.00080)	0.00057	0.00004	(0.00050, 0.00066)
(II)	0.00072	0.00005	(0.00063, 0.00083)	0.00057	0.00007	(0.00045, 0.00072)
(IIIa)	0.00073	0.00011	(0.00053, 0.00095)	0.00059	0.00013	(0.00035, 0.00088)
(IIIb)	0.00073	0.00012	(0.00052, 0.00098)	0.00059	0.00015	(0.00035, 0.00091)
(IVa)	0.00073	0.00011	(0.00053, 0.00094)	0.00058	0.00013	(0.00035, 0.00087)
(IVb)	0.00073	0.00011	(0.00053, 0.00095)	0.00058	0.00014	(0.00035, 0.00088)
(V)	0.00073	0.00011	(0.00053, 0.00096)	0.00059	0.00014	(0.00035, 0.00089)
(VI)	0.00066	0.00009	(0.00049, 0.00086)	0.00050	0.00011	(0.00031, 0.00074)

Table 2. Continued

Method	$m_{60,2030} = 0.00349$			$m_{60,2050} = 0.00165$		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	0.00354	0.00057	(0.00257, 0.00477)	0.00169	0.00037	(0.00108, 0.00256)
(II)	0.00357	0.00079	(0.00228, 0.00533)	0.00176	0.00067	(0.00080, 0.00342)
(IIIa)	0.00355	0.00065	(0.00240, 0.00497)	0.00171	0.00044	(0.00098, 0.00271)
(IIIb)	0.00355	0.00066	(0.00240, 0.00498)	0.00171	0.00045	(0.00099, 0.00273)
(IVa)	0.00357	0.00068	(0.00239, 0.00505)	0.00172	0.00047	(0.00096, 0.00275)
(IVb)	0.00355	0.00067	(0.00235, 0.00495)	0.00171	0.00046	(0.00097, 0.00274)
(V)	0.00355	0.00068	(0.00237, 0.00506)	0.00171	0.00046	(0.00097, 0.00280)
(VI)	0.00352	0.00061	(0.00246, 0.00488)	0.00170	0.00046	(0.00094, 0.00274)

Method	$m_{80,2030} = 0.03530$			$m_{80,2050} = 0.02116$		
	Sample mean	Sample s.d.	95% prediction interval	Sample mean	Sample s.d.	95% prediction interval
(I)	0.03555	0.00388	(0.02862, 0.04375)	0.02142	0.00320	(0.01582, 0.02860)
(II)	0.03565	0.00538	(0.02638, 0.04722)	0.02180	0.00555	(0.01283, 0.03480)
(IIIa)	0.03567	0.00440	(0.02761, 0.04485)	0.02155	0.00374	(0.01494, 0.02946)
(IIIb)	0.03562	0.00448	(0.02734, 0.04471)	0.02155	0.00380	(0.01483, 0.02978)
(IVa)	0.03567	0.00458	(0.02731, 0.04529)	0.02151	0.00393	(0.01462, 0.02995)
(IVb)	0.03556	0.00453	(0.02732, 0.04482)	0.02146	0.00387	(0.01471, 0.02967)
(V)	0.03560	0.00458	(0.02738, 0.04559)	0.02152	0.00390	(0.01466, 0.03005)
(VI)	0.03520	0.00433	(0.02735, 0.04466)	0.02120	0.00410	(0.01418, 0.02972)

It should be noted that (I) incorporates only the random error of projecting the mortality index and ignores the existence of parameter error, which unavoidably lead to an understatement of variability in mortality movements. In contrast, all the other methods also take parameter error into account, but through different ways: (II) focuses on the estimation error of the drift term of the random walk process, while (III) to (VI) allow for the estimation error of the original model parameters. Undoubtedly, (III) to (VI) offer a more precise and sophisticated mechanism for integrating different sources of errors. Nevertheless, (II) has two potential advantages over the others. First, its computation time is much shorter and the programming required is less demanding. Second, as old-age improvement has become more important, and since (II) generates future death rates at old ages and mortality index with greater variability, this method tends to give a higher estimate of variability for such aggregate measures as life expectancy and present value of an annuity at retirement ages. It means that the resulting allowance for uncertainty would be more conservative, which may be preferred by certain pension and annuity providers. In summary, the optimal choice depends on the requirements of the analysis and the experience and preference of the user. Ideally, if time permits and resources are available, a practitioner could test different simulation methods and compare the results, so as to gain a better understanding of the underlying uncertainty of mortality changes.

Finally, there are three more things to note. First, as discussed above, the difference between (IIIa) and (IIIb) is minimal because the estimated scale parameter is close to 1. (It was calculated to be

1.386 for UK male pensioner data in Renshaw & Haberman, 2008). This may not be the case for other countries or populations. Second, we suggest earlier that the close agreement of the figures from (III) to (V) may be an indication of the good fit of the Poisson LC model (equation (1)) to the data. We suspect that for the situation where the model fit is unsatisfactory (e.g. there are unusual patterns in the residuals), the results produced by different simulation methods may differ considerably. In those cases, the differences may be reduced if some modifications are made to the model to first improve the data fitting. Third, though in this paper, the projection is performed for 42 years based on only 24 years of data, this work is for demonstration purposes only. In practical applications, the convention is that the maximum length of the projection period is taken approximately the same as the length of the fitting period (e.g. Booth *et al.*, 2002).

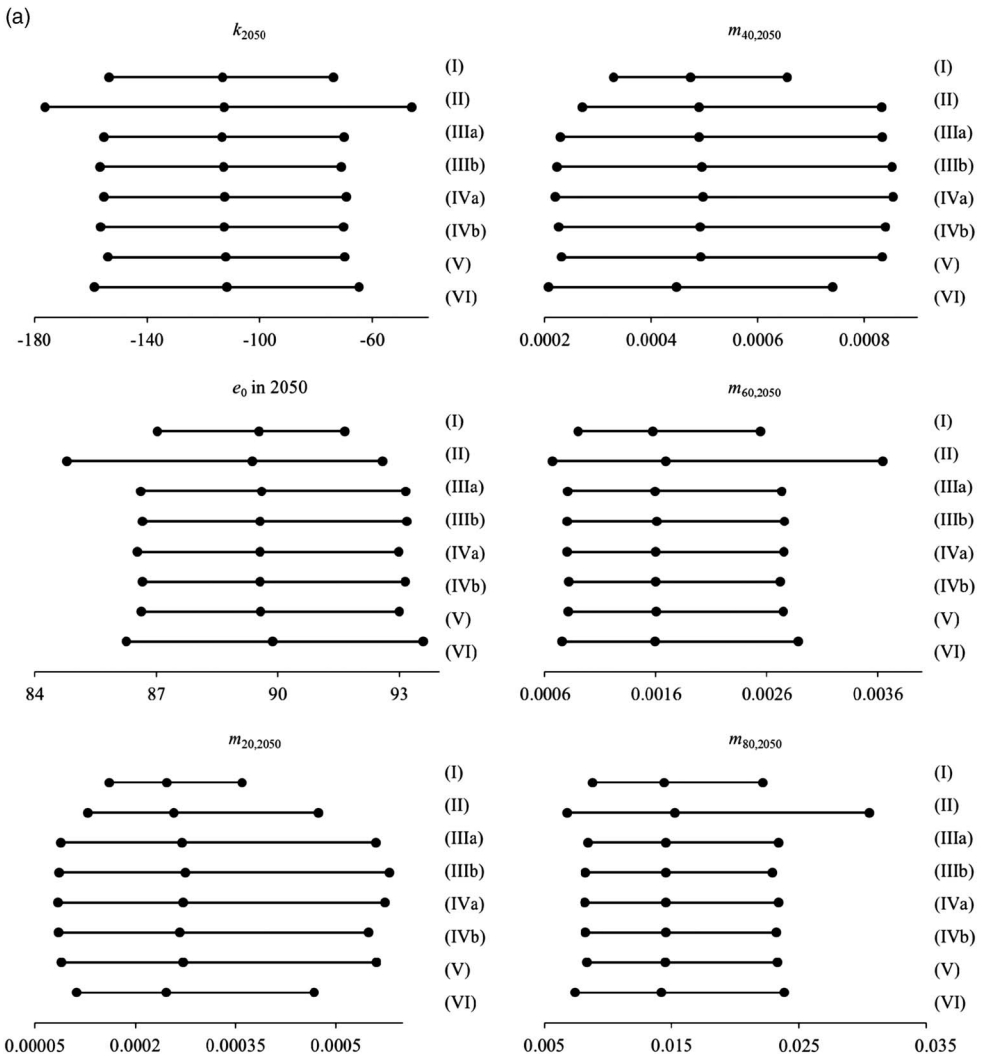


Figure 4. (a) 2.5th percentiles, sample means, and 97.5th percentiles for future mortality index, life expectancy at birth, and death rates in 2050 (females).

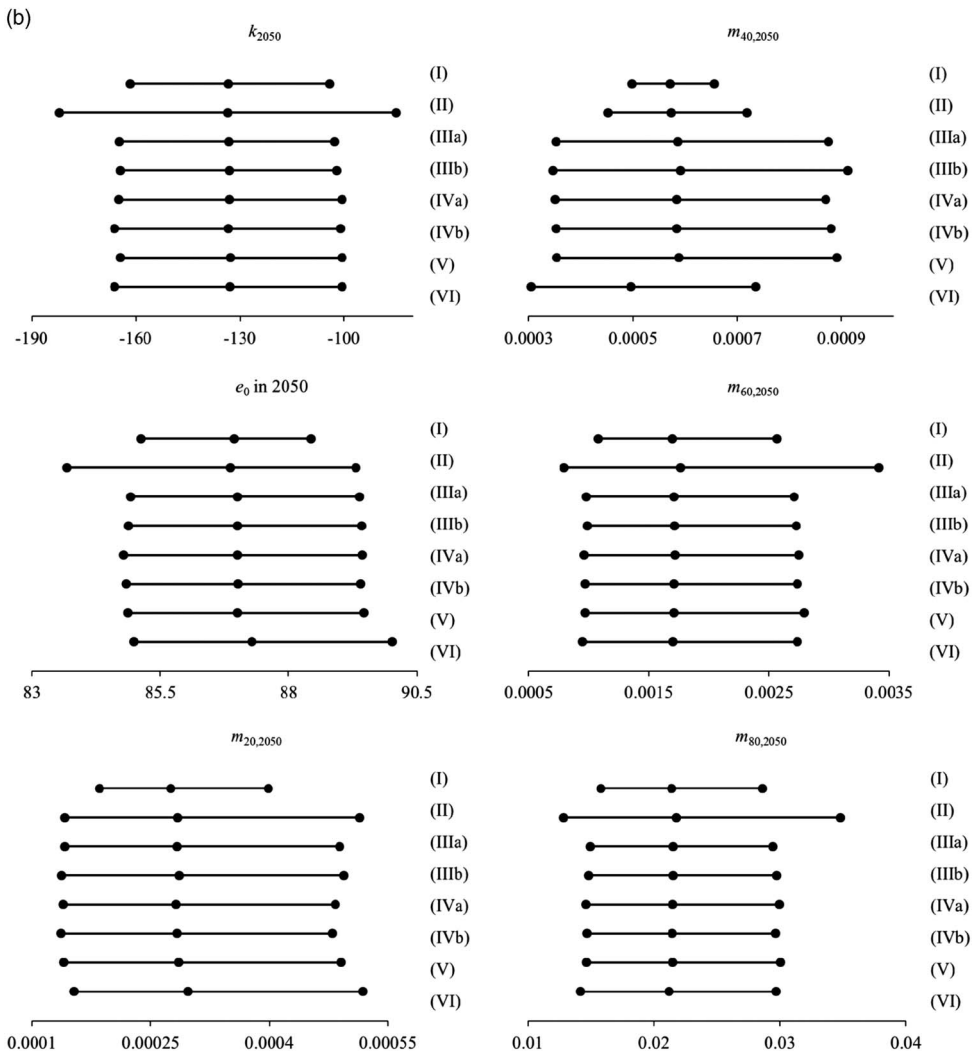


Figure 4. (b) 2.5th percentiles, sample means, and 97.5th percentiles for future mortality index, life expectancy at birth, and death rates in 2050 (males).

4. Concluding Remarks

In this paper, we provide a quantitative comparison of six simulation strategies for mortality projection using the Poisson LC model. These strategies include considering the error term of the mortality index only, including the parameter error of the drift term, simulating the parameters from the multivariate normal distribution, bootstrapping the number of deaths from the Poisson distribution, bootstrapping the deviance residuals of the fitted model, and Bayesian MCMC simulation. We apply these different methods to New Zealand mortality data and find that (a) the first method produces the lowest variability estimates; (b) the second method largely gives the highest variability estimates; and (c) the last four methods lead to similar results. We recognise that the

second method involves much less programming and computation demand than the last four methods and that the final choice depends on the specific requirements and preference of the user. We also emphasise the importance of having a satisfactory model fit before carrying out any simulation of future values.

For future research, we are exploring the possibility to further incorporate the model error. One option is to make use of the Bayesian framework to consider two or more models at the same time and set a prior probability for each model, as suggested in Cairns (2000). It would also be informative to conduct a similar quantitative study for other mortality model structures and data sets. Moreover, D'Amato *et al.* (2011) recently proposed the use of stratified sampling for handling heterogeneity between clusters within a population. It would be interesting to see how this approach could improve the residual bootstrapping process in the cases where there exists dependence between the residuals for certain sub-groups of the population under study. Other work considering the dependence structure in the residuals includes Debón *et al.* (2008, 2010) and D'Amato *et al.* (2012).

Acknowledgements

The author thanks the editor and the anonymous referees for their valuable comments and suggestions. The author is also grateful to Nick Parr and Leonie Tickle of Macquarie University for their generous support and guidance on the thesis.

References

- Australian Prudential Regulation Authority. (2013). Prudential Standard GPS 320. Available online at the address <http://www.apra.gov.au/GI/PrudentialFramework/Documents/GPS-320-Actuarial-and-Related-Matters-January-2013.pdf> (accessed 18 March 2014).
- Booth, H., Maindonald, J. & Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, *56*, 325–336.
- Brouhns, N., Denuit, M. & Van Keilegom, I. (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal*, *3*, 212–224.
- Brouhns, N., Denuit, M. & Vermunt, J.K. (2002a). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, *31*, 373–393.
- Brouhns, N., Denuit, M. & Vermunt, J.K. (2002b). Measuring the longevity risk in mortality projections. *Bulletin of the Swiss Association of Actuaries*, 105–130.
- Cairns, A.J.G. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, *27*, 313–330.
- CEIOPS. (2010). QIS5 technical specifications. Available online at the address http://ec.europa.eu/internal_market/insurance/docs/solvency/qis5/201007/technical_specifications_en.pdf (accessed 18 March 2014).
- Coale, A. & Guo, G. (1989). Revised regional model life tables at very low levels of mortality. *Population Index*, *55*, 613–643.
- Czado, C., Delwarde, A. & Denuit, M. (2005). Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, *36*, 260–284.
- D'Amato, V., Di Lorenzo, E., Haberman, S., Russolillo, M. & Sibillo, M. (2011). The Poisson log-bilinear Lee-Carter model: applications of efficient bootstrap methods to annuity analyses. *North American Actuarial Journal*, *15*(2), 315–333.
- D'Amato, V., Haberman, S., Pisco, G. & Russolillo, M. (2012). Modelling dependent data for longevity projections. *Insurance: Mathematics and Economics*, *51*, 694–701.

- Debón, A., Martínez-Ruiz, F. & Montes, F. (2010). A geostatistical approach for dynamic life tables: the effect of mortality on remaining lifetime and annuities. *Insurance: Mathematics and Economics*, **47**, 327–336.
- Debón, A., Montes, F., Mateu, J., Porcu, E. & Bevilacqua, M. (2008). Modelling residuals dependence in dynamic life tables: a geostatistical approach. *Computational Statistics and Data Analysis*, **52**, 3128–3147.
- Efron, B. & Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- Human Mortality Database. (2012). Human Mortality Database, University of California, Berkeley, USA and Max Planck Institute for Demographic Research, Germany. Available online at the address www.mortality.org (accessed 3 August 2012).
- Kogure, A., Kitsukawa, K. & Kurachi, Y. (2009). A Bayesian comparison of models for changing mortalities toward evaluating longevity risk in Japan. *Asia-Pacific Journal of Risk and Insurance*, **3**(2), 1–21.
- Kogure, A. & Kurachi, Y. (2010). A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. *Insurance: Mathematics and Economics*, **46**, 162–172.
- Koissi, M., Shapiro, A.F. & Hognas, G. (2006). Evaluating and extending the Lee-Carter model for mortality forecasting: bootstrap confidence interval. *Insurance: Mathematics and Economics*, **38**, 1–20.
- Lee, R. (2000). The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal*, **4**(1), 80–93.
- Lee, R.D. & Carter, L.R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, **87**(419), 659–671.
- Li, J. (2010). Projections of New Zealand mortality using the Lee-Carter model and its augmented common factor extension. *New Zealand Population Review*, **36**, 27–53.
- Li, J. (2013). A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies*, **67**(1), 111–126.
- Li, J. (2014). An application of MCMC simulation in mortality projection for populations with limited data. *Demographic Research*, **30**(1), 1–48.
- Li, N., Lee, R. & Tuljapurkar, S. (2004). Using the Lee-Carter method to forecast mortality for populations with limited data. *International Statistical Review*, **72**(1), 19–36.
- Renshaw, A.E. & Haberman, S. (2008). On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee-Carter modelling. *Insurance: Mathematics and Economics*, **42**, 797–816.
- Spiegelhalter, D., Thomas, A., Best, N. & Lunn, D. (2003). WinBUGS user manual. Available online at the address www.mrc-bsu.cam.ac.uk/bugs (accessed 15 January 2009).
- Wang, D. & Lu, P. (2005). Modelling and forecasting mortality distributions in England and Wales using the Lee-Carter model. *Journal of Applied Statistics*, **32**(9), 873–885.