



DEMOGRAPHIC RESEARCH

A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 30, ARTICLE 1, PAGES 1-48

PUBLISHED 8 JANUARY 2014

<http://www.demographic-research.org/Volumes/Vol30/1/>

DOI: 10.4054/DemRes.2013.30.1

Research Article

An application of MCMC simulation in mortality projection for populations with limited data

Jackie Li

© 2014 Jackie Li.

*This open-access work is published under the terms of the Creative Commons Attribution NonCommercial License 2.0 Germany, which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit.
See <http://creativecommons.org/licenses/by-nc/2.0/de/>*

Table of Contents

1	Introduction	2
2	The modified Lee-Carter method for limited data	3
3	Use of Bayesian Lee-Carter models for limited data	5
4	An illustration with Taiwanese mortality data	8
5	Application to Chinese mortality data	31
6	Concluding remarks	39
7	Acknowledgments	40
	References	41
	Appendix	43

An application of MCMC simulation in mortality projection for populations with limited data

Jackie Li¹

Abstract

In this paper, we investigate the use of Bayesian modeling and Markov chain Monte Carlo (MCMC) simulation, via the software WinBUGS, to project future mortality for populations with limited data. In particular, we adapt some extensions of the Lee-Carter method under the Bayesian framework to allow for situations in which mortality data are scarce. Our approach would be useful for certain developing nations that have not been regularly collecting death counts and population statistics. Inferences of the model estimates and forecasts can readily be drawn from the simulated samples. Information on another population resembling the population under study can be exploited and incorporated into the prior distributions in order to facilitate the construction of probability intervals. The two sets of data can also be modeled in a joint manner. We demonstrate an application of this approach to some data from China and Taiwan.

¹ Nanyang Technological University, Singapore. E-mail: JackieLi@ntu.edu.sg.

1. Introduction

The Lee-Carter method (Lee and Carter 1992) and its various extensions (e.g., Lee 2000) have been shown to perform satisfactorily for several countries' mortality data when the data volume is sufficient. For many developing nations, however, there is no periodic collection and systematic management of death data, or annual figures are only available for recent periods, which are not long enough to produce sensible forecasts using the usual Lee-Carter approaches. To address this problem, Li, Lee, and Tuljapurkar (2004) have adapted the original Lee-Carter method to two countries with a limited number of data points at uneven intervals. Based on an a priori assumption that the mortality index follows a random walk with drift, the authors derived formulae for estimating the corresponding parameters with as few as three data points. They also suggested "borrowing" missing information from another population with characteristics similar to those of the population under study, but they did not further develop the borrowing strategy. Our approach here builds on these ideas. With limited data, we adopt two extensions (Gaussian and Poisson errors) of the Lee-Carter structure, and arbitrarily model the mortality index by a random walk with drift without sufficient observations of the underlying process. In particular, we explore how information regarding the variability of the death rates can possibly be "borrowed" from a "similar" population to facilitate the construction of probability intervals. We also attempt to co-model the population under study and the reference population.

The modeling and analysis in this paper are performed under the Bayesian paradigm. Bayesian modeling has some distinct advantages in the context of mortality projection. First, certain prior or reference information, such as the experience of another population, can be incorporated into the modeling process in a formal manner. Second, both the log-bilinear structure and the random walk can be taken into account simultaneously within the same framework, unlike in the rather incoherent two-step procedure of the original Lee-Carter method. Furthermore, with the help of modern computing power, it is possible to implement computer-intensive methods, such as Markov chain Monte Carlo (MCMC) simulation, which can alleviate the problems associated with the analytical intractability of Bayesian models that have often arisen in the past. This simulation technique provides a convenient way to obtain distributions and construct probability intervals for the model estimates and forecasts, in which both process error and parameter error are allowed for.

Some previous work on Bayesian mortality modeling was done by Czado, Delwarde, and Denuit (2005), who implemented a Poisson log-bilinear model with a deterministic trend model for the mortality index to forecast mortality and tested it on French male data. Pedroza (2006) formulated the Lee-Carter method as a state-space model, using Gaussian error terms and a random walk with drift for the mortality index,

to forecast U.S. male mortality. Kogure, Kitsukawa, and Kurachi (2009) considered both the Gaussian and Poisson formulations and compared three different model structures (random walk with drift, deterministic trend model, and stationary model) for the mortality index using Japanese male data. While these authors provided clear demonstrations of their approaches and examples using sufficient datasets (over 30 years), our focus here is on those situations with a very limited amount (e.g., only a few years) of data. In the following sections, we apply the proposed approach to some mortality data from China and Taiwan.

Despite the above-mentioned advantages, the implementation of Bayesian modeling and MCMC simulation is often technically and computationally demanding, as it involves dealing with complex mathematics and scientific programming. The high costs associated with acquiring the necessary knowledge may dissuade some practitioners from experimenting with Bayesian techniques, and may discourage wider applications. One current alternative is to use the specialized software WinBUGS (Spiegelhalter et al. 2003), in which the programming language is relatively easy to handle, and a straightforward specification of many Bayesian models is possible. If greater flexibility is needed, experts may under certain circumstances be able to develop their own MCMC algorithms from scratch. For example, Girosi and King (2008) developed a sophisticated Bayesian framework to incorporate covariates and prior information in order to improve mortality forecasts. But for many practical applications, WinBUGS offers a much more accessible platform to practitioners who are interested in performing Bayesian modeling, but who are unfamiliar with the underlying details. For the kind of models studied in this paper, only a few dozen lines of WinBUGS codes are generally needed. This programming efficiency is highly convenient and it greatly enhances the applicability and practicality of Bayesian modeling in mortality studies. In this work, we present an application of WinBUGS that is used to carry out MCMC simulation for our analysis; the codes used are noted in the appendix. Interested readers may also refer to Chapter 9 of Bijak (2011) for more information about Bayesian computing in practice.

2. The modified Lee-Carter method for limited data

The original method proposed by Lee and Carter (1992) involves the main structure below:

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

in which $m_{x,t}$ is the central death rate at age x in year t , a_x describes the overall mortality pattern across age, b_x represents the sensitivity of the log death rate to changes in the mortality index k_t , and $\varepsilon_{x,t}$ is the error term with mean zero and variance σ_ε^2 . The a_x parameters are first estimated by averaging $\ln m_{x,t}$ over time t . The b_x and k_t parameters are then computed by applying singular value decomposition (SVD) to the matrix with components $\{\ln m_{x,t} - a_x\}$, subject to two constraints $\sum_x b_x = 1$ and $\sum_t k_t = 0$. Finally, the k_t parameters are re-calculated in such a way that the fitted number of deaths and the actual number of deaths are equal for each year. The key strength of this method lies in its straightforwardness, as well as the fact that it produces a highly linear time series of k_t for the data of several countries (e.g., Lee and Miller 2001). This linear feature implies that a random walk with drift would be appropriate for modeling the k_t series:

$$k_t = \mu + k_{t-1} + e_t, \quad (2)$$

where μ is the drift term and e_t 's are independent and identically distributed (iid) error terms with mean zero and variance σ_k^2 . The estimated drift term $\hat{\mu}$ is often negative and corresponds to a broadly linear decline of the mortality index. Future mean values are then projected as $\hat{k}_t = \hat{\mu} + \hat{k}_{t-1}$. This simple model for k_t is adopted in most applications in the literature.

To allow for a situation in which there are only a limited number of data points, probably at uneven intervals, Li, Lee, and Tuljapurkar (2004) modified the random walk process for k_t above. These authors first made a strong assumption that the mortality index follows a random walk with drift, even without support from a sufficient sample size. They argued that this assumption can be justified by the observed patterns of mortality decline in many other countries. For example, Tuljapurkar, Li, and Boe (2000) showed that the mortality index is highly linear for the G7 countries for the period from 1950 to 1994. Let $t_0, t_1, t_2, \dots, t_n$ be the points of time when mortality data are collected, and n can be as small as two, i.e., three data points. Li, Lee, and Tuljapurkar (2004) deduced from (2) that for $1 \leq h \leq n$,

$$k_{t_h} - k_{t_{h-1}} = \mu(t_h - t_{h-1}) + e_{t_{h-1}+1} + e_{t_{h-1}+2} + \dots + e_{t_h}, \quad (3)$$

and estimated the parameters by

$$\hat{\mu} = \frac{k_{t_n} - k_{t_0}}{t_n - t_0} ; \quad (4)$$

$$\hat{\sigma}_k^2 = \frac{\sum_{h=1}^n (k_{t_h} - k_{t_{h-1}} - \hat{\mu}(t_h - t_{h-1}))^2}{t_n - t_0 - \frac{\sum_{h=1}^n (t_h - t_{h-1})^2}{t_n - t_0}} ; \quad (5)$$

$$\widehat{\text{Var}}(\hat{\mu}) = \frac{\hat{\sigma}_k^2}{t_n - t_0} . \quad (6)$$

Future values of the mortality index and the log death rates are then sequentially simulated as (for $t > t_n$)

$$k_t = k_{t_n} + \left(\hat{\mu} - \sqrt{\widehat{\text{Var}}(\hat{\mu})} \eta \right) (t - t_n) + e_{t_n+1} + e_{t_n+2} + \dots + e_t ; \quad (7)$$

$$\ln m_{x,t} = \ln m_{x,t_n} + b_x (k_t - k_{t_n}) , \quad (8)$$

in which it is assumed that $\eta \sim N(0, 1)$ and $e_s \sim N(0, \sigma_k^2)$ (for $s > t_n$) are independent. It is also suggested that the variance of $\hat{\sigma}_k$ can be estimated and incorporated into (7).

3. Use of Bayesian Lee-Carter models for limited data

Based largely on the model setting in Kogure, Kitsukawa, and Kurachi (2009), and using notation similar to that of the previous section, we consider two alternative model specifications. The Gaussian and Poisson error structures, with the prior distributions, are listed below:

$$\ln m_{x,t_h} = a_x + b_x k_{t_h} + \varepsilon_{x,t_h} \quad \text{or} \quad (\text{Gaussian error structure}) \quad (9)$$

$$D_{x,t_h} \sim \text{Poisson} \left(E_{x,t_h} \exp(a_x + b_x k_{t_h}) \right) ; \quad (\text{Poisson error structure}) \quad (10)$$

$$k_{t_h} - k_{t_{h-1}} = \mu(t_h - t_{h-1}) + e_{t_{h-1}+1} + e_{t_{h-1}+2} + \dots + e_{t_h} ; \text{ (modified as in (3) above) } \quad (11)$$

$$\varepsilon_{x,t_h} \sim N(0, \sigma_\varepsilon^2) ; \quad \text{(error term in (9))} \quad (12)$$

$$e_{t_h} \sim N(0, \sigma_k^2) ; \quad \text{(error term in (11))} \quad (13)$$

$$a_x \sim N(0, \sigma_a^2) ; \quad \text{(prior distribution of } a_x \text{)} \quad (14)$$

$$b_x \sim N\left(\frac{1}{n_a}, \sigma_b^2\right) ; \quad \text{(prior distribution of } b_x \text{)} \quad (15)$$

$$\mu \sim N(\mu_0, \sigma_\mu^2) ; \quad \text{(prior distribution of } \mu \text{)} \quad (16)$$

$$\sigma_\varepsilon^{-2} \sim \text{Gamma}(\alpha_\varepsilon, \beta_\varepsilon) ; \quad \text{(prior distribution of } \sigma_\varepsilon^2 \text{)} \quad (17)$$

$$\sigma_k^{-2} \sim \text{Gamma}(\alpha_k, \beta_k) , \quad \text{(prior distribution of } \sigma_k^2 \text{)} \quad (18)$$

where data are collected in years $t_0, t_1, t_2, \dots, t_n$; m_{x,t_h} , D_{x,t_h} , and E_{x,t_h} are the central death rate, number of deaths, and (known) exposure at age x in year t_h ; σ_a^2 , σ_b^2 , and σ_μ^2 are the variances of the prior distributions; μ_0 is the mean of the drift term μ ; n_a is the number of age groups; and α_ε , β_ε , α_k , and β_k are the parameters of the prior distributions of the error terms' (inverse) variances. Given the variances σ_ε^2 and σ_k^2 , it is assumed that the error terms ε_{x,t_h} 's are iid across age and time, and that e_{t_h} 's are iid over time. The two constraints $\sum_x b_x = 1$ and $\sum_{h=0}^n k_{t_h} = 0$ are also set here to ensure that the parameters are identifiable and that the simulated results converge. Note that it is necessary to choose between using (9), (12), and (17) (Gaussian); or (10) (Poisson).

The parameters a_x , b_x , and k_{t_h} in the Gaussian and Poisson error structures (9) and (10) are in line with those in (1). If relevant exposures and death counts data are available, it may be preferable to use (10) rather than (9), as the Poisson assumption is a natural choice for counting the number of deaths. In contrast, the homoscedastic error terms (12) of the Gaussian assumption may be problematic, as the log death rates at older ages often have higher variability than those at younger ages. One way to tackle

this problem is to set different variances for different age groups, as was suggested in Pedroza (2006), if the data volume is adequate.

The model setting above involves two main modifications to the initial Lee-Carter method. First, without much support from the data being studied, the mortality index is arbitrarily modeled by a random walk with drift, as in Li, Lee, and Tuljapurkar (2004). The future values of k_t 's (simulated as $k_t = \mu + k_{t-1} + e_t$) and $m_{x,t}$'s for $t > t_n$ are then generated in the simulation process. Apart from the ample evidence of the linearity of many other populations, it would, from a pragmatic point of view, be difficult to justify the use of more complicated models (e.g., higher-order ARIMA) if there are only a few data points in the time series. Second, the log-bilinear structure and the random walk are treated in a more coherent manner under the Bayesian framework than in the original two-step estimation procedure. Another advantage is that since the process distributions and the prior distributions are all handled within the same framework, both process error and parameter error are automatically allowed for when the posterior distributions and the probability intervals are deduced.

Note that, by using the trick in Li, Lee, and Tuljapurkar (2004), the distributions of the log death rates in (9) (or the number of deaths in (10)) are specified only for those few years when data are actually collected. Accordingly, the random walk with drift is formulated as in (3), in which the sum of the error terms is modeled as a single normal random variable itself (see Appendix). From an empirical perspective, we realize that, with WinBUGS, if all of the years during the period under consideration are fully specified in the original way instead and the missing data are treated as variables, the number of variables involved increases significantly, and it becomes very difficult for the simulation process to reach convergence. For example, for the Taiwanese data studied in the next section, more than 1,600 additional variables are needed for the missing years if the latter (original) approach is taken. Relevant exposures data are also lacking if the Poisson error structure (10) is used. Hence, we choose to specify only those years with data, as in Li, Lee, and Tuljapurkar (2004). In this way, computation efficiency is much improved and the degree of convergence is satisfactory. On the other hand, by developing a new MCMC algorithm, the multiple imputation technique may be adopted to allow for the missing values across time, as was suggested in Pedroza (2006).

Compared to Kogure, Kitsukawa, and Kurachi (2009), we use more diffuse prior distributions. The variances σ_a^2 and σ_b^2 are taken as 10 times the sample variances of \hat{a}_x 's and \hat{b}_x 's (by SVD), respectively. The terms α_e and α_k are set to be 2.01. Note that taking these close to two gives a large variance, since $\text{Var}(\sigma_k^2) = \beta_k^2 / (\alpha_k - 1)^2 (\alpha_k - 2)$. For the other parameters, we follow Kogure, Kitsukawa, and Kurachi (2009) and make use of (4), (5), and (6).

Under the Bayesian framework, the key step is to find the posterior distribution of the unknown parameters and quantities given the data. The posterior density function is derived from $f(\boldsymbol{\theta} | \mathbf{D}) \propto f(\mathbf{D} | \boldsymbol{\theta}) f(\boldsymbol{\theta})$, where f denotes a density function, $\boldsymbol{\theta}$ represents the unknown parameters and quantities, and \mathbf{D} stands for the data. However, due to the complexity of the models considered, it is difficult to obtain an explicit expression for the posterior density. A practical solution is to utilize MCMC simulation, in which samples are simulated from a Markov chain that has the stationary distribution equal to the posterior distribution. Inferences can then be made from the simulated distribution. The MCMC simulation technique used in WinBUGS is called Gibbs sampling. As was discussed above, the model specification is straightforward in the WinBUGS platform. For example, the Poisson error structure (10) can simply be coded as “ $d[i,j] \sim \text{dpois}(\text{lambda}[i,j])$ ” and “ $\text{lambda}[i,j] <- e[i,j] * \exp(a[i] + b[i] * k[j])$ ”, and also the prior distribution (16) as “ $\mu \sim \text{dnorm}(\mu0, \text{invsigma2}.\mu)$ ”, which are fairly self-explanatory in their meaning. More details on programming and implementation can be found in the appendix and in Spiegelhalter et al. (2003).

The initial values for the simulation are assigned randomly by a built-in function in WinBUGS. We also try using the SVD estimates and the values given by (4) and (5) as the initial values, and the results are similar.

4. An illustration with Taiwanese mortality data

The data on the number of deaths and exposures (by single age with two sexes combined) from Taiwan from 1970 to 2008 are collected from the Human Mortality Database (HMD). For illustration purposes, we apply the Poisson error structure (10) to the data of 1970, 1980, and 1990, as if the data were only available for these three particular years. The death rates are projected for the period from 1991 to 2008 under the Bayesian framework as described previously, and the projected figures are then compared with the actual observations.

Figure 1 plots the log death rates against age for 1970, 1980, and 1990, as well as the survival curve. We can see that mortality has generally declined over time, and that the improvement has been uneven at different ages. There has also been a continual rectangularization of the survival curve. The observed patterns of these data form an initial basis for the projection from 1991 to 2008. Accordingly, the top panels of Figure 2 show the parameter estimates (i.e., sample means) of a_x , b_x , and k_{i_h} (markers), and the actual values of life expectancy at birth (the zigzag increasing trend), with their projections (solid lines) and 95% prediction intervals (inner pair of dashed lines), which were generated by applying the Bayesian model to the data of 1970, 1980, and 1990.

Note that only the narrowest pair of prediction intervals are of concern for now. For comparison, the bottom panels of Figure 2 set out the matching results produced from using the full set of data from 1970 to 1990 and the ML (maximum likelihood) estimates. It is interesting to observe that although the a_x and b_x parameter estimates are by nature less stable in the former case (with only three years of data) than in the latter (with 21 years of data), the patterns of the two cases are broadly similar. Moreover, since the estimation of the drift term depends largely on the starting and ending values k_{t_0} and k_{t_n} , we can see that there is not much difference in the projected mortality index and in the projected life expectancy between the two cases. The implication is that the extra 18 years of data of the latter case do not really add much content to the computation of the mean forecasts, given the same length of the fitting period and a highly linear series of the mortality index. With as few as three data points, sensible mean forecasts could still be made.

Nevertheless, when the variability of the model forecasts is of concern, there are two potential drawbacks to using only a few years of data of one single population spanning a period of only around 20 years. First, the few years of data may not be sufficient for providing a reasonable estimation of the variability of the mortality index. As shown in the top panel of Figure 2 (page 13), the estimated k_{1970} , k_{1980} , and k_{1990} lie in an almost perfectly straight line. The resulting value given by (5) is then very small, and does not truly reflect the underlying variability of the k_{t_n} series. Second, since the focus is on one population only and the fitting period is not very long, the possibility of a major shift in the trend in mortality improvement could be overlooked. The life expectancy plot in the top panel of Figure 2 (page 14) reveals the potential significance of these problems. For the earlier part of the projection period (from 1991), the actual life expectancy trend moves outside the 95% prediction intervals a few times; for the later period, there is a significant underestimation of life expectancy and the prediction bounds completely fail to capture the observed trend. The width of the bounds is only about one year at 2008. At the single age level, Figure 3 also shows that the actual log death rates (the zigzag decreasing trends) turn out to have declined more quickly than had been projected (solid lines) at several ages, and the 95% prediction intervals (inner pair of dashed lines) are unable to cover the actual values in a number of years.

Figure 1: Log death rates and survival curve – Taiwan

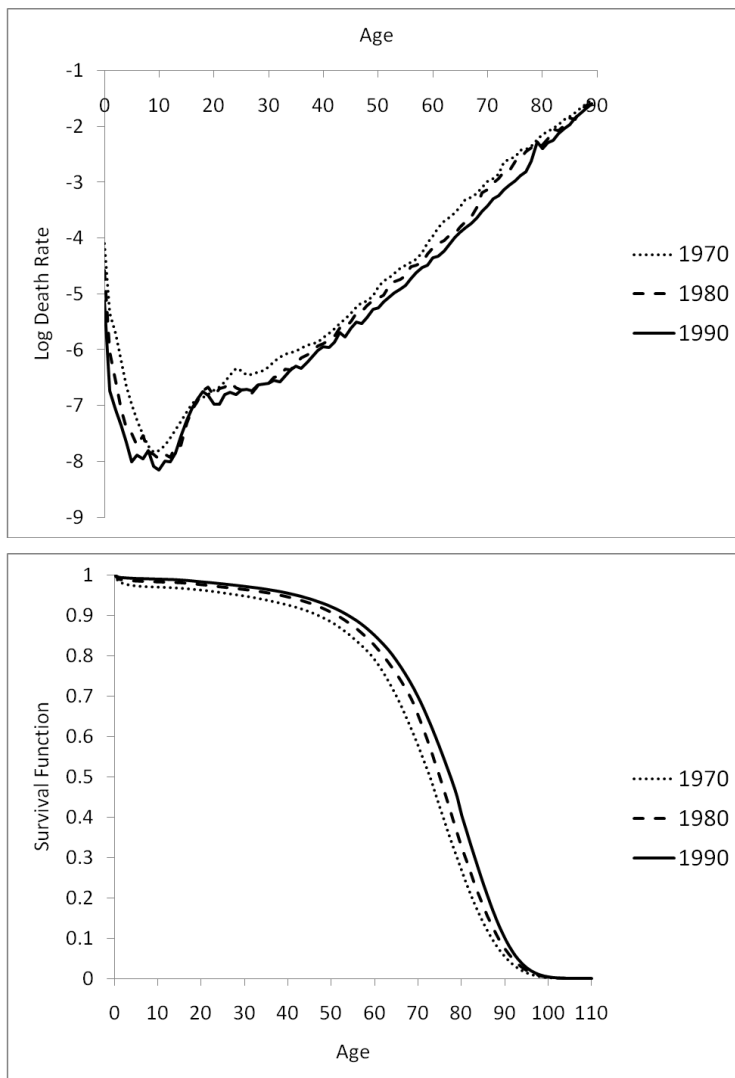


Figure 2: Parameter estimates of a_x , b_x , and k_{t_h} and life expectancy at birth – only 3 years of Taiwan data of 1970, 1980, and 1990 (top) vs. 21 years of Taiwan data from 1970 to 1990 (bottom)

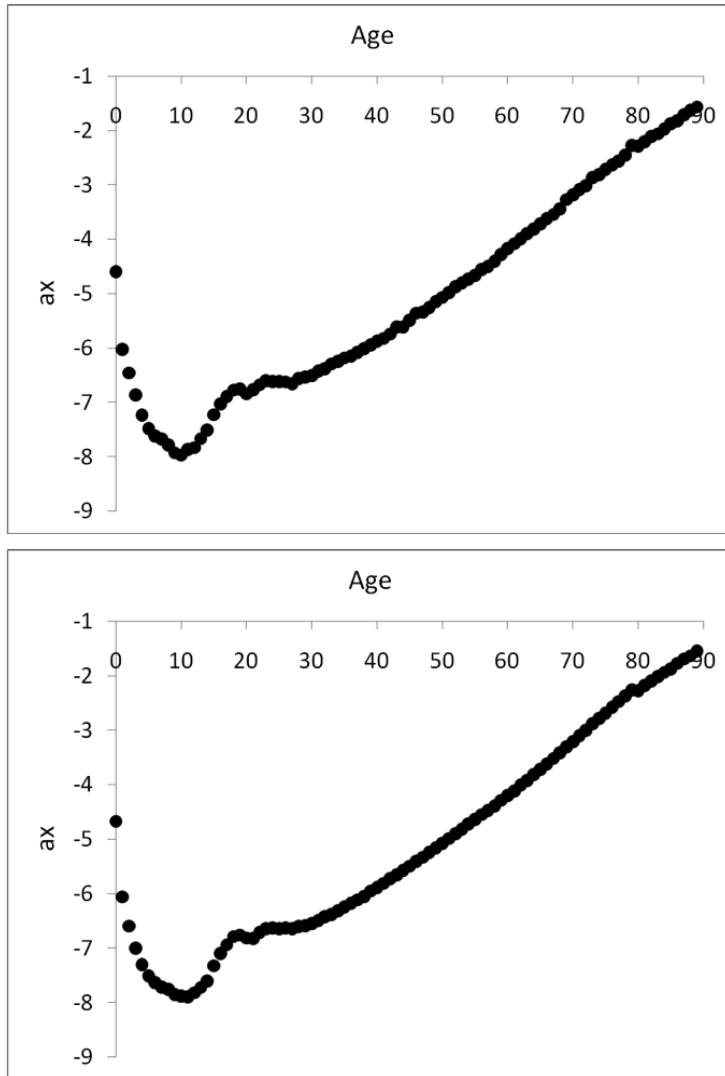


Figure 2: (Continued)

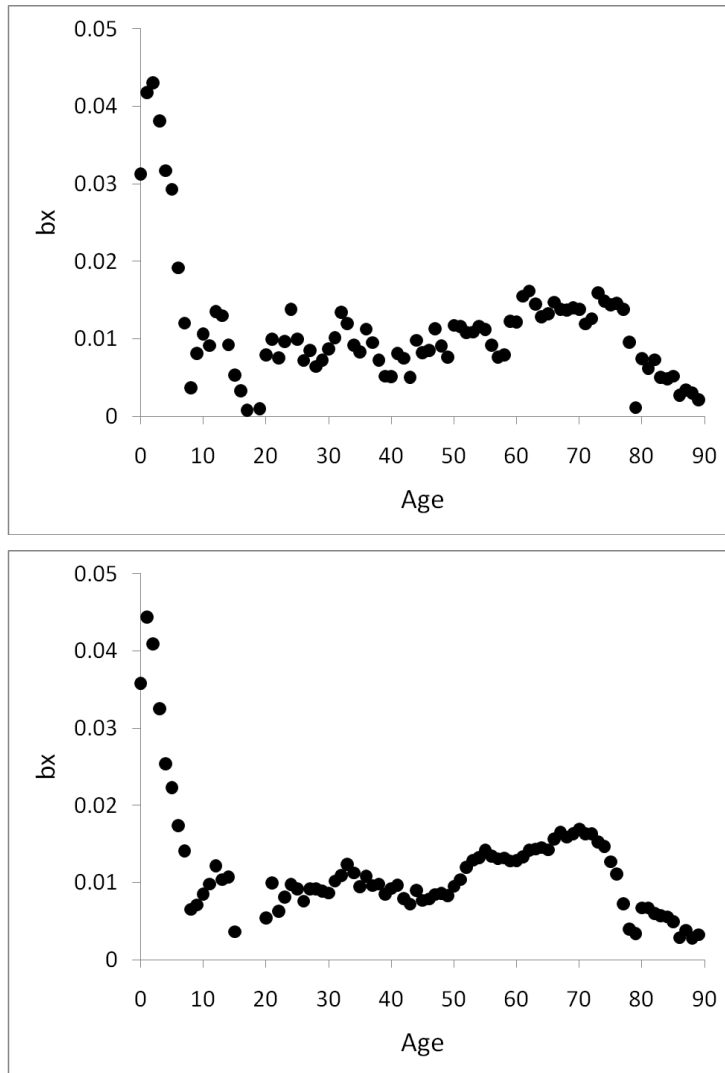


Figure 2: (Continued)

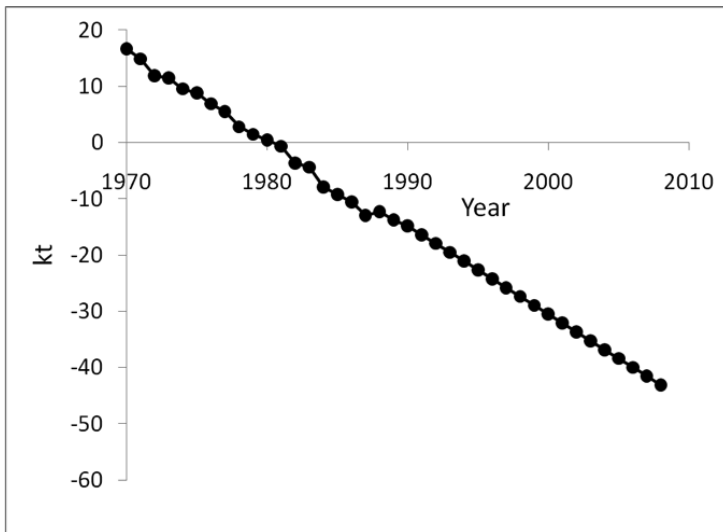
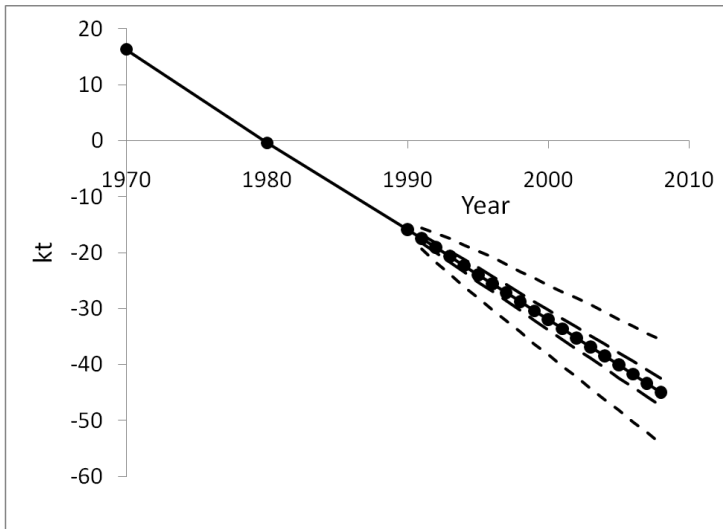
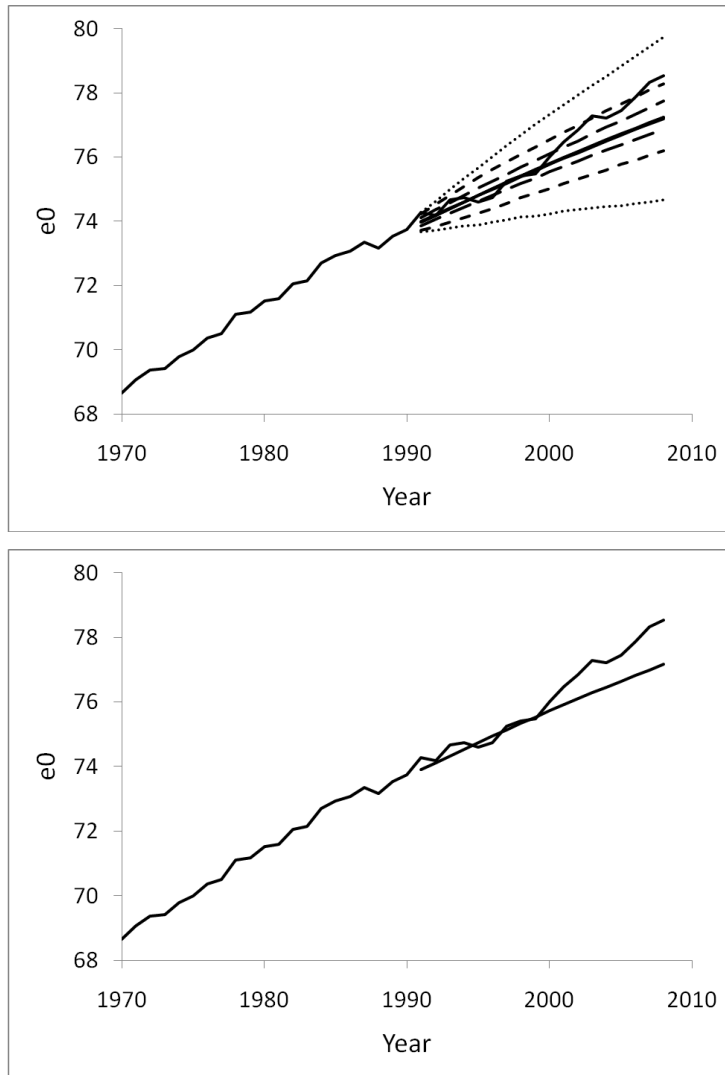


Figure 2: (Continued)



Note: For the projections on pages 13 and 14, the solid straight lines represent the sample means / projected values. The inner pair of dashed lines refers to the 95% prediction intervals without adjustment. The middle pair of dashed lines and the outer pair of dotted lines refer to the intervals after adjusting for σ_k^2 and then Δ_x respectively.

To deal with these problems, certain references can be drawn from another population which has more data available and is believed to possess similar attributes. Continuing with the hypothetical example above with only three years of Taiwanese data, we make use of Japanese data from 1970 to 1990 (from the HMD) and examine the estimated mortality index and the rate of change in mortality rates. We have chosen Japan as the source of extra information because historical factors, cultural interactions, trade, and geographical proximity connect Japan and Taiwan. In addition, Japan is the world leader in life expectancy, and useful insights could be gleaned from its experience. We start by attempting to “borrow” information that will allow us to adjust the variability of the death rates (and also to incorporate the possibility of a major shift). But this information is not for modifying the mean values, which are based solely on the few years of Taiwanese data. As was discussed previously, these few data points would still form a reasonable basis for projecting the mean values if the mortality index is assumed to be highly linear.

When comparing the Taiwanese data (three years) with the Japanese data (21 years), the first thing we discover is that the estimated value of $|\sigma_k/\mu|$ of the latter, computed from using (4) and (5), is 4.44 times of that of the former. (In fact, the similar ratio comparing the full set of Taiwanese data to only the three years considered is 4.18.) The implication is that if the variability of the Japanese mortality index offers a reasonable indication of the underlying variability of the Taiwanese index, the “true” value of σ_k^2 for the Taiwanese case should be around 20 times the value that was initially estimated based on the three years of data used. Accordingly, we revise the relevant parameters of the prior distributions and fit the Bayesian model again. As expected, there is virtually no change in the mean forecasts after these adjustments are made. As illustrated in the top panel of Figure 2 (page 14), however, the observed trend of life expectancy now lies mostly within the new, wider prediction intervals (middle pair of dashed lines), except for a few years in the later part of the projection period. Note that the width of the new bounds is around two years at 2008.

Figure 3: Actual log death rates vs. projected values with 95% prediction intervals – Taiwan

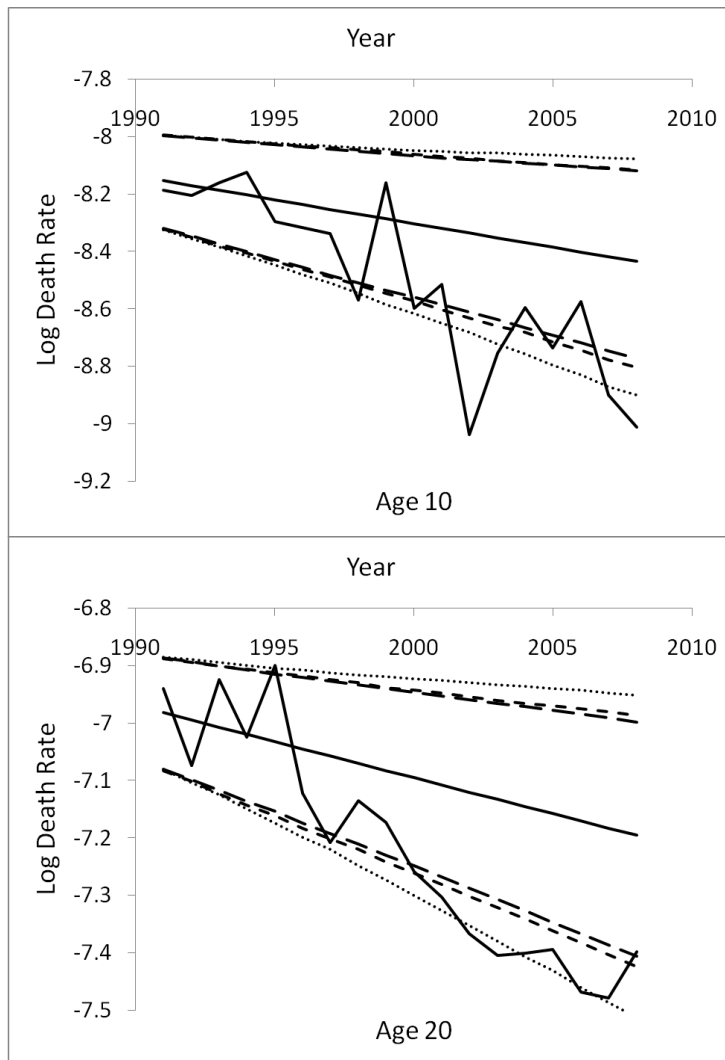


Figure 3: (Continued)

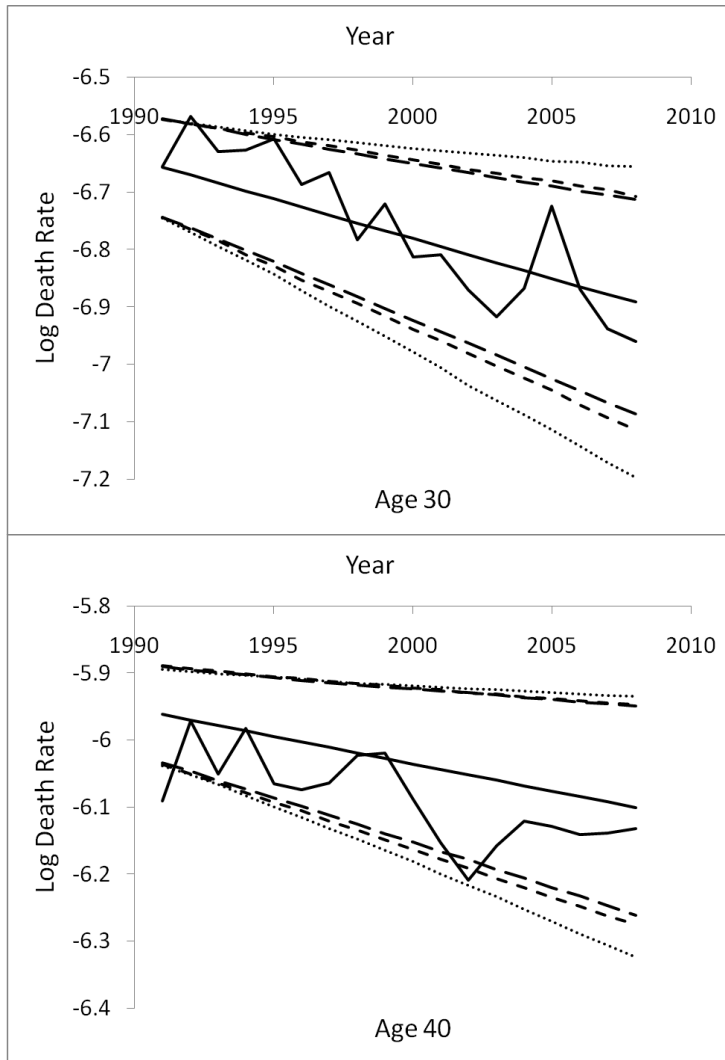


Figure 3: (Continued)

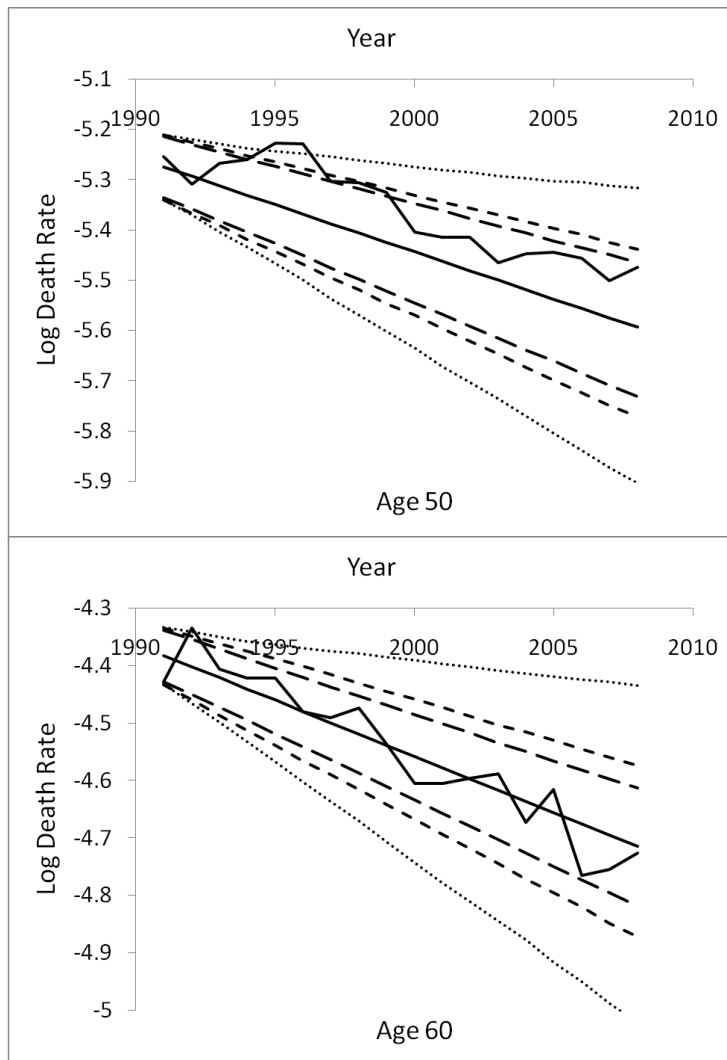
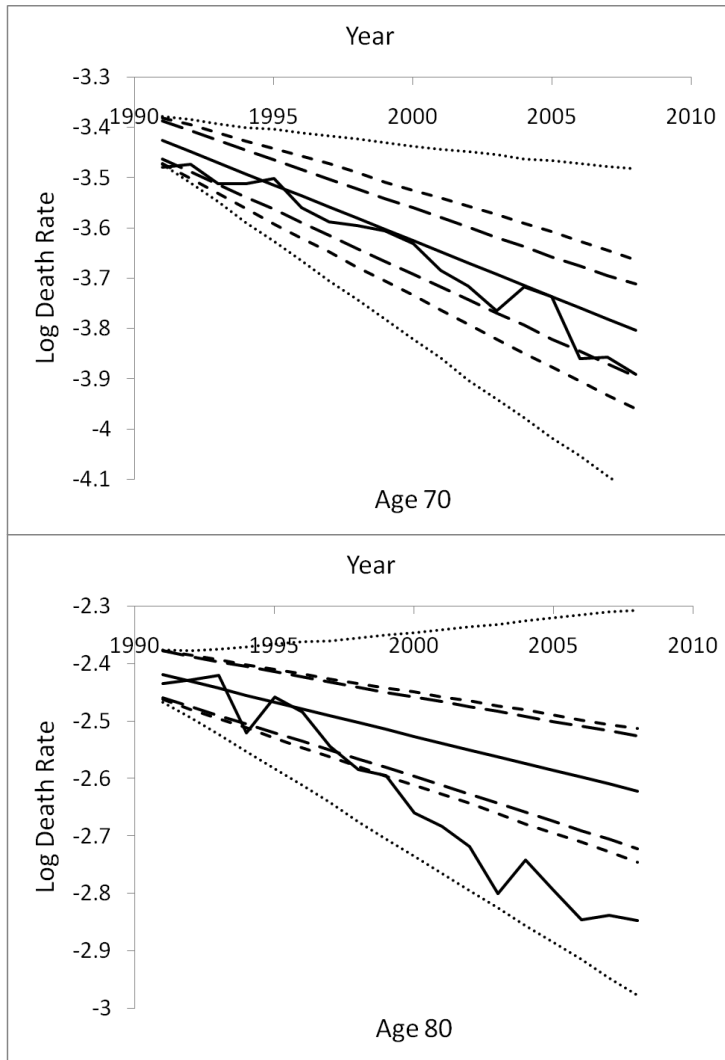


Figure 3: (Continued)



Note: The zigzag decreasing trends represent the actual values, and the solid straight lines represent the projected values. The inner pair of dashed lines refers to the 95% prediction intervals without adjustment. The middle pair of dashed lines and the outer pair of dotted lines refer to the intervals after adjusting for σ_k^2 and then Δ_x respectively.

As an alternative, we have also tested co-modeling the two datasets and aligning their σ_k/μ ratios. In particular, (10) to (11), (13) to (16), and (18) are applied to the two datasets simultaneously (with two different sets of notation for differentiation). For the Taiwanese data, the following modifications are made:

$$\mu \sim \text{N}\left(\mu_0, \frac{(\mu_0 r)^2}{t_n - t_0}\right); \quad (\text{prior distribution of } \mu) \quad (19)$$

$$\sigma_k^2 = (\mu r)^2, \quad (20)$$

in which r is the σ_k/μ ratio of the Japanese data. Effectively, the information of the underlying variability of the Japanese mortality index “flows through” the Bayesian mechanism to influence the assessment of the variability of the Taiwanese mortality index. The resulting prediction intervals are found to be of a similar width (not shown here) to those produced by the above approach of revising the constant parameters of the priors, and the computation time increases because we are dealing with a larger data volume.

Despite the improvement above in dealing with the overall variability by adjusting the priors, the effects at the single age level are still not that satisfactory. Figure 3 illustrates that the new prediction bounds (middle pair of dashed lines) are only marginally wider than previously for the younger ages; and while the differences are more prominent for the older ages, the unexpectedly large drop in death rates for ages 80 and over (not shown here) causes a significant breach of the lower bound. It is the latter case of older ages that increasingly contributes to rising longevity: based on our calculations on the Taiwanese data, those aged 50 and above are responsible for 51% of the increase in life expectancy from 1970 to 1980, and this proportion rises to 65% for the period from 1980 to 1990 (75% for 1990-2008). In fact, we find that the key contributors have been moving to progressively higher age groups over time for many populations. It is important to make a special allowance for these old ages where necessary when seeking to make a more reliable forecast of future mortality and life expectancy.

The corresponding results via the initial Li, Lee, and Tuljapurkar (2004) approach are also provided in Figure 4 for further comparison. Though the two sets of results are obtained from using different error structures (Poisson vs Gaussian) and are not directly comparable, we can see that the Bayesian prediction intervals in Figure 3 (inner and middle pairs of dashed lines) are generally wider than those shown in Figure 4 (inner and outer pairs of dashed lines). This difference shows why it is advantageous that both process error and parameter error are taken into account naturally within the Bayesian

paradigm, particularly when data are limited. Table 1 also provides some numerical results of the two cases for illustration.

The second thing we realize is that the annual rate of decline of death rates from 1970 to 1990 is on average about 1.5% higher for Japan than for Taiwan, and this advantage is spread fairly evenly across different age groups. It is possible for other populations to learn from Japan's experience and catch up with its pace to some extent (at least in certain age groups) in the longer run. Hence, to further improve the results, in addition to the adjustments above (revising the priors), we modify the projection formula $k_t = \mu + k_{t-1} + e_t$ with the following:

$$k_t^{(x)} = \mu + k_{t-1}^{(x)} + e_t + \Delta_x, \quad (\text{specific to age } x) \quad (21)$$

in which $\Delta_x \sim N(0, \sigma_x^2)$ is independent of the other variables and represents a “shock” to the original drift term μ . In this way, the mean forecasts remain more or less the same, but the variability can be adjusted so as to allow for the possibility of a major shift in mortality decline at certain ages. For demonstration purposes, we arbitrarily set the values of σ_x^2 so that the 99th percentile of $b_x \Delta_x$ equals the difference in the annual rate of decline of death rates between Japan and Taiwan from 1970 to 1990. Roughly speaking, we make an a priori assumption that, in general, it is possible for Taiwan to make up some of this difference after 1990, but that there is only a 1% chance of Taiwan completely catching up with the staggering pace of improvement observed in Japan. The corresponding results are also shown in Figures 2 and 3. We can now see that most of the observed values, including those for ages 80 and over, lie well within the revised prediction bounds (outer pair of dotted lines), although the intervals seem to be somewhat too wide for some of the older ages, particularly for the upper bound. The width of the bounds for life expectancy is about five years at 2008. With more information about the population or from other relevant sources, it may be possible to refine the prediction intervals by, for example, restricting the addition of Δ_x to only specific ages, or truncating the normal distribution of Δ_x from above.

Figure 4: Actual log death rates vs. projected values with 95% prediction intervals via the Li, Lee, and Tuljapurkar (2004) approach – Taiwan

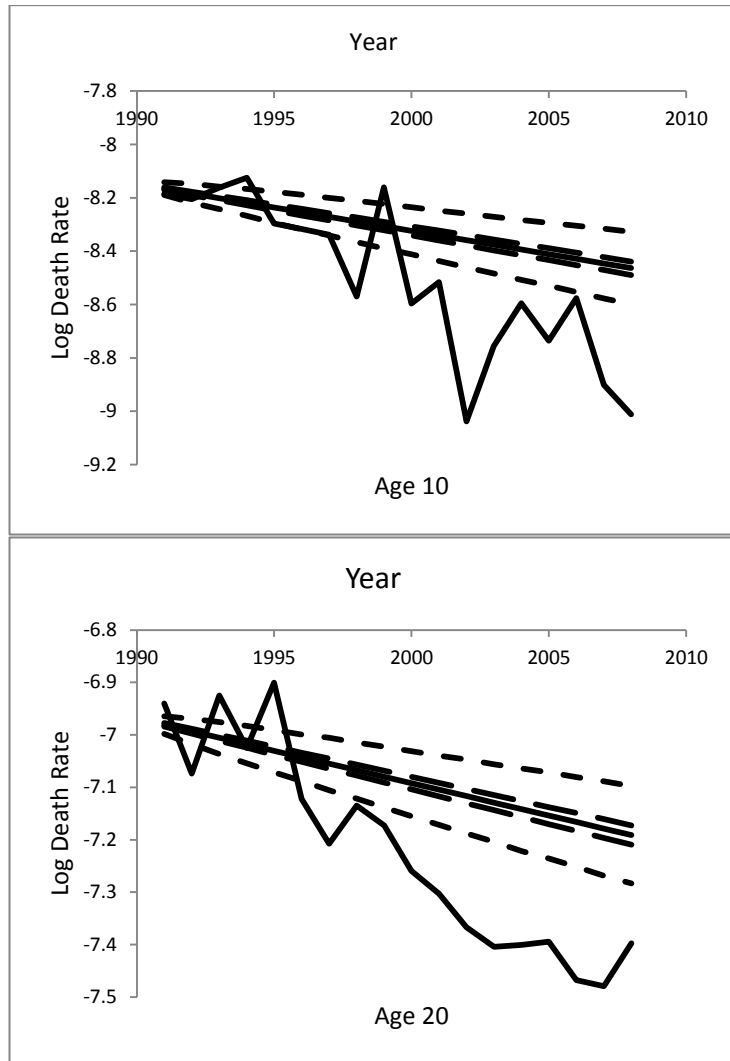


Figure 4: (Continued)

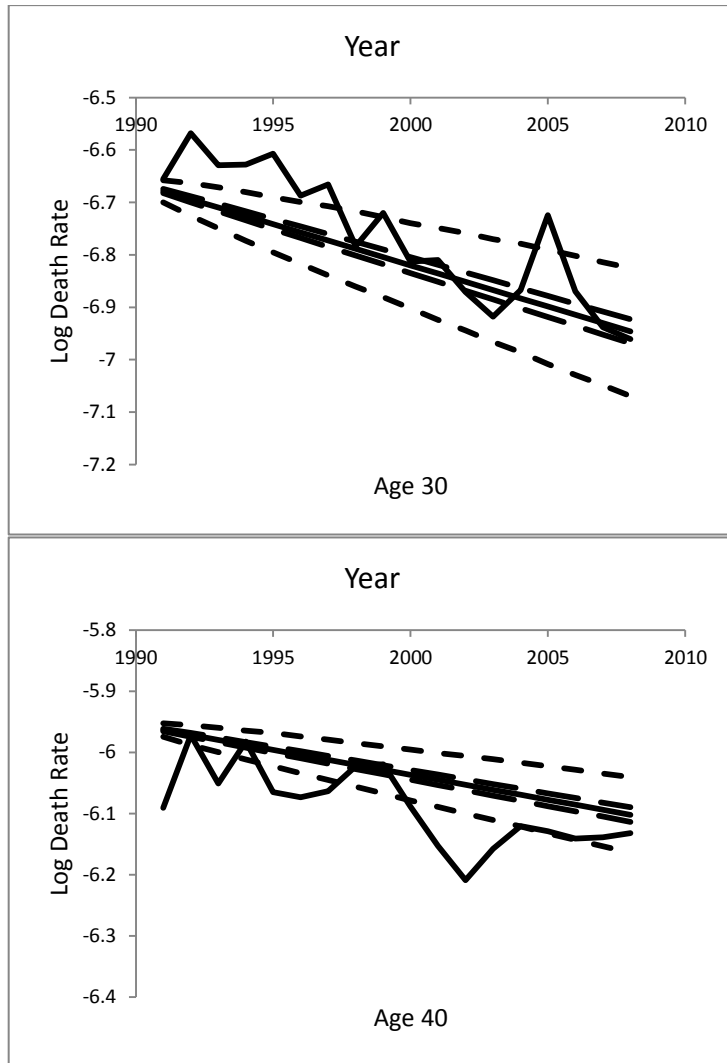


Figure 4: (Continued)

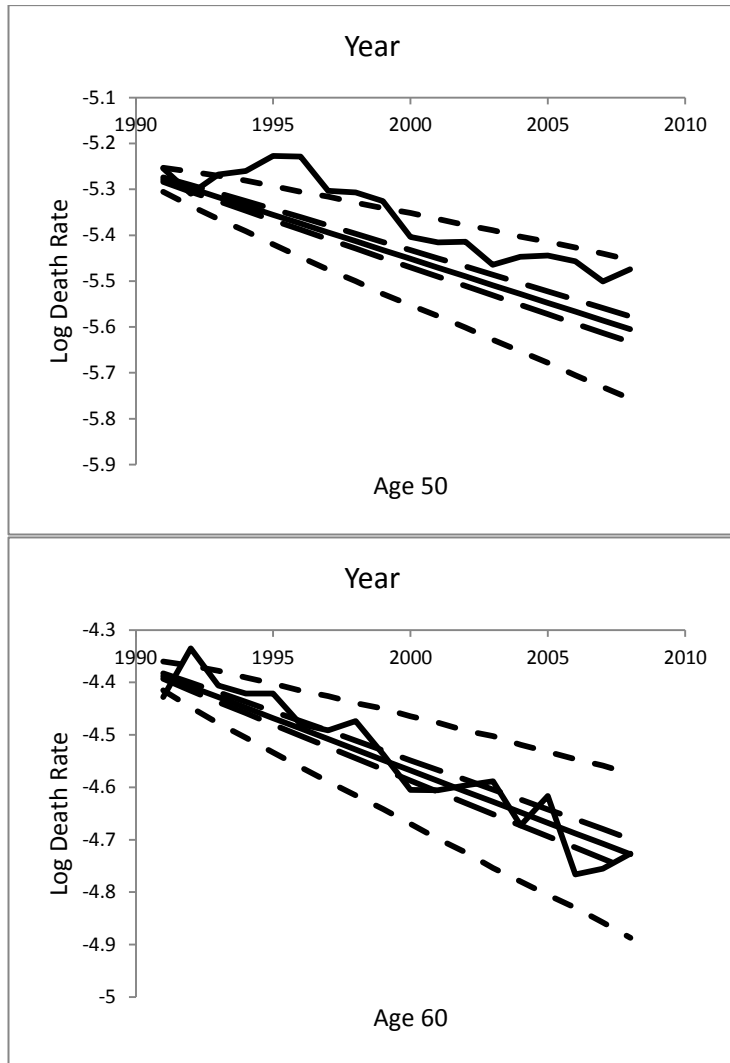
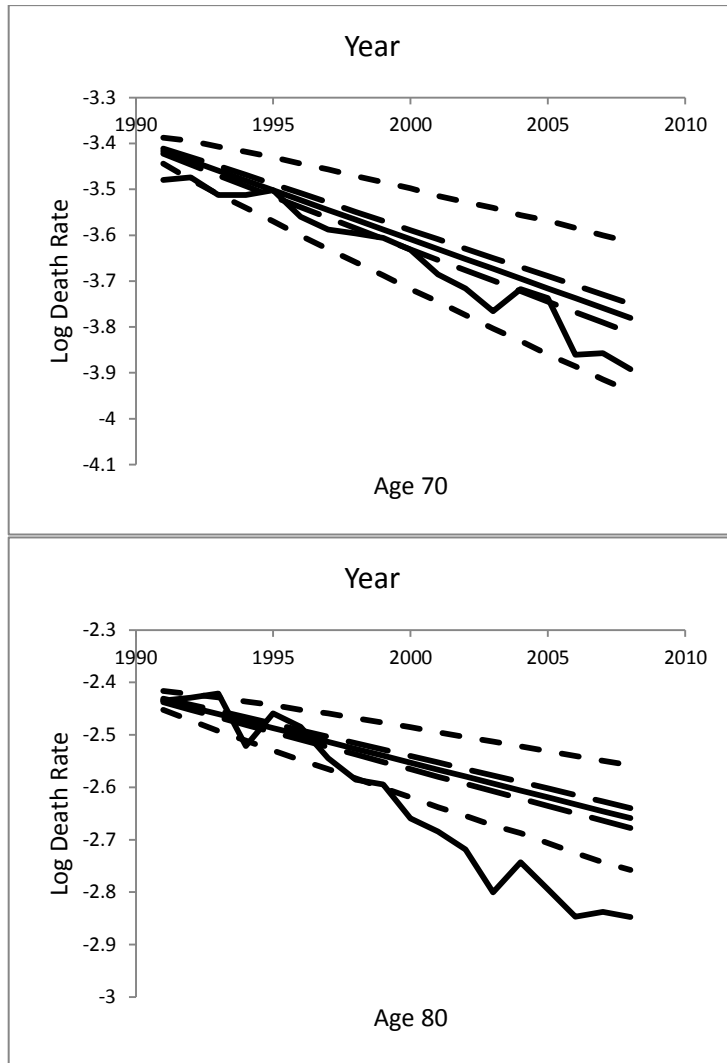


Figure 4: (Continued)



Note: The zigzag decreasing trends represent the actual values, and the solid straight lines represent the projected values. The inner (outer) pair of dashed lines refers to the 95% prediction intervals before (after) adjusting for σ_k^2 .

Table 1: 95% prediction intervals of the projected log death rates in 2008 simulated from the Bayesian approach and the Li, Lee, and Tuljapurkar (2004) approach – Taiwan

Age	Bayesian approach		Li, Lee, and Tuljapurkar (2004)	
	Not adjusted	Adjusted	Not adjusted	Adjusted
10	(-8.77,-8.12)	(-8.80,-8.12)	(-8.49,-8.44)	(-8.60,-8.33)
20	(-7.41,-7.00)	(-7.42,-6.99)	(-7.21,-7.17)	(-7.28,-7.10)
30	(-7.09,-6.71)	(-7.11,-6.71)	(-6.97,-6.92)	(-7.07,-6.82)
40	(-6.26,-5.95)	(-6.27,-5.95)	(-6.11,-6.09)	(-6.16,-6.04)
50	(-5.73,-5.46)	(-5.77,-5.44)	(-5.63,-5.58)	(-5.76,-5.45)
60	(-4.82,-4.61)	(-4.87,-4.57)	(-4.76,-4.70)	(-4.89,-4.57)
70	(-3.90,-3.71)	(-3.96,-3.66)	(-3.81,-3.75)	(-3.94,-3.62)
80	(-2.72,-2.53)	(-2.75,-2.51)	(-2.68,-2.64)	(-2.76,-2.56)

From this rather hypothetical example, we can see that the Bayesian framework is flexible in allowing for different extents of variability when constructing prediction intervals for the model forecasts. The experience of another population can be readily incorporated into the modeling process by adjusting the parameters of the prior distributions, modeling the two sets of data simultaneously, or adding a “shock” component to the drift term of the mortality index. As was mentioned earlier, we suppose that the supplementary data are relevant only for the underlying variability, but not for the mean values.

On the other hand, prior information or subjective beliefs regarding the mean of the drift term may also be used to help determine the mean forecasts if the value given by (4) is deemed to be too optimistic or pessimistic. (In effect, ignoring the error terms, $b_x \mu_0$ refers to the “average” annual rate of change of death rates at age x . Since $\sum_x b_x = 1$, μ_0 / n_a then roughly refers to the overall annual rate of change of death rates.) In such cases, it may be possible to adjust μ_0 directly, or, alternatively, to co-model the limited dataset with a sufficient, related one jointly. One option is to adopt the idea in Carter and Lee (1992) and modify (10) as follows:

$$D_{x,t,i} \sim \text{Poisson}\left(E_{x,t,i} \exp(a_{x,i} + b_{x,i} k_t)\right); \quad (\text{Poisson error structure}) \quad (22)$$

in which there is a common mortality index k_i for the two populations, and the new subscript i indicates the particular population being referred to, e.g., $i = 1$ for Taiwan and $i = 2$ for Japan. As the volume of the Japanese data is much larger than that of the Taiwanese data used, the estimation of the common mortality index is largely determined by the former, which would influence both the mean and the variability of the projected mortality decline of the latter. Compared to all of the previous adjustments, incorporating a common mortality index provides a more coherent and systematic way to spread the information of the reference population to the population with limited data. But in this way, the projected mean values would also be affected. Figure 5 shows that while the projected mortality improvement and prediction intervals are in line with the actual figures for the older ages, there is significant overestimation of mortality decline for the younger ages. These effects are in line with our earlier observation that the death rates of Japan (1970-1990) decreased at a faster pace than those of Taiwan for different age groups. To use this co-modeling approach properly, a decision must be made about whether the population under investigation would follow the past trends of the reference population, and if so, for which age groups this would occur.

Figure 5: Actual log death rates vs. projected values with 95% prediction intervals using a common mortality index – Taiwan

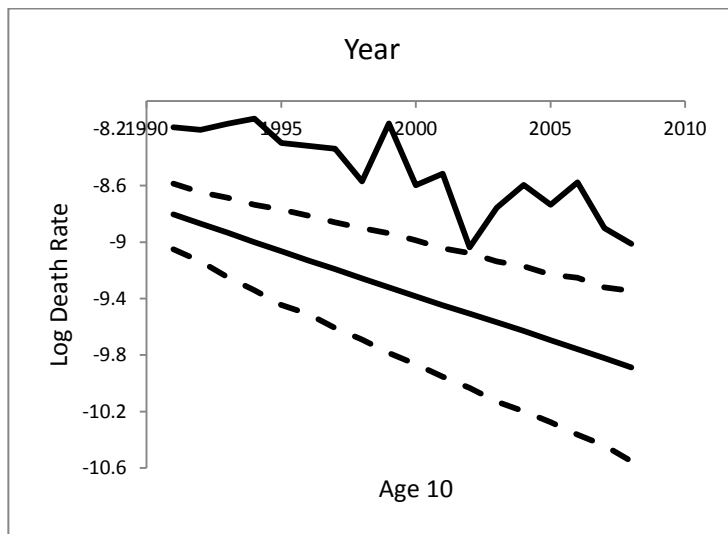


Figure 5: (Continued)

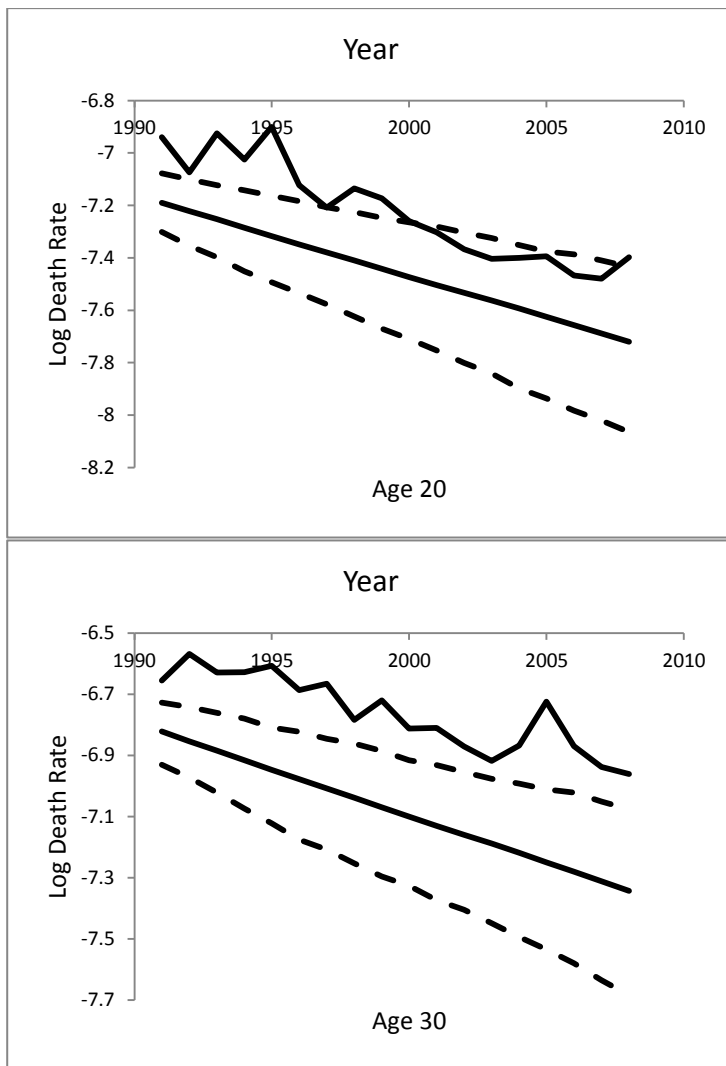


Figure 5: (Continued)

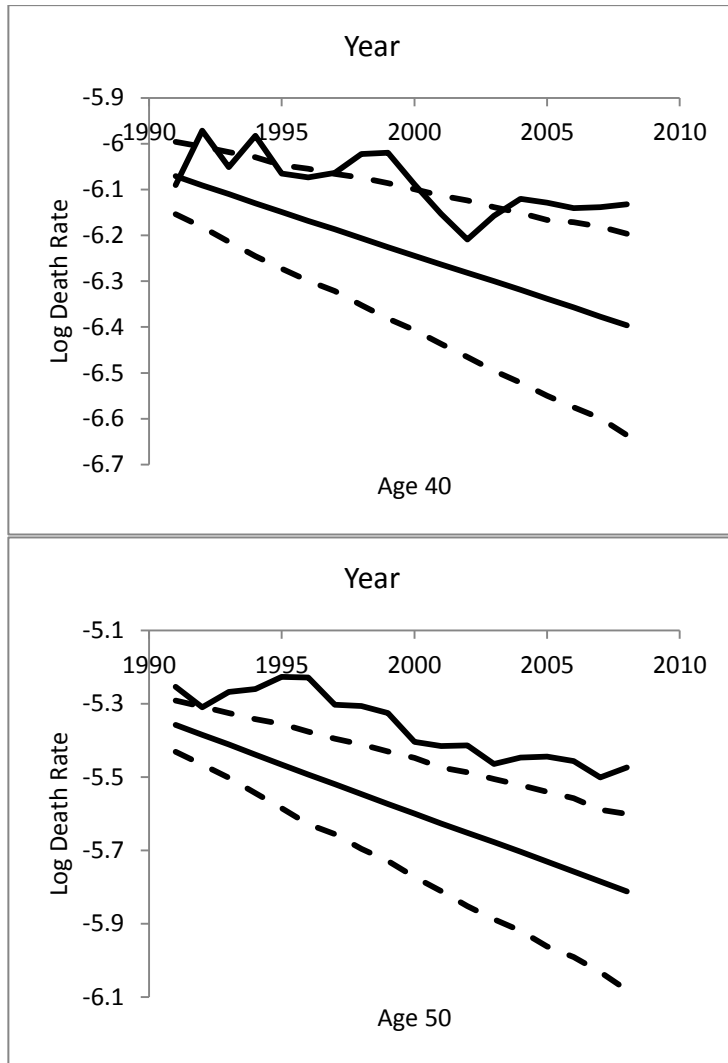


Figure 5: (Continued)

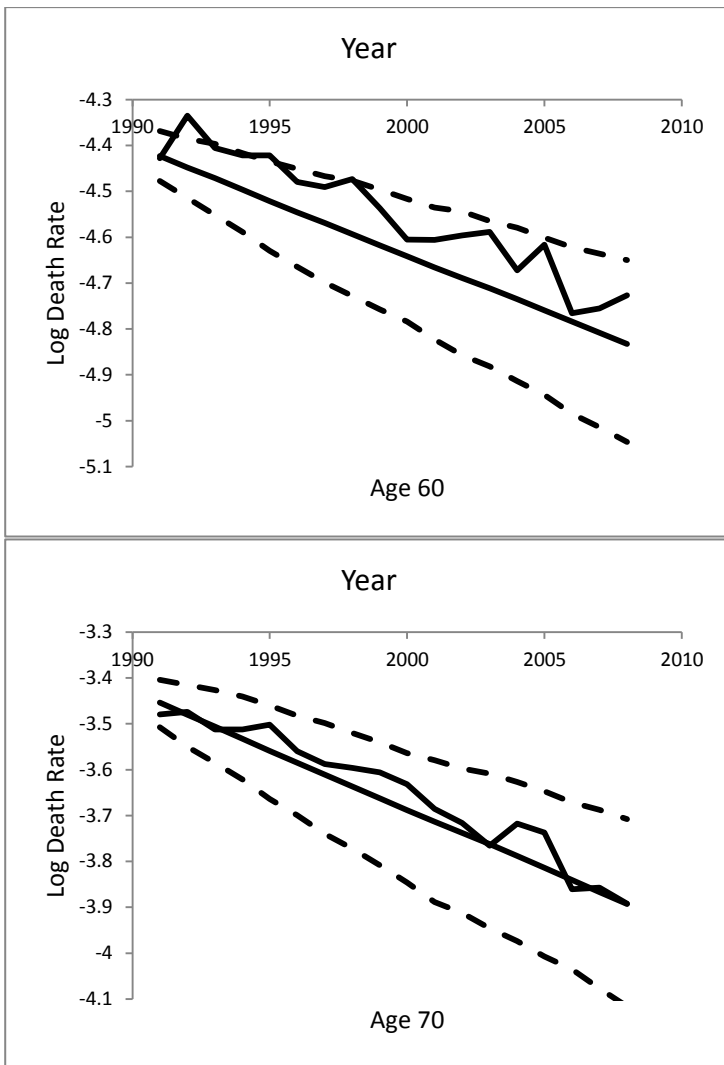
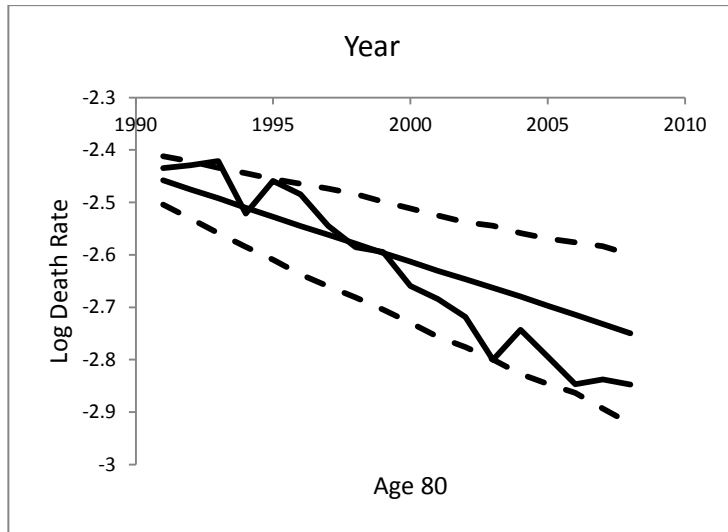


Figure 5: (Continued)

Note: The zigzag decreasing trends represent the actual values, and the solid straight lines represent the projected values. The dashed lines refer to the 95% prediction intervals.

5. Application to Chinese mortality data

The death rates of five-year age groups (by sex) of China from 1964 to 2000 are obtained from Banister and Hill (2004). The data are available for four periods: 1964-1982, 1982-1990, 1990-2000, and 1999-2000. Due to data limitations and for ease of computation, these death rates are treated here as having been collected in 1973, 1986, 1995, and 1999, respectively. The Gaussian error structure (9) is applied to these four years of data, since the numbers of deaths and exposures are not provided along with the death rates. The future death rates and life expectancy at birth are then projected for 31 years from 2000 to 2030. We also exploit the previous Taiwanese and Japanese data to provide additional information for the underlying variability of the Chinese data in a manner similar to the approach used in the last section.

The log death rates and the survival curve are plotted separately for females and males in Figure 6. As we can see, mortality has decreased substantially for both sexes during the period, and a rectangularization of the survival curve has taken place. While these four years of data determine the mean forecasts for 2000 to 2030, other experience can be used to help assess whether the underlying variability is sufficiently

allowed for. We compare the Chinese data with the whole set of Taiwanese data and find that their values of $|\sigma_k/\mu|$ are indeed very close; the latter is only around 1.1 times larger. This observation suggests that an adjustment of σ_k^2 does not seem necessary here. Moreover, we realize that the differences in the annual rate of the decline of death rates between the two populations vary for different ages, with the rate being about 0.7% higher for China at ages zero to 59, but 0.4% higher for Taiwan at ages 60 and above. If the comparison is made with the Japanese data instead, the two figures become 0.4% and 1%, respectively. As was discussed earlier, the contribution of the older ages to rising longevity has become increasingly important over time. We therefore adjust the projection formula again as in (21) for ages 60 and above only, using the 1% difference based on the Japanese data.

Figure 7 displays the parameter estimates (markers) and observed life expectancy (the zigzag increasing trend), with their projections (solid lines) and 95% prediction intervals (dashed lines before adjustment; dotted lines after adjustment). For comparison, the projected values of life expectancy from UNdata (circles) are also shown. We can see that for both females and males, the UNdata figures are lower than our projections, and move closer to the lower bounds over time. In particular, our projected values at 2030 are 81 for females and 77 for males, compared to the UNdata figures of 79 and 75, respectively. For both sexes, the width of the bounds at 2030 is four years before adjustment and is around six years after adjustment. As a further comparison, Taiwanese life expectancy figures (crosses) are also calculated and included in Figure 7. It is interesting to note that the projected values of life expectancy of China and Taiwan move almost in parallel over the period, with the latter being six years higher than the former for females, and more than three years higher for males.

Figure 6: Log death rates and survival curve – China, females (top) and males (bottom)

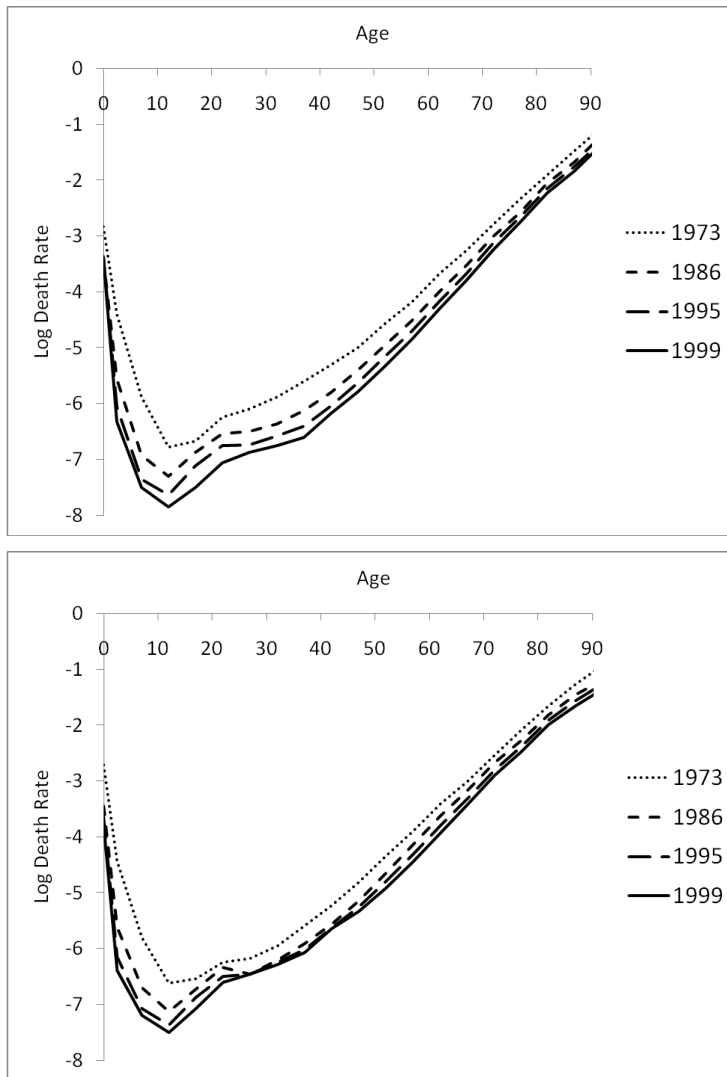


Figure 6: (Continued)

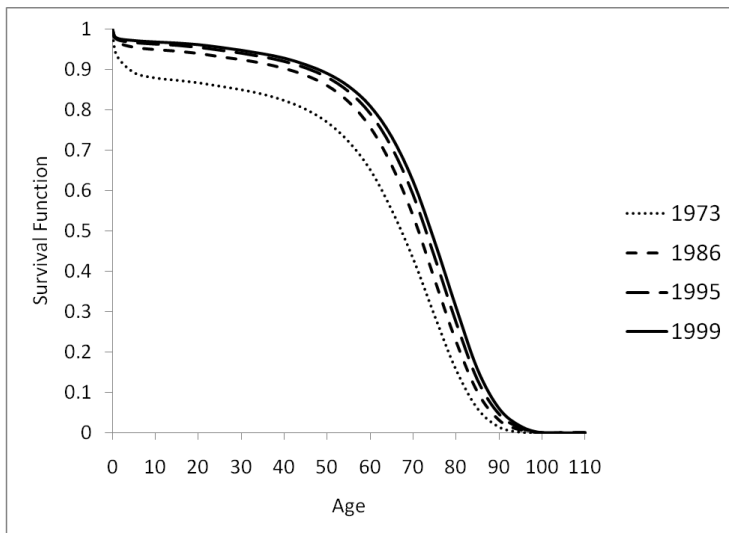
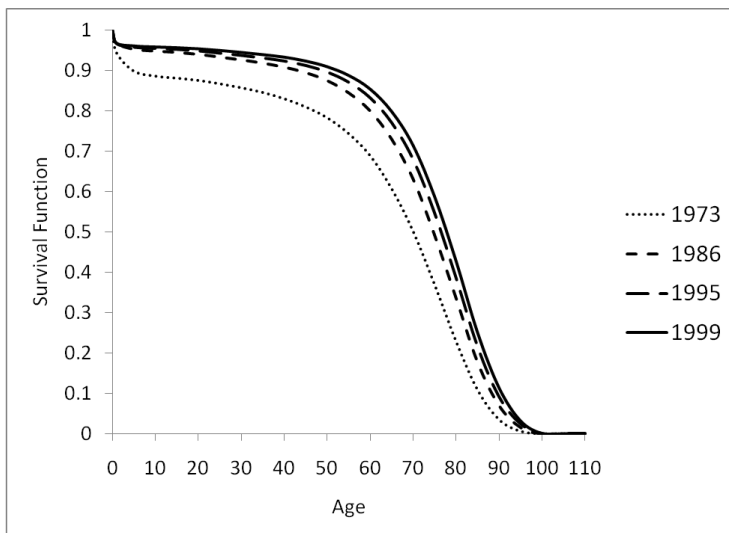


Figure 7: Parameter estimates of a_x , b_x , and k_{t_h} and life expectancy at birth – China, females (top) and males (bottom)

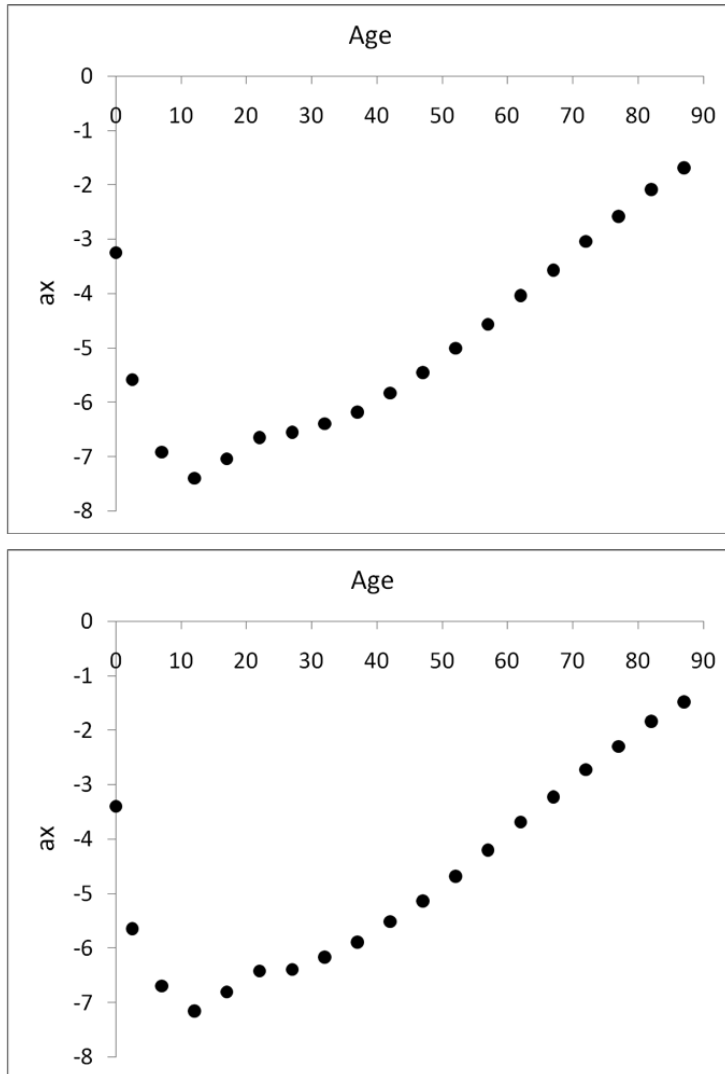


Figure 7: (Continued)

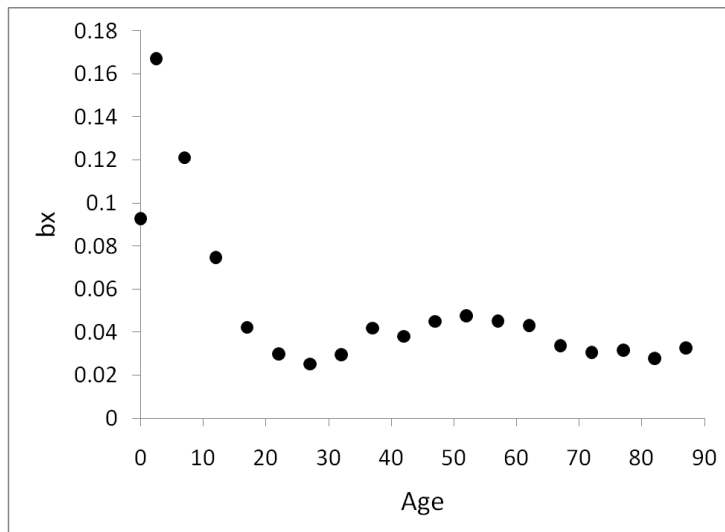
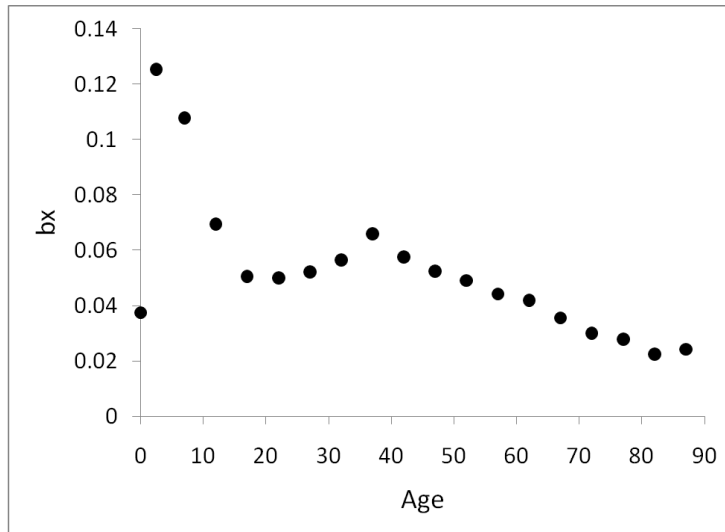


Figure 7: (Continued)

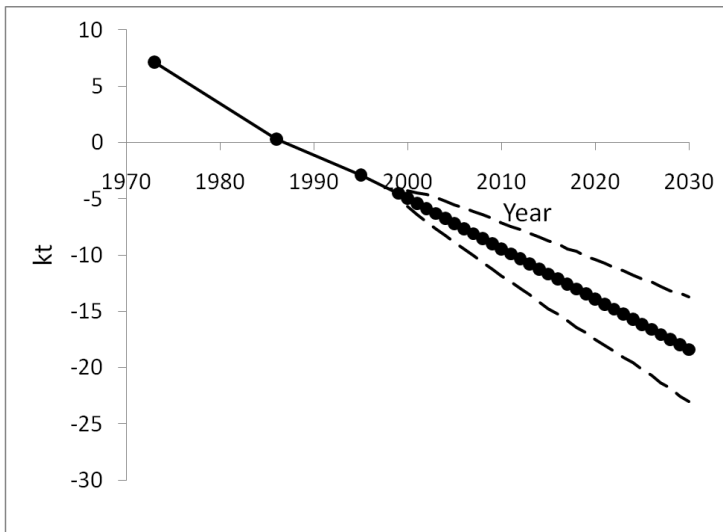
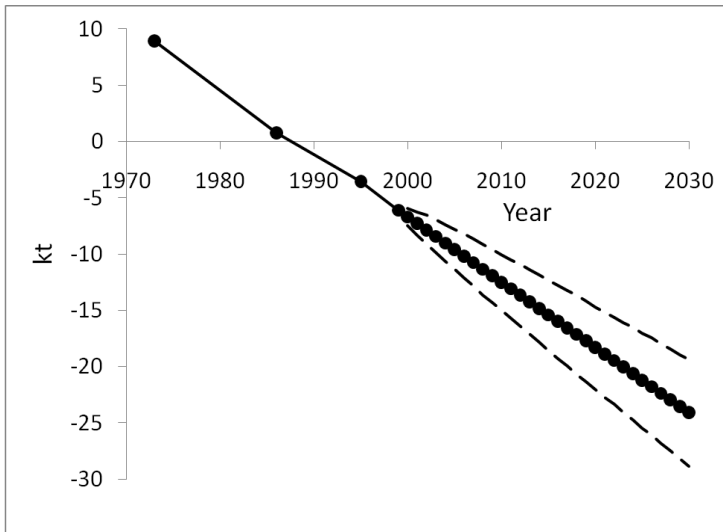
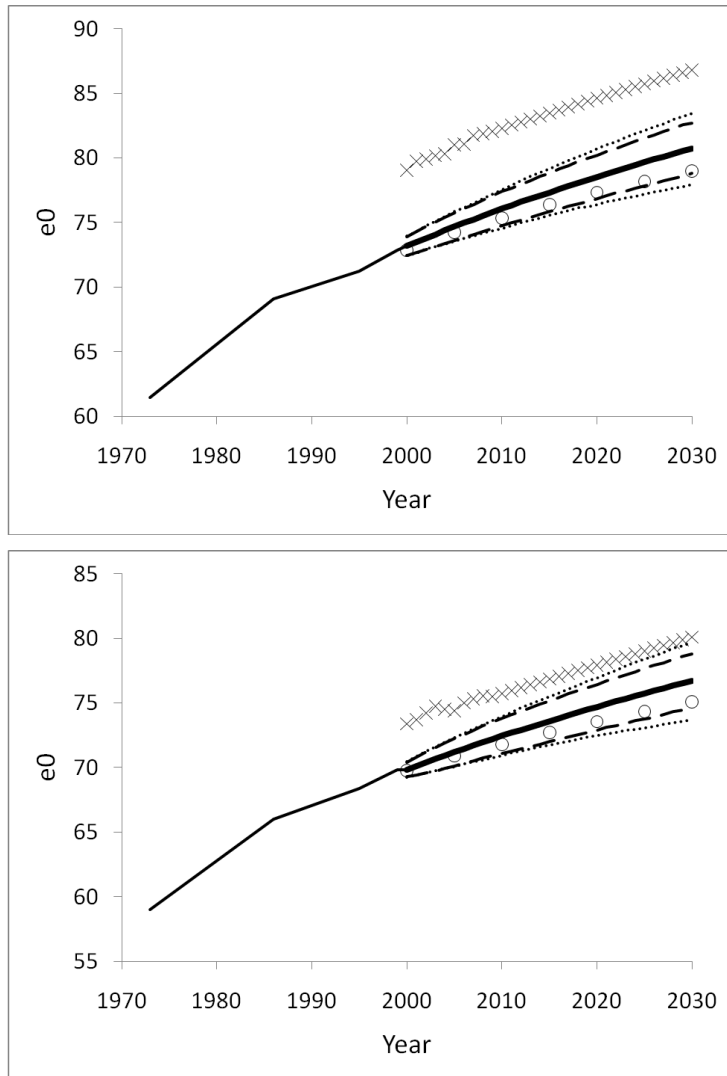


Figure 7: (Continued)



Note: For the projections on pages 37 and 38, the solid straight lines represent the sample means / projected values. The dashed (dotted) lines refer to the 95% prediction intervals before (after) adjustment. The UNdata Chinese figures and the calculated Taiwanese figures are shown as circles and crosses respectively (page 38).

6. Concluding remarks

In this paper, we examine a Bayesian approach to adapting some extensions of the Lee-Carter method in order to deal with situations in which mortality data are available for a few years only. Under the Bayesian framework, we employ MCMC simulation to project future mortality and life expectancy and construct probability intervals for the projected values. The specialized software WinBUGS used in this work provides an efficient way to program Bayesian models and perform MCMC simulation. Specifically, we demonstrate how other mortality experience can be utilized to help determine the underlying variability of the population under study, through revising the parameters of the prior distributions, modeling two populations simultaneously, and adding a “shock” component to the drift term in the projection formula. We also experiment with the use of a common mortality index for co-modeling two datasets. The simulated results of the mean values and prediction intervals based on some Taiwanese and Chinese data look reasonable in general. We realize that even with only a few years of data, the parameter estimates display proper patterns, and sensible mean values can still be projected assuming a linear mortality index. Our approach here would be useful for certain developing nations, helping them to better understand the potential risk of underestimating longevity and allowing them to undertake more informed planning of public policies and social benefits. It also provides a flexible way for insurance companies that have operations in new markets to properly assess the longevity risk embedded in their products.

Future research will include co-modeling two or more populations using different model structures within the Bayesian paradigm. To cite a recent example, Cairns et al. (2011) incorporated the age-period-cohort model into a Bayesian framework designed for modeling a large population jointly with a small sub-population. One major challenge for academics and practitioners is how to strike a balance between the sophistication and the practicality of using Bayesian models. On one hand, while it is desirable for experts to develop their own computation algorithms and to build more flexible models for fully Bayesian modeling, the process is time-consuming and technically demanding. On the other hand, WinBUGS offers a user-friendly platform for formulating Bayesian models, but this software does not support certain overly complicated model specifications. Further research is required to design a form of Bayesian analysis which is both flexible and ready for use in practice.

7. Acknowledgments

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 for their generous support and guidance. This research work was conducted as part of the author's thesis under Macquarie University, Australia.

References

- Banister, J. and Hill, K. (2004). Mortality in China 1964-2000. *Population Studies* 58(1): 55-75. doi:10.1080/0032472032000183753.
- Bijak, J. (2011). *Forecasting international migration in Europe: A Bayesian view*. Springer. doi:10.1007/978-90-481-8897-0.
- 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(3): 373-393. doi:10.1016/S0167-6687(02)00185-3.
- Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., and Khalaf-Allah, M. (2011). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin* 41(1): 29-59.
- Carter, L.R. and Lee, R.D. (1992). Modeling and forecasting US sex differentials in mortality. *International Journal of Forecasting* 8(3): 393-411. doi:10.1016/0169-2070(92)90055-E.
- Coale, A. and Guo, G. (1989). Revised regional model life tables at very low levels of mortality. *Population Index* 55(4): 613-643. doi:10.2307/3644567.
- Czado, C., Delwarde, A., and Denuit, M. (2005). Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics* 36(3): 260-284. doi:10.1016/j.insmatheco.2005.01.001.
- Giroi, F. and King, G. (2008). *Demographic forecasting*. Princeton University Press.
- Human Mortality Database (HMD) (2010). University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). <http://www.mortality.org>.
- Kogure, A., Kitsukawa, K., and 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). doi:10.2202/2153-3792.1036.
- Lee, R. (2000). The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal* 4(1): 80-91. doi:10.1080/10920277.2000.10595882.
- Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association* 87(419): 659-671.

- Lee, R. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography* 38(4): 537-549. doi:10.1353/dem.2001.0036.
- Li, N., Lee, R., and Tuljapurkar, S. (2004). Using the Lee-Carter method to forecast mortality for populations with limited data. *International Statistical Review* 72(1): 19-36. doi:10.1111/j.1751-5823.2004.tb00221.x.
- Pedroza, C. (2006). A Bayesian forecasting model: Predicting U.S. male mortality. *Biostatistics* 7(4): 530-550. doi:10.1093/biostatistics/kxj024.
- Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). *WinBUGS User Manual*. <http://www.mrc-bsu.cam.ac.uk/bugs>.
- Tuljapurkar, S., Li, N., and Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. *Nature* 405: 789-792. doi:10.1038/35015561.
- UNdata (2011). <http://data.un.org>.

Appendix

In this appendix, we present the WinBUGS codes used for the Taiwanese example above, and provide some technical details of the MCMC simulation process. The codes are listed as follows (whole left column and then right column):

```

model;
mu~dnorm(mu0,invsigma2.mu)
invsigma2~dgamma(alpha,beta)

{
  for(j in 1:18){
    for(i in 1:90){
      for(j in 1:3){
        d[i,j]~dpois(lambda[i,j])
        lambda[i,j]<-e[i,j]*exp(a[i]+b[i]*k[j])
      }
      for(i in 2:90){
        a[i]~dnorm(0,invsigma2.a)
        b[i]~dnorm(0.01111,invsigma2.b)
      }
      a[1]~dnorm(0,invsigma2.a)
      b[1]<-1-sum(b[2:90])

      k[2]<-10*mu+k[1]+w[1]
      r1<-invsigma2/10
      w[1]~dnorm(0,r1)

      k[3]<-10*mu+k[2]+w[2]
      r2<-invsigma2/10
      w[2]~dnorm(0,r2)

      k[1]<-(0-30*mu-w[1]-sum(w[1:2]))/3
    }
  }
  for(j in 1:18){
    k[j+3]<-mu+k[j+2]+w[j+2]
    for(i in 1:90){
      m[i,j]<-exp(a[i]+b[i]*k[j+3])
    }
    for(j in 3:20){
      w[j]~dnorm(0,invsigma2)
    }
    for(j in 1:18){
      x[j]<-log(m[90,j])-log(m[89,j])
      y[j]<-(log(m[90,j]+0.55)-log(m[89,j])-21*x[j])/(-210)
      for(i in 91:110){m[i,j]<-m[90,j]*exp(x[j]*(i-90)-y[j]*(i-90)*(i-89)/2)}
      lx[1,j]<-1
      for(i in 2:111){lx[i,j]<-lx[i-1,j]*exp(-m[i-1,j])}
      for(i in 1:110){ae0[i,j]<-lx[i,j]*(1-0.5*(1-exp(-m[i,j])))}
      ae0[111,j]<-lx[111,j]*(1-0.5)
      e0[j]<-sum(ae0[1:111,j])
    }
  }
}

```

The notation used in the codes can readily be matched against (10) to (11), (13) to (16), and (18). More specifically, we denote the following (i and j referring to x and t_h respectively; $t_0 = 1970$, $t_1 = 1980$, $t_2 = 1990$):

$$d[i,j] = D_{x,t_h}, \quad e[i,j] = E_{x,t_h}, \quad (i = 1-90 \text{ refer to } x = 0-89; j = 1-3 \text{ refer to } t_h = 1970, 1980, 1990)$$

$$a[i] = a_x, \text{ invsigma2.a} = \sigma_a^{-2}, \quad b[i] = b_x, \text{ invsigma2.b} = \sigma_b^{-2}, \quad (i = 1-90 \text{ refer to } x = 0-89)$$

$$k[j] = k_{t_h}, \quad (j = 1-21 \text{ refer to } t_h = 1970, 1980, 1990 \text{ and then } t = 1991-2008)$$

$$w[1] = e_{t_0+1} + e_{t_0+2} + \dots + e_{t_1} \sim N(0, (t_1 - t_0)\sigma_k^2), \quad r[1] = (t_1 - t_0)^{-1}\sigma_k^{-2},$$

$$w[2] = e_{t_1+1} + e_{t_1+2} + \dots + e_{t_2} \sim N(0, (t_2 - t_1)\sigma_k^2), \quad r[2] = (t_2 - t_1)^{-1}\sigma_k^{-2},$$

$$w[3] = e_{t_2+1}, \quad w[4] = e_{t_2+2}, \dots, \quad w[20] = e_{t_2+18}, \quad \text{invsigma2} = \sigma_k^{-2},$$

$$\mu = \mu, \quad \mu_0 = \mu_0, \quad \text{invsigma2.mu} = \sigma_\mu^{-2}, \quad \alpha = \alpha_k, \quad \beta = \beta_k,$$

$$m[i,j] = m_{x,t}, \quad (i = 1-110 \text{ refer to } x = 0-109; j = 1-18 \text{ refer to } t = 1991-2008)$$

in which the two constraints $\sum_x b_x = 1$ and $\sum_{h=0}^2 k_{t_h} = 0$ are imposed by transforming them into:

$$b_0 = 1 - \sum_{x \neq 0} b_x;$$

$$k_{t_0} = -((t_1 + t_2 - 2t_0)\mu + 2(e_{t_0+1} + e_{t_0+2} + \dots + e_{t_1}) + (e_{t_1+1} + e_{t_1+2} + \dots + e_{t_2}))/3.$$

Note that as the death rates are more volatile from age 90 onwards, we only include the data of ages zero to 89 in the modeling process. Accordingly, we use an approach similar to that of Coale and Guo (1989) to extend the death rates to age 109 for each year t (≥ 1991), assuming $m_{109,t} = m_{89,t} + 0.55$ based on our observations of other populations. The method is as below:

$$\ln m_{89,t} - \ln m_{88,t} = x_t, \quad \ln m_{90,t} - \ln m_{89,t} = x_t - y_t, \quad \ln m_{91,t} - \ln m_{90,t} = x_t - 2y_t, \dots,$$

$$\ln m_{109,t} - \ln m_{108,t} = x_t - 20y_t, \quad (\text{note: } x_t \text{ here is a rate of change, but does not mean age } x)$$

which imply that

$$y_t = -(\ln m_{109,t} - \ln m_{88,t} - 21x_t)/210, \quad \ln m_{90,t} = \ln m_{89,t} + x_t - y_t,$$

$$\ln m_{91,t} = \ln m_{89,t} + 2x_t - 3y_t, \dots,$$

$$\ln m_{109,t} = \ln m_{89,t} + 20x_t - 210y_t,$$

in which x_t and y_t are denoted as $x[j]$ and $y[j]$ in the codes ($j = 1-18$ referring to $t = 1991-2008$). Then for a particular year t , we estimate life expectancy at birth as:

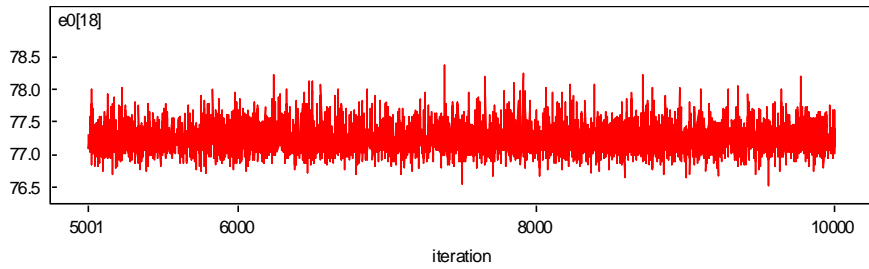
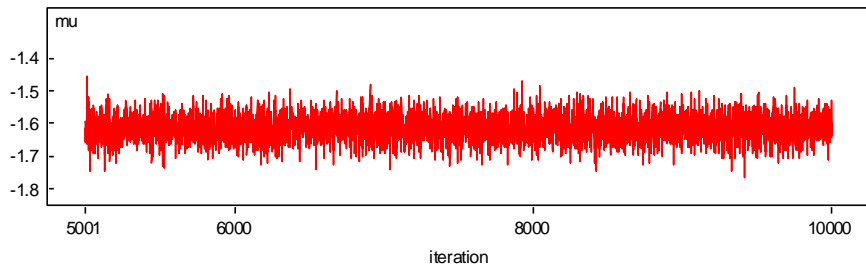
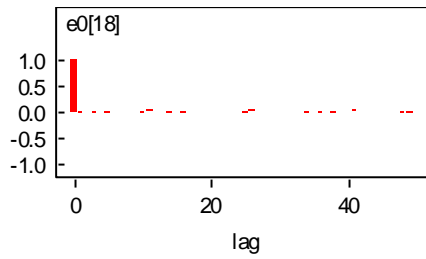
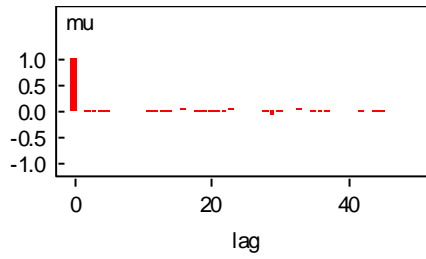
$$e_0 = \int_0^{\infty} {}_s p_0 ds \approx {}_{0.5} p_0 + {}_{1.5} p_0 + {}_{2.5} p_0 + \dots + {}_{110.5} p_0 = {}_{0.5} p_0 + {}_{0.5} p_1 \times_1 p_0 + {}_{0.5} p_2 \times_2 p_0 + \dots + {}_{0.5} p_{110} \times_{110} p_0,$$

where ${}_{s+1} p_0 \approx {}_s p_0 \times \exp(-m_{s,t})$ (integer s) and ${}_{0.5} p_x \approx 1 - 0.5(1 - \exp(-m_{x,t}))$. The

term ${}_s p_0$ is denoted as $lx[i,j]$ and e_0 as $e0[j]$ in the codes ($i = 1-111$ referring to $s = 0-110$; $j = 1-18$ referring to $t = 1991-2008$).

In the MCMC simulation process, we discard the first 5,000 iterations, and store every 100th iteration afterwards to contribute to the required statistics. In effect, 5,000 samples are collected for computing the parameter estimates and constructing probability intervals. In this way, we find that the autocorrelations between the simulated samples are minimal and that the level of convergence is satisfactory. For example, the autocorrelation plots (top) and the history plots (bottom) of the drift term and life expectancy at 2008 are shown below. Convergence is further assured by the fact that all of the Monte Carlo errors are less than 2% of the sample standard deviations. In an earlier work, Cairns et al. (2011) also take every 50th iteration in their MCMC simulation in order to reduce the degree of autocorrelation between successive samples. Empirically, we also realize that when the prior distributions of μ and σ_k^2 are adjusted to allow for higher variability, the thinning of 100 iterations may need to be increased to minimize the autocorrelations (particularly between the simulated samples of the drift term μ).

To check the reasonableness of the simulation results, the MCMC parameter estimates are compared with the ML (maximum likelihood; see Brouhns, Denuit, and Vermunt 2002) estimates (using the same set of data), and they are found to be in agreement with each other. More precise prior choices are also tested, such as using the ML estimates of σ_a^2 (or σ_b^2) and setting $\alpha_k = 2.1$. The sample means of the results are basically not affected, and the resulting changes in the sample variances are immaterial (unless the prior distributions of μ and σ_k^2 are adjusted). Moreover, we have tested both the ML estimates and the different values generated by the WinBUGS built-in function as the initial values. The corresponding results are very similar, and convergence is readily achieved.



In the codes, instead of simulating from a Poisson distribution, the future death rates are generated from using $m_{x,t} = \exp(a_x + b_x k_t)$. This procedure serves as a practical alternative since the “future” exposures in the projection period (1991-2008) are supposed to be “unknown” at the end of the fitting period (1970-1990). To examine the significance of this matter, we treat the actual exposures of the projection period as “estimated” figures and take them as “given.” We then model the future number of deaths with a Poisson distribution and carry out the simulation again. The resulting prediction intervals for the log death rates are only marginally wider than previously, and there is effectively no change in the prediction intervals for life expectancy. These results are in line with the findings in Lee and Carter (1992), which suggest that for life expectancy forecasts, it is reasonable to focus on the errors of the mortality index but not those of the age-specific death rates, because of substantial cancellation across different ages. (We also test the Gaussian error structure (9) on the Chinese data with and without assuming an error term for the future death rates, and observe that the differences in the prediction intervals for the death rates and life expectancy are trivial.) It appears that if the purpose is to forecast life expectancy, this is a practical and convenient way to simply generate the future death rates with the formula above. On the other hand, if more accurate prediction bounds are required for the death rates in the short run, the future exposures may first be estimated in some way and then taken as given. The future number of deaths in the projection period can then be modeled with a Poisson distribution, consistent with the fitting period.

There is a final note regarding the revised projection formula $k_t^{(x)} = \mu + k_{t-1}^{(x)} + e_t + \Delta_x$. As was discussed, the term Δ_x represents a “shock” to the drift term μ and allows for the possibility of a major shift in mortality decline. For convenience, the Δ_x ’s of different ages are assumed to be perfectly correlated in the modeling process; i.e., a major shift, if it happens, affects all ages in conjunction (although to different extents).

