Network for Studies on Pensions, Aging and Retirement

Netspar DISCUSSION PAPERS

Hong Li, Anja De Waegenaere and Bertrand Melenberg

The Choice of Sample Size for Mortality Forecasting

A Bayesian Learning Approach

DP 11/2014-072

The Choice of Sample Size for Mortality Forecasting: A Bayesian Learning Approach

Hong Li[a,1], Anja De Waegenaere[a,b], Bertrand Melenberg[a,b,2]

[a]Department of Econometrics and Operations Research, Tilburg University
[b]Netspar

November 27, 2014

Abstract

Forecasted mortality rates using mortality models proposed in the recent literature are sensitive to the sample size. In this paper we propose a method based on Bayesian learning to determine model-specific posterior distributions of the sample sizes. In particular, the sample size is included as an extra parameter in the parameter space of the mortality model, and its posterior distribution is obtained based on historical performance for different forecast horizons up to 20 years. Age- and gender-specific posterior distributions of sample sizes are computed. Our method is applicable to a large class of linear mortality models. As illustration, we focus on the first generation of the Lee-Carter model and the Cairns-Blake-Dowd model. Our method is applied to U.S. and Dutch data. For both countries we find highly concentrated posterior distributions of the sample size that are gender- and age-specific. In the out-of-sample forecast analysis, the Bayesian model outperforms the original mortality models with fixed sample sizes in the majority of cases.

Keywords: Lee-Carter model, Cairns-Blake-Dowd model, Gibbs sampling

---

[1]Corresponding author. Email: h.li@tilburguniversity.edu.

[2]The authors are grateful to David Blake, Mitja Stadje, Patrick Tuijp, Tim Boonen, the seminar participants at Longevity 9 and Netspar Pension Day, and an anonymous referee for their helpful comments. All errors are our responsibility. This work is part of the research programme Managing Longevity Risk, which is financed by the Netherlands Organisation for Scientific Research (NWO).

1 Introduction

Compared with the extensive literature on mortality forecasting, relatively little is known about the choice of sample size for the existing mortality models. Most mortality models proposed in the past twenty years are based on a "linear extrapolation" approach, meaning that the evolution of the modeled quantities (such as the log of central death rates or the logit[3] of death probabilities) is assumed to follow a linear trend. Two most widely used models in this class are the ones proposed by Lee and Carter (1992) (the Lee-Carter model) and Cairns, Blake, and Dowd (2006) (the CBD model).[4] Various extensions to these two models have been proposed, including Brouhns et al. (2002), Li and Lee (2005), Renshaw and Haberman (2006), to name a few. See Cairns et al. (2009), who summarize several most popular models in the linear extrapolation class.

An immediate result of using linear extrapolation models is that the forecasted mortality rates are highly sensitive to the choice of sample size. This phenomenon has been documented by many authors. In their original paper, Lee and Carter (1992) estimate the model based on U.S log mortality data from 1900 to 1989. They argue that the length of the sample size is not critical as long as it is longer than 20 years. However, Lee and Miller (2001) restrict the sample size to start from 1950, and a better fit is obtained. Baran et al. (2007) forecast the mortality rate of the Hungarian population from 2004 to 2040 using the Lee-Carter model based on mortality data from 1949 to 2003. As a result, the forecasted mortality rates for males from age 45 to 55 are increasing, which does not seem to be reasonable. However, when they apply the Lee-Carter model with data from 1989 to 2003, both forecasted mortality for males and females are decreasing. Cairns et al. (2006) apply the CBD model to the England & Wales male data, and get significantly different forecasts when using data from 1961 to 2002 than when using data from 1981 to 2002. See Pitacco et al. (2009) and van Berkum et al. (2013) for more thorough summaries.

In this paper, we aim to answer the following question: If one wishes to fit some linear mortality model to some data set, what sample size should one choose? To answer this question, we propose a Bayesian learning approach to determine model-specific posterior distributions of the sample size for a class of linear mortality models.[5] In particular, the sample size is included as

[3] We define logit $x = \log \frac{x}{1-x}$ for $x \in (0, 1)$.

[4] Besides the linear extrapolation models, there are other classes of mortality models, see, for example, Börger (2009), Börger et al. (2013), Miltersen and Persson (2005), Bauer et al. (2010), Plat (2011), and Börger (2010). However, these models have more complex structures and are not considered in this paper.

[5] There is an extensive literature on the topic of Bayesian sample size determination, such as Adcock (1997), De Santis (2004), Wang and Gelfand (2002), and Weiss (1997), to name a few. However, the focus of these studies is different from our focus, including, for example, determining the minimal sample size such that a hypothesis test has enough

an extra parameter into the parameter space of the mortality model and its (conditional) posterior distribution is updated based on the historical performance for different forecast horizons. Briefly speaking, we first choose a set of sample sizes, estimate the mortality model based on each of the sample sizes, then perform the out-of-sample forecast for different forecast horizons using each model. A sample size receives a larger posterior probability mass if the corresponding model yields a larger likelihood in the out-of-sample forecast; this can be seen as a weighted average of the in-sample fit and the out-of-sample forecast performance.

First of all, our paper contributes to the literature on sample size determination for mortality forecasting. Despite its importance, only few studies focus on this topic. Booth et al. (2002) provide tests for the optimal sample size of the Lee-Carter model. Denuit and Goderniaux (2005) model the time effect in the Lee-Carter model as a linear function of time, and look for a sample size such that the adjustment coefficient R^2 is maximized. O'Hare and Li (2012) and van Berkum et al. (2013) study structural changes in mortality rates using several mortality models. In these studies, the authors first estimate a particular model, and then detect structural breaks in the implied latent time process. A common feature of the existing literature is that it looks for a sample size for which the implied time effect process is close enough to linear so that the underlying model specification is considered to be valid.

Our method has several advantages compared with the existing studies. First, our method allows for an age-specific posterior distribution of the sample size, while using data of all ages in the estimation. All the aforementioned studies determine the optimal sample size by focusing at the underlying latent time process. Since this time process is an aggregation of the time effects of the different ages, no age-specific information can be elicited from it. In our model, age- and gender-specific posterior distributions of the sample size can be naturally computed, while data for both genders and multiple ages are used in the estimation and learning process. Second, we look at the choice of sample size in a probabilistic way, and thus avoid sharp transitions in the use of data. In O'Hare and Li (2012) and van Berkum et al. (2013), structural breaks are allowed to exist. If such a structural break is detected, then only data after this break should be used and data before the new break will be discarded. In contrast, our method results in a smoother transition in the use of data when facing a structural break (or other nonlinearities): When using a Bayesian learning method, a break gains more and more posterior weight only when more and more subsequent observations support its existence. (We provide more concrete illustrations via simulation in Section 5.) Third, by choosing different out-of-sample forecast horizons in the updating process, we may gain

power or a desired estimation precision can be achieved.

better insights in the suitability of different mortality models. Insensitivity of the posterior distribution of the sample size to the forecast horizon can be seen as an indicator that the mortality model is suitable as a forecast model: The selected sample seems to support the linearity assumption of the model. However, sensitivity of the posterior distribution of the sample size to the forecast horizon suggests that the linearity assumptions does not fit the data well, so that the mortality model is not likely to be suitable as a forecast model.

This paper also contributes to the literature of Bayesian mortality modeling. As will be discussed later, our method consists of two parts: Estimating the underlying mortality model for a fixed sample size, and estimating the posterior distribution of the sample size. There are many existing studies focusing on the first part, i.e., the Bayesian analysis of mortality models for a fixed sample size. For example, Pedroza (2002), Pedroza (2006), Czado et al. (2005), Reichmuth and Sarferaz (2008), Kogure et al. (2009), and Kogure and Kurachi (2010) study the Bayesian modelling of the Lee-Carter model and its several extensions. Moreover, Cairns et al. (2011) propose a Bayesian estimation method for a two population mortality model. For the Bayesian mortality modelling, given a fixed sample size, our method extends the state space formulation proposed by Pedroza (2002) and Pedroza (2006) to a more general setup. Also, correlations among multiple populations are explicitly allowed for. Our formulation is applicable to various existing linear mortality models, such as most model specifications studied in Cairns et al. (2009). (More details are discussed in Section 4.)

Although our method has a wide applicability, we mainly focus on the original Lee-Carter and the CBD model with the underlying time effect process modeled by a random walk with drift. The reason is that the simplicity of these model specifications allows us to demonstrate our approach in a clearer way. We apply our method to U.S. and Dutch data, and find quite concentrated posterior distributions for most combinations of age, gender, forecast horizon, and country for both mortality models, where these posterior distributions turn out to be age- and gender-specific. For the Dutch case we also compare the out-of-sample forecast performance of the Bayesian models and the original models with the sample size selection based on the optimal Deviance Information Criterion (DIC, see (Spiegelhalter et al. 2002)). In particular, the mean squared errors of the out-of-sample forecasts from 2000 to 2009 are compared. We find that the Bayesian models outperform the original models in the majority of the cases that we consider.

The remainder of the paper is organized as follows. In the next section a brief introduction to the Lee-Carter model and the CBD model is provided. In Section 3 we discuss the importance of sample size determination in more details. In Section 4 a formal description of the Bayesian model is given. In Section 5 we evaluate the performance of our method with simulated data. Section 6 presents the empirical results. Key results of the out-of-sample forecast is reported in Section 7. Section 8 presents the conclusion. The Appendix contains some technical details.

2 Mortality models

In this section we introduce the notations used in this paper and give a brief description of the Lee-Carter (Lee and Carter 1992) and the CBD model (Cairns et al. 2006).

2.1 Notation

We construct an index, i, belonging to an index set I, with $i = (x, t, g)$, where x, t, and $g \in \{m, f\}$ denote age, time, and gender, respectively. In other words, the index $i = (x, t, g)$ represents the group of individuals aged x at time t with gender g. For any $i = (x, t, g)$:[6]

- m_i is the crude central death rate. Specifically, $m_i = \frac{D_i}{E_i}$, where D_i is the observed number of deaths for i, and E_i the corresponding exposure.

- q_i is the one year death probability, which can be approximated by the crude death rates via (see, for example, Pitacco et al. 2009)

$$q_i \approx 1 - \exp(-m_i). \tag{2.1}$$

- $\mathcal{M}$ is the set of the stochastic mortality models considered.

- $\mathcal{J} \subset \mathbb{N}$ is the set of sample sizes. For any $J \in \mathcal{J}$, J is the length of the sample period upon which we estimate a model. For example, if the end year is 1999, then $J = 30$ means mortality data from year 1970 to 1999, and $J = 20$ means mortality data from year 1980 to 1999. We denote $J_{min} = \min\{J \in \mathcal{J}\}$ and $J_{max} = \max\{J \in \mathcal{J}\}$.

- Θ_M is the set of parameters for the stochastic mortality model $M \in \mathcal{M}$. Moreover, we define $\Theta_M = \tilde{\Theta}_M \times \mathcal{J}$, where $\tilde{\Theta}_M$ is the space of parameters except for the sample size. The structure of Θ_M is model-specific, i.e., its composition varies with the stochastic mortality model. In the Appendix, we specify the parameter sets for the Lee-Carter model and for the CBD model.

[6] We use notations such as $q_{x,t,g}$ and q_i interchangeably, according to the specific context. There is no difference in their meaning.

2.2 The stochastic mortality models

We focus on two most widely used stochastic mortality models: the Lee-Carter model (Lee and Carter 1992) and the CBD model (Cairns et al. 2006). Therefore, we have $\mathcal{M} = \{LC, CBD\}$. In particular, we restrict ourselves to the case where the underlying latent time process follows a random walk with drift.[7] For each model, we fit the model to both the male and female population, where correlation between these two populations are explicitly modeled. A brief description of the two models is provided next.

Lee-Carter model

The Lee-Carter model postulates that, for any $i = (x, t, g)$, the central death rate m_i is given by

$$\ln m_i = \alpha_{x,g} + \beta_{x,g}\kappa_{t,g} + \varepsilon_i,\ \varepsilon_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2_{x,g}), \tag{2.2}$$

where ε_i is the corresponding error term, and $\mathcal{N}(0, \sigma^2_{x,g})$ is the normal distribution with mean 0 and variance $\sigma^2_{x,g}$. We model the two time varying parameters $\boldsymbol{\kappa}_t = [\kappa_{t,m}, \kappa_{t,f}]'$ as a random walk with drift process

$$\boldsymbol{\kappa}_t = \boldsymbol{d} + \boldsymbol{\kappa}_{t-1} + \boldsymbol{\omega}_t,\ \boldsymbol{\omega}_t \overset{i.i.d}{\sim} \mathcal{N}(0, \Sigma_\omega), \tag{2.3}$$

where $\boldsymbol{d} = [d_m, d_f]'$, $\boldsymbol{\omega}_t = [\omega_{t,m}, \omega_{t,f}]'$, and Σ_ω is the covariance matrix of $\boldsymbol{\omega}_t$. In order to estimate the model, some normalization has to be imposed. In this paper, we make the same normalization as in the original paper:

$$\sum_{x \in N} \beta_{x,g} = 1,\ \text{and} \sum_{t \in T_J} \kappa_{t,g} = 0,$$

for $g \in \{m, f\}$, where N and T_J is the set of ages and years, respectively, where the subscript J indicates the dependence of the time set on the underlying sample size.

CBD model

For any $i = (x, t, g)$, the CBD model (Cairns et al. 2006) postulates that the logit of the one year death probability q_i is given by

$$\text{logit } q_i = \log(\frac{q_i}{1 - q_i}) = \kappa^{(1)}_{t,g} + \kappa^{(2)}_{t,g} x + \varepsilon_i,\ \varepsilon_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2_{x,g}), \tag{2.4}$$

where we assume that the time varying parameters $\boldsymbol{\kappa}_t = [\kappa^{(1)}_{t,m}, \kappa^{(2)}_{t,m}, \kappa^{(1)}_{t,f}, \kappa^{(2))}_{t,f}]'$ follow a random walk with drift process

$$\boldsymbol{\kappa}_t = \boldsymbol{d} + \boldsymbol{\kappa}_{t-1} + \boldsymbol{\omega}_t,\ \boldsymbol{\omega}_t \overset{i.i.d}{\sim} \mathcal{N}(0, \Sigma_\omega), \tag{2.5}$$

[7]This restriction is mainly for expositional purposes. In Section 4 we discuss the class of models to which the Bayesian sample size selection method can be applied.

where $\boldsymbol{d} = [d_{1,m}, d_{2,m}, d_{1,f}, d_{2,f}]'$, $\boldsymbol{\omega}_t = [\omega_{m,t}^{(1)}, \omega_{m,t}^{(2)}, \omega_{f,t}^{(1)}, \omega_{f,t}^{(2)}]'$, and Σ_ω is the covariance matrix of $\boldsymbol{\omega}_t$.

3 Sample size selection

If the mortality trend is linear, the choice of sample size is simple: We should use the largest sample size we can (from a statistical point of view). However, when the mortality trend is not linear, the problem becomes more complicated. For example, if the mortality data consists of two clear linear segments with different slopes, i.e., there is an obvious structural break, then we should probably only use the data after the break. However, if the mortality trend is, for example, accelerating, or there seems to be a structural break near the end of the sample, i.e., only few observations are available after the suspectable break, then it is not obvious which sample size one should use. In such cases, it is not clear, compared with the more recent data, how relevant the older data is to forecasting the future. A sample size yielding better in-sample fit does not necessarily produce better forecasts.

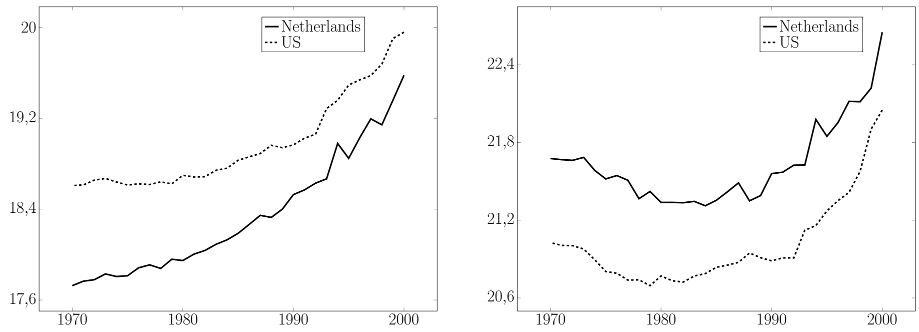
3.1 Importance of sample size

In this subsection, we illustrate the potential importance of the choice of sample size. Figure 1 displays the best estimates of the expected remaining lifetime for the U.S. and the Dutch population of age 65 in 2010 and 2009, respectively.[8] The estimates are based on the Lee-Carter model using different sample sizes. In the figure we plot the starting year of the sample on the horizontal axis. In each case, the end year of the sample is $T = 2010$ for U.S. data, and $T = 2009$ for Dutch data. Hence, the corresponding sample size J equals $J = T-$ *starting year* $+1$. We can see that the estimated expected remaining lifetimes for both genders in both countries increase significantly as the sample size decreases. In particular, as the sample size decreases from 40 to 10 years, the estimated expected remaining lifetime increases by about 7.2% and 6.7% for the U.S. and the Dutch male population, and 4.7% and 2.7% for the female population, respectively. We see that the forecasts of the Lee-Carter model are sensitive to the underlying sample size. Therefore, the role of sample size determination is not negligible.

3.2 Existing methods to select a sample size

Booth et al. (2002) propose a method to find the optimal sample size for the Lee-Carter model, where the optimal sample size identifies the largest

[8] The end years we use are the most recent years for which data is available for these two countries. The best estimates of the expected remaining lifetime are obtained by extrapolating the $\boldsymbol{\kappa}$ process, setting the error terms equal to zero.



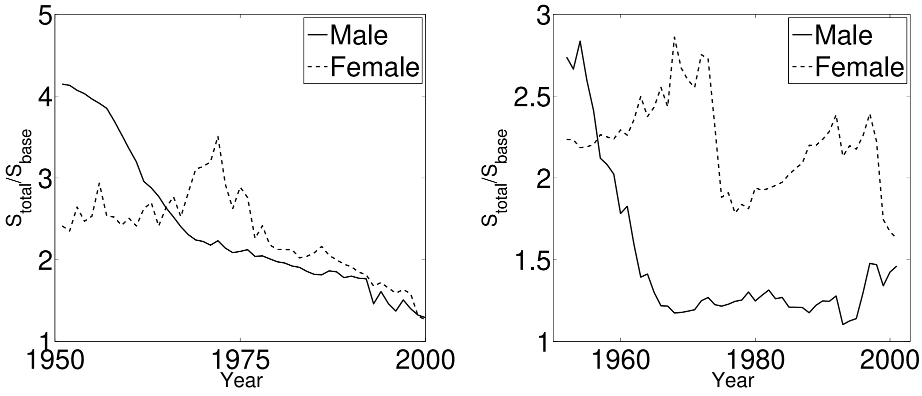
(a) Male (b) Female

Figure 1: The expected remaining lifetime for the U.S. and the Dutch population of age 65 in 2010 and 2009, respectively. The left figure reports the result for the male population and the right figure for the female population. The calculations are based on the Lee-Carter model using different sample sizes. The x-axis denotes the *starting year* of the sample. The end year for the U.S. and the Dutch population are 2010 and 2009, respectively.

sample size for which the estimated κ introduced in Section 2 is reasonably close to a linear process. In particular, the authors first compute for each sample size a measure to quantify the total lack of fit caused by treating the κ process as a fitted linear process, S_{total}, together with the base lack of fit caused by using the estimated κ-s, S_{base}. Since S_{total} includes additional lack of fit compared to S_{base}, the ratio S_{total}/S_{base} will be larger than or equal to one for all sample sizes, and the closer this ratio is to one, the closer the κ process is to linear.

Booth et al. (2002) apply their method to Australian data, and find a clear choice of starting year at around 1968. We apply their method to Dutch and U.S. data. The results are reported in Figure 2. From the figure, we see a sharp drop of the ratio at 1965 for the U.S. male. However, for U.S. females and both Dutch males and females, the choice of the sample size is not obvious. For the U.S. female, a sharp drop happens at 1974, but the ratio is not stable afterwards. For the Dutch population, the ratios are increasing with the sample size, indicating that the decrease of the mortality rate is probably accelerating for both genders. Therefore, for the U.S. female and the Dutch data, clear conclusions regarding the choice of sample size cannot be drawn from the method of Booth et al. (2002), and other methods of determining the sample size seem to be needed.

An alternative approach would be to choose the sample size that yields the minimal DIC (Deviance information criterion) ratio. DIC is a Bayesian version of the AIC (Akaike information criterion) and BIC (Bayesian information criterion), see, e.g., Spiegelhalter et al. (2002). Similar to AIC and BIC, it offers a relative estimate of the information lost when a given model



(a) Netherlands

(b) US

Figure 2: The application of the method in Booth et al. (2002) to the Dutch and the US data. Data from age 21 to 89 and year 1950 to 2009 is used for the Netherlands, and from year 1950 to 2010 is used for the U.S.. The horizontal axis denotes the *starting year* of the sample.

is used to represent the process that generates the data. As such, it provides a means for model selection. The details of this approach are explained in the Appendix. In Section 7, we will compare the out-of-sample forecast accuracy when sample size is determined based on DIC, to the out-of-sample forecast accuracy resulting from the Bayesian sample selection method that we propose in this paper.

4 The Bayesian model

In this section, we give a brief introduction to the Bayesian learning approach that we propose. The method allows to determine age- and gender-specific posterior distributions of the sample size. In the first subsection, we present the basic idea. In the second subsection, we discuss our approach in more detail.

Throughout the remainder of the paper, we will use the notation $[t_1 : t_2] = \{t_1, t_1+1, \cdots, t_2-1, t_2\}$, for $t_1 \leq t_2 \in \mathbb{N}$. Let mortality model $M \in \mathcal{M}$ be given, and suppose that we have at our disposal a set of mortality data indexed by $[T_0 : T]$, i.e., T_0 is the starting year of our sample, and T is the last year for which data is available. For each $t \in [T_0 : T]$, we denote by y_t the vector of mortality quantities at time t, with components $y_{x,t,g}$, where $x \in N$ and $g \in \{m, f\}$.[9] Moreover, we let Y_t be the observed mortality

[9]The mortality quantities will be model dependent: y_t includes the log of the central death rates for the Lee-Carter model and the logit of the one year death probabilities for the CBD model.

quantities up to time t, i.e., $Y_t \equiv \{y_{T_0}, y_{T_0+1}, ..., y_t\} = y_{[T_0:t]}$.

4.1 The basic idea

In this subsection we present the basic idea of our approach. We first choose a set of sample sizes, a prior distribution over this set of sample sizes, and we select the latest part of the in-sample data, $y_{[\widetilde{T}:T]}$ for $\widetilde{T} \leq T$, to be used for Bayesian learning. For any given model and forecast horizon μ we then update the posterior distribution over the set of sample sizes for each time $t \geq \widetilde{T}$. In this updating we make use of the predictive posterior likelihood of each of the components of y_t. This predictive posterior likelihood is estimated using the sample that is μ periods before time t, with sample size belonging to the set of sample sizes. More precisely, for any given mortality model $M \in \mathcal{M}$, the procedure is as follows:

1. Choose:

   (a) a set of sample sizes, $\mathcal{J}$,[10]

   (b) a forecast horizon $\mu \in \mathbb{N}$,

   (c) a prior distribution over the set $\mathcal{J}$ for each x, g, and μ,

   (d) a value for $\widetilde{T} \in [T_0 : T]$; $\widetilde{T}$ is the first year in which the prior distributions of the model parameters and the sample size will be updated.

   These choices should satisfy

$$J_{max} + \mu \leq \widetilde{T} - T_0 + 1. \tag{4.1}$$

   Condition (4.1) ensures that for all sample sizes in the set $\mathcal{J}$, both the data that is needed to estimate the model, and the data that is needed for the Bayesian updating of the prior distributions in year $\widetilde{T}$ using a forecast horizon of μ years, are included in the sample period $[T_0 : \widetilde{T}]$.

2. Apply the Bayesian learning approach. This learning approach is illustrated in Figure 3.

   (a) For each $J \in \mathcal{J}$, estimate (the posterior distribution of the parameters of) the model in year $\widetilde{T}$, based on the sample $y_{[\widetilde{T}-\mu-J+1:\widetilde{T}-\mu]}$. See line (a) in Figure 3.

   (b) Derive the posterior predictive distribution of each of the components $y_{x,\widetilde{T},g}$ of the vector $y_{\widetilde{T}}$, and update the posterior distribution of J, based on the corresponding likelihood with respect to $y_{x,\widetilde{T},g}$.

[10] The set of sample sizes can in principle depend on the underlying mortality model. However, for the purpose of comparison, we impose the same set of sample sizes for each mortality model and for the same data set.

(c) Forward by one period. Estimate the model based on the sample $y_{[\widetilde{T}-\mu-J+2:\widetilde{T}-\mu+1]}$ for each $J \in \mathcal{J}$, and update the posterior distribution of J in year $\widetilde{T}+1$, based on the corresponding likelihood of each of the components of $y_{\widetilde{T}+1}$. See line (b) in Figure 3.

(d) Continue the process until the final step, where the update is based on the posterior predictive likelihood of each of the components of y_T. See line (c) in Figure 3.

To illustrate the procedure, we give a concrete example. Let $[\tilde{T} : T] = [1991 : 2010]$, i.e., consider the Bayesian learning approach where a period of 20 years is used for the updating procedure. Also, let the forecast horizon be $\mu = 20$ and let the set of sample sizes be $\mathcal{J} = \{10, 11, \cdots, 39\}$. Hence, the updating process starts at $\tilde{T} = 1991$. To derive the posterior predictive distribution of the components of y_{1991}, we estimate the model based on the sample $y_{[1962:1971]}$ for $J = 10$, based on the sample $y_{[1961:1971]}$ for $J = 11$, and so on up to $J = 39$ based on the sample $y_{[1933:1971]}$. We use these estimates to derive the posterior predictive distribution of the components of $y_{1971+\mu} = y_{1991}$, and to update the posterior distribution of $J \in \mathcal{J}$. We then repeat this procedure for the cases where the end year of the sample period used for estimating the model is 1972, 1973, and so on until 1991 in the final step.

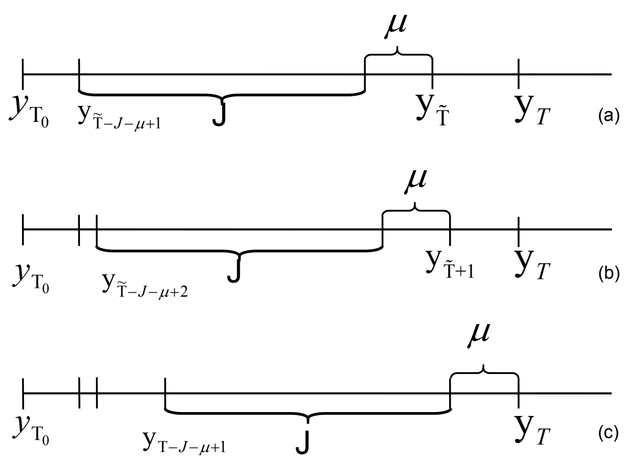


Figure 3: A brief illustration of the updating procedure of the method.

Our method can be performed using only one μ, but the use of multiple μ-s can give extra flexibility. For example, the posterior distribution for a smaller μ can be used if one is interested in short-term forecasting, and the posterior distribution for a larger μ can be used if one is interested in long-term forecasting. Alternatively, the use of different μ-s can serve as a diagnostic test. If for a model M the posterior distributions of the sample size turn out to be insensitive to different μ-s, this might be seen as evidence supporting the use of model M as a forecast model. In contrast, if the posterior distributions are sensitive to the change of μ, namely, the model seems to be too sensitive to the nonlinearities in the data, then it might not be suitable to be used as a forecast model.

4.2 The posterior sample size distribution

In this subsection, we discuss how, starting from a prior distribution over the set of sample sizes, the posterior distribution can be determined in each step of the Bayesian learning process as described in the previous subsection.

For each $M \in \mathcal{M}$, we can write the joint posterior distribution of the parameters $\theta_M = (\tilde{\theta}_M, J)$ in year $t \geq \widetilde{T}$, given any age x, gender g, and forecast horizon μ as

$$p(\tilde{\theta}_M, J|Y_t, x, g, \mu, M) = p(\tilde{\theta}_M|Y_t, J, x, g, \mu, M) \cdot p(J|Y_t, x, g, \mu, M), \quad (4.2)$$

where we shall assume $p(\tilde{\theta}_M|Y_t, J, x, g, \mu, M) = p(\tilde{\theta}_M|Y_t, J, \mu, M)$, i.e., the same for all x and g. Our primary interest is in $p(J|Y_t, x, g, \mu, M)$. This posterior density is estimated by using an iterative updating algorithm. Suppose that we are at time t and we know $p(J|Y_{t-1}, x, g, \mu, M)$. Given J, we estimate the conditional posterior distribution of $\tilde{\theta}_M$ using data $y_{[t-\mu-J+1:t-\mu]}$. This posterior distribution is denoted by $p(\tilde{\theta}_M|Y_{t-1}, J, \mu, M)$ (the same for all x and g). Then, we update $p(J|Y_{t-1}, x, g, \mu, M)$ to $p(J|Y_t, x, g, \mu, M)$ using Bayes' formula

$$p(J|Y_t, x, g, \mu, M) \propto p(y_{x,t,g}|Y_{t-1}, J, x, g, \mu, M) \cdot p(J|Y_{t-1}, x, g, \mu, M), \quad (4.3)$$

where

$$\begin{aligned} &p(y_{x,t,g}|Y_{t-1}, J, x, g, \mu, M) \\ &= \int_{\tilde{\Theta}_M} p(y_{x,t,g}|\tilde{\theta}_M, Y_{t-1}, x, g, \mu, M) \cdot p(\tilde{\theta}_M|Y_{t-1}, J, \mu, M) d\tilde{\theta}_M. \end{aligned} \quad (4.4)$$

Here, $p(y_{x,t,g}|\tilde{\theta}_M, Y_{t-1}, x, g, \mu, M)$ stands for the density function of $y_{x,t,g}$ according to model M, given the information contained in Y_{t-1} up to time $t-\mu$, and given the parameter vector $\tilde{\theta}_M$.

We continue the updating procedure until we obtain $p(J|Y_T, x, g, \mu, M)$. At every $t \geq \widetilde{T}+1$, we treat the posterior distribution over J at $t-1$ as the corresponding prior distribution over J at t for every x, g, and μ. For the first year in which the density is updated, i.e., for $t = \widetilde{T}$, $p(J|Y_{t-1}, x, g, \mu, M)$ in (4.3) is replaced by the prior distribution $p(J|x, g, \mu, M)$.

In the practical applications, we replace (4.4) by

$$\widehat{p}_L(y_{x,t,g}|Y_{t-1}, J, x, g, \mu, M) \propto \frac{1}{L}\sum_{\ell=1}^{L} p(y_{x,t,g}|\tilde{\theta}_M^{(\ell)}, Y_{t-1}, J, x, g, \mu, M), \quad (4.5)$$

where we use L random draws $\tilde{\theta}_M^{(\ell)}$, $\ell = 1, \cdots, L$ from the distribution $p(\tilde{\theta}_M|Y_{t-1}, J, \mu, M)$. Since this distribution is the posterior distribution of the rest parameters given the sample size J, standard Bayesian models and techniques can be used to to draw from this distribution. In the Appendix we present a detailed sampling procedure for the models used in this paper. In the remainder of this subsection we present a general class of state space models, incorporating most existing linear mortality models (including the ones we apply), which can be used in our approach. This is the class of so-called Time Varying-Vector Auto Regressive (TV-VAR) models, see, for example, Damien et al. (2013) for a more detailed overview.

Suppose we have at our disposal a set of mortality quantities, $y_{[T_0:T]}$, where y_t is a $2n$-dimensional column vector for each t (containing n ages for both gender), and we would like to estimate (the posterior distribution of) $\tilde{\theta}_M$ in year $\tilde{T}$. Given J and μ, we use data $y_{[\tilde{T}-\mu-J+1:\tilde{T}-\mu]}$ in the estimation. Furthermore, assume that the y_t process can be modeled as

$$y_t = B_t X_t + F_t' z_t + \varepsilon_t, \;\; \varepsilon_t \sim N(0, \Sigma_\varepsilon), \quad (4.6)$$

$$z_t = \sum_{i=1:p} G_{t,i} z_{t-i} + \omega_t, \;\; \omega_t \sim N(0, \Sigma_{\omega_t}), \quad (4.7)$$

where, moreover, ε_t and ω_t are allowed to be correlated with each other for every $t \in [\tilde{T} - \mu - J + 1 : \tilde{T} - \mu]$, but where ε_t and ω_t are assumed to be independent of ε_s and ω_s with $s \in [\tilde{T} - \mu - J + 1 : t - 1]$. Model (4.6) – (4.7) includes, for example, all model specifications studied in Cairns et al. (2011) except for M2A and M3B.[11] In model (4.6) – (4.7), the parameter vector $\tilde{\theta}_M$ is a vector consisting of the components of B_t, X_t, F_t, $[G_{t,i}]_{i=1:p}$, Σ_{ω_t}, z_t, and Σ_{ε_t}, for $t \in [\tilde{T} - \mu - J + 1 : \tilde{T} - \mu]$. Specifically, X_t is often a time-invariant matrix, such as the α_x-s in the Lee-Carter model; z_t is a vector representing the period effects and the cohort effects; B_t, F_t, and G_t are the corresponding parameters that could either be matrices or vectors, and are often time-invariant. Equation (4.7) is the vector-latent autoregressive process for the period effects and the cohort effects. The dimensions of X_t, z_t, B_t, F_t, and $\sum_{i=1:p} G_{t,i}$ depend on the underlying model specifications.

Take the Lee-Carter model for example:[12] In this case $p = 1$, $B_t = I_n$

[11] The specifications M2A and M3B in Cairns et al. (2011) include a moving averaging part in the cohort effect process, and are thus not included in the TV-VAR formulation.

[12] See the Appendix for the CBD-model.

for all t, $X_t = [a_{1,m}, a_{2,m}, ..., a_{n,m}, a_{1,f}, a_{2,f}, ..., a_{n,f}]'$ for all t,

$$F_t = \begin{bmatrix} b_{1,m} & b_{2,m} & ... & b_{n,m} & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & b_{1,f} & b_{2,f} & ... & b_{n,f} \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \end{bmatrix}$$

for all t, $z_t = [\kappa_{t,m}, \kappa_{t,f}, d_m, d_f]'$,

$$G_{t,1} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

for all t, and $\omega_t = [\omega_{t,m}, \omega_{t,f}]'$. Also, ε_t is the vector of error terms, Σ_ε is a diagonal matrix, and $\Sigma_{\omega_t} = \Sigma_\omega$ for all t.

In general, model (4.6) – (4.7) can be estimated using numerical algorithms. The estimation can be done in two steps. First, given an appropriate prior distribution, the posterior distributions of z_t at each time t can be obtained by backward filtering, see West and Harrison (1997) and Damien et al. (2013). Second, we can derive the conditional posterior distributions of the other parameters. We can then estimate these, together with the posterior distribution of the z_t-s, by the Gibbs sampler. This estimation algorithm is applicable to most existing mortality models. An example can be found in Pedroza (2006), who fits model (4.6) – (4.7) to a single gender Lee-Carter model.[13]

5 Some Illustrations

In this section, we apply our method to three specific hypothetical scenarios to see what posterior distributions of the sample size we would obtain in each of these scenarios. In the first two scenarios we impose a linear structural break in the data – a break in the middle of the sample and a break near the end of the sample, respectively. In the literature, structural breaks are detected. For example, van Berkum et al. (2013) test multiple structural breaks in the Dutch male data for the years [1950 : 2008] using the Lee-Carter model, and detect breakpoints at the year 1970 and 2002. However, such structural breaks might be hard to distinguish from an accelarating trend. Therefore, we consider an accelerating trend in mortality rates as a third scenario.

The basic idea is as follows. We simulate hypothetical true data, using a certain model specification. We then apply the Bayesian learning approach to determine the posterior sample size distribution. We allow for cases where, in the Bayesian learning approach, the user uses a model specification that is different from the model specification that was used

[13] In the Appendix we present this sampling procedure of the models used in this paper.

to simulate the data. For expositional purposes, we present results only for the case where the data is simulated based on the CBD model specification, whereas the Bayesian learning approach assumes the Lee-Carter model.[14] When simulating the hypothetical true data, we impose trend breaks and an accelerating trend via a parabola structure by adjusting the specification of the $\kappa^{(1)}$ process in the CBD model. We study what posterior distribution of sample size our method would suggest for the Lee-Carter model in each of these situations. We index the simulated data by the years $[T_0 : T] = [1 : 70]$, and use a period of 20 years to update the posterior distributions, i.e., $\left[\widetilde{T}, T\right] = [51 : 70]$. We choose $\mathcal{J} = \{10, 11, ..., 50\}$.

5.1 Early linear trend break

Figure 4a shows the simulated one year death probabilities based on the CBD model for Dutch male age classes from age 51 to age 60 with a linear structural change at year 30. The starting year of the simulated data is 2009. In particular, we impose the following structure on $\kappa^{(1)}_{t,m}$:

$$\begin{aligned} \kappa^{(1)}_{m,t+1} &= \kappa^{(1)}_{m,t} + d_1 + \omega_t, \text{ if } t \leq 30, \text{ and} \\ \kappa^{(1)}_{m,t+1} &= \kappa^{(1)}_{m,t} + 4 \times d_1 + \omega_t, \text{ if } t > 30, \end{aligned} \tag{5.1}$$

with $\omega_t \overset{i.i.d}{\sim} N(0, \sigma^2_\omega)$. The parameter values for d_1 and σ_ω are obtained by fitting the original CBD model to the Dutch male data aged 0 to 90 and years 1970 to 2009. This yields $\hat{d}_1 = -0.0204$ and $\hat{\sigma}^2_\omega = 0.00063$. We update the posterior distribution using simulated data from [51 : 70]. The results are shown in the first row in Figure 5. The horizontal axis denotes the starting year. The sample size corresponding to a certain starting year and forecast horizon μ is computed in the same way as in Figure 1: $J = 70 - \mu -$ *starting year* $+1$. For example, for $\mu = 10$, the starting year 30 is equivalent to a sample size of 31 years (30 to 60). We see that the posterior distributions become flatter and flatter, and the peak of the posterior distribution shifts from around 30 to around 40 as μ decreases from 20 to 1. The reason is that we update the posterior distribution based on the likelihood with respect to data μ years ahead. Therefore, for a smaller μ it is more difficult for the model to distinguish between random variations and the underlying trend in the data, and thus the posterior distributions are less informative. Moreover, data before year 30 receives only small probability mass for all μ-s, meaning that data before the break is considered not informative and is therefore discarded. In other words, the existence of the structural break is detected by our method.

[14] We have also considered cases where the true data is generated with Lee-Carter, and the Bayesian learning procedure assumes CBD, as well as cases in which there is no model misspecification. In each of these cases, the results are similar to the ones presented in this section.

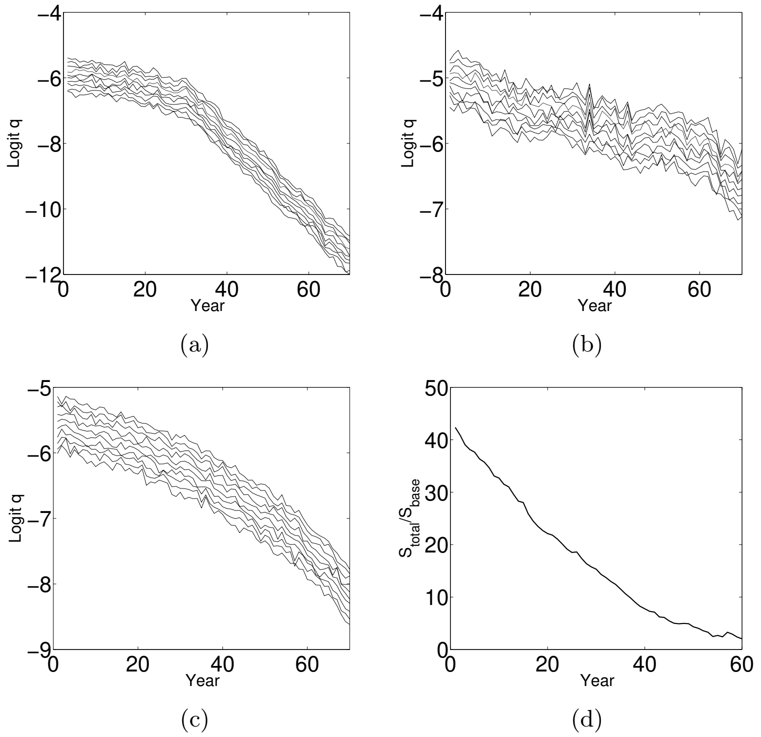


Figure 4: Panel 4a and 4b display the simulated data based on the CBD model with a linear structural break at year 30 and 60, respectively. Panel 4c display the simulated data with an accelerating trend. The parameters of the CBD model are estimated using age 0 to 90 and years 1970 to 2009. The starting values of the simulated data are Dutch males aged 51 to 60 in 2009. Panel 4d reports the results of the method from Booth et al. (2002) to the non-linear simulated data.

5.2 Late linear trend break

Figure 4b shows the simulated one year death probabilities based on the model specification as in (5.1), but now with the break date at 60. Again, we update the posterior distribution using the sample [51 : 70], which means that we have only 10 observations after the trend break.

The results are shown in the second row in Figure 5. Since we have only 10 observations after the break, the forecast horizons 15 and 20 years are not informative. The interpretation of the results is as follows. For $\mu = 10$, the end of sample in the last update is year 60. In other words, no data after the break is used. In this case, any choice of sample size yields a similar future mortality trend. However, models based on larger sample sizes yield a better in-sample fit, and thus receive a larger probability mass. For $\mu = 5$ and $\mu = 1$, the end-of-sample in the final update is after the break, namely 65 and 69, respectively. Therefore, the model based on a shorter sample size yields more precise forecasts. From the results, we see that shorter sample sizes receive more and more probability mass as we change from $\mu = 10$ to $\mu = 1$, meaning that our method detects the existence of the structural break. However, sample sizes of 40 to 50 years still receive most probability mass, because very short sample sizes may yield a poor fit.

The result in this case study indicates how our method treats a recent structural break. When a structural break is detected, our method would not immediately suggest to use only data after the break. Instead, the posterior distribution shifts gradually in favor of the break as more and more supportive data is observed. When there are enough observations after the break, as in Section 5.1, the posterior distributions become more robust to different μ-s, and suggest that only data after the break should be used.

5.3 Accelerating mortality trend

Figure 4c shows the simulated data using a CBD model with nonlinear mortality trends. In particular, we impose

$$\kappa^{(1)}_{m,t+1} = a \times \left(\kappa^{(1)}_{m,t}\right)^2 + \kappa^{(1)}_{m,t} + d_1 + \omega_t,\ \omega_t \overset{i.i.d}{\sim} N(0, \sigma^2_\omega), \tag{5.2}$$

with, similarly, $d_1 = -0.0204$ and $\sigma^2_\omega = 0.00063$. Also, we set $a = -0.0075$. The value of a is chosen so that the simulated data has a visual indication of an accelerating trend. Again, we update the posterior distribution based on the simulated data [51 : 70].

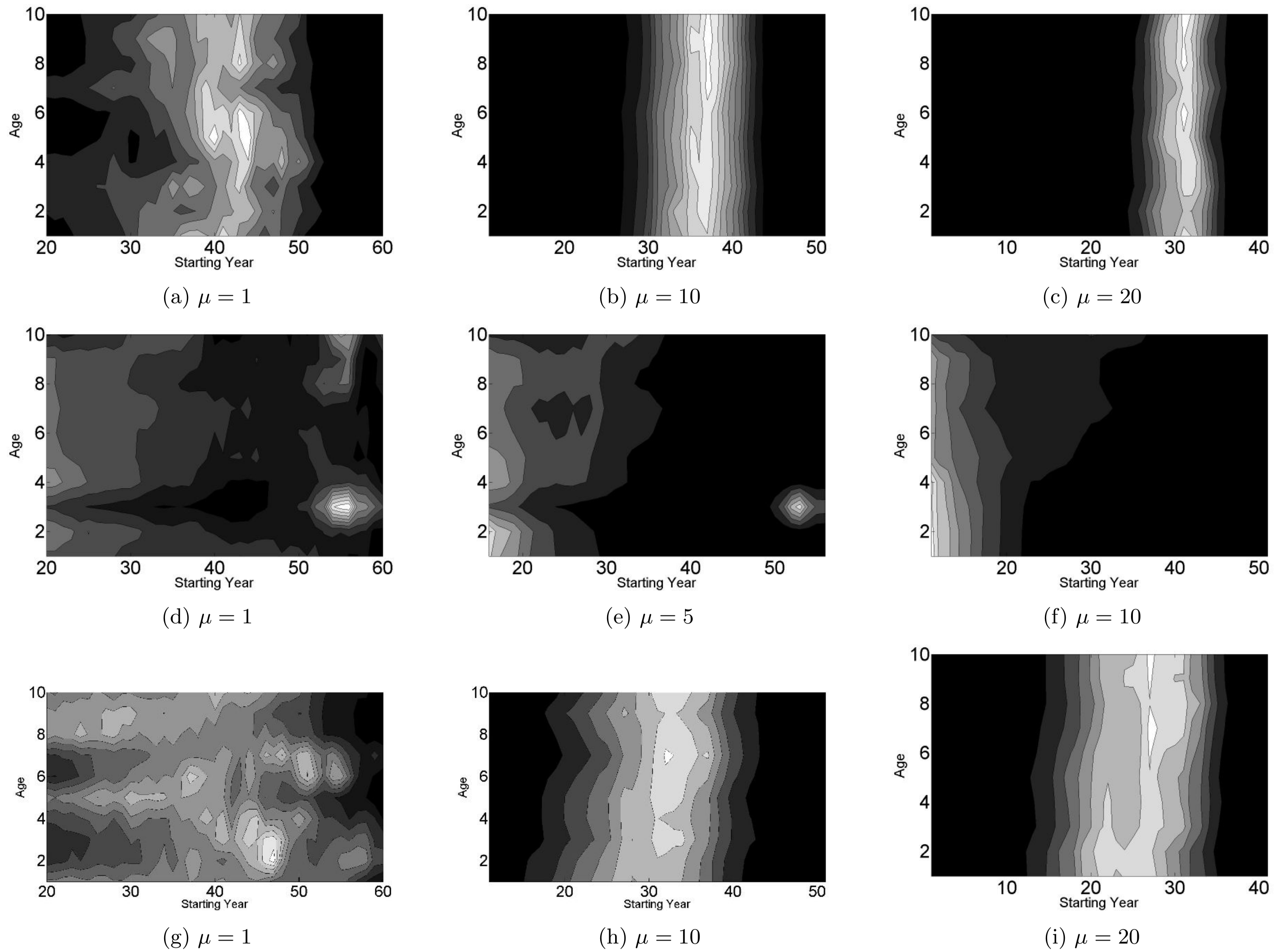


Figure 5: The heat map of the posterior distributions for the Lee-Carter model with $\mu = \{1, 10, 20\}$ for the simulated data with, from top to bottom, a linear trend break at year 30, year 60, and a non-linear trend break based on the CBD model, respectively. Black corresponds to very low probability, light grey corresponds to low probability, and dark grey corresponds to high probability.

In the case of an accelerating mortality trend, all linear models are misspecified. Figure 4d reports the application of the method from Booth et al. (2002) to the data in this case. We see that the ratio S_{total}/S_{base} (as defined in Section 3.2) increases substantially with sample size, thus the Booth's method would suggest the use of the shortest sample size possible in this case. The use of a very short sample size is not desirable, since it may produce estimations and forecasts that are very sensitive to noises in the data. To the contrary, the results in the bottom row in Figure 5 show that posterior distributions with clear and concentrated peaks are obtained for all forecast horizons except for $\mu = 1$, and the posterior means of the sample size are of reasonable lengths (around 30 years for $\mu = 10$ and 25 years for $\mu = 20$). The reason is that, in our method, models with smaller sample sizes in general yield more precise point forecasts but poorer fits (larger variances, etc), and our method is able to identify the sample sizes which yield the best balance between the in-sample fit and the out-of-sample forecast performance. However, compared with Section 5.1, we see that the posterior distribution are flatter in this situation, indicating that the accelerating trend indeed leads to more ambiguous choice of sample size.

6 Empirical results

In this section, we derive posterior distributions for the sample size for the mortality rates of the Dutch and the U.S. population. Specifically, we use the one year mortality rates from 1911 to 2009 for the Dutch population and 1933 to 2010 for the U.S. population, i.e., $T_0 = 1911$ for the Dutch case and $T_0 = 1933$ for the U.S. case. For both data sets, we use ages 21 to 89 to fit the Lee-Carter model, and age 60 to 89 to fit the CBD model. Because the Dutch mortality data is available for a longer period, we choose the updating period to be 1980 to 1999, and preserve the last 10 years for the out-of-sample forecast analysis (presented in the next section). However, the U.S. mortality data is not long enough, so we use 1991 to 2010 for the updating, and no out-of-sample forecast is performed. Thus, $[\widetilde{T} : T]$ is $[1980 : 1999]$ for Dutch data and is $[1991 : 2010]$ for U.S data. Moreover, we set $\mathcal{J}_{Dutch} = \{10, 11, ..., 50\}$, $\mathcal{J}_{US} = \{10, 11, ..., 39\}$. For both countries, we determine posterior distributions for the sample size based on forecast horizons $\mu \in \{1, 5, 10, 15, 20\}$. Due to the limitation of space, we report only the results for male data.[15]

6.1 Uniform prior

Figure 6 reports the conditional posterior distribution of the sample size for Dutch male data. The figure shows that the posterior distributions of

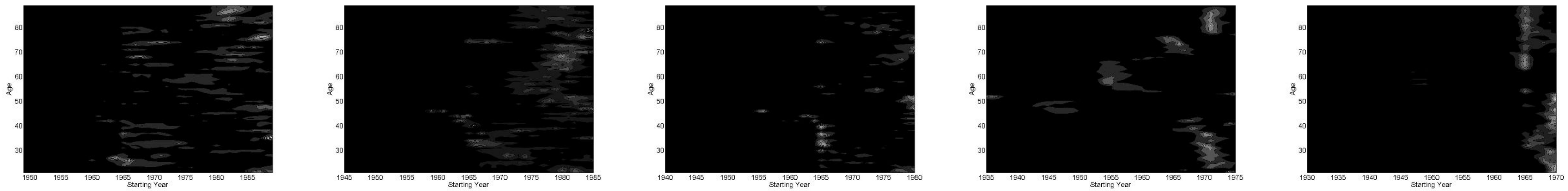
[15] The results for female data are available upon request from the corresponding author.

the sample sizes can be different for different age and model combinations. Similar to the simulation case studies, we see that in general the posterior distributions become more and more concentrated as μ increases.

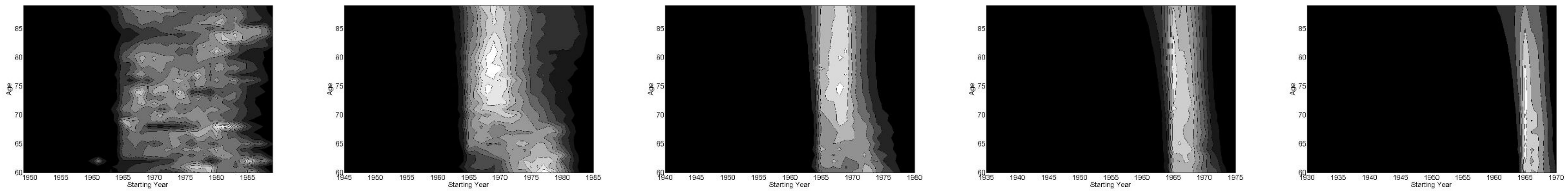
For the CBD model, we see that the posterior distributions are consistent for all μ-s: most probability mass falls within the starting years $[65 : 70]$, especially for $\mu \in \{5, 10, 15, 20\}$. Moreover, the starting years with high probabilities are similar across ages. For the Lee-Carter model, we see that data before the year 1965 receives almost no probability mass, which is in line with the CBD model. However, the posterior distributions for the Lee-Carter model are much more sensitive to the change of μ-s.

Figure 7 reports the posterior distributions for the U.S. male population. For the CBD model, the starting years which receive high posterior probabilities are earlier for the older ages. This result holds for all μ-s. Moreover, the peaks of the posterior distributions for the CBD model move forward as μ decreases, instead of being relatively insensitive to changes in μ as in the Dutch case. One possible explanation is that there exists a cohort effect in the US data. In particular, younger ages have different mortality patterns than older ages, and these patterns are better captured by more recent data. The CBD specification that we use does not model the cohort effect explicitly, and thus treats it as a kind of nonlinearity. As a result, the peak of the posterior distributions might be more sensitive to the changes in μ, which indicates a larger degree of model misspecification. In fact, the existence of a cohort effect in the U.S. data is also confirmed by Cairns et al. (2009).

From Figures 6 and 7, we see that, especially for the Lee-Carter model, the posterior distributions of the sample size are age-specific. This observation is in contrast to the existing sample size determination studies, where typically a single sample size is chosen to estimate the mortality model for all ages. In fact, if such age heterogeneity exists, then one may overlook relevant age-specific information when choosing a single sample size. As an illustration we consider the method by Booth et al. (2002) who determine an optimal sample size from the κ process of the Lee-Carter model. As a sensitivity test, we apply this method to three sub-age groups of the U.S. population. The three groups are $[21 : 40]$, $[41 : 60]$, and $[61 : 90]$. Results are reported in Figure 8. We see that while the optimal starting years for the female population are consistent for all three sub-age groups, the optimal starting years for the male population are different. In particular, the optimal starting years for the age group $[41 : 60]$ and $[61 : 90]$ are in mid 1970s and 1960s, indicating that the older ages have an earlier optimal starting year. Moreover, the optimal starting year for the age group $[21 : 40]$ cannot be clearly determined for the male population. Therefore, it seems to make sense to be careful when determining a single starting year for a wide range of ages.

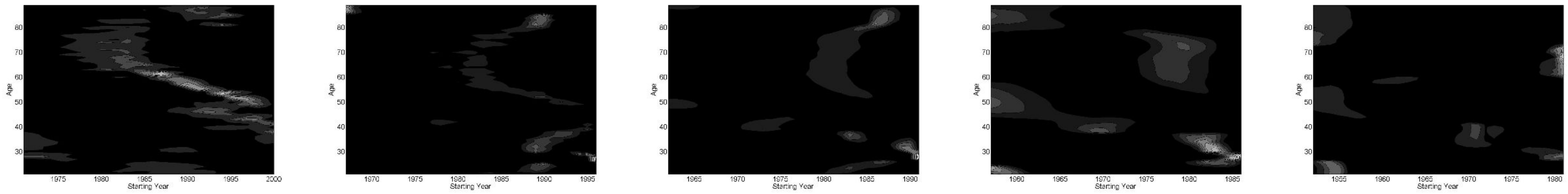


(a) Lee-Carter, $\mu = 1$ (b) Lee-Carter, $\mu = 5$ (c) Lee-Carter, $\mu = 10$ (d) Lee-Carter, $\mu = 15$ (e) Lee-Carter, $\mu = 20$

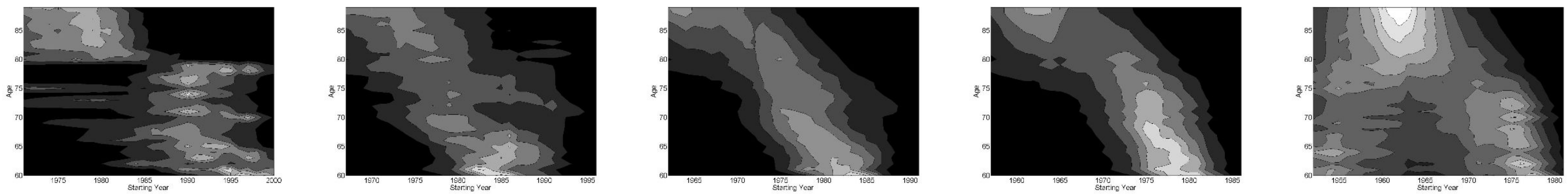


(f) CBD, $\mu = 1$ (g) CBD, $\mu = 5$ (h) CBD, $\mu = 10$ (i) CBD, $\mu = 15$ (j) CBD, $\mu = 20$

Figure 6: The heat map of the posterior distributions of the sample size for $\mu = \{1, 5, 10, 15, 20\}$ for the Dutch male population. The first row is the result for the Lee-Carter model, and the second row for the CBD model. The horizontal axis denotes the starting year. The sample size corresponding to a certain *starting year* and μ is computed as: *sample size* $= 1999 - \mu -$ *starting year* $+1$. Black corresponds to very low probability, light grey corresponds to low probability, and dark grey corresponds to high probability.

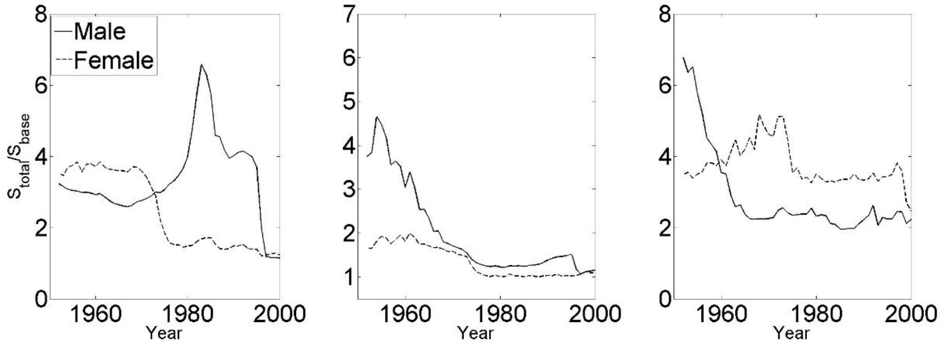


(a) Lee-Carter, $\mu = 1$ (b) Lee-Carter, $\mu = 5$ (c) Lee-Carter, $\mu = 10$ (d) Lee-Carter, $\mu = 15$ (e) Lee-Carter, $\mu = 20$



(f) CBD, $\mu = 1$ (g) CBD, $\mu = 5$ (h) CBD, $\mu = 10$ (i) CBD, $\mu = 15$ (j) CBD, $\mu = 20$

Figure 7: The heat map of the posterior distributions of the sample size for $\mu = \{1, 5, 10, 15, 20\}$ for the US male population. See Figure 6 for further details.



(a) U.S. Age [21 : 40] (b) U.S. Age [41 : 60] (c) U.S. Age [61 : 90]

Figure 8: The application of the method in Booth et al. (2002) to the sub age groups of the U.S. data. The three sub age groups are [21 : 40], [41 : 60], and [61 : 90], respectively.

6.2 Sensitivity with respect to the prior distribution

In this section, we examine the effect of the choice of prior distribution on the corresponding posterior distributions. Due to limitation of space, we report the results of three forecast horizons, 1, 10, and 20 years, for the Lee-Carter model using Dutch data.[16] The results for forecast horizons 5 and 15 years are similar.

For the sensitivity analysis, we consider five sets of prior distributions. Specifically, for each given (x, g, μ, M), we consider $p(J_{i+1}|x, g, \mu, M) = a \times p(J_i|x, g, \mu, M)$, where $J_{i+1} = J_i + 1$ for all i with $J_1 \leq J_i \leq J_{max-1}$, and $a \in \{\frac{1}{1.2}, \frac{1}{1.1}, 1, 1.1, 1.2\}$. In other words, we look at prior distributions for which the prior probability mass increases (decreases) with the sample size. The choices of a allow for great discrepancy among sample sizes. In particular, we have $p(J_{max}|x, g, \mu, M) \approx 50p(J_{min}|x, g, \mu, M)$ when $a = 1.1$, and $p(J_{max}|x, g, \mu, M) \approx 1764p(J_{min}|x, g, \mu, M)$ when $a = 1.2$.

For each model-population-forecast-horizon combination, we

1. obtain the posterior distribution with different prior distributions for each age, then take the average posterior distribution among all ages;
2. compute the posterior mean, 5% quantile, and 95% quantile for the averaged posterior distributions;
3. compare the sets of posterior mean and quantiles from all prior distributions.

[16]We report the results for the Dutch case since Dutch mortality data appears to be more non-linear than the U.S. data, and is more sensitive to the choice of prior distribution. The results for the Lee-Carter model using U.S. data, and for the CBD model using both data sets are available from the corresponding author upon request.

In principle, we can compare the set of posterior mean and quantiles for all single ages. However, the results are qualitatively the same for all ages, and we only report results for the average posterior distribution for the sake of conciseness.

The results are reported in Table 1. The first observation is that, for all forecast horizons, as smaller sample sizes receive larger (smaller) prior probability mass, the posterior distribution shifts towards (away from) them. The shift direction of the posterior distribution is consistent with the change of prior distribution. Second, the degree of shift of the posterior distributions is rather small compared to the change of the prior distributions. In particular, as a changes from $\frac{1}{1.2}$ to 1.2 ($p(J_{max}|x, g, \mu, M)$ becomes about 311 thousand times bigger), the posterior mean and quantiles of starting years change only from $(1962, 1969, 1981)$ to $(1977, 1987, 1990)$. This analysis indicates that the sample period is large enough to make the posterior distributions rather robust with respect to the change of prior distributions. Moreover, as the forecast horizon, μ, increases, the posterior distribution becomes more stable. Take the sample example discussed above, but with $\mu = 20$. The posterior distribution changes now from $(1957, 1962, 1967)$ to $(1964, 1968, 1971)$. The degree of change is even much smaller than the change when $\mu = 1$. From the sensitivity analysis, we see that, given the sample period used, the posterior distributions are rather robust to the choice of prior distributions investigated.

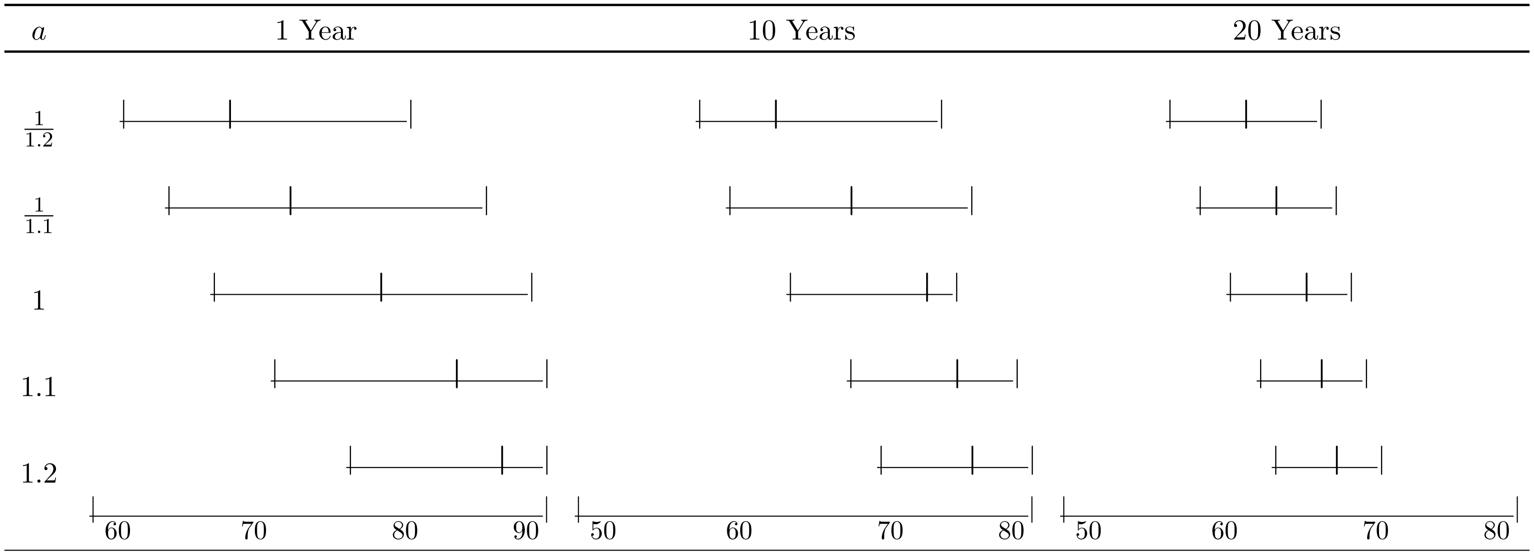


Table 1: The posterior mean, 5%, and 95% quantiles of the posterior distributions of sample sizes with different priors. Lee-Carter, Dutch population.

7 Out-of-sample forecast

After the posterior distributions $p(\tilde{\theta}_M|Y_t, J, \mu, M)$ and $p(J|Y_t, x, g, \mu, M)$ are obtained, the corresponding age and gender specific posterior predictive distribution, averaged over all sample sizes, can be derived. For example, for given (Y_t, μ, M), we have for each (x, g)

$$p(y_{x,t+1,g}|Y_t, x, g, \mu, M) = \sum_{J\in\mathcal{J}} \int_{\tilde{\Theta}_M} p(y_{x,t+1,g}|\tilde{\theta}_M, Y_t, x, g, \mu, M) p(\tilde{\theta}_M|Y_t, J, \mu, M) p(J|Y_t, x, g, \mu, M) d\tilde{\theta}_M. \tag{7.1}$$

Take the Lee-Carter model as an example. Suppose we have $k = 1, 2, ..., K$ draws of $\tilde{\theta}_{LC}$ for each J and μ. Then we have

$$\boldsymbol{\kappa}_{t+1}^{(k)}(J, \mu, LC) \sim \mathcal{N}(\boldsymbol{\kappa}_t^{(k)} + \boldsymbol{d}^{(k)}, \Sigma_\omega^{(k)}), \tag{7.2}$$

and

$$\boldsymbol{y}_{x,t+1}^{(k)}(J, \mu, LC) \sim \mathcal{N}(\boldsymbol{\alpha}_x^{(k)} + \boldsymbol{\beta}_x^{(k)} \boldsymbol{\kappa}_{t+1}^{(k)}, \Sigma_{\varepsilon_x}^{(k)}), \tag{7.3}$$

where all parameters in the above equations are drawn from $p(\tilde{\theta}_M|Y_t, J, \mu, LC)$, and where the notation (J, μ, LC) emphasizes the dependence of the parameters upon the specific model, sample size, and forecast horizon.[17] The posterior mean of $\boldsymbol{y}_{x,t+1}(J, \mu, LC) = (y_{x,t+1,m}(J, \mu, LC), y_{x,t+1,f}(J, \mu, LC))'$ is thus approximated by

$$\hat{y}_{x,t+1,g}(J, \mu, LC) = \frac{1}{K} \sum_{k=1}^{K} y_{x,t+1,g}^{(k)}(J, \mu, LC), \quad g \in \{m, f\}, \tag{7.4}$$

and the weighted posterior mean, $\hat{y}_{x,t+1,g}(\mu, LC)$, is approximated by

$$\hat{y}_{x,t+1,g}(\mu, LC) = \sum_{J\in\mathcal{J}} p(J|y_t, x, g, \mu, LC) \times \hat{y}_{x,t+1,g}(J, \mu, LC), \quad g \in \{m, f\}. \tag{7.5}$$

where we "integrate" out the sample size. We calculate the posterior predictive distribution similarly for the CBD model.

For the Bayesian model, the mean squared error for period $[T+1 : T+S]$ is calculated as

$$\mathrm{MSE}(x, g, \mu, M) = \frac{1}{S} \sum_{t=1}^{S} (\hat{y}_{x,T+t,g}(\mu, M) - y_{x,T+t,g})^2, \tag{7.6}$$

where $y_{x,T+t,g}$ is the observed mortality data at $T+t$ for (x, g), and $\hat{y}_{x,T+t,g}(\mu, M)$ is the posterior mean of $y_{x,T+t,g}$ according to model M and forecast horizon μ.

[17] In the sequel, notation such as (μ, LC) indicates an analogous dependence.

We perform the out-of-sample forecast for Dutch data for the period [2000 : 2009], using both the Bayesian models and the original models with the sample size yielding the minimal Deviance Information Criterion (DIC) ratio. The details of the DIC approach are explained in the Appendix.

The sample size which yields the minimal DIC ratio is considered as the optimal one, and is denoted by J_{opt}. In particular, an optimal sample size is computed for each (g, M) combination.[18] The comparison proceeds as follows. For each (g, M) combination, we compute the mean squared error,

$$\mathrm{MSE}(J_{opt}, x, g, M) = \frac{1}{S}\sum_{t=1}^{S}(\hat{y}_{x,T+t,g}(J_{opt}, g, M) - y_{x,T+t,g})^2. \tag{7.7}$$

In Equation (7.7), $\hat{y}_{x,T+i,g}(J_{opt}, g, M)$ is the posterior mean of $y_{x,T+i,g}$ generated by model M and the sample size with the minimal DIC, J_{opt}. The MSE-s generated in (7.7) are then compared with those generated by the Bayesian models.

The J_{opt}-s are reported in Table 2. We see that the J_{opt}-s are quite similar for all model-gender combinations. Moreover, they are rather small: They range from 10 years to 14 years. Therefore, it seems that there exist non-linearities in the data, and the fits of the linear Lee-Carter and CBD model deteriorate as the sample size increases. This observation is in line with Figure 2a, where the fit of the method in Booth et al. (2002) to Dutch data is plotted.

In Figure 9, we report the comparison of the normalized DIC ratios and the average posterior distributions for Dutch males data with $\mu = 1$, 10, and 20 using the Lee-Carter model. The normalized DIC ratios are computed by first taking the negative of the DIC ratios, then normalize them to sum up to 1. In this way, in contrast to the original DIC ratios, the higher the normalized ratio, the better the corresponding sample size. We see that, in all scenarios, the normalized DIC ratios are much flatter than the corresponding posterior distributions, meaning that no particular sample size is clearly better than others. This analysis indicates that, if we would like to determine an optimal sample size based on DIC, the result is likely to be much more sensitive to the possible noise in the data and the chosen sample period compared with the posterior means generated by the Bayesian models. The results for the CBD model are similar and therefore omitted.

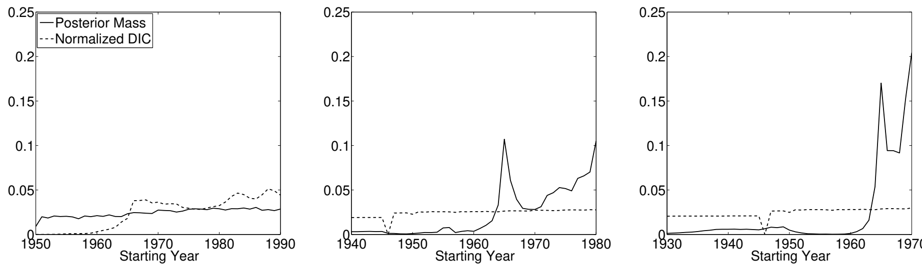
Finally, from Figure 10, we see that the out-of-sample forecast performance from the models with J_{opt}-s are dominated by the Bayesian model in the majority of cases. For the females population, the Bayesian CBD model performs quite well. In particular, it produces the smallest MSE-s for age 55 and 65, with all μ-s. For the males population, both Bayesian models generate very small MSE-s for the age 85 with all μ-s. The performance of the

[18]The forecast horizon does not play a role in the DIC method.

J_{opt}	male	female
LC	1986	1990
CBD	1990	1990

Table 2: The optimal starting year under the DIC criterion for Dutch data. In each entry, the first number is the optimal starting year for the males population, and the second number the optimal starting year for the females population.

Bayesian models are between the two models with J_{opt}-s with $\mu \in \{1, 10\}$, and are slightly worse than these two models with $\mu = 20$. However, the posterior distributions of the starting years derived with $\mu = 20$ is more suitable for longer forecast horizons. Therefore, the out-of-sample performance of 10 years does not necessarily mean that the posterior distribution is not suitable to be used for forecasting further into the future.

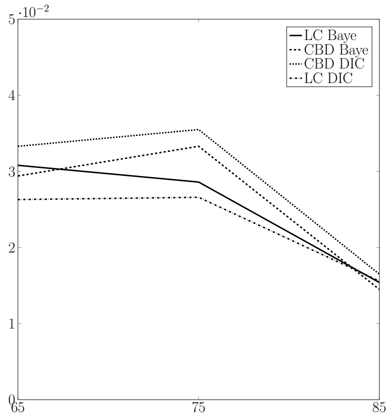


(a) Lee-Carter, $\mu = 1$ (b) Lee-Carter, $\mu = 10$ (c) Lee-Carter, $\mu = 20$

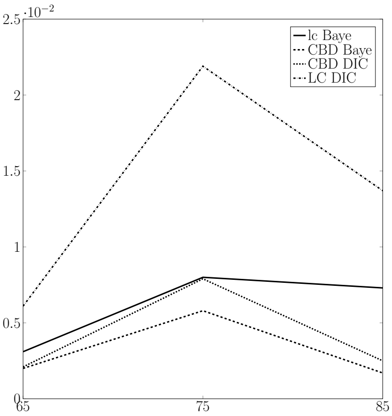
Figure 9: The comparison of the normalized DIC ratios (dashed line) and the posterior distributions of sample size (solid line) for $\mu \in \{1, 10, 20\}$ for the Dutch males population using the Lee-Carter model. We first take the negative of the DIC ratios, and then normalized them so that they sum up to 1.

8 Conclusion

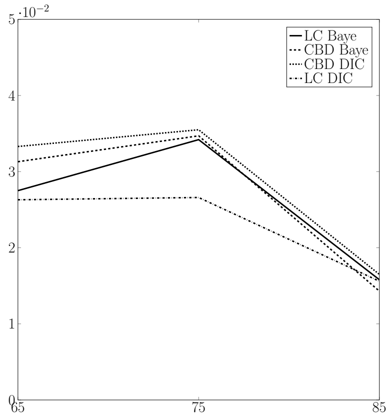
In this paper, we studied the sample size determination of mortality forecasting. In particular, the sample size is incorporated as an extra parameter into the parameter space of the Lee-Carter model and the CBD model, and its conditional posterior distribution is updated based on historical performance for different forecast horizons. Our method is applicable to many linear mortality models studied in the literature. We focused on the first generation Lee-Carter and CBD model with the underlying latent time effect modeled by a random walk with drift: The simplicity of these two models allows us to illustrate our method more clearly.



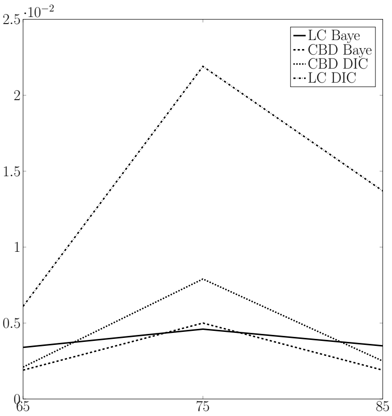
(a) Male, $\mu = 1$



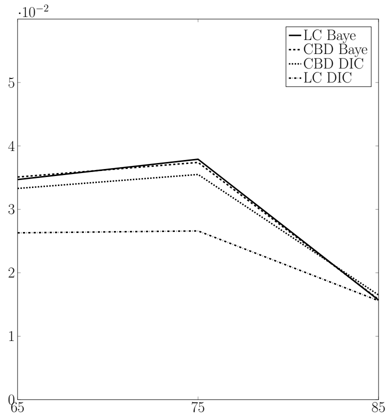
(b) Female, $\mu = 1$



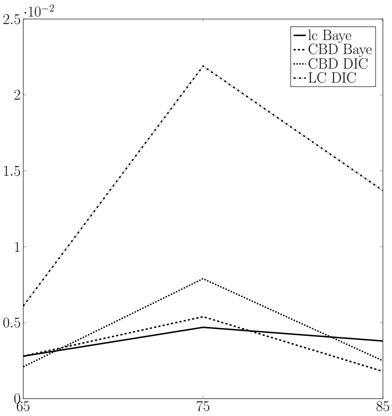
(c) Male, $\mu = 10$



(d) Female, $\mu = 10$



(e) Male, $\mu = 20$



(f) Female, $\mu = 20$

Figure 10: The mean squared errors of the out-of-sample forecasts for the Dutch population aged $\{65, 75, 85\}$ for the years 2000 to 2009. The mean squared errors are calculated for the Lee-Carter and the CBD model based on the sample size with the minimal DIC ratios (LC DIC and CBD DIC), as well as the Bayesian models (LC Baye and CBD Baye). All mean squared errors are calculated in terms of logit q.

We applied our method to Dutch and U.S. data, and found quite concentrated posterior distributions of the sample size for most age, gender, and forecast horizon combinations for both models and data sets. In particular, the posterior distributions of the CBD model are more insensitive to the changes in forecast horizons than the ones from the Lee-Carter model, indicating that the CBD model specification might be more robust to nonlinearities in the data. Moreover, the posterior distributions for both models are age- and gender-specific, meaning that using a single sample size for all ages might not be optimal. In particular, for the CBD model we found that the posterior distribution of the sample size is more concentrated around earlier starting years for older ages. One possible reason is that there exists a cohort effect in the U.S. data: people in younger ages have different mortality patterns than people in older ages, and these patterns are reflected in more recent data. This finding is in line with Cairns et al. (2009).

We carried out an out-of-sample forecast analysis using Dutch data from 2000 to 2009. In particular, we compared the out-of-sample performance using the Bayesian models and the original models based on the sample size that gives the minimal Deviance Information Criterion ratio. We found that the forecasts from the Bayesian models are more precise than the forecasts from the original models in over half of the cases. Therefore, the forecasts of the Bayesian model seem to be more robust than the ones based on the optimal sample size under the traditional Deviance Information Criterion.

There are several directions for future research. First, it could be interesting to apply our method to other mortality models, such as the models studied in Cairns et al. (2011). Second, when more mortality models are included, it could be interesting to compare the performance of each model-sample size combination, for example, using Bayes factors (Kass and Raftery 1995). More insight of the mortality models and the choice of sample size could be gained by doing this analysis. Third, a further model averaging could be done: We can average the forecasts produced by all model-sample size combinations with weights equal to the corresponding likelihood (Hoeting et al. 1999). By doing this, more robust forecasts might be obtained.

Appendices

A The Gibbs sampler for the mortality models

A.1 Lee-Carter model

For each J and μ, $\tilde{\theta}_{LC} = (\boldsymbol{\kappa}', \boldsymbol{d}', \boldsymbol{\alpha}', \boldsymbol{\beta}', \text{vech}(\Sigma_\varepsilon)', \text{vech}(\Sigma_\omega)')'$. In particular, we let Σ_ε be diagonal for all Bayesian models, so $\text{vech}(\Sigma_\varepsilon)'$ is the vector of σ_{ε_x}-s in all cases. Noninformative marginal prior distributions are chosen:

$p(\boldsymbol{\alpha}', \boldsymbol{\beta}') \propto 1$, $p(\sigma^2_{\varepsilon_{x,g}}) \propto \frac{1}{\sigma^2_{\varepsilon_{x,g}}}$ for each x and g, and $p(\boldsymbol{d}', \Sigma_\omega) \propto |\Sigma_\omega|^{-\frac{3}{2}}$. Moreover, the marginal prior distributions are assumed to be independent of each other.

We illustrate the estimation of (the posterior distribution of) $\tilde{\theta}_{LC}$ in year $\tilde{T}$. The sample used in the estimation in this case is $y_{[\tilde{T}-\mu-J+1:\tilde{T}-\mu]}$. The Gibbs sampling proceeds as follows:

1. First we simulate the state process $\boldsymbol{\kappa}$. Denote by D_t the time t information set for $t \in [\tilde{T} - \mu - J + 1 : \tilde{T} - \mu]$, $\boldsymbol{\kappa}_t | D_{t-1} \sim N(a_t, R_t)$ the prior of $\boldsymbol{\kappa}_t$ at time t, $\boldsymbol{\kappa}_t | D_t \sim N(\boldsymbol{m}_t, C_t)$ the posterior of $\boldsymbol{\kappa}_t$ at time t, and $\boldsymbol{y}_t | D_{t-1} \sim N(\boldsymbol{f}_t, Q_t)$ the one step forecast distribution. Then, given starting values $\boldsymbol{m}_{\tilde{T}-\mu-J}$ and $C_{\tilde{T}-\mu-J}$, the updating equations for time t are

$$\begin{aligned} \boldsymbol{a}_t =& \boldsymbol{m}_{t-1} + \boldsymbol{d}, \\ R_t =& C_{t-1} + \Sigma_\omega, \\ Q_t =& F' R_t F + \Sigma_\varepsilon, \\ A_t =& R_t F Q_t^{-1}, \\ \boldsymbol{f}_t =& \boldsymbol{\alpha} + F \boldsymbol{a}_t, \\ \boldsymbol{m}_t =& \boldsymbol{a}_t + A_t(\boldsymbol{y}_t - \boldsymbol{f}_t), \\ C_t =& R_t - A_t Q_t A_t'. \end{aligned} \tag{A.1}$$

   In system (A.1), $\boldsymbol{y}_t$ is the log central death rate vector at time t, and

$$F = \begin{pmatrix} \boldsymbol{\beta}_m' & 0 \\ 0 & \boldsymbol{\beta}_g' \end{pmatrix},$$

   where $\boldsymbol{\beta}_m$ and $\boldsymbol{\beta}_g$ are column vectors representing the $\boldsymbol{\beta}$-s for male and female, respectively.

2. Fit the updating system (A.1) for each $t \in [\tilde{T} - \mu - J + 1 : \tilde{T} - \mu]$, then draw $\boldsymbol{\kappa}_{\tilde{T}-\mu}$ from $\mathcal{N}(\boldsymbol{a}_{\tilde{T}-\mu}, R_{\tilde{T}-\mu})$. For the rest of the t-s, draw $\boldsymbol{\kappa}_t$ from

$$\boldsymbol{\kappa}_t | \boldsymbol{\kappa}_{t+1}, y \sim \mathcal{N}(\boldsymbol{h}_t, H_t),$$

   where $\boldsymbol{h}_t = \boldsymbol{a}_t + B_t(\boldsymbol{\kappa}_{t+1} - \boldsymbol{a}_{t+1})$, and $H_t = R_t - B_t R_{t+1} B_t'$, and $B_t = C_t R_{t+1}^{-1}$.

3. For every age x and gender g, draw $\sigma^2_{\varepsilon_x}$ from

$$\sigma^2_{\varepsilon_x} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, y \sim \text{Inv-Gamma}(\frac{J}{2}, \frac{\sum_{t=\tilde{T}-\mu-J+1}^{\tilde{T}-\mu} (y_{x,t,g} - \alpha_{x,g} - \beta_{x,g}\kappa_{t,g})^2}{2}).$$

4. For each gender g, let $X = (\iota, \boldsymbol{\kappa}_g)$, where ι is a $J \times 1$ vector of ones, and $\boldsymbol{\kappa}_g$ is the vector of $\kappa_{t,g}$-s. Then for each age x, draw $\alpha_{x,g}$ and $\beta_{x,g}$ from

$$(\alpha_{x,g}, \beta_{x,g})|y, \boldsymbol{\kappa}_g, \sigma^2_{\varepsilon_{x,g}} \sim \mathcal{N}((X'X)^{-1}X'y, \sigma^2_{\varepsilon_{x,g}}(X'X)^{-1}).$$

5. Draw d from

$$\boldsymbol{d}|\boldsymbol{\kappa}, \Sigma_\omega \sim \mathcal{N}(\frac{\boldsymbol{\kappa}_{\tilde{T}-\mu} - \boldsymbol{\kappa}_{\tilde{T}-\mu-J+1}}{J}, \frac{\Sigma_\omega}{J}).$$

6. Draw Σ_ω from

$$\Sigma_\omega^{-1}|\boldsymbol{\kappa}, \boldsymbol{d} \sim \text{Wishart}(J-1, \frac{\hat{\Sigma}_\omega^{-1}}{J}),$$

where $\hat{\Sigma}_\omega = \frac{1}{J}\sum_{t=\tilde{T}-\mu-J+1}^{\tilde{T}-\mu}(\boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1} - \boldsymbol{d})(\boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1} - \boldsymbol{d})'$.

We continue the iteration until convergence is reached. In this paper, we set the iteration times to be 2,000, and convergence is reached in all cases under the criterion proposed in Gelman and Rubin (1992). The initial value of $m_{\tilde{T}-\mu-J}$ and $C_{\tilde{T}-\mu-J}$ are set equl to $\bar{\boldsymbol{\kappa}}$ and $\hat{\Sigma}_\omega$, respectively, obtained from a weighted least squared estimation of model (2.2) - (2.3) using the dataset $y_{[\tilde{T}-\mu-J+1:\tilde{T}-\mu]}$.

A.2 CBD model

For each J and μ, $\tilde{\theta}_{CBD} = (\boldsymbol{\kappa}', \boldsymbol{d}', \text{vech}(\Sigma_\varepsilon)', \text{vech}(\Sigma_\omega)')'$. Similar to the Lee-Carter model, assume that y_t is a $2n$-dimensional column vector for each t. In the context of model (4.6) - (4.7), we have $p = 1$, $B_t = 0$ and $X_t = 0$ for all t,

$$F_t = \begin{bmatrix} 1 & 1 & ... & 1 & 0 & 0 & ... & 0 \\ x_1 & x_2 & ... & x_n & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 1 & 1 & ... & 1 \\ 0 & 0 & ... & 0 & x_1 & x_2 & ... & x_n \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \end{bmatrix}$$

for all t, $z_t = [\kappa^{(1)}_{t,m}, \kappa^{(2)}_{t,m}, \kappa^{(1)}_{t,f}, \kappa^{(2)}_{t,f}, d^{(1)}_m, d^{(2)}_m, d^{(1)}_m, d^{(2)}_m,]'$,

$$G_{t,1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

for all t, and $\omega_t = [\omega_{t,m}^{(1)}, \omega_{t,m}^{(2)}, \omega_{t,f}^{(1)}, \omega_{t,f}^{(2)}]'$. Also, ε_t is the vector of error terms, Σ_ε is a diagonal matrix, and $\Sigma_{\omega_t} = \Sigma_\omega$ for all t.

Again, we choose the noninformative marginal prior distributions $p(\sigma^2_{\varepsilon_{x,g}}) \propto \frac{1}{\sigma^2_{\varepsilon_{x,g}}}$ for each x and g, and $p(\boldsymbol{d}', \Sigma_\omega) \propto |\Sigma_\omega|^{-\frac{3}{2}}$. The marginal prior distributions are assumed to be independent of each other.

Also in this case we illustrate the estimation of (the posterior distribution of) $\tilde{\theta}_{LC}$ in year $\tilde{T}$. The Gibbs sampling procedure for the CBD model is as follows.

1. For $t \in [\tilde{T} - \mu - J + 1 : \tilde{T} - \mu]$, denote by $\boldsymbol{\kappa}_t | D_{t-1} \sim N(a_t, R_t)$ the prior of $\boldsymbol{\kappa}_t$ at time t, $\boldsymbol{\kappa}_t | D_t \sim N(\boldsymbol{m}_t, C_t)$ the posterior of $\boldsymbol{\kappa}_t$ at time t, and $\boldsymbol{y}_t | D_{t-1} \sim N(\boldsymbol{f}_t, Q_t)$ the one step forecast distribution.[19] The updating equations for time t are, given the starting values $\boldsymbol{m}_{\tilde{T}-\mu-J}$ and $C_{\tilde{T}-\mu-J}$

$$\begin{aligned} \boldsymbol{a}_t =& \boldsymbol{m}_{t-1} + \boldsymbol{d}, \\ R_t =& C_{t-1} + \Sigma_\omega, \\ Q_t =& F' R_t F + \Sigma_\varepsilon, \\ A_t =& R_t F Q_t^{-1}, \\ \boldsymbol{f}_t =& F \boldsymbol{a}_t, \\ \boldsymbol{m}_t =& \boldsymbol{a}_t + A_t(\boldsymbol{y}_t - \boldsymbol{f}_t), \\ C_t =& R_t - A_t Q_t A_t'. \end{aligned} \tag{A.2}$$

   In system (A.2),

$$F = \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix},$$

   and $X = [\boldsymbol{\iota}, \boldsymbol{x}]'$, where $\boldsymbol{x} = [x_1, x_2, ..., x_n]'$ is the vector of ages. In our case, $\boldsymbol{x} = [61, 62, ..., 89]'$.

2. Fit the updating system (A.2) for all t, and draw the $\boldsymbol{\kappa}$ process in the same way as in the Lee Carter model.

3. For every age x and gender g, draw $\sigma^2_{\varepsilon_x}$ from

$$\sigma^2_{\varepsilon_x} | \boldsymbol{\kappa}, \boldsymbol{y} \sim \text{Inv-Gamma}\left(\frac{J}{2}, \frac{\sum_t (y_{x,t,g} - \kappa_{t,g}^{(1)} + \kappa_{t,g}^{(2)} x)^2}{2}\right).$$

4. Draw $\boldsymbol{d}$ from

$$\boldsymbol{d} | \boldsymbol{\kappa}, \Sigma_\omega \sim \mathcal{N}\left(\frac{\boldsymbol{\kappa}_{\tilde{T}-\mu} - \boldsymbol{\kappa}_{\tilde{T}-\mu-J+1}}{J}, \frac{\Sigma_\omega}{J}\right).$$

[19] The prior distribution of $\boldsymbol{\kappa}_{\tilde{T}-\mu-J+1}$ is set in the same way as for the Lee-Carter model.

5. Draw σ_ω^2 from

$$\Sigma_\omega^{-1}|\boldsymbol{\kappa}, \boldsymbol{d} \sim \text{Wishart}(J-1, \frac{\hat{\Sigma}_\omega^{-1}}{J}),$$

where $\hat{\Sigma}_\omega = \frac{1}{J}\sum_{t=\tilde{T}-\mu-J+1}^{\tilde{T}-\mu}(\boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1} - \boldsymbol{d})(\boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1} - d)'$.

We continue this procedure until convergence is reached. The initial values $m_{\widetilde{T}-\mu-J}$ and $C_{\widetilde{T}-\mu-J}$ are the estimates of $\overline{\kappa}$ and $\widehat{\Sigma}_\omega$, applying the estimation method in Cairns, Blake, and Dowd (2006), using the dataset $y_{[\tilde{T}-\mu-J+1:\tilde{T}-\mu]}$.

B The DIC approach

Define the deviance as

$$D(\tilde{\theta}_M|J, g, M) = -2\log(p(Y_T|\tilde{\theta}_M, J, g, M)) + C \qquad \text{(B.1)}$$

for each sample size, gender, and model combination, where $\tilde{\theta}_M$ is the parameter drawn from the corresponding posterior distribution, $p(Y_T|\tilde{\theta}_M, J, g, M)$ is the likelihood function, and C is a constant which will cancel out when comparing different models. For the computation of the DIC ratios, we let $J \in \{10, 11, ..., 50\}$ in line with the Bayesian models. Moreover, we use Dutch mortality data from 1950 to 1999 to compute the likelihoods. In other words, $p(Y_T|\tilde{\theta}_M, J, M)$ in (B.1) is the likelihood of $y_{[1999-J+1:1999]}$ based on model M.

Following Spiegelhalter et al. (2002), we define

$$p_D(J, g, M) = E_{\tilde{\theta}_M}[D(\tilde{\theta}_M|J, M)] - D(\hat{\tilde{\theta}}_M|J, M),$$

where $E_{\tilde{\theta}_M}[.]$ is calculated using the posterior distribution of $\tilde{\theta}_M$ given (J, M), and $\bar{\tilde{\theta}}_M$ is the posterior mean of $\tilde{\theta}_M$. The Deviance Information Criterion is then defined as

$$\text{DIC}(J, g, M) = E_{\tilde{\theta}_M}[D(\tilde{\theta}_M|J, g, M)] + p_D(J, g, M). \qquad \text{(B.2)}$$

References

Adcock, C. (1997). Sample size determination: a review. *Journal of the Royal Statistical Society: Series D (The Statistician) 46*(2), 261–283.

Baran, S., J. Gáll, M. Ispány, and G. Pap (2007). Forecasting hungarian mortality rates using the lee-carter method. *Acta Oeconomica 57*(1), 21–34.

Bauer, D., M. Börger, and J. Ruß (2010). On the pricing of longevity-linked securities. *Insurance: Mathematics and Economics 46*(1), 139–149.

Booth, H., J. Maindonald, and L. Smith (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population studies 56*(3), 325–336.

Börger, M. (2009). The Volatility of Mortality. *Deutscher SCOR-Preis für Aktuarwissenschaften 2008: Zusammenfassung eingereichter Arbeiten 8*, 25.

Börger, M. (2010). Deterministic shock vs. stochastic value-at-risk: an analysis of the Solvency II standard model approach to longevity risk. *Blätter der DGVFM 31*(2), 225–259.

Börger, M., D. Fleischer, and N. Kuksin (2013, 11). Modeling the mortality trend under modern solvency regimes. *ASTIN Bulletin FirstView*, 1–38.

Brouhns, N., M. Denuit, and J. Vermunt (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics 31*(3), 373–393.

Cairns, A., D. Blake, and K. Dowd (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance 73*(4), 687–718.

Cairns, A. J., D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, and M. Khalaf-Allah (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics 48*(3), 355–367.

Cairns, A. J., D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, A. Ong, and I. Balevich (2009). A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal 13*(1), 1–35.

Cairns, A. J. B., D. Blake, K. Dowd, G. D. Coughlan, and M. Khalaf-Allah (2011). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin-Actuarial Studies in non LifeInsurance 41*(1), 29.

Czado, C., A. Delwarde, and M. Denuit (2005). Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics 36*(3), 260–284.

Damien, P., P. Dellaportas, N. G. Polson, and D. A. Stephens (2013). *Bayesian Theory and Applications*. Oxford University Press.

De Santis, F. (2004). Statistical evidence and sample size determination for bayesian hypothesis testing. *Journal of statistical planning and inference 124*(1), 121–144.

Denuit, M. and A. Goderniaux (2005). Closing and projecting life tables using log-linear models. *Bulletin of the Swiss Association of Actuaries 1*(4), 29–49.

Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457–472.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: a tutorial. *Statistical science*, 382–401.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the american statistical association 90*(430), 773–795.

Kogure, A., K. Kitsukawa, and Y. Kurachi (2009). A Bayesian comparison of models for changing mortalities toward evaluating longevity risk in Japan. *Asia-Pacific Journal of Risk and Insurance 3*(2).

Kogure, A. and Y. Kurachi (2010). A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. *Insurance: Mathematics and Economics 46*(1), 162–172.

Lee, R. and L. Carter (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 659–671.

Lee, R. and T. Miller (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography 38*(4), 537–549.

Li, N. and R. Lee (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography 42*(3), 575–594.

Miltersen, K. R. and S.-A. Persson (2005). Is mortality dead? stochastic forward force of mortality rate determined by no arbitrage. Technical report, Working Paper, University of Bergen.

O'Hare, C. and Y. Li (2012). Identifying structural breaks in stochastic mortality models. *Available at SSRN 2192208*.

Pedroza, C. (2002). *Bayesian Hierarchical Time Series Modeling of Mortality Rates*. Ph. D. thesis, Havard University.

Pedroza, C. (2006). A Bayesian forecasting model: predicting US male mortality. *Biostatistics 7*(4), 530–550.

Pitacco, E., M. Denuit, S. Haberman, and A. Olivieri (2009). *Modelling longevity dynamics for pensions and annuity business*. Oxford University Press, USA.

Plat, R. (2011). One-year value-at-risk for longevity and mortality. *Insurance: Mathematics and Economics 49*(3), 462–470.

Reichmuth, W. H. and S. Sarferaz (2008). Bayesian demographic modeling and forecasting: an application to US mortality. Technical report, SFB 649 discussion paper.

Renshaw, A. E. and S. Haberman (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics 38*(3), 556–570.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(4), 583–639.

van Berkum, F., K. Antonio, and M. Vellekoop (2013). Structural changes in mortality rates with an application to Dutch and Belgian data.

Wang, F. and A. E. Gelfand (2002). A simulation-based approach to bayesian sample size determination for performance under a given model and for separating models. *Statistical Science 17*(2), 193–208.

Weiss, R. (1997). Bayesian sample size calculations for hypothesis testing. *Journal of the Royal Statistical Society: Series D (The Statistician) 46*(2), 185–191.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models*. Springer.