Prediction intervals for Lee-Carter-based mortality forecasts

Nico Keilman, nico.keilman@econ.uio.no
Dinh Quang Pham, dqp@ssb.no

June 2006

Abstract

Mortality predictions derived from the model proposed by Lee and Carter are uncertain for three reasons. The predictions are based on
1. observed death counts and empirical death rates. These are subject to Poisson variability.
2. estimates of the parameters of the Lee-Carter-model
3. extrapolated values of the model's time index.

Authors who have calculated prediction intervals for future death rates and life expectancies have usually considered the third source of uncertainty, but many have neglected the first or the second source, or both. The extent to which each source contributes to overall uncertainty is an empirical issue.

We use simulation to construct prediction intervals around future death rates and life expectancies computed by means of the Lee-Carter-model, considering all three sources. We determine the relative contribution of the second and the third source, and analyse the impact of the first source on prediction intervals. The method is applied to mortality data for men and women in Norway for the period 1900-2004.

1. Sources of uncertainty

The model proposed by Lee and Carter (1992) for age-specific mortality has gained some popularity in the past decade. The model gives an accurate description of historical death rates in various countries; see for example Tuljapurkar et al. (2000) who successfully applied the model to data for the period 1950-1994 from the seven economically most developed countries (G7). The model also provides a simple way of projecting age-specific mortality into the future. The reason is that in many applications, the only parameter to be extrapolated, the so-called time index, shows a near linear development as a function of time.

The model, sometimes in a modified form, has been applied for predicting age-specific mortality and life expectancies in the USA (Lee and Carter 1992; Lee and Miller 2001), the G7-countries (Tuljapurkar et al. 2000), Austria (Carter and Prskawetz 2001; Prskawetz and Carter 2001), Australia (Booth et al. 2002; De Jong and Tickle 2006), Sweden (Lundström and Qvist 2004; Tuljapurkar 2005) and surely a number of other countries. Since future mortality is unknown, the predictions are uncertain. Therefore, an important issue relates to the correct characterization of prediction intervals around future death rates and life expectancies, in particular in connection with probabilistic forecasts.

Below we will argue that mortality predictions based on the model of Lee and Carter (1992) - LC henceforth- are uncertain for three reasons: The predictions are based on
1. observed death counts and empirical death rates. These are subjected to Poisson variability.
2. estimates of the parameters of the LC model
3. extrapolated values of the model's time index.

Authors who have calculated prediction intervals for future death rates and life expectancies have usually taken the third source into account, at least partly (e.g. Lee and Carter 1992; Tuljapurkar et al 2000; Tuljapurkar 2005). However, many authors have neglected the first or the second source, or both. Exceptions are Wilmoth (1993), Brouhns et al. (2005), and Koissi et al. (2006) for Poisson variability, and Lee and Carter (1992), Brouhns et al. (2005), and Koissi et al. (2006) for the estimation uncertainty. We do not know of analyses that take account of all three sources of uncertainty simultaneously. The extent to which each source contributes to overall uncertainty is an empirical issue.

LC computed prediction intervals around future life expectancies for the US. They predicted a life expectancy in 2065 for the two sexes combined equal to 86.1 years, with a 95 per cent prediction interval between 80.9 and 90.2 years. This prediction interval accounted for the prediction uncertainty in the time index only. In a sensitivity analysis, they found that this prediction uncertainty dominated over other forms of uncertainty in the long run. However, in their analysis they ignored estimation uncertainty in estimates of the time index and Poisson variability in the empirical death rates. In addition, they extrapolated empirical death rates for ages beyond 84, but did not evaluate the precision of those extrapolations. Therefore, we suspect that their prediction intervals were too narrow.

The purpose of the present paper is to compute prediction intervals around future death rates and life expectancies computed by means of the LC model, considering all three sources. We determine the relative contribution of the second and the third source, and analyse the impact of the first source on prediction intervals. The method is applied to annual mortality data for men and women by single ages in Norway for the period 1900-2004. The predictions, both the medians, expected values, and the prediction intervals, have been used as a source of information when Statistics Norway prepared its 2005-based population forecast.

2. Approach

LC modelled the log of the death rate for a certain age x and calendar year t ($m_{x,t}$) as a function of age-specific parameters a_x and b_x, and a time-specific parameter k_t (the time index), in the following way:

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

The model explains age–specific mortality as a combination of a common age pattern a_x, which is independent of time, and a time index k_t, which is the same for all ages, and which describes the decrease of mortality over time. In practice, however, the decrease is not the same for all age groups, and hence a multiplicative parameter b_x modifies the time index for ages x. When there are ω+1 age groups (x = 0, 1, 2, …, ω) and T+1 calendar years (t = 0, 1, 2, …, T), the (ω+1)x(T+1)-matrix of death rates $m_{x,t}$ is summarized by $2\omega + T + 3$ parameters. In practice, there are fewer parameters than death rates. Therefore, the model does not describe the death rates one hundred per cent accurately, and hence it contains a residual $\varepsilon_{x,t}$. We assume that these residuals are independent and normally distributed with constant variance.

LC estimated the parameters a_x as the average of the empirical log-rates: $\hat{a}_x = \frac{1}{T}\sum_t \ln(m_{x,t})$.

Next, they estimated b_x and k_t by Singular Value Decomposition (SVD) of the matrix of centred log-rates $\ln(m_{x,t})$ - $\hat{a}_x$, with the normalizing constraints that the sums of all b_x-values and k_t -values equal one and zero, respectively.

Several authors (Lee and Miller 2001; Carter and Prskawetz 2001; Prskawetz and Carter 2001; Booth et al, 2002) have argued that the empirical residuals $\hat{\varepsilon}_{x,t}$ display a systematic age or time pattern. Indeed, when we compared observed life expectancy for Norway with the life expectancy predicted by model (1), we found that the predicted increase for men was too slow in recent years. There were also systematic patterns in the residuals across ages. These are two examples where the model fit is not good enough, and one may improve the model by including an extra term:

(2) $\ln(m_{x,t}) = a_x + b_{1x} k_{1t} + b_{2x} k_{2t} + \varepsilon_{x,t}$.

One can estimate this model by applying SVD to the matrix of centred log-rates, extracting just two components. The estimated components $b_{1x}\, k_{1t}$ and $b_{2x}\, k_{2t}$ will be independent.

Generalizing further, the model fits perfectly with N components, where N is the rank of the matrix that contains the death rates broken down by age and time. This leads to

$$(3)\ \ln(m_{x,t}) = a_x + b_{1x}k_{1t} + b_{2x}k_{2t} + \ldots + b_{Nx}k_{Nt}.$$

In this perspective, the original LC model uses one component, and out-of-sample prediction requires extrapolating just one time index.

The method that results in LC based mortality rates and life expectancies for future years involves a great number of steps, mixing estimation and prediction. The expressions for the expected values of future mortality rates and life expectancies are straightforward. However, analytical expressions for prediction intervals are unknown because of the complexity of the method, and we have not attempted to derive them. Instead, we have used bootstrapping to find prediction intervals, similar to Lee and Carter (1992 Appendix B), but we have attempted to take account of all three sources of uncertainty. First, we will give a verbal description of our algorithm. Section 3 contains mathematical and statistical details.

The following steps are involved in the algorithm that results in prediction intervals around future mortality. Steps 1-3 apply to point predictions for relevant variables, while steps 4-6 give the methods for obtaining interval predictions. The algorithm is the same for men and women.

1. Estimate age-specific death rates for each year.

The estimates are based on observed death counts for each calendar year, and observed population counts at 31st December of each year, both broken down by age. Two features merit special attention. First, death counts for a given year are broken down by age on 31st December of that year. Hence, we follow birth cohorts throughout the year and adopt the so-called period-cohort observational plan for death counts, death rates, and age-specific parameters (the C-P interval of Preston et al. 2001, p. 31). Our perspective is different from that of a life table, in which cohorts are followed from one exact age to the next (the A-C interval of Preston et al.). The reason for adopting the period-cohort perspective is that the future rates have been used in Statistics Norway's 2005-based population forecast, which is of the cohort-component type. Forecast models of this type follow birth cohorts from one point in time to the next. Second, we assume a constant force of mortality and uniformly distributed immigration over each unit age-time interval (Van Imhoff 1990). Rates estimated on the basis of a constant force of mortality have somewhat better statistical properties than the more

usual ("actuarial") rates, which are estimated using an assumption of uniformly distributed deaths over the unit interval (Hoem and Funck-Jensen 1982). For short age intervals and time intervals, the numerical differences between the two types of rates are small.

2. Estimate the LC model

We have taken Poisson variability of the age-specific death rates into account by using Weighted Least Squares (WLS) estimation, with weights equal to the inverse of the estimated variances for the death rates. This method gives relatively more weight to ages and times with high mortality. This step results in estimates of the parameters a_x, b_{ix}, and k_{it} for each component i (i = 1, 2 … N) together with an estimate of the residual variance.

3. Extrapolate the time indices and predict age-specific mortality

Extrapolate the time factors k_{1t} ... k_{Nt} of the LC model by an appropriate time series model. Predict age-specific mortality rates and derive future life expectancies.

4. Draw ("simulate") values of the residual for each age-time interval, and estimate the LC model S_ε times, using the method of step 2

For each age-time interval, the LC model fitted in Step 2 gives rise to a residual. These residuals are assembled in a table with ω+1 rows and T+1 columns. Simulate a new table with residuals by drawing, for each age x and year t, a new residual from a randomly chosen row and a randomly chosen column from the original table. Add these simulated residuals to the fitted values for $\ln(m_{x,t})$ from step 2, to obtain simulated death rates. Combine these simulated death rates with observed population numbers to obtain simulated numbers of deaths. The result of this bootstrapping procedure is a table of dimension (ω+1) times (T+1) with simulated age- and time-specific deaths. Repeat this simulation step of the residual S_ε times. Estimate the LC model S_ε times.

The result of estimating the LC model S_ε times is a set of S_ε parameter estimates for a_x, b_{1x} ... b_{Nx} and k_{1t} ... k_{Nt}. Compute, for each parameter estimate, empirical standard errors across the S_ε simulations.

5. For each of the S_ε simulations of step 4, extrapolate the time indices S_k times, and predict age-specific mortality S_k times

For the first of the S_ε simulations of step 4, use the time series model identified in step 2 to extrapolate the k_{1t} ... k_{Nt}. Predict age-specific mortality rates. Simulate prediction intervals around mortality rates by drawing the parameters of the time series model from their estimated distributions S_k times. Repeat this step for all S_ε simulations of step 4. The result is a set of $S_\varepsilon.S_k$ simulations for future mortality rates.

6. Collect the results for all simulations, and compute empirical predictive distributions for age-specific mortality and for the life expectancy in future years.

The setup with S_ε simulations for the residual of the Lee Carter model and S_k simulations for the time series extrapolation of the time index allows us to decompose the prediction intervals based on $S_\varepsilon.S_k$ simulations into a contribution from the Lee Carter model and one from the time series extrapolation.

The bootstrapping procedure in step 4 can be done in various ways. LC sampled from the age columns of the matrix of residuals, because their b_x estimates showed modest negative correlations across ages. Koissi et al. (2006) and Brouhns et al. (2005) sampled cell wise for separate ages and years. We experimented with sampling of age columns, of time rows, and of cells for separate ages and years.

As an alternative for the WLS-estimation in step 2, one could also estimate model parameters by Maximum Likelihood, based on a Poisson model for the numbers of deaths by age and calendar year; see for instance Koissi et al. (2006) and Brouhns et al. (2005). Given age and calendar year, the Poisson parameters are the death rate and the exposure time (or the product of these two). Both the death rate and the exposure time depend on the LC model. This results in an expression for the log-likelihood, the complexity of which increases with the number of components in the LC model. Koissi et al. (2006) and Brouhns et al. (2005) assume fixed exposure times (given age and calendar year), and only one component. We have used up to five components, and assumed exposure times that depend on the LC model. Although the log-likelihood still would be tractable, we prefer WLS-estimation, since it gives a direct means of comparing results for two situations. When death rates are interpreted as given and Poisson variability is ignored, one selects WLS-weights equal to one. When death rates are interpreted as estimates, one selects WLS-weights equal to the inverse of the variances. Note that the WLS-estimation in step 2 can be considered a first-degree approximation of Maximum Likelihood estimation, in which each death rate equals the expected number of deaths per person and per unit of time in a Poisson model (Van Imhoff 1991).

3. Illustration Norway 1900-2050

We have analysed mortality for Norwegian men and women separately. We have used population numbers by age and sex, and annual death counts in one-year age intervals since 1900. Longer time series for Norwegian mortality do exist, but the quality of the data

before 1900 is much less than that in the 20th century. Moreover, it is questionable whether the same model can be used to describe mortality patterns over such a long period. Lee (2000) questions this for the case of the USA on substantive grounds. Empirically we found for the case of Norway that our preferred model (see below) fits the data less well during the first half of the 20th century than during the second half.

3.1 Estimation of the death rate and exposure time assuming a piece wise constant intensity for mortality and a uniform distribution for immigration

Notation:

$L(x,t)$	number of persons at 31 December of year t, who are aged x, $x \geq 0$. Age x is measured in completed years. In other words, $L(0,t)$ represents the number of persons who were born during year t.
$D(x,t)$	number of persons who die during year t. Age x (in completed years) is the age at 31 December of year t.
$m(x,t)$	death rate for persons who die during year t and who would have been x years old at 31 December of year t.
$I(x,t)$	net immigration during year t of persons aged x at 31 December of year t.
$E(x,t)$	exposure time during year t of persons aged x at 31 December of year t.
$B(t)$	number of live births during year t.

The bookkeeping equations are

$$L(x,t) = L(x\text{-}1,t\text{-}1) - D(x,t) + I(x,t),\ x > 0, \text{ and} \tag{4}$$

$$L(0,t) = B(t) - D(0,t) + I(0,t). \tag{5}$$

The purpose is to compute the death rate $m(x,t)$ and the exposure time $E(x,t)$ for all x and t. The death rate is the dependent variable of the LC model. The exposure time is required in the Poisson bootstrap of Section 3.2. We distinguish two cases: $x > 0$ and $x = 0$.

A. $x > 0$

We assume that the force of mortality is constant during year t. In addition, we assume that net immigration is uniformly distributed over year t. In that case we can write (Van Imhoff 1990)

$$L(x,t) = L(x\text{-}1,t\text{-}1).\exp(-m(x,t)) + \{1 - \exp(-m(x,t))\}.I(x,t)/m(x,t) \tag{6}$$

This is a non linear equation in $m(x,t)$. We had data for $L(x,t)$ and $D(x,t)$ for all $x > 0$ and all t from 1900 to 2004. First, we computed net immigration from expression (4). Next, we used Newton Raphson iteration to compute $m(x,t)$ from expression (6). The starting value for the iterations was the actuarial death rate $2D(x,t)/[L(x\text{-}1,t\text{-}1) + L(x,t)]$.

Given the rate $m(x,t)$ and deaths $D(x,t)$, exposure time $E(x,t)$ follows from $E(x,t) = D(x,t)/m(x,t)$.

B. $x = 0$

Using the same assumptions as under A, we have (Van Imhoff 1990)

(7) $L(0,t) = \exp(1 - m(0,t)).\{B(t) + I(0,t)\}/m(0,t).$

To compute $m(0,t)$, we have used a similar procedure. First, expression (2) gave the sum $B(t)+I(0,t)$. Next, Newton Raphson iteration of expression (7) resulted in $m(0,t)$. Finally, we found $E(0,t)$ as $D(0,t)/m(0,t)$.

A death rate computed this way can be interpreted as the Maximum Likelihood estimate of the underlying piecewise constant force of mortality. The asymptotic variance is estimated as $m^2(x,t)/D(x,t)$ (Hoem and Funck Jensen 1982). By the Delta method we find that the variance of $\ln(m(x,t))$ is approximately equal to $1/D(x,t)$. Figure 1 gives the age-specific death rates for men and women at ten-year intervals.

[Figure 1. Death rates by age and sex, 1900, 1910, 1920, …, 2000]

3.2 Estimation of the LC model with WLS

We modelled the death rates $m(x,t)$ computed in Section 3.1 as a Lee-Carter model with N components

(3) $\ln(m_{x,t}) = a_x + b_{1x}k_{1t} + b_{2x}k_{2t} + \ldots + b_{Nx}k_{Nt} + \varepsilon_{x,t}.$

Estimation of the parameters of model (3) by means of Least Squares or Singular Value Decomposition assumes homoscedastic residuals $\varepsilon_{x,t}$. But the variance of $\ln(m_{x,t})$ is not constant across age or time. Therefore we applied Weighted Least Squares estimation with weights $D(x,t)$ (Wilmoth 1993). Thus our estimation procedure gives much weight to ages and periods in which there are many deaths. This is statistically justified by the fact that such weights are approximately equal to estimates of $1/\mathrm{Var}[\ln(m_{x,t})]$. An additional practical advantage is that this procedure automatically omits age/time combinations for which there are no deaths.

For N components, the first-order conditions for the minimization problem result in a non-linear system of $2N+1$ equations in the $2N+1$ parameters $a_x, b_{1x}, k_{1t}, b_{2x}, k_{2t}, ..., b_{Nx}, k_{Nt}$. We used the SVD estimates for the parameters as starting values in the iterative algorithm to find the WLS estimates. In each iteration, we checked the normalization constraints $\sum a_x = 0$ and $\sum b_{i,x} = 1$ (i=1,2,…,N) and adjusted parameter values proportionally.

In our application to Norwegian mortality data, two components were sufficient to obtain homoscedastic residuals, both for men and women. The model with two components explains 95 per cent of the variance in the empirical death rates for men, and 97 per cent of the variance for women. By way of comparison, both SVD estimation and Ordinary Least Squares estimation required $N = 5$ components to obtain homoscedastic residuals.

Figures 2-6 give parameter estimates for men and women. The first time index in Figure 3 shows that the mortality decline for Norwegian women since 1900 was stronger than that for men. Infectious diseases, which hit children and young adults in the beginning of the period, became less important. The age pattern in Figure 4 illustrates this also. Between the mid-1950s and mid- 1980s, the mortality decline for men stagnated almost entirely. However, the decline continued in recent years, and it was stronger than that of women. Figures 5 and 6 illustrate that the general development has to be modified. Since the mid-1960s, mortality for men between 12 and 47 years of age, and for women under 65 declined more slowly than the patterns in Figures 3 and 4 suggest, because the products $b_2(x).k_2(t)$ are positive for these groups in this period. Coronary heart diseases were an important death cause in this period.

[Figure 2 Estimates of $a(x)$]
[Figure 3 Estimates of $k_1(t)$]
[Figure 4 Estimates of $b_1(x)$]
[Figure 5 Estimates of $k_2(t)$]
[Figure 6 Estimates of $b_2(x)$]

3.3 Time series modelling of the time factors k_{1t} and k_{2t} and point predictions for mortality

First, we used time series models to compute point predictions for the time factors k_{1t} and k_{2t} obtained by the method of Section 3.2. These point predictions are described in this section. This section gives the point predictions for the death rates and the life expectancies. In Sections 3.4-3.6 we shall turn to prediction intervals.

For men, we obtained a good fit for the time factors using the following AR(1)/Random Walk model with exponential/linear time trend (standard errors in parentheses):

$$(8) \qquad k_1(t) = 0.98k_1(t\text{-}1) - \underset{(0.00021)}{0.00093}(t\text{-}1899)^{1.8} + \delta_1(t)$$

$$k_2(t) = k_2(t\text{-}1) - \underset{(0.00041)}{0.00078}(t\text{-}1899) + \delta_2(t)$$

Model (8) explains 99 per cent of the variance in $k_1(t)$ over the years 1900-2004, and 87 per cent of the variance in $k_2(t)$ over the same period. The coefficient 0.98 and the power 1.8 in the expression for $k_1(t)$ are not estimated, but chosen after some experimentation. A model with coefficient equal to 1 and power equal to 2 showed the same good fit to the time-series of $k_1(t)$ as model (8), but it predicted an unrealistic age pattern with increasing death rates for men around 25 years of age after 2020. The reason is the increase in (positive) values for $k_2(t)$ (Figure 5) combined with positive $b_2(x)$-values for men aged 15-45 (Figure 6), which increase death risks for this age group. This effect disappears with a coefficient slightly lower than 1. The power of 1.8 was necessary to avoid a situation in which death rates for men are lower than those for women after 2020. $\delta_1(t)$ and $\delta_2(t)$ are normally distributed independent residuals.

For women, we found the following model

$$(9) \qquad k_1(t) = 0.99k_1(t\text{-}1) - \underset{(0.00075)}{0.0040}(t\text{-}1899)^{1.5} + \delta_1(t)$$

$$k_2(t) = 0.95k_2(t\text{-}1) + \underset{(0.00532)}{0.00336}(t\text{-}1899) + \delta_2(t)$$

Model (9) explains 99.5 per cent of the variance in $k_1(t)$ for women between 1900 and 2004, and 92.6 per cent of the variance in $k_2(t)$. The estimate 0.00336 is not significantly different from zero at the 10 per cent level, and hence we could have omitted the linear time trend from the equation for $k_2(t)$. This would have resulted in a fall in $k_2(t)$ by five per cent per year. But Figure 5 shows an increase since the mid-1960s. There is no reason why this trend suddenly should be reversed.

We have used models (8) and (9) to predict the time factors to 2050. Together with the age profiles $a(x)$, $b_1(x)$, and $b_2(x)$ they result in predicted death rates for men and women. For

future years, we used weighted averages of the age profiles $b_1(x)$, and $b_2(x)$ for men and women in order to avoid divergence between male and female mortality. When we used the estimated age profiles for each sex separately, we noted a much stronger fall in women's mortality than that of men in the future. This is not realistic, given the ongoing convergence between life expectancies of men and women in Norway. To solve this problem, we assumed that the future $b_1(x)$ for men would consist of 20 per cent of the historical $b_1(x)$ profile for men and 80 per cent of the historical profile for women (Figure 4). The future $b_1(x)$ profile for women consists of 90 per cent of their historical profile, and 10 per cent of that of men. For the future $b_2(x)$ profile of men, we used again weights of 20 per cent and 80 per cent. For women, we used 95 per cent and 5 per cent. These weights were obtained after some experimentation, in which the aim was to avoid divergence between male and female mortality.[1] Finally, we smoothed the age profiles thus obtained before applying them in the extrapolations, in order to avoid irregularities in predicted death rates.[2] Figure 7 summarizes the predicted death rates in the form of life expectancies at birth for the years to 2050. The underlying death rates for some selected years are illustrated in Figure 8.

[Figure 7. Life expectancy at birth, 1900-2050]
[Figure 8. Death rates for men and women, 2000, 2010, …, 2050]

Figure 7 shows a slight reduction of the gap between male and female mortality. In 2004, the life expectancy difference was 4.9 years; in 2050 the model predicts a difference of 3.9 years. Death rates fall for both sexes (Figure 8), due to the reduction in the time factors.

3.4 Poisson Bootstrapping

The LC model gives rise to a residual $\varepsilon_{x,t}$ for each combination of age and calendar year. Denote by $\varepsilon^s_{x,t}$ the sth bootstrap sample from the table $[\varepsilon_{x,t}]$, $x = 0,1, \ldots 100$, $t = 1900, 1901, \ldots, 2004$. We applied non-parametric bootstrapping, in which we obtained replications by sampling $\varepsilon^s_{x,t}$–values with replacement from the initial data $\varepsilon_{x,t}$ (Efron and Tibshirani 1998). This way we simulated S_ε tables with residuals. Next we computed simulated deaths as $D^s(x,t) = m(x,t)\exp(\varepsilon^s_{x,t})E(x,t)$, where both the death rate $m(x,t)$ and the exposure time $E(x,t)$ are determined by the parameter estimates $a_x, b_{1x}, k_{1t}, b_{2x}, k_{2t}$.[3] This resulted in S_ε tables with

[1] Nan and Lee (2005) also report divergence problems when LC parameters are extrapolated separately for sub populations. They propose using a common time factor, or a common age profile, or both (in addition to factors and profiles that are specific for each sub population).
[2] De Jong and Tickle (2006) include a smoother directly in the estimation procedure of their LC model. We have not used this approach, because a model with smoothed age profiles deteriorates the fit in terms of the life expectancy for the early decades of the 20th century in our data set.
[3] Since the number of deaths is unknown here, we computed exposure time as

simulated deaths $D^s(x,t)$, $s = 1, 2, \ldots, S$. For each of the S_ε tables, we estimated LC parameters using the methods of Section 3.1 (replacing $D(x,t)$ by $D^s(x,t)$, but keeping the observed values of $L(x,t)$) and Section 3.2 (with N equal to 2). This resulted in S_ε sets of parameter estimates $a_x^s, b_{1x}^s, k_{1t}^s, b_{2x}^s, k_{2t}^s$. We used those parameter estimates to compute S_ε life tables for men and women for each of the years 1900-2004. Figure 9 plots in-sample predictions for the life expectancy at birth for men and women based on the empirical death rates, in the form of 80 per cent prediction intervals based on $S_\varepsilon = 1000$ simulations.[4] The intervals are the result of residual variance and uncertain parameter estimates in the LC model. Since the predictions apply to the period 1900-2004, extrapolation of the time factors was not necessary, and hence the time series models (8) and (9) do not contribute to the intervals for the in-sample predictions. For most of the period, the intervals are about one year wide. This agrees with the findings by Koissi et al. (2006) who investigated uncertainty in life expectancy resulting from estimation uncertainty of the LC parameters. The wide intervals during the first half of the 20th century indicate that the model fit was bad during that period. There are two explanations, both caused by the Weighted Least Squares algorithm. First, the algorithm gives more weight to the second half of the century (when annual numbers of deaths increased from 30000 to 45000) than to the first half (when the numbers fluctuated between 30000 and 35000). Second, numbers of deaths became more concentrated at higher ages during the century, because child mortality decreased considerably. Thus, the algorithm gives less weight to children and young adults. At the same time the life expectancy at birth is strongly determined by mortality at young ages.

[Figure 9. Life expectancy at birth, men and women, 80 % prediction intervals]

3.5 Prediction intervals for age-specific mortality

For each of the S_ε sets of time factors obtained by the method of the previous section, we used the time series models given in expressions (8) and (9) to compute prediction intervals around future values of k_{1t}^s and k_{2t}^s. We used simulation to obtain prediction intervals for k_{1t}^s and k_{2t}^s. Residuals $\delta_i(t)$ and coefficients for the time trends were drawn from their respective estimated normal distributions. For each of the S_ε sets of time factors

---

$E(x,t) = \{L(x\text{-}1,t\text{-}1) - I(x,t)/m(x,t)(1 - \exp(\text{-}m(x,t))\}/m(x,t) + I(x,t)/m(x,t)$.

[4] The constant intensity assumption and the period-cohort observation intervals for the death rates result in the following expressions for the life table variables. Let $m(x)$, $q(x)$, and $E(x)$, represent the death rate, the death probability, and the exposure time between ages x and $x+1$ in the life table. Let $l(x)$ and $e(x)$ represent the number of persons alive and the remaining life expectancy at age x, with $l(0)=1$. Then $q(x)=1-\exp(\text{-}m(x))$ for $x>0$, $q(0)=1-\exp(\text{-}m(0)/2)$, $l(x)=l(x\text{-}1)(1\text{-}q(x))$, $E(x)=l(x)q(x)/m(x)$, and $e(x)=\sum E(y)/l(x)$, where the sum runs from $y=x$ to $y=\omega$.

k_{1t}^s and k_{2t}^s (t = 1900-2004), we simulated S_k future time series of k_{1t}^s and k_{2t}^s (t = 2005-2050). The result was $S_\varepsilon.S_k$ sets of predictions for k_{1t}^s and k_{2t}^s, which, by the LC model, resulted in $S_\varepsilon.S_k$ sets of death rates predictions.

Figure 10 gives 80 per cent prediction intervals for male and female mortality in 2050, while 80 per cent prediction intervals around the life expectancy at birth are illustrated in Figure 11. Both graphs are based on S_ε = 100 and S_k =300 simulations; in all 30,000 simulations.

[Figure 10. 80 per cent prediction intervals for age-specific mortality of men and women in 2050]
[Figure 11. 80 per cent prediction intervals for the life expectancies of men and women, 2005-2050]

The 80 per cent prediction intervals for the life expectancy in 2050 are 5.6 (men) and 5.2 (women) years wide. The widths of the corresponding 95 per cent intervals are 9.5 and 8.4 years, respectively (figures not shown here). Intervals presented by Tuljapurkar et al (2000) for the G7 countries (Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States) are much smaller than ours are. Their 95 per cent intervals of combined-sex life expectancy at birth in 2050 range from a minimum of 3.3 years for Canada to a maximum of 8.9 years for the UK[5]. These intervals result from a Lee-Carter model for age-specific mortality for the two sexes combined, assuming a random walk with drift model for the (single) time index. The authors used an abridged life table with five-year age classes up to 80-84. Ages 85 years and higher were lumped into one age class (except for Japan). The age and sex aggregation, which reduces random fluctuations, may have caused these relatively narrow intervals[6].

3.6 Decomposing prediction intervals

We have decomposed the prediction intervals in Figures 10 and 11 into two parts. One part of each interval is due to the fit of the Lee Carter model, and the other part is due to the time series extrapolation of the time indices. We have assumed that these two parts are

[5] We multiplied the 90 per cent interval bounds reported by Tuljapurkar et al. (2000) by 1.96/1.645=1.19, assuming a normal distribution.
[6] The impact of age grouping can be illustrated as follows. For the US, Tuljapurkar et al (2000) find a 95 per cent interval that is 5.2 years wide in 2050. Their model (a Lee-Carter model with a single time index modelled as a random walk with drift) applied to *unabridged* combined-sex life tables gives a corresponding interval that is 8.3 years wide (interpolated between 2035 and 2065 in Figure 2.5 of Lee and Tuljapurkar, 2001).

additive and independent. The sum of the two parts is not exactly equal to width of the interval obtained in the previous section; the difference is due to an interaction term. This interaction term is caused by two facts. First, the additivity assumption is not correct, because of the non-linear nature of the prediction interval and the parts thereof. Second, the two parts are probably not independent. In other words, a bad fit of the Lee-Carter model will probably go together with an imprecise forecast of the time indices.

We selected $S_\varepsilon = 100$ and $S_k = 1$ to find the part of the prediction interval caused by the fit of the Lee-Carter model. Similarly, by choosing $S_\varepsilon = 1$ and $S_k = 300$ we found the part caused by the time series extrapolation of the time indices. Figure 12 compares these two parts with the results of the 30,000 simulations ($S_\varepsilon = 100$ and $S_k = 300$) of Section 3.5.

[Figure 12. Width of 80 per cent prediction interval for male life expectancy, and constituent components]

In-sample predictions during the years 1900-2004 do not include any extrapolation of the time indices. Hence the only component that constitutes the prediction interval is due to the fit of the Lee Carter model. Beginning in 2005, the time indices contribute increasingly larger parts to the prediction intervals. A small interaction term is negative in the first few years, but beginning in 2028 it is positive. This shows that simulations with large $\varepsilon_{x,t}$-residuals and large standard errors for the estimated $b_{i,x}$ and $k_{i,t}$-parameters are associated with rather uncertain long-term predictions for the $k_{i,t}$-parameters. Figure 13 shows that in the short run, the fit of the LC model explains a considerable share of the width of the prediction interval. But this share decreases rapidly, and after 15 years into the future it stabilizes at about 20 per cent.

[Figure 13. Relative contribution of LC uncertainty and time series uncertainty to width of 80% prediction interval for life expectancy of men
Note: shares do not sum to unity because of positive (stacked column < 1) or negative (stacked column >1) interaction]

For women, the results are qualitatively the same, but the share due to time series uncertainty stabilizes at a lower level, around 70 per cent. At the same time, the (positive) share due to interaction is roughly of the same magnitude as the share due to LC uncertainty.

The conclusion is that long-run prediction intervals are too narrow, in case one only accounts for uncertainty caused by the extrapolation model of the time indices. When uncertainty due to the fit of the LC model is taken into account as well, prediction intervals become wider. In the Norwegian data set we find that long-run prediction intervals for the life expectancy are

too narrow by 100*20/80 = 25 per cent for men, and 100*30/70 = 40 per cent for women, when LC uncertainty is ignored.

Figure 14 shows similar shares as Figure 13 does, but here the variable of interest is the predicted death rate for men in 2050. We note that again, by and large, the time index contributes about 80 per cent to the prediction interval when both uncertainty sources are taken into account, but there are two exceptions. For men aged between 15 and 50, there is some positive interaction, which reduces the share of the time index. At ages 95 or higher, LC uncertainty becomes more important. This is due to the fact that the fit of the LC model is rather bad at those extreme ages. At age 95, the 80 per cent prediction intervals for the death rate is 0.06 wide, increasing to a width of 0.20 at age 100 (with point predictions increasing from 0.29 to 0.44 over this age interval). Between ages 70 and 90, where most deaths take place, the interval width increases from 0.008 to 0.05.

[Figure 14. Relative contribution of LC uncertainty and time series uncertainty to width of 80% prediction interval for death rates of men in 2050
Note: shares do not sum to unity because of positive (stacked column < 1) or negative (stacked column >1) interaction]

3.7 The effect of Poisson variance

Each empirical death rate (see Figure 1) is the estimate of the parameter in a simple Poisson model. Hence one has to take estimation uncertainty into account. We used Weighted Least Squares regression of model (3), with weights equal to the inverse value of the estimation variances to accomplish this. This method gives little weight to years or ages in which the empirical death rate has a large standard error.

In order to investigate the effect of Poisson variance, we re-estimated model (3) with N=2 using Ordinary Least Squares. Thus we assumed homoscedastic residuals. We used models (8) and (9) for time index extrapolations. In-sample predictions based on 100*300 = 30,000 simulations, which are influenced by the fit of the Lee Carter model and not by the extrapolation of the time indices, showed rather strong effects for men. Compared to the prediction intervals based on WLS in Figure 9, the OLS-based life expectancy intervals for men that ignore Poisson variance were roughly 2-4 times as large. For women, OLS-based life expectancy intervals had about the same width for the years 1900-1950, and they were roughly 30 per cent wider for later years (until 2004). Life expectancy forecasts for the years 2005-2050 (in terms of median and mean values across simulations) were up to 2 years lower for women, but up to 7 years lower for men. This is explained by the emphasis that WLS

gives to recent years with sharp increases in the life expectancy, in particular for men. Indeed, the OLS-estimates for the time index $k_1(t)$ for men do not fall as quickly after 1985 as the WLS-estimates do. The prediction intervals for men became excessively wide – for women they were about 30 % wider than corresponding WLS-based intervals.[7]

Thus we conclude that in case one ignores Poisson variance and extracts just a few components, the estimates for the death rates become biased with too much weight on periods/ages with large variance in the rates. The predictions will be wrong accordingly. In the Norwegian data set this was particularly the case for men. We found that WLS requires only two components to be extracted, rather than five or more in the case of OLS. The assumption of homoscedastic residuals applied in OLS is violated, unless one includes many components.

4. Summary

We have analysed Lee-Carter based prediction intervals for age-specific mortality and for the life expectancy of men and women. Mortality predictions based on this method are uncertain for three reasons. The predictions are based on

1. estimates of the parameters of the Lee-Carter-model;
2. extrapolated values of the model's time index;
3. observed death counts and empirical death rates. These are subject to Poisson variability.

We have described a simulation method that decomposes prediction intervals into a contribution due to the first factor and a contribution due to the second factor. We have applied this method to annual mortality data for men and women by one-year age group in Norway during the years 1900-2004. Ignoring factor no. 2 above implied long-run prediction intervals for the life expectancy that were too narrow by 25 per cent for men and by 40 per cent for women.

Ignoring Poisson variability (factor nr. 3 above) led to too low life expectancy forecasts; forecasts for 2050 were two years too low for women and seven years too low for men. In-sample predictions led to excessively wide prediction intervals, in particular for men.

[7] Time indices based on OLS are different from those in Figures 3 and 5, and hence models (8) and (9) are not necessarily appropriate for extrapolation purposes. For reasons of comparison, we have retained them. When we assumed a random walk with drift for the time index in a model with N=1 and estimated by OLS (as many authors have done), we found that in-sample predictions of male life expectancy were structurally too low in recent years.

References

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

Brouhns, N., M. Denuit and I. van Keilegom (2005) Bootstrapping the Poisson log-bilienear model for mortality forecasting. *Scandinavian Actuarial Journal* 3, 212-224.

Carter, L.R. and A. Prskawetz (2001) Examining structural shifts in mortality using the Lee-Carter method, Rostock, MPIDR Working Paper WP-2001-007.
http://www.demogr.mpg.de/Papers/Working/wp-2001-007.pdf

De Jong, P. and L. Tickle (2006) Extending Lee-Carter mortality forecasting. *Mathematical Population Studies* 13(1) 1-18.

Efron, B. and R. Tibshirani (1998) *An introduction to the Bootstrap*. Boca Raton: CRC Press.

Hoem, J.M. and U. Funck Jensen (1982) Multistate life table methodology: A probabilist critique. Pp. 155-264 in K. Land and A. Rogers (eds.), *Multidimensional Mathematical Demography*. New York: Academic Press.

Koissi, M.-C., A. Shapiro and G. Högnäs (2006) Evaluating and extending the Lee-Carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics* 38, 1-20.

Lee, R. (2000) The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal* 4(1), 80-91.

Lee, R. D. and L. R. Carter (1992) Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association* 87(419) 659-675.

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

Lee, R. and S. Tuljapurkar (2001) Population forecasting for fiscal planning: Issues and Innovation. Pp. 7-57 in A. Auerbach and R. Lee (eds.). *Demographic change and fiscal policy*. Cambridge University Press.

Lundström, H. and J. Qvist (2004) Mortality forecasting and trend shifts: An application of the Lee-Carter model to Swedish mortality data. *International Statistical Review* 72(1) 19-36.

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

Preston, S., P. Heuveline and M. Guillot (2001) *Demography: Measuring and Modelling Population Processes.* Oxford: Blackwell Publishers.

Prskawetz, A. and L. R. Carter (2001) Untersuchung struktureller Brüche in der Mortalitätsentwicklung in Österreich unter Benutzung der Lee-Carter-Methode. *Demographische Informationen* 2001, 69-80

Tuljapurkar, S., N. Li and C. Boe (2000) A universal pattern of mortality decline in the G7 countries. *Nature* 405, 15 June 2000, 789-792.

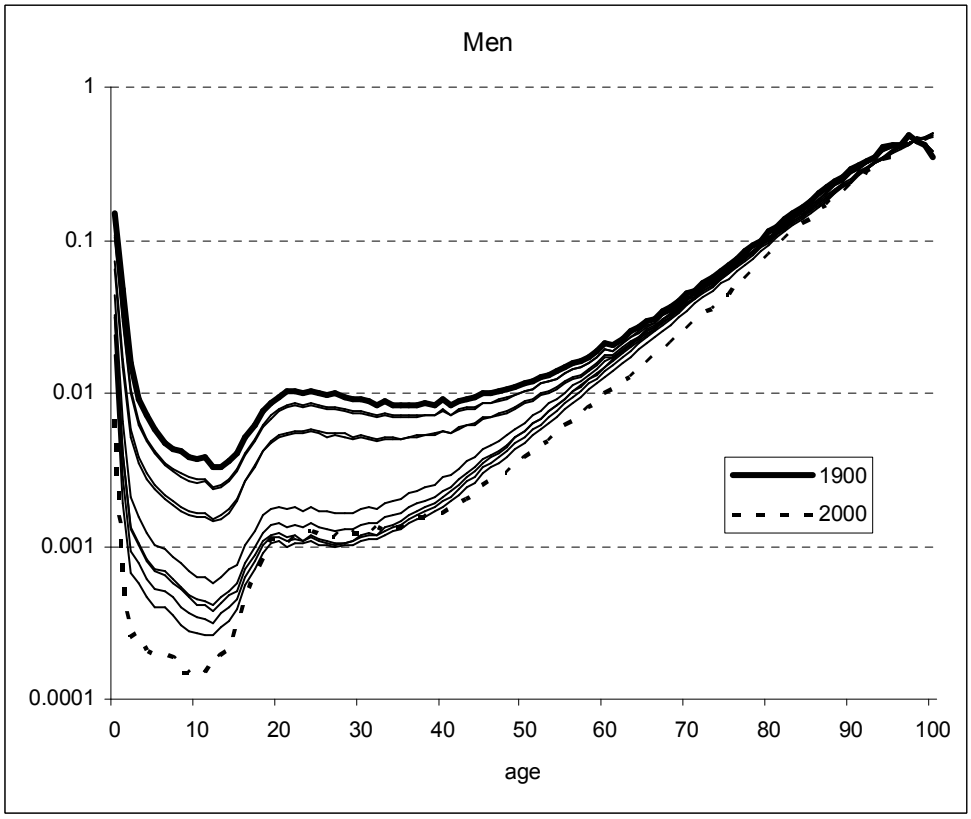
Tuljapurkar, S. (2005) Stochastic forecasts of mortality, population, and pension systems. Pp. 65-77 in N. Keilman (ed.) *Perspectives on mortality forecasting Volume II: Probabilistic Models*. Stockholm: Swedish Social Insurance Agency (Social Insurance Studies no. 2).

Van Imhoff, E. (1990) The exponential multidimensional demographic projection model. *Mathematical Population Studies*, 2(3) 171-182.

Van Imhoff, E. (1991) Profile: A program for estimating the coefficients of demographic age-intensity profiles. NIDI Report nr. 15. The Hague: Netherlands Interdisciplinary Demographic Institute.

Wilmoth, J.R. (1993) Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Technical Report, Berkeley: Department of Demography, University of California at Berkeley.

Figure 1. Death rates by age and sex, 1900, 1910, 1920, …, 2000



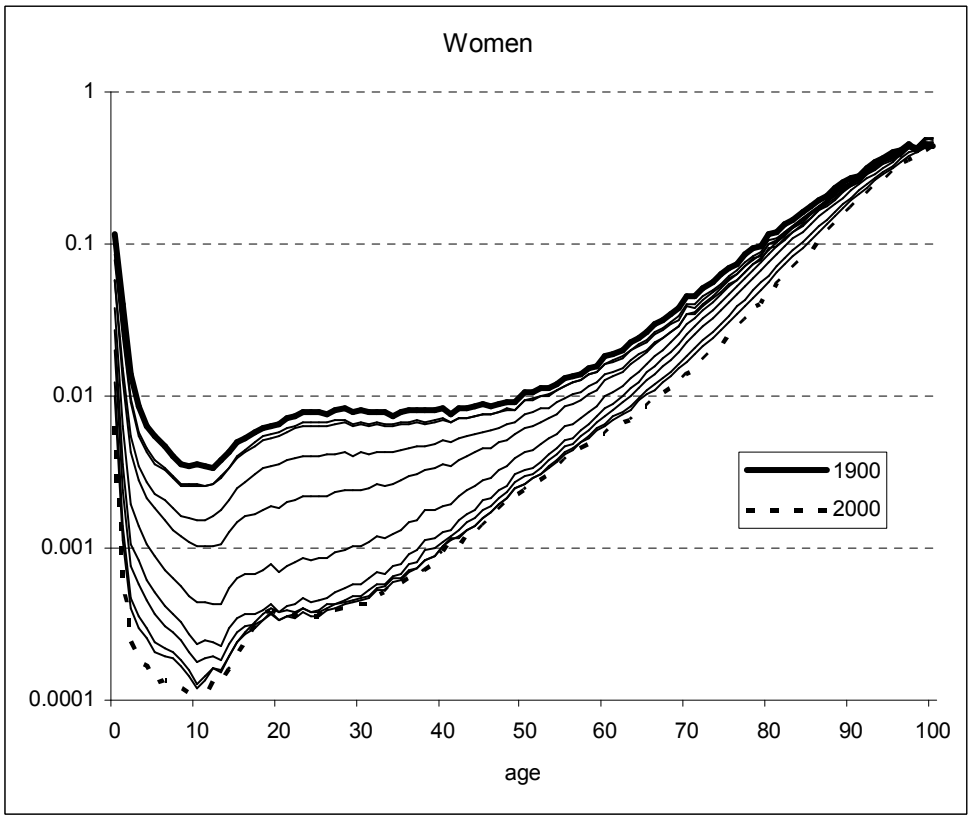


Figure 2 Estimates of $a(x)$

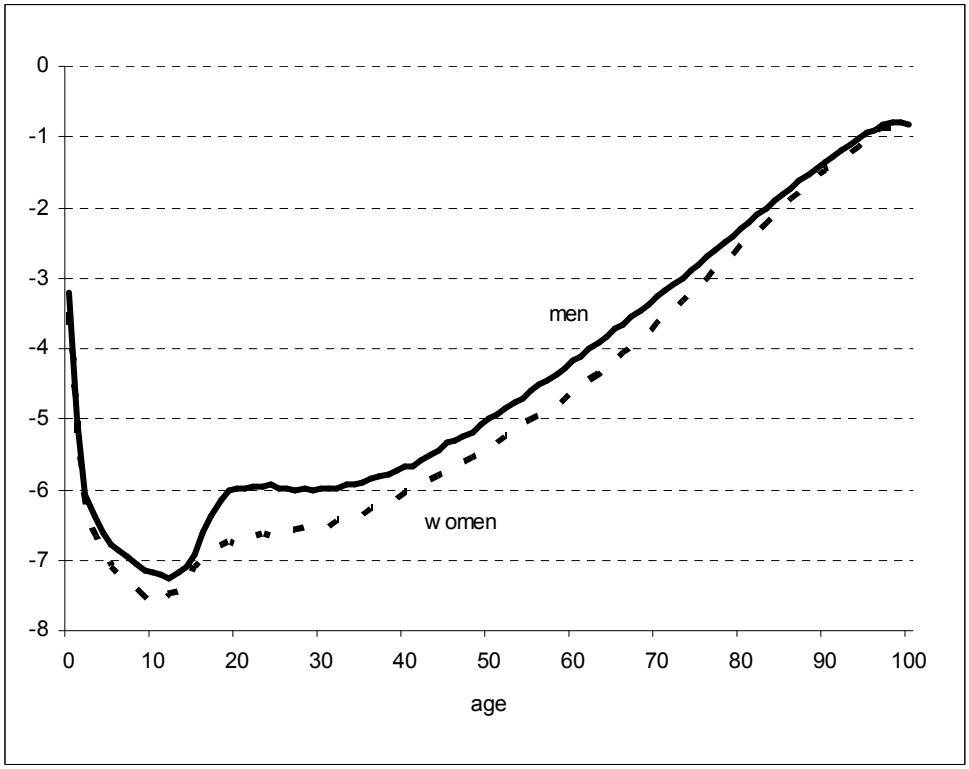


Figure 3 Estimates of $k_1(t)$

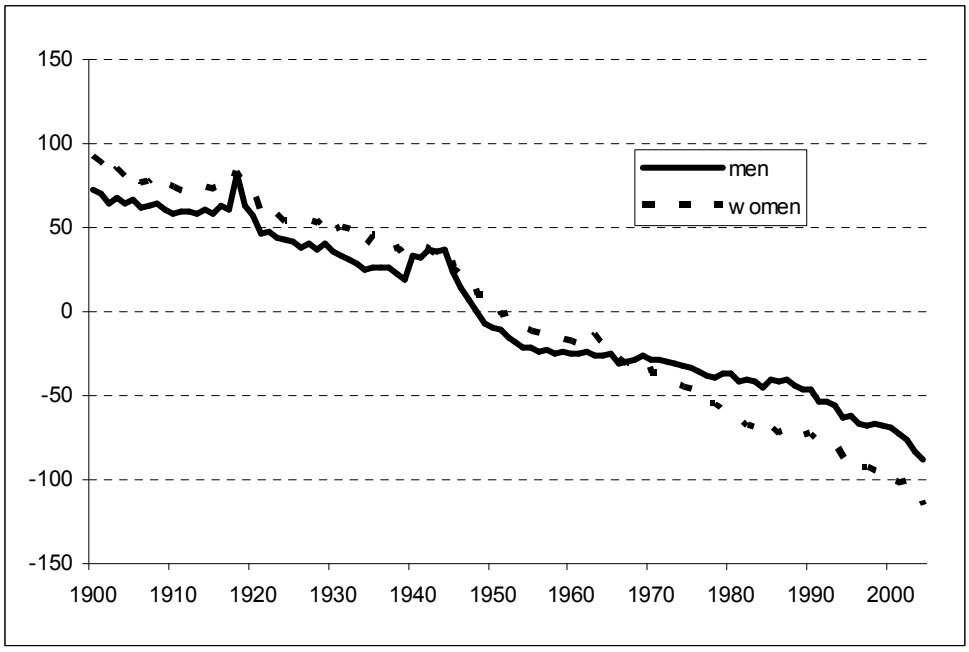


Figure 4 Estimates of $b_1(x)$

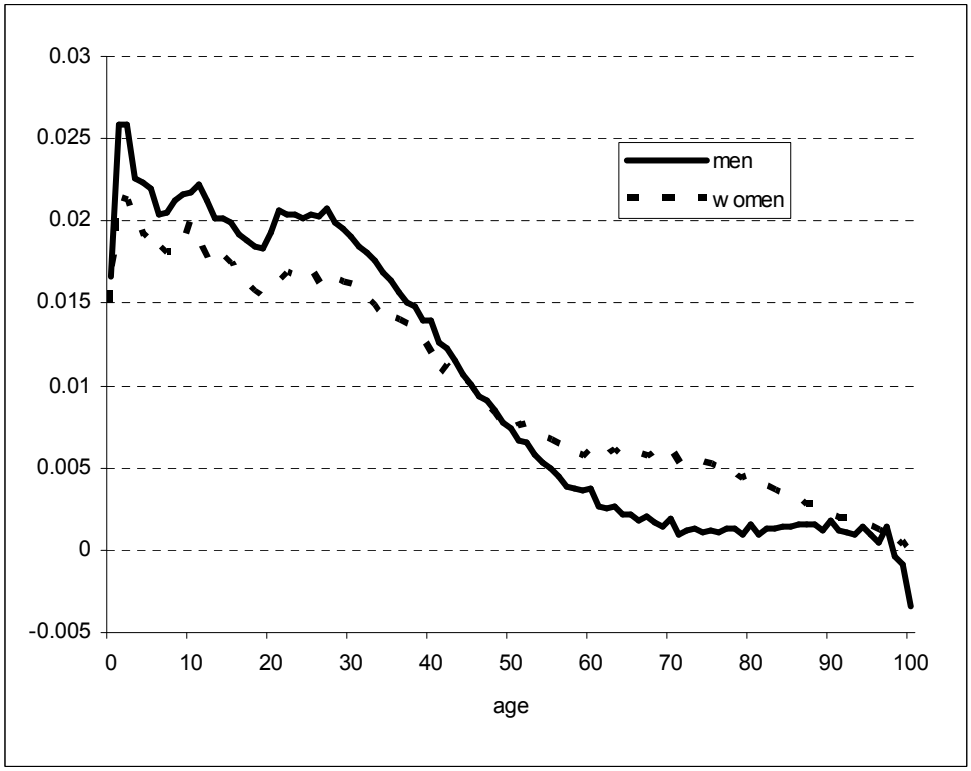


Figure 5 Estimates of $k_2(t)$

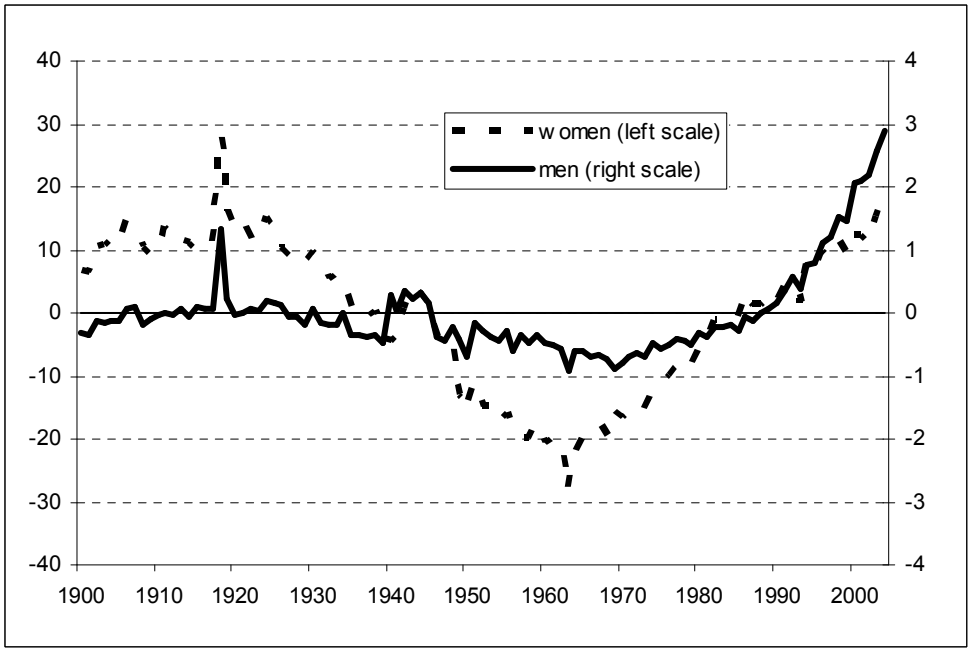


Figure 6 Estimates of $b_2(x)$

Figure 7. Life expectancy at birth, 1900-2050

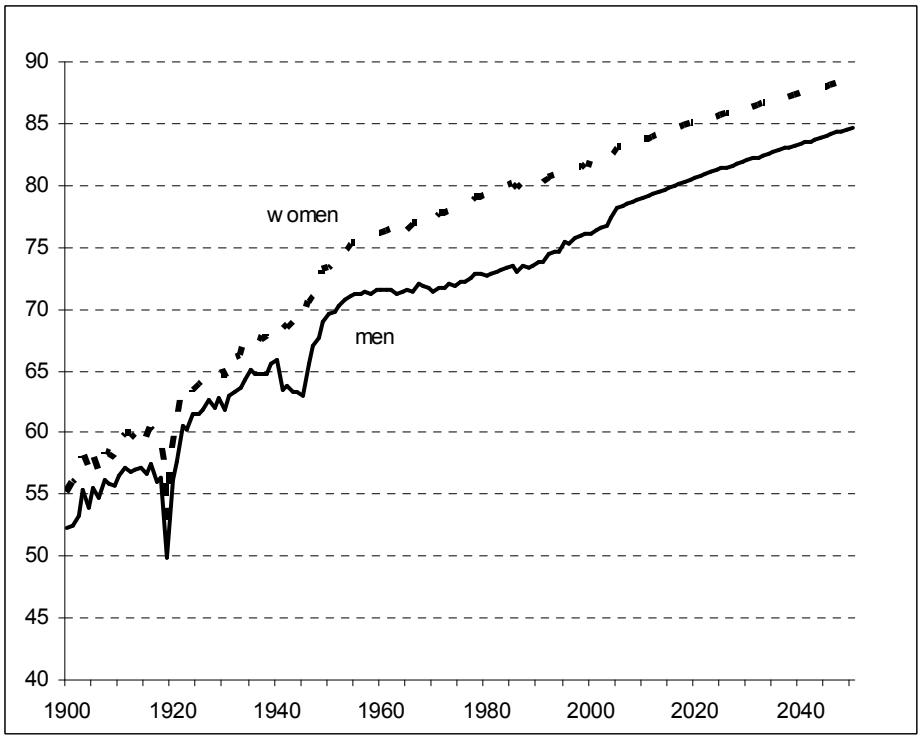
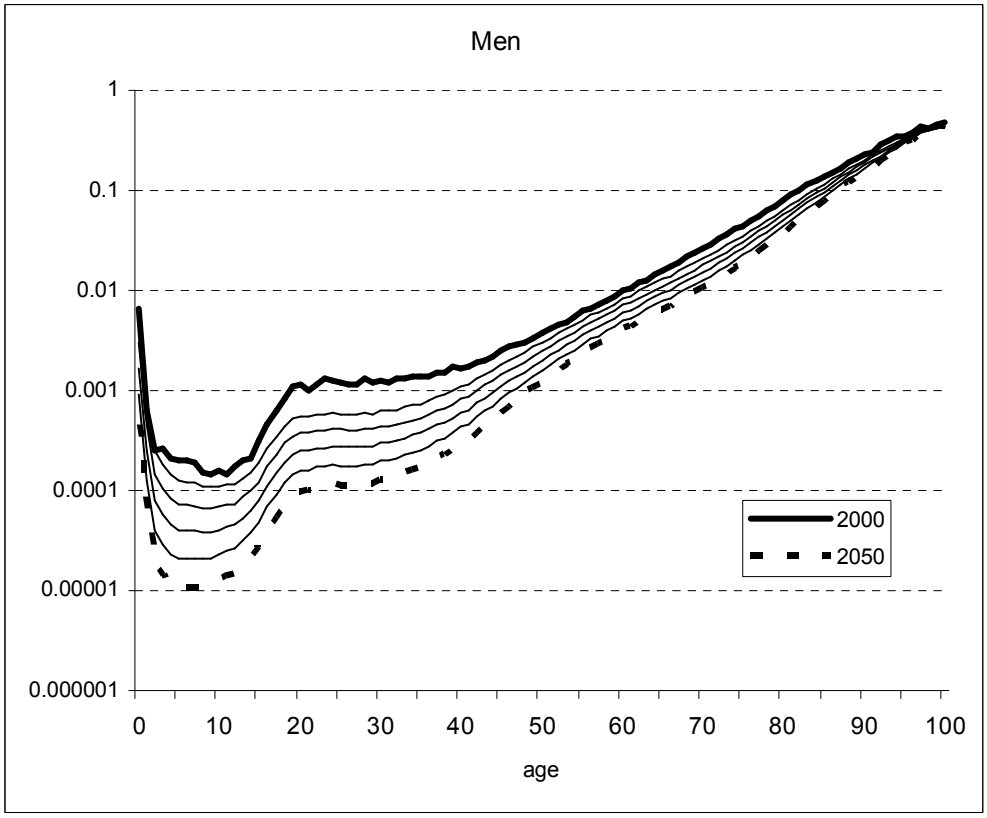


Figure 8. Death rates for men and women, 2000, 2010, …, 2050



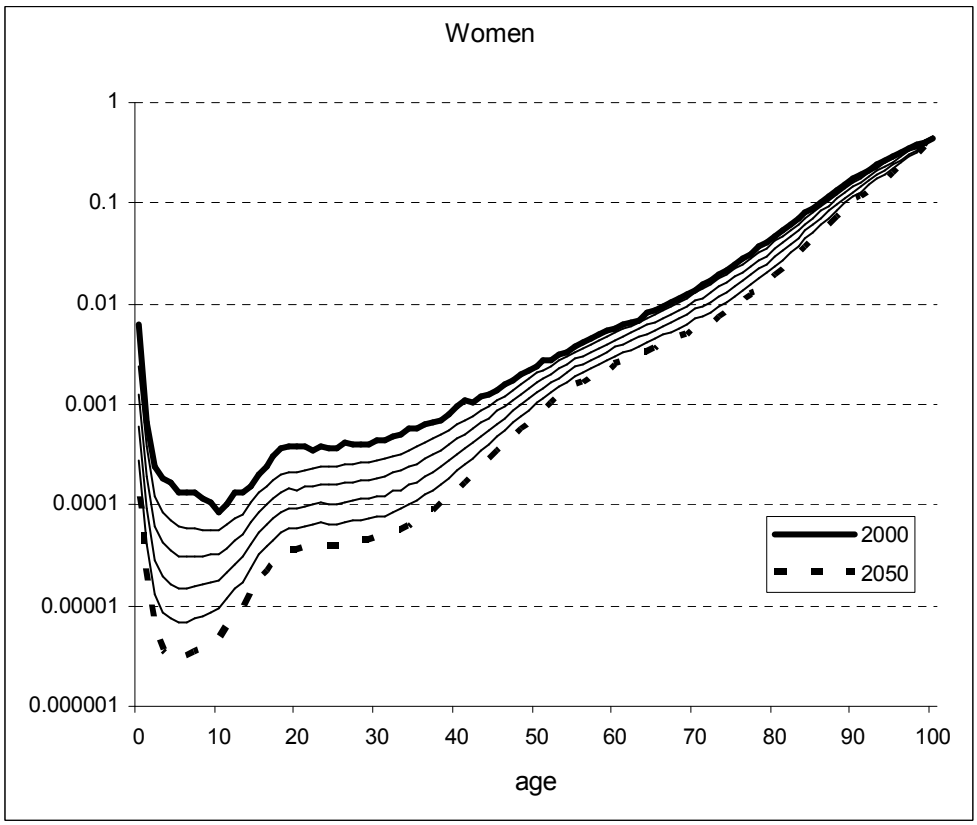


Figure 9. Life expectancy at birth, men and women, 80 % prediction intervals

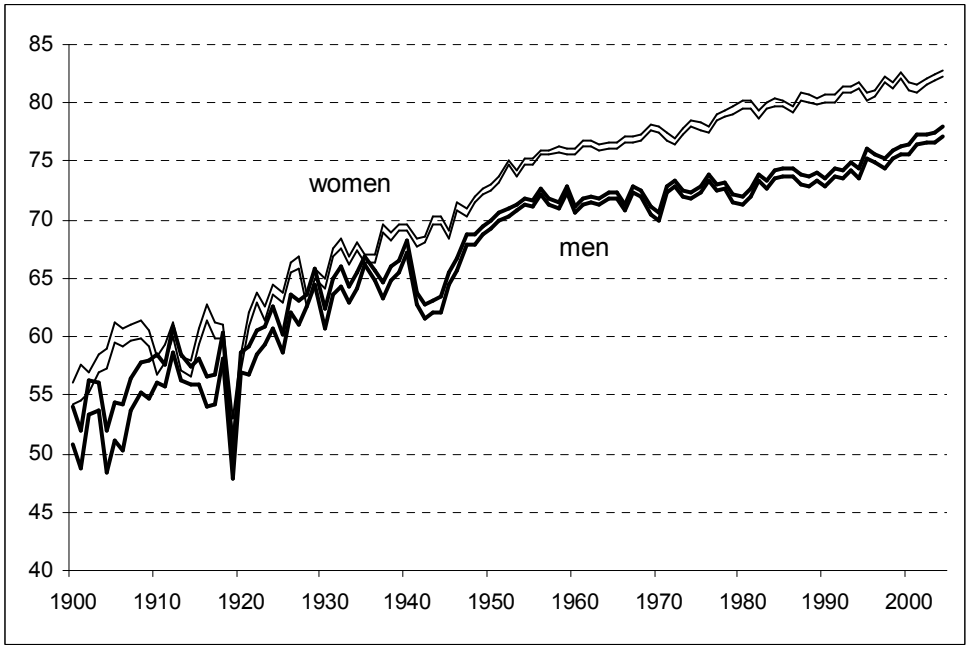


Figure 10. 80 per cent prediction intervals for age-specific mortality of men and women in 2050

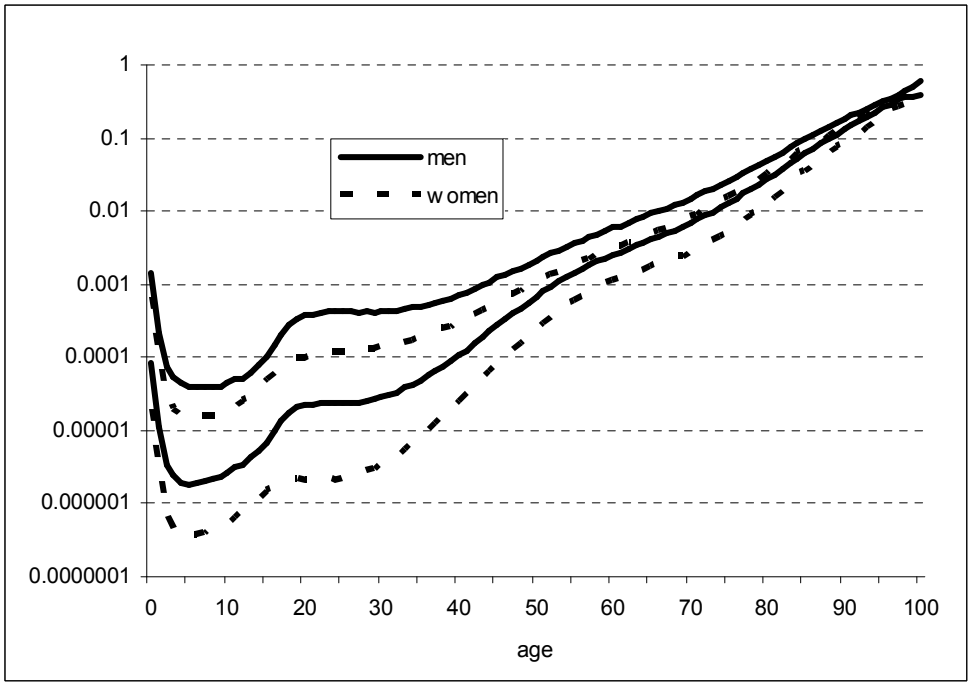


Figure 11. 80 per cent prediction intervals for the life expectancies of men and women, 2005-2050

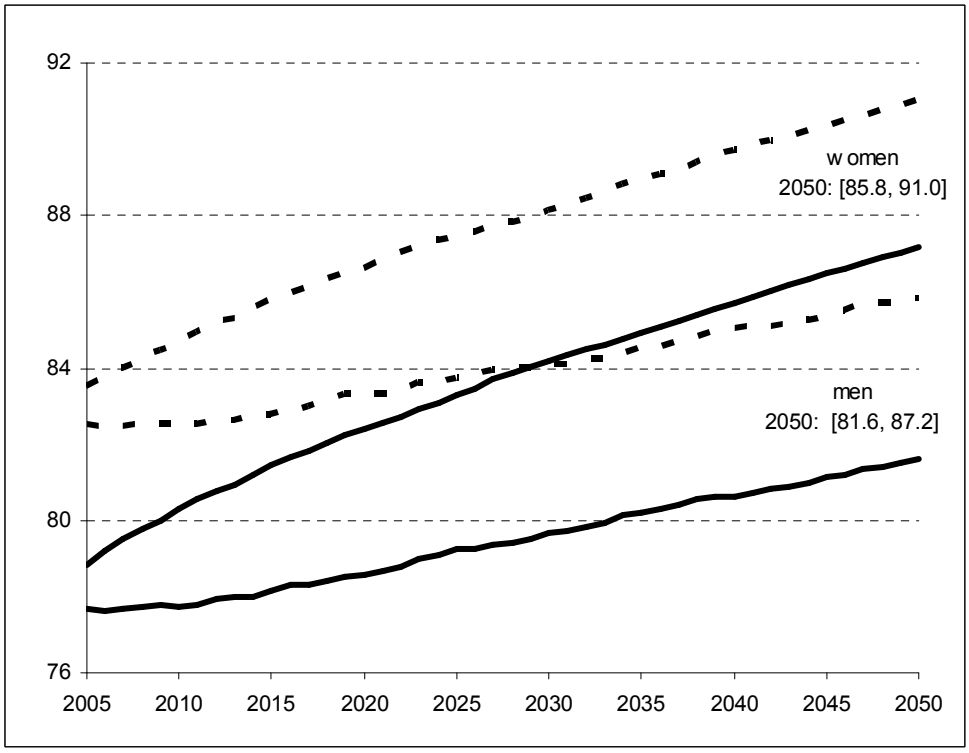


Figure 12. Width of 80 per cent prediction interval for male life expectancy, and its components

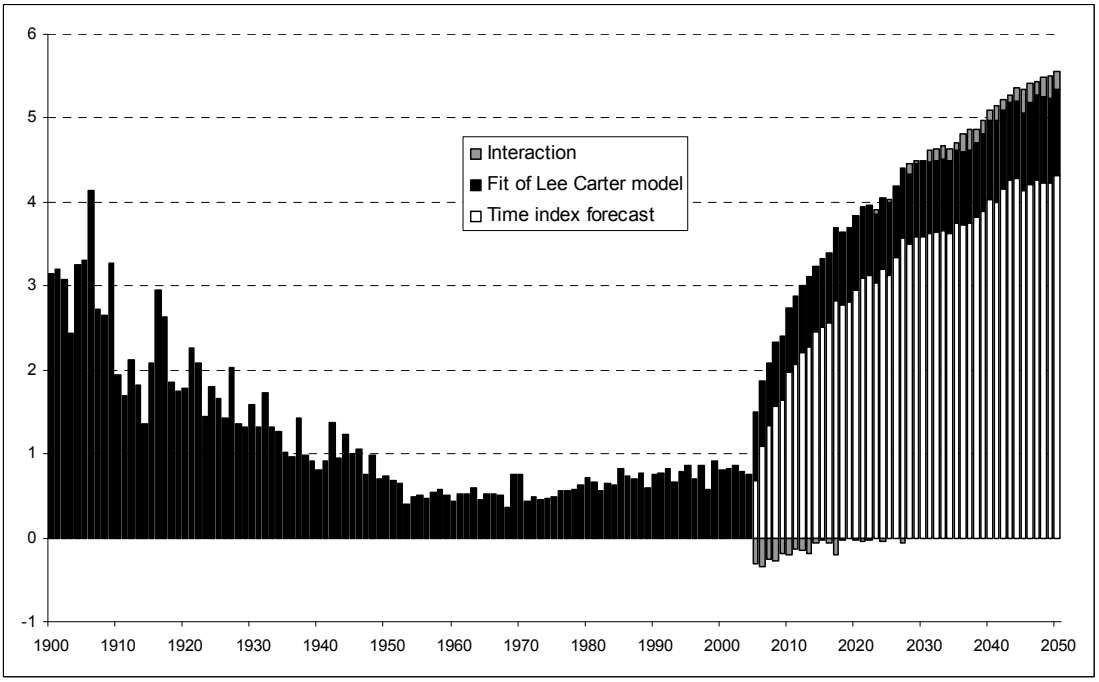


Figure 13. Relative contribution of LC uncertainty and time series uncertainty to width of 80% prediction interval for life expectancy of men
Note: shares do not sum to unity because of positive (stacked column < 1) or negative (stacked column >1) interaction

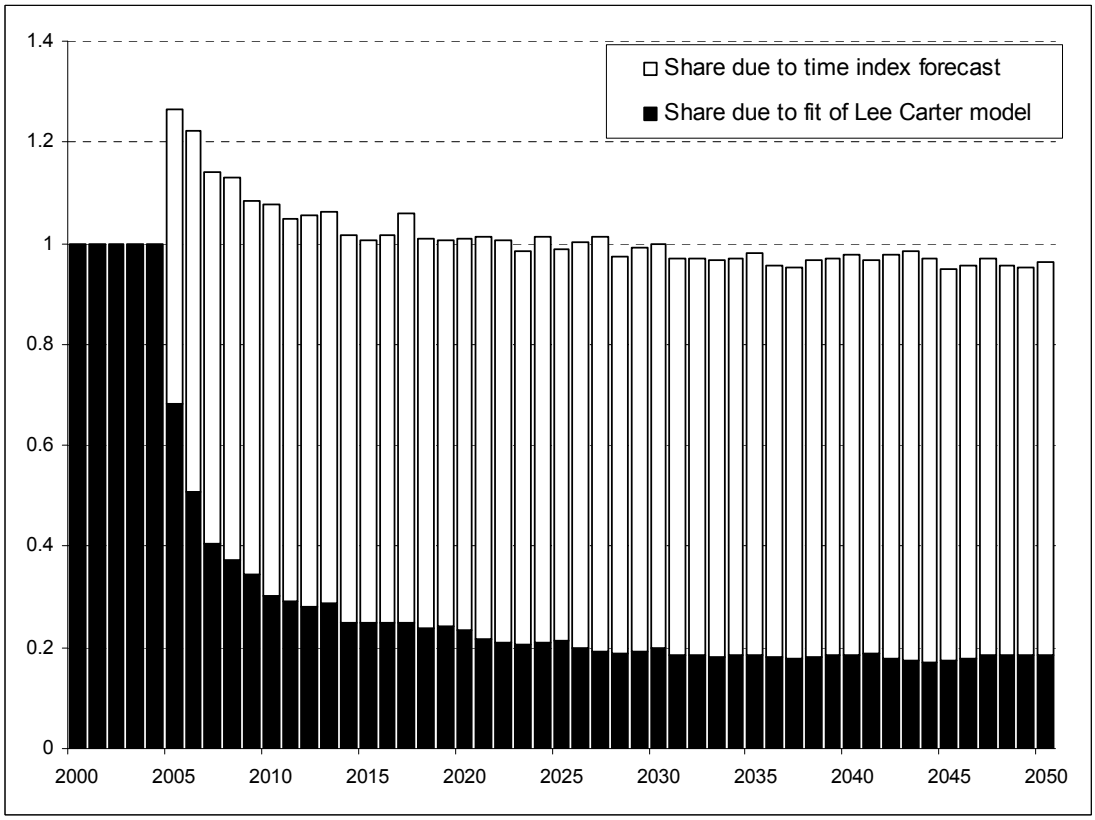


Figure 14. Relative contribution of LC uncertainty and time series uncertainty to width of 80% prediction interval for death rates of men in 2050
Note: shares do not sum to unity because of positive (stacked column < 1) or negative (stacked column >1) interaction