



ELSEVIER

Available online at www.sciencedirect.com

International Journal of Forecasting 24 (2008) 323–342

*international journal
of forecasting*www.elsevier.com/locate/ijforecast

Stochastic population forecasts using functional data models for mortality, fertility and migration

Rob J. Hyndman^{a,*}, Heather Booth^b^a *Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia*^b *Australian Demographic & Social Research Institute, The Australian National University, Canberra ACT 0200, Australia*

Abstract

Age–sex-specific population forecasts are derived through stochastic population renewal using forecasts of mortality, fertility and net migration. Functional data models with time series coefficients are used to model age-specific mortality and fertility rates. As detailed migration data are lacking, net migration by age and sex is estimated as the difference between historic annual population data and successive populations one year ahead derived from a projection using fertility and mortality data. This estimate, which includes error, is also modeled using a functional data model. The three models involve different strengths of the general Box–Cox transformation chosen to minimise out-of-sample forecast error. Uncertainty is estimated from the model, with an adjustment to ensure that the one-step-forecast variances are equal to those obtained with historical data. The three models are then used in a Monte Carlo simulation of future fertility, mortality and net migration, which are combined using the cohort-component method to obtain age-specific forecasts of the population by sex. The distribution of the forecasts provides probabilistic prediction intervals. The method is demonstrated by making 20-year forecasts using Australian data for the period 1921–2004. The advantages of our method are: (1) it is a coherent stochastic model of the three demographic components; (2) it is estimated entirely from historical data with no subjective inputs required; and (3) it provides probabilistic prediction intervals for any demographic variable that is derived from population numbers and vital events, including life expectancies, total fertility rates and dependency ratios.

© 2008 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

Keywords: Fertility forecasting; Functional data; Mortality forecasting; Net migration; Nonparametric smoothing; Population forecasting; Principal components; Simulation

1. Introduction

Stochastic methods of population forecasting are rapidly gaining recognition. Fully probabilistic population forecasts have the major advantage of probabilistic

consistency among all forecast variables, including derived indices (Lee & Tuljapurkar, 1994). The Netherlands official statistical agency now uses stochastic methods, and other countries, such as the US, are also adopting them in producing official forecasts.

Forecasts of the size and structure of the population are central to social and economic planning. Not least of the demographic challenges facing developed countries is the rapid ageing of the population. The

* Corresponding author.

E-mail addresses: Rob.Hyndman@buseco.monash.edu.au (R.J. Hyndman), Heather.Booth@anu.edu.au (H. Booth).

major driver of this ageing process is the fertility fluctuations of the past, notably the post-war baby boom and the subsequent decline, but the declining mortality is also significant. One response to population ageing has been an increase in immigration to make up for past shortfalls in births (United Nations, 2001). Immigration has thus become a major driver of population change, and in some countries amounts to 50% of the number of births; Australia and Spain are examples. Stochastic population forecasting must take proper account of all three of these components of demographic change; each must be forecast.

Several different approaches to stochastic demographic forecasting have been developed in recent years; for a review see Booth (2006). The most widely used are those that involve some form of extrapolation, often using time series methods. Functional data methods fall into this category, but they have only recently been adopted in demographic forecasting (Hyndman & Ullah, 2007). Functional data methods have the advantage of providing a flexible framework that can be used for all three demographic processes. This paper applies functional data models to forecasting mortality, fertility and net international migration in national population forecasting.

1.1. Stochastic demographic forecasting

Stochastic methods of mortality forecasting have received considerable attention. The most prominent method is the Lee–Carter method (Lee & Carter, 1992), which now has several variants and extensions. Lee and Miller's (2001) variant is widely used. A variant proposed by Booth, Maindonald, and Smith (2002) has been shown to be at least as accurate as Lee–Miller in the short term (Booth, Hyndman, Tickle, & de Jong, 2006; Booth, Tickle, & Smith, 2005). Further developments have incorporated a heteroscedastic Poisson error structure (e.g., Brouhns, Denuit, & Vermunt, 2002; Wilmoth, 1993), extended applicability to mortality reduction factors (Renshaw & Haberman, 2003a) and examined the use of more than one term (Booth et al., 2002; Renshaw & Haberman, 2003b). Parallel approaches within the GLM framework have also been developed (Renshaw & Haberman, 2003c). Two recent extensions involve incorporating non-parametric smoothing into the model. De Jong and Tickle (2006) combine spline smoothing and estimation via the Kalman filter to fit a

generalized version of the Lee–Carter model, while Hyndman and Ullah (2007), following the functional data paradigm, propose smoothing the mortality curves for each year using constrained regression splines prior to fitting a model using principal components decomposition. These methods are compared by Booth et al. (2006).

Methods for forecasting fertility are less well developed. Fertility has proved difficult to forecast due to structural change, and estimates of uncertainty are highly dependent on the particular model. Using a method parallel to the Lee–Carter method, Lee (1993) found it necessary to pre-specify the long-term mean value of total fertility and to impose limits to reduce the width of the prediction interval (see also Lee, 1999). A principal components approach was also employed by Bozok and Bell (1987), who used the first four components and multivariate ARIMA methods. Hyndman and Ullah (2007) applied a similar method as part of a functional data approach, and this method is further developed in this paper.

Methods for forecasting migration are the least developed and are often extremely simple (George & Perreault, 1992). The main reason for this stems from a lack of time series of reliable, detailed data that accurately represent actual migration flows. A solution to the lack of data is to estimate net migration as the difference between the increment in population size and the natural increase using the demographic growth–balance equation. For subnational population forecasting involving internal migration forecasts, this method is often the only approach available because data are not collected; see also Miller (2003) and Miller and Lee (2004). The use of net migration does not address concerns that forecast accuracy depends on disaggregation according to different groups or types of immigrants and emigrants (Hilderink et al., 2002; Rogers, 1990). However, De Beer (1997) found consistency between time series forecasts of total immigration, emigration and net migration. Disaggregation by age of total net migration forecasts has been achieved using a reduced version of the multi-exponential model (Keilman & Pham, 2004).

The forecast components are used to generate stochastic population forecasts through stochastic population renewal using the cohort-component method of population projection (Preston, Heuveline, & Guillot, 2001, pp.119–129). The population forecast is achieved either analytically using the stochastic Leslie matrix (Alho & Spencer, 1985; Lee & Tuljapurkar, 1994;

Sykes, 1969), or more simply by Monte Carlo simulation to produce a distribution of possible outcomes. For both approaches, it is necessary to specify the mean (or median), variance–covariance structure and distributional form for each demographic component.

1.2. Aims

This paper aims to apply functional data models and time series methods to forecasting mortality, fertility and net international migration for use in national population forecasting. These forecast components are combined using the cohort-component method and Monte Carlo simulation to produce probabilistic population forecasts by age and sex. The use of extrapolative methods presupposes that the trends of the past will be continued into the future. This assumption has often proved to be a better basis for forecasting than either structural modelling involving exogenous variables or methods based on expectation (Booth, 2006).

The method is illustrated using Australian data for 1921–2004 with a forecast horizon of 20 years. As complete and reliable data for international migration are lacking, annual net migration is estimated using the growth-balance equation. Although this estimate includes errors in vital registration data and annual population estimates, the resulting series is superior to the available migration data in its coverage of years, events and single years of age.

In the following section, we discuss the data requirements of our approach, and explain how the derived data are obtained. Section 3 describes the functional data models that we fit to each of the components (mortality, fertility and net migration). In Section 4, we discuss how to simulate future sample paths for each of these components, thereby obtaining simulated projections of age-specific population numbers by sex. We apply the method to Australian data in Section 5 to obtain twenty-year probabilistic forecasts of the population by age and sex. Some conclusions and discussion, including extensions of this approach, are contained in Section 6.

2. Data requirements

We use the following data: age-sex-specific birth and death numbers for each calendar year, age-sex-

specific population numbers at 1 January of each year, and age–sex-specific exposures to risk (i.e., populations at 30 June) for each year. We use the following notation:

$B_t(x)$ = Births in calendar year t to females of age x ;
 $D_t(x)$ = Deaths in calendar year t of persons of age x ;
 $P_t(x)$ = Population of age x at 1 January of year t ;
 $E_t(x)$ = Population of age x exposed to risk at 30 June of year t ;

where $x=0, 1, 2, \dots, p-1, p^+$ and $t=1, \dots, n$. Here, p^+ denotes the open-ended upper age group. The superscripts M and F will be used to refer to males and females respectively.

Let $m_t(x)=D_t(x)/E_t(x)$ be the age–sex-specific central death rates in calendar year, t and let $f_t(x)=B_t(x)/E_t^F(x)$ be the age-specific fertility rates in calendar year t . Births are divided by sex using ρ , the sex-ratio at birth.

Net migration, denoted G_t , is estimated using the demographic growth-balance equation. Thus:

$$\begin{aligned} G_t(x, x+1) &= P_{t+1}(x+1) - P_t(x) + D_t(x, x+1) \\ &\quad x=0, 1, 2, \dots, p-2; \\ G_t(p-1^+, p^+) &= P_{t+1}(p^+) - P_t(p^+) - P_t(p-1) \\ &\quad + D_t(p-1^+, p^+); \\ G_t(B, 0) &= P_{t+1}(0) - B_t + D_t(B, 0); \end{aligned}$$

where $D_t(x, x+1)$ refers to deaths in calendar year t of persons aged x at the beginning of year t , $D_t(p-1^+, p^+)$ refers to deaths in calendar year t of persons aged $p-1$ and older at the beginning of year t , and $D_t(B, 0)$ refers to deaths in calendar year t of births during year t ; and similarly for net migration, $G_t(x, x+1)$, $G_t(p-1^+, p^+)$ and $G_t(B, 0)$. In other words, these are cohort rhomboids in the Lexis diagram. Deaths are estimated using the standard life table approach of population projection. See Preston et al. (2001) for further details.

The estimation of net migration in the last year of the fitting period requires population data for 1 January, year $n+1$. Note that as $G_t(B, 0)$ refers to the migration of those born in year t , it is smaller than $G_t(0, 1)$. The $G_t(x, x+1)$ values are forecast; this corresponds directly to the cohort approach of population dynamics.

3. Functional data modelling approach

To obtain forecasts of each component, we first develop functional time series models for the five sex-

specific components: $m_t^F(x)$, $m_t^M(x)$, $f_t(x)$, $G_t^F(x, x+1)$, and $G_t^M(x, x+1)$. The five models will then be used in the simulation of the future population. We follow the approach of Hyndman and Ullah (2007) to model each of these components.

3.1. Functional data models

Let $y_t^*(x)$ denote the quantity being modelled — either mortality rates, fertility rates, or net migration numbers for age x in year t . First, we use a Box and Cox (1964) transformation of $y_t^*(x)$ to allow for variation that increases with the value of $y_t^*(x)$. (That is, the variability in rates is greater when the rates are higher.) Thus

$$y_t(x) = \begin{cases} \frac{1}{\lambda} \left([y_t^*(x)]^\lambda - 1 \right) & \text{if } 0 < \lambda \leq 1; \\ \log_e(y_t^*(x)) & \text{if } \lambda = 0. \end{cases}$$

The value of λ determines the strength of the transformation, with $\lambda = 1$ giving effectively no transformation. Because net migration involves negative values of $y_t^*(x)$, we use $\lambda = 1$ in that case.

Then we assume the following model for the transformed quantity $y_t(x)$:

$$y_t(x) = s_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$s_t(x) = \mu(x) + \sum_{k=1}^K \beta_{t,k}\phi_k(x) + e_t(x),$$

where $s_t(x)$ is an underlying smooth function of x , $\varepsilon_{t,x}$ are iid standard normal variates, and $\sigma_t(x)$ allows the variance to change with age and time. That is, $s_t(x)$ is a smooth function of age which we observe with error. The second equation describes the dynamics of $s_t(x)$ evolving through time. In this equation, $\mu(x)$ is the mean of $s_t(x)$ across years, $\{\phi_k(x)\}$ is a set of orthogonal basis functions calculated using a principal components decomposition, and $e_t(x)$ is the model error, which is assumed to be serially uncorrelated. The dynamics of the process are controlled by the time series coefficients $\{\beta_{t,k}\}$, which are assumed to behave independently of each other (this follows from using a principal components decomposition). There are three sources of randomness in our model: $\varepsilon_{t,x}$ represents the random variation in births, deaths or migrants from the relevant distribution (e.g., Poisson or normal); $e_t(x)$

represents the residual error in modelling $s_t(x)$ using a finite set of basis functions; and there is also the randomness inherent in the time series model for each $\{\beta_{t,k}\}$, which drives the dynamic changes in the smooth curves $\{s_t(x)\}$.

This model was first proposed by Hyndman and Ullah (2007) for mortality and fertility rates, although they used log transformations instead of the more general Box–Cox transformation. It has also been used by Erbas, Hyndman, and Gertig (2007) for forecasting breast cancer mortality rates.

As Hyndman and Ullah (2007) pointed out, the model is a generalization of the well-known Lee and Carter (1992) model for forecasting mortality rates. In the Lee–Carter approach, $y_t^*(x)$ denotes mortality rates and $\lambda = 0$, so that $y_t(x)$ represents the log mortality for year t and age x . The Lee–Carter method does not assume smoothness, so $\sigma_t(x) = 0$ and $y_t(x) = s_t(x)$. Then $\mu(x)$ is estimated as the average of $y_t(x)$ across years, $K = 1$, and $\phi_1(x)$ and $\beta_{t,1}$ are computed from the first principal component of the matrix of $[y_t(x) - \hat{\mu}(x)]$. Forecasts are obtained by fitting a time series model to $\beta_{t,1}$; in practice this is almost always a random walk with drift.

In this paper, we extend the method of Hyndman and Ullah (2007) by using a more general transformation, and by applying the model to net migration numbers, mortality rates and fertility rates. We also modify the method Hyndman and Ullah used to calculate the forecast variance to allow for better calibration with the observed data.

Hyndman and Ullah proposed robust estimation of the model terms in order to handle the effect of epidemics and wars on mortality data. We will avoid this additional complexity in modelling Australian data by restricting the mortality data to 1950 onwards.

The modelling steps (described in detail by Hyndman & Ullah, 2007) are:

1. Estimate smooth functions $s_t(x)$ using nonparametric regression applied to $y_t(x)$ for each year t . (In our application to Australian data, we use weighted penalized regression splines.)
2. Estimate $\mu(x)$ as the mean of $s_t(x)$ across years.
3. Estimate $\beta_{t,k}$ and $\phi_k(x)$, $k = 1, \dots, K$, using a principal components decomposition of $[y_t(x) - \hat{\mu}(x)]$.
4. Estimate time series models for $\beta_{t,k}$, $k = 1, \dots, K$. We use exponential smoothing state space models.

The value of K must be specified. Hyndman and Ullah (2007) proposed selecting K to minimize the mean integrated squared forecast error. Since then, we have found that the method is insensitive to the choice of K , provided K is large enough. That is, there is little cost (apart from computing time) in choosing a large K , whereas a K that is too small may result in poor forecast accuracy. Consequently, in this analysis we choose $K=6$ for all components; this seems to be larger than any of the components really require.

The observational variance, $\sigma_t^2(x)$, depends on the nature of the data. For deaths, we estimate the observational variance from $y_t^*(x)=m_t(x)$, assuming that deaths are Poisson distributed (Brillinger, 1986) with a mean parameter $m_t(x)E_t(x)$. Thus, $y_t^*(x)$ has approximate variance $E_t^{-1}(x)m_t(x)$, and the variance of $y_t(x)$ (via a Taylor approximation) is

$$\sigma_t^2(x) \approx [m_t(x)]^{2\lambda-1} E_t^{-1}(x).$$

For births, we assume a Poisson distribution (Keilman, Pham, & Hetland, 2002) with mean $f_t(x)E_t^F(x)$, which gives

$$\sigma_t^2(x) \approx [f_t(x)]^{2\lambda-1} E_t^{-1}(x).$$

For migration data, we make no distributional assumptions, and we estimate $\sigma_t^2(x)$ using a nonparametric regression of $[y_t(x)-s_t(x)]^2$ against x .

The success of the model depends on how well the bivariate surface $\{s_t(x)-\mu(x)\}$ can be approximated by the sum of a few products of univariate functions of time (t) and age (x). So far, we have applied this model to mortality data from about twenty populations, and to fertility and migration data from Australia. This experience suggests that the model is good at producing point forecasts, but not quite so good at estimating forecast variance. Consequently, below we propose an adjustment to the forecast variance implied by the model.

3.2. Functional forecasts

Suppose we have data up to time $t=n$, and we wish to estimate future values of $y_t(x)$ for $t=n+1, \dots, n+h$ and all x . Let $\hat{\beta}_{n,k,h}$ denote the h -step ahead forecast of $\beta_{n+h,k}$, let $\hat{y}_{n,h}(x)$ denote the h -step ahead forecast of

$y_{n+h}(x)$ and let $\hat{s}_{n,h}(x)$ denote the h -step ahead forecast of $s_{n+h}(x)$. Then

$$\hat{y}_{n,h}(x) = \hat{s}_{n,h}(x) = \hat{\mu}(x) + \sum_{k=1}^K \hat{\beta}_{n,k,h} \hat{\phi}_k(x). \quad (1)$$

A forecast of $y_t^*(x)$ is found through back-transformation.

Following Hyndman and Ullah (2007), we can give the following expression for forecast variance

$$\begin{aligned} V_h(x) &= \text{Var}[s_{n+h}(x)|\mathcal{I}, \Phi] \\ &= \hat{\sigma}_{\mu}^2(x) + \sum_{k=1}^K u_{n+h,k} \hat{\phi}_k^2(x) + v(x), \end{aligned} \quad (2)$$

where $\mathcal{I} = \{y_t(x_i)\}$ denotes all observed data, $u_{n+h,k} = \text{Var}(\beta_{n+h,k}|\beta_{1,k}, \dots, \beta_{n,k})$ can be obtained from the time series model, $\hat{\sigma}_{\mu}^2(x)$ (the variance of the smooth estimate $\hat{\mu}(x)$) can be obtained from the smoothing method used, and $v(x)$ is estimated by averaging $\hat{e}_t^2(x)$ for each x . Thus, the smoothing error is given by the first term, the error due to predicting the dynamics is given by the second term, and the third term gives the error due to the unexplained dynamic variation. After the observational error is also included, we obtain

$$\text{Var}[y_{n+h}(x)|\mathcal{I}, \Phi] = V_h(x) + \sigma_t^2(x).$$

Note that correlations between ages are dealt with naturally in this formulation due to the smooth functions of age (x). Also, correlations between years are handled by the time series models for the coefficients $\beta_{t,1}, \dots, \beta_{t,K}$.

For low values of h , we can check the validity of $V_h(x)$ by computing the in-sample empirical forecast variance

$$W_h(x) = \frac{1}{n-h-m+1} \sum_{t=m}^{n-h} [s_{t+h}(x) - \hat{s}_{t,h}(x)]^2,$$

where m is the smallest number of observations used to fit a model. In practice, considerable differences between $W_h(x)$ and $V_h(x)$ can occur. Consequently, we use the following adjusted variance expression:

$$\begin{aligned} \text{Var}[y_{n+h}(x)|\mathcal{I}, \Phi] &= V_h(x)W_1(x)/V_1(x) \\ &\quad + \sigma_t^2(x). \end{aligned} \quad (3)$$

This adjusts the variance so that the one-step forecast variance matches the in-sample empirical one-step forecast variance. It is assumed that the same multiplicative adjustment is applicable at higher forecast horizons. In our empirical analysis, we have found that the adjustment results in smaller variances at most ages for mortality and fertility forecasts.

Note that we could check for bias in the point forecasts similarly by averaging $s_{t+h}(x) - \hat{s}_{t,h}$, but empirical results suggest that the model provides excellent forecasts, so that this bias is negligible.

4. Stochastic cohort simulation from functional data models

Population sample paths are simulated using the cohort-component method adapted to permit random observational error. The algorithm, given in the Appendix, shows the order in which the calculations need to be carried out. The base population is the observed population at 1 January in year $n+1$.

For each of $m_t^F(x)$, $m_t^M(x)$, $f_t(x)$, $G_t^F(x, x+1)$ and $G_t^M(x, x+1)$, we simulate a large number of future sample paths by using the time series models to generate random sample paths of $\beta_{t,k}$ for $t=n+1, \dots, n+h$ conditional on $\beta_{1,k}, \dots, \beta_{n,k}$. We also generate random values of $e_t(x)$ by bootstrapping the estimated values. We then apply the variance adjustment described in Eq. (3); the adjusted value of $s_{n+h}(x)$ is

$$\tilde{s}_{n+h}(x) = \hat{\mu}(x) + \sum_{k=1}^K \hat{\beta}_{n,k,h} \hat{\phi}_k(x) + \left[s_{n+h}(x) - \hat{\mu}(x) - \sum_{k=1}^K \hat{\beta}_{n,k,h} \hat{\phi}_k(x) \right] \sqrt{W_1(x)/V_1(x)}.$$

In order to take into account the births and deaths among annual migrants, the population is first adjusted for net migration: migrants are assumed to spend, on average, half of the year exposed to events. Cohort net migration at the beginning of year $n+h$ is computed using the simulated smoothed mean $\tilde{s}_{n+h}(x)$ with resampling of the errors $y_t(x) - s_t(x)$. Half of the simulated annual net migration, $G_t(x, x+1)$, is added to the population aged x on 1 January of year t , and half is added to the population aged $x+1$ at the end of year t . We denote the population adjusted for the first half of migration by R . For net migration of infants born in

year t , half of $G_t(B, 0)$ is added to births in year t and half is added to the population aged 0 at the end of year t . At the oldest ages, $G_t(p-1^+, p^+)$ is assumed to be equally divided between migrants aged $p-1$ and p^+ at 1 January of year t .

Mortality is applied next. For each sex, we take a simulated sample path of $\{m_t(x)\}$, $t=n+1, \dots, n+h$. In order to obtain random numbers of deaths, the mid-year population is needed; however, this depends on the number of deaths. This circularity is dealt with as follows. Using the life table based on simulated values of $\{m_t(x)\}$ and $R_t(x)$, we obtain the expected number of cohort deaths. We use these deaths to estimate R_{t+1} , and hence the mid-year population. The product of this mid-year population and $\{m_t(x)\}$ defines the Poisson distribution from which $D_t(x)$ is randomly drawn. Deaths in year t at each pair of adjacent ages are averaged to obtain cohort deaths, and hence R_{t+1} . The addition of the second half of cohort net migration gives the population on 1 January of year $t+1$, P_{t+1} .

The number of births is then obtained using the simulated fertility rate $f_t(x)$, and the population adjusted for half of net migration. Births are assumed to be Poisson with mean $f_t(x)[R_t^F(x) + R_{t+1}^F(x)]/2$ for $x=15, \dots, 49$. A random draw from this distribution determines $B_t(x)$; these are summed to give a simulated value of B_t . Male (and hence female) births are allocated binomially with probability $\rho/(\rho+1)$. We ignore any variation in ρ , as the effect is small compared to the other sources of variation.

Mortality is applied to births using the procedure described above (see the Appendix). The randomly-drawn number of deaths, $D_t(0)$, is divided into deaths to births in year t , $D_t(B, 0)$, and deaths at age 0 to births in year $t-1$ using a separation factor estimated as the ratio of expected deaths between birth and age 0 and expected deaths at age 0 in year t . This separation factor is also used to obtain cohort deaths at age 0 to 1 as the weighted average of $D_t(0)$ and $D_t(1)$.

Hence, using simulated births, deaths and migrants, we generate the population for the next year. This is repeated for years $t=n+1, \dots, n+h$, and the whole procedure is repeated for each population sample path. Thus, we obtain a large number of future sample paths of age–sex-specific population and vital event numbers which can be used to estimate, with uncertainty, any demographic variable that is derived from

population numbers and vital events, including life expectancies, total fertility rates and dependency ratios.

5. Application to Australia

Most of the data were obtained from the [Human Mortality Database \(2006\)](#) and consist of central death rates, start-year and mid-year populations by sex and age in single years for 0–99 and 100+ years for 1921–2003 (the start-year population for 2004 is also used in estimating net migration). The base population for the forecast refers to 1 January 2004. Age-specific fertility rates by single years of age for 15–49 for 1921–2003 were obtained from the Australian Demographic Data-Bank (located at the Australian Demographic and Social Research Institute, Australian National University).

To avoid difficulties with war years, the 1918 Spanish influenza epidemic, and structural change over the course of the twentieth century, we only use mortality data from 1950 onwards. When computing net migrant numbers, we only use data from 1973 onwards, as the population numbers are less reliable before that.

We follow the modelling framework outlined in Section 3. For mortality, we choose $\lambda=0$, which is consistent with other studies (e.g., [Booth et al., 2002](#); [Hyndman & Ullah, 2007](#); [Lee & Carter, 1992](#), etc.). As described by [Hyndman and Ullah \(2007\)](#), we constrain the fitted curves to be monotonically increasing for $x>65$.

For fertility, there seems to be no consensus on the best transformation to use. [Hyndman and Ullah \(2007\)](#) use logarithms; [Lee \(1993\)](#) uses no transformation on age-specific fertility rates, but a logistic transformation (defined on the interval $[0,4]$) on the total fertility rate. We chose the value of $\lambda=0.4$ as it gave relatively small out-of-sample forecast errors (on the untransformed scale) and the narrowest prediction intervals when applied to the Australian data for 1921–1993. The sex ratio at birth was set to be $\rho=1.0545$, based on current Australian data.

For migration, the data are both positive and negative, so we do not use any transformation; that is, $y_t^*(x)=y_t(x)$.

For all models, $K=6$ basis functions are used. This is larger than any of the components seem to require. As noted previously, the method is insensitive to the choice of K , provided that K is large enough; in other

words, additional basis functions do not decrease forecast accuracy.

The time series models used are additive state space models for exponential smoothing, as described in by [Hyndman, Koehler, Snyder, and Grose \(2002\)](#). We used a mixture of local-trend and damped-trend models with additive errors for mortality and fertility, and local-level models for migration. For mortality, these are the same model choices as were used by [Booth et al. \(2006\)](#). The parameters were optimized by minimizing the average of the squared in-sample h -step forecast errors for $h=1, 2, \dots, 8$.

5.1. Results

Estimates of net migration for the period 1973–2003 are shown in [Fig. 1](#). It can be seen that there is a high degree of variation at all but the oldest ages around a jagged age-specific mean.

The variation in these estimated numbers is a combination of the variation in migration, errors in the age-specific population numbers, errors in estimated age-specific deaths, and errors in the estimated number of births. As population numbers are Estimated Resident Populations (ERPs) produced by the Australian Bureau of Statistics (ABS) rather than independent empirical counts, they are the product of models, and assumptions and will be subject to (unknown) systematic biases as well as random error. Such biases may have contributed to the jagged mean age distribution, particularly at ages 35 to 60 where the male and female age patterns are almost identical; it is also likely that digit preference occurs in the reported ages in the population numbers, deaths, and possibly the original migration data used in official estimates of intercensal population size. These biases have undoubtedly produced the elevated estimates at age 80. These irregularities will largely be removed by smoothing. Positive net migration at very old ages is also likely to be the result of bias; [Wilson and Bell \(2004\)](#) also found that the ERP inflates the population at ages 90+. The high volatility in estimated net migration between birth and age 0 on 1 January of the following year is likely to be an artifact of ABS estimation procedures, rather than a reflection of the true variability in migration at this age.

While there is a high degree of variation, an age pattern can readily be discerned. The male and female patterns are similar. The mean suggests that the level of female net migration is higher than that for males at

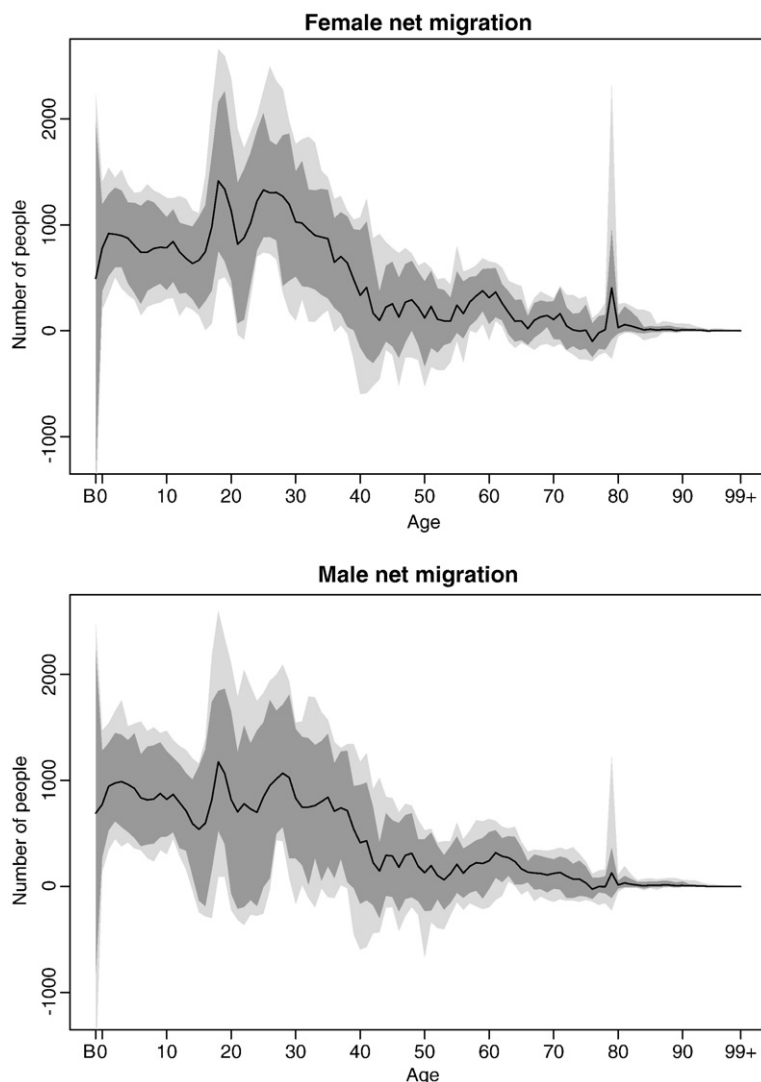


Fig. 1. Net migration estimates, 1973–2003. The light-shaded region shows 2.5% and 97.5% quantiles; the dark-shaded region shows 10% and 90% quantiles; and the solid line shows the mean. On the age-axis, B represents the interval between birth and age 0 on 1 January in the following year.

young adult ages; this is a result of male-dominated emigration rather than of female-dominated immigration. The number of children is also relatively high, a result of Australian migration policies concerning family migration. For both males and females, the first peak at young adult ages (at age 18) is due to overseas students, while the second (at age 28 for males and 25 for females) is due to labour migration and spouses. A further retirement-related peak occurs at around age 60. Compared with the labour-dominated migration

model of Rogers and Castro (1981) (see also Rogers, Castro, & Lea, 2005), these distributions are relatively flat across the age range.

Fig. 2 shows the first basis function and the first coefficient for all five models. The first term accounts for at least 94% of the variation in mortality, 65% for fertility and 51–60% for net migration. Taking female mortality as an example, the coefficient indicates a fairly steady decline in mortality over time, while the basis function indicates that the decline has been faster at very

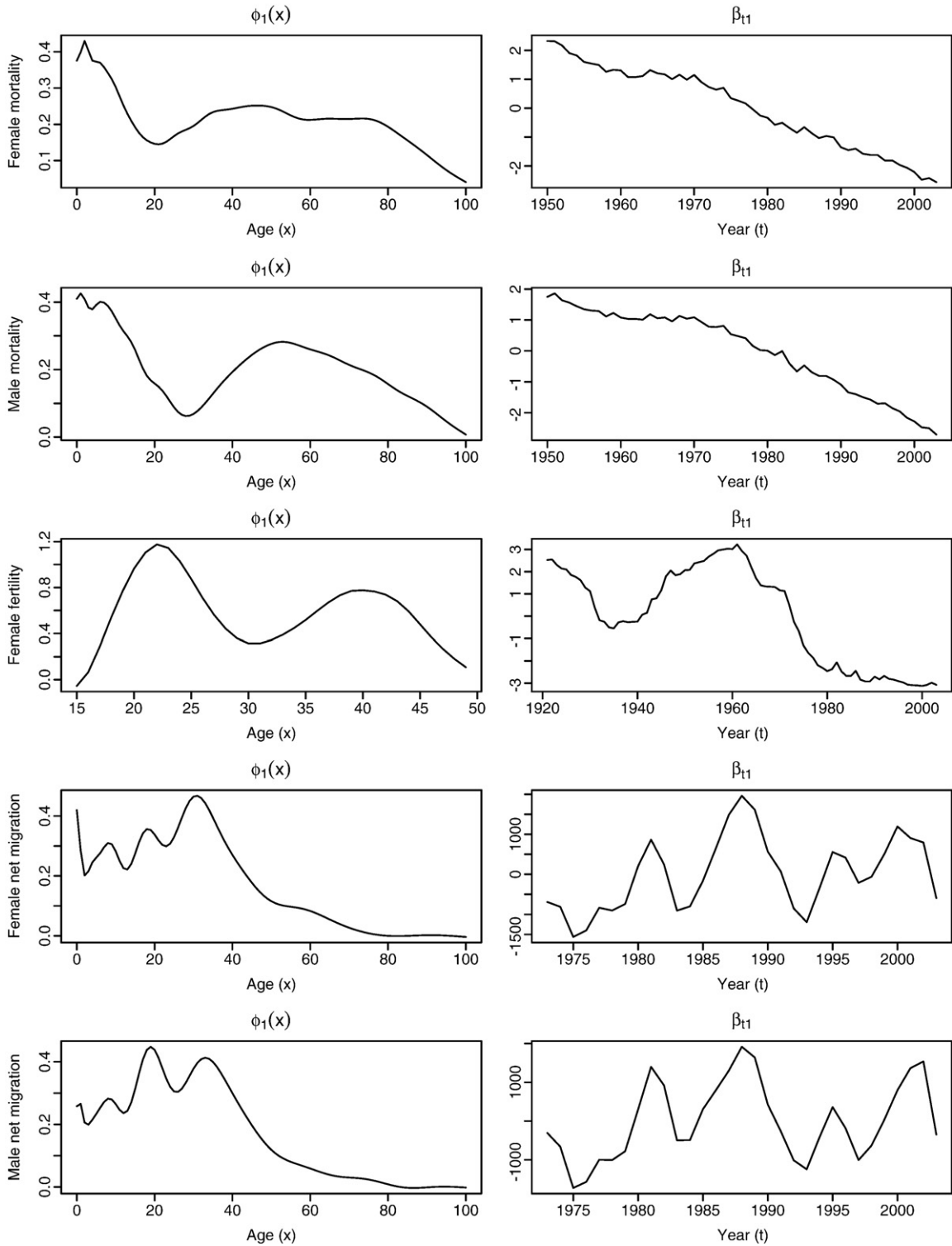


Fig. 2. The first basis function and coefficient for each component of the Australian data.

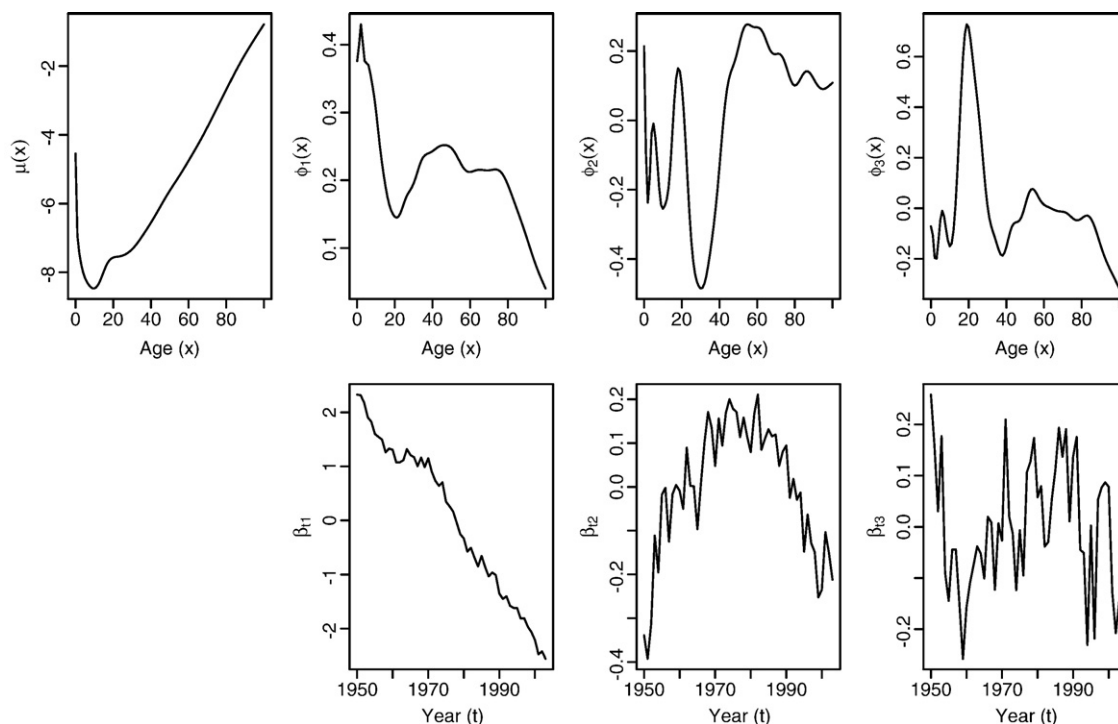


Fig. 3. Fitted basis functions and coefficients for Australian female log death rates.

young ages and at ages 40–80. Fig. 3 shows a further two terms for this model, which modify the dominant first term (the mean age pattern is also shown).

Fig. 4 shows forecast log death rates by sex for the first and last years of the forecast period (2004 and 2023), with 80% prediction intervals. Also shown are the observed rates for 2003. The forecast decline in male mortality is less rapid at young adult ages (20–40 years) than at older ages (40–80), reflecting recent trends. The 2023 point forecast rates for females at ages 0–1 and 7–14 slightly exceed those for males, but the prediction intervals are wide at these points.

Forecast fertility rates are shown in Fig. 5. The relatively recent change in the second derivative at ages 19 and 20, as seen in the 2003 data, is a more marked feature in the forecast, resulting in particularly low rates in the early to mid-twenties. While such a marked age pattern of fertility may at first appear extreme, it is not without precedent: an even more marked effect has been observed in Ireland (Chandola, Coleman, & Hiorns, 1999). This feature is emerging in several

populations, including the United Kingdom (Chandola et al., 1999) and Germany (Betz & Lipps, 2004). It is likely to arise from heterogeneity with respect to marital status (as Chandola et al. suggest) or socio-economic circumstances: one group experiencing early fertility which will decline only slowly (possibly towards a threshold beyond which rates will not decline) and another group experiencing a rapid decline in early fertility. However, the message contained in the very wide prediction intervals in 2023 should not be ignored.

Net migration forecasts are shown in Fig. 6. The volatility in the data is reflected in the wide prediction interval which swamps the point forecast. The point forecasts are identical for every year of the forecast, as the point forecasts of all coefficients are constant over the forecast horizon. The prediction intervals increase with the forecast horizon. The forecast annual net migration is 44,200 females and 43,200 males.

Forecast life expectancy by sex is shown in Fig. 7. By 2023, the forecast values are 87.9 years with an 80% prediction interval of 85.2 to 90.3 for females,

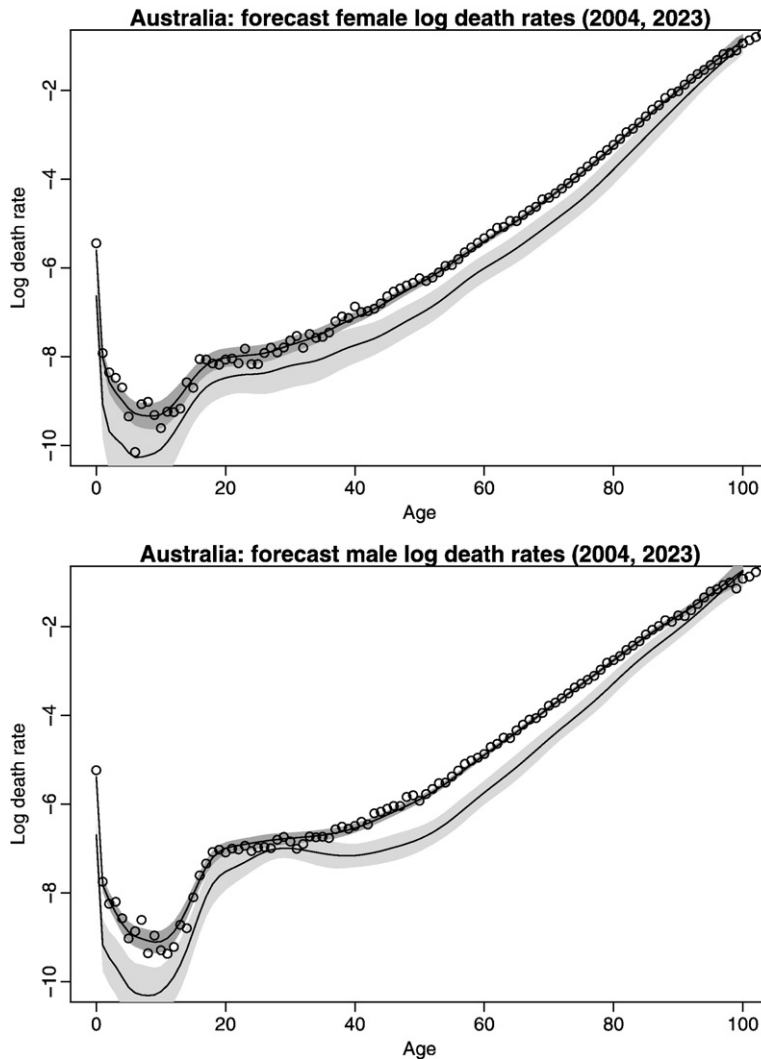


Fig. 4. Forecast log death rates for 2004 (upper line) and 2023 (lower line), along with 80% prediction intervals. The actual mortality rates for 2003 are also shown as circles.

and 84.3 years with an interval of 81.8 to 86.6 for males. Male life expectancy forecasts increase at a faster rate than those for females for the next 20 years. The forecast total fertility (Fig. 8) shows a slightly increasing (but insignificant) trend, with a large prediction interval of 0.90 to 2.61 in 2024, reflecting the large uncertainty in forecasting fertility.

The final population forecasts for males and females in 2024 are shown as a population pyramid with 80% prediction intervals in Fig. 9, along with the 2004 base population. These are based on 10,000 simulated

sample paths. The broad prediction intervals at young ages reflect the greater uncertainty in forecast fertility, while the uncertainty at ages 20–50 is largely due to migration.

In Fig. 10, our forecasts are compared with the population projections given in Australian Bureau of Statistics (2003) which use a base population of 30 June 2002, and those given in Australian Bureau of Statistics (2006) which use a base population of 30 June 2004. This graph demonstrates that consecutive ABS series are remarkably inconsistent in their probability coverage,

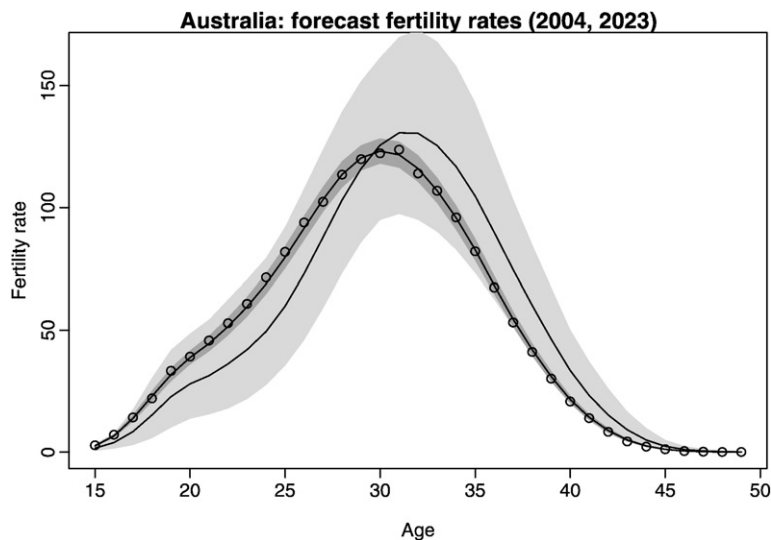


Fig. 5. Forecast fertility rates for 2004 (inside dark grey region) and 2023 (inside light grey region), along with 80% prediction intervals. Actual fertility rates for 2003 are also shown as circles.

with the most recent projections probably showing an upward bias and a severe underestimate of the lower tail of the future population distribution. Most users tend to interpret series A and C as something like a prediction interval. Yet this analysis shows that the probability of being above and below series A and C is unequal both across forecast horizons and, in particular, between releases of the projections. Consequently, the ABS series are likely to be misleading to users.

6. Comments and conclusions

The above analysis has demonstrated that functional data models can be applied successfully to forecasting mortality rates, fertility rates and migration numbers. This modelling framework is highly adaptable. The same basic model applies, allowing for the different characteristics of the three demographic components through the use of different transformations. The Box–Cox transformation has proved useful. Though this transformation is commonly used in statistics, it has rarely been applied in demographic modelling. For mortality, the preferred transformation is the logarithm, which coincides with convention. For fertility, a weaker transformation ($\lambda=0.4$) is preferred;

this use of the Box–Cox transformation is a substantial improvement over previous research, where it has been necessary to impose judgmental limits in order to constrain the prediction intervals to plausible values.

The forecasts produced here are based on the first six principal components. The additional components may serve to incorporate relatively recent changes in pattern. For mortality, the use of several components is the main difference between our method and the Lee–Carter method (which uses only the first component, and also involves an adjustment). The extra principal components allow more accurate forecasting of age-specific death rates (Hyndman & Ullah, 2007), though typically at least 90% of the variation is explained by the first component. The use of smoothing prior to modelling results in forecast age patterns that are relatively smooth compared with the jagged age distributions of forecasts produced by the Lee–Carter method (Giroi & King, 2006).

The use of several components is more important for fertility, where the first component explains a smaller proportion of the variation (65% in the current example). The emergence of relatively low rates in the early to mid-twenties is the combined effect of the first and second components; this pattern has been forecast to evolve to a diminishing extent over time.

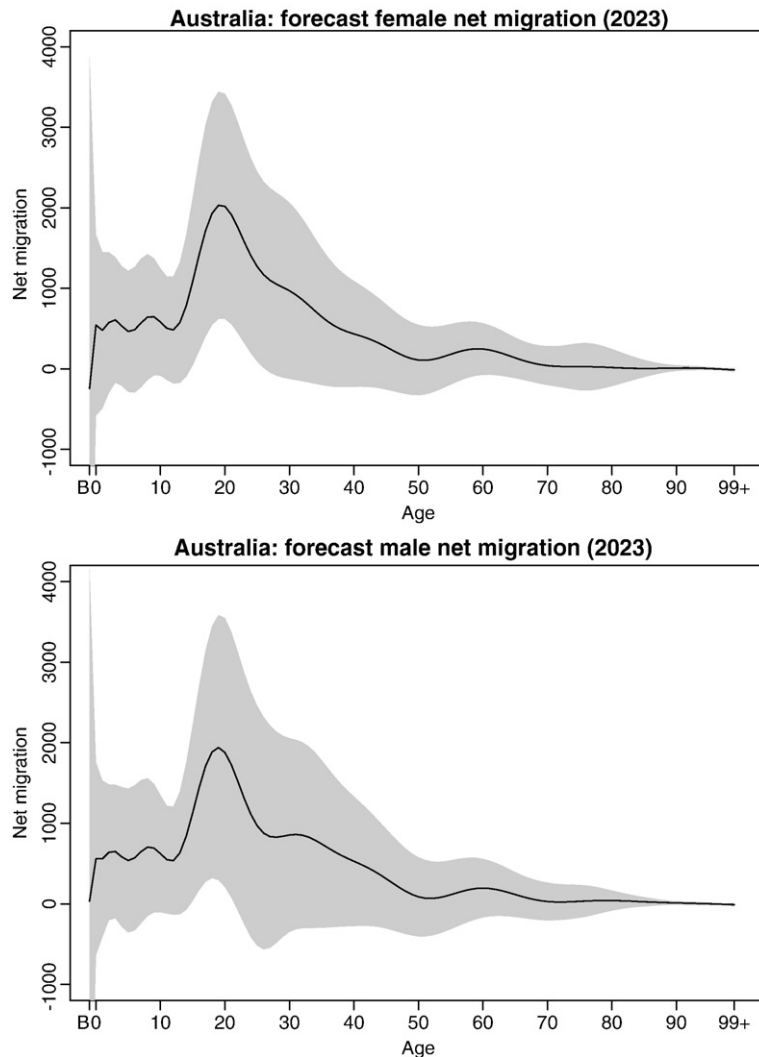


Fig. 6. Forecast net migration numbers for 2023, along with 80% prediction intervals. All other years are almost identical due to the low correlation between years.

For net migration, the numbers by age and sex estimated from the growth-balance equation include net errors in the recorded populations and in vital events. Errors arising from digit preference and random error are removed by smoothing. While forecast net migration will include systematic errors not removed by smoothing, such error is unlikely to be significant in Australian data. For the purpose of forecasting population, the inclusion of such error will not bias the forecast if past errors can be assumed

to be constant for each age over time. In countries with high levels of illegal immigration, where the estimation of net age–sex-specific migration is highly preferable to the use of recorded data, this assumption may not be appropriate and care should be exercised in forecasting. For example, if population coverage has increased over time, the trend in net migration will be overestimated.

In the Australian case, the age–sex-specific net migration forecast is a simple constant for each age and

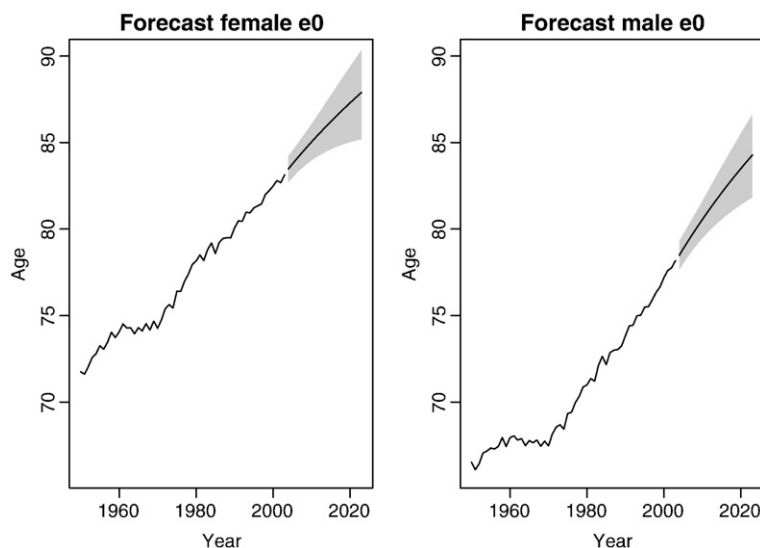


Fig. 7. Forecasts of e_0 (life expectancy at age 0) for 2004–2023.

sex. Though the level of migration has fluctuated over the last 30 years, there is no significant trend in either direction; this can be attributed to the general stability of migration policies and economic factors. In countries where such factors have been less stable,

for example, Spain, which has changed from a country of emigration to a country of immigration in the last few decades, the forecasting of net migration will undoubtedly be more difficult. One avenue for the improvement of migration forecasting is to take into

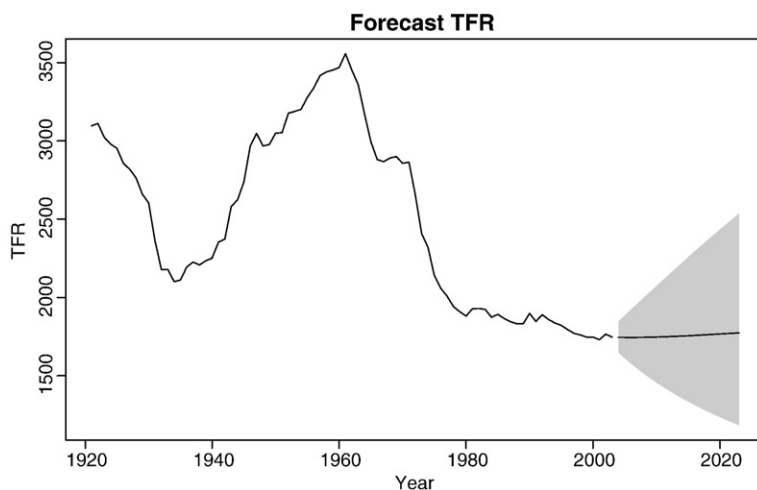


Fig. 8. Forecasts of total fertility rate (TFR) for 2004–2023.

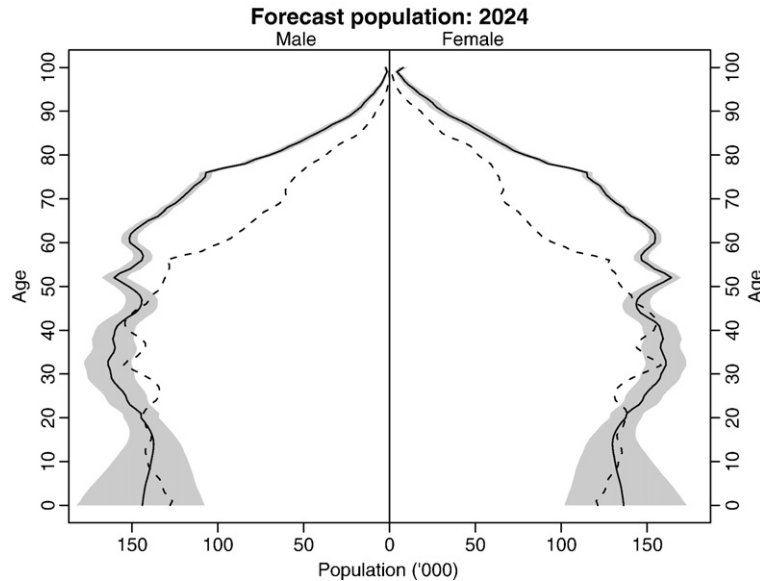


Fig. 9. Forecast population pyramid for 2024, along with 80% prediction intervals. The actual population pyramid for 2004 is shown using dashed lines.

account correlations between past fertility and current migration in the context of population ageing; another is to add exogenous variables, such as unmet demand for labour. Further research is needed to examine and implement these possibilities.

The forecast life expectancy in 2023 is 87.9 years for females and 84.3 for males. This corresponds to an increase of 3.3 years for males for the first decade of

forecasts (close to the historic rate of increase since 1970), but dropping to 2.8 for the second decade of forecasts. For females, the forecasts show an increase of 2.5 years for the first decade of forecasts (close to the historic rate in the last decade of observation) and 2.1 years for the second decade. For females, the forecast and historical increases are also close to the “best-practice” linear increase of 2.5 years per decade

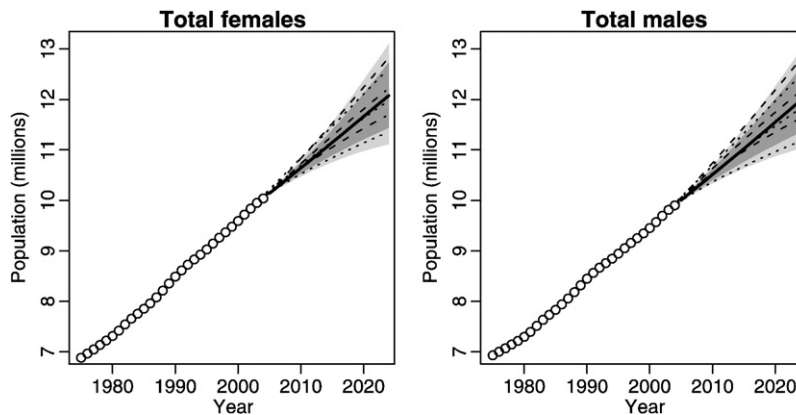


Fig. 10. Twenty-year forecasts of the total population for each sex, along with 80% and 95% prediction intervals. The dashed lines show the projections from [Australian Bureau of Statistics \(2006\)](#), series A, B and C. The dotted lines show the projections from [Australian Bureau of Statistics \(2003\)](#), series A, B and C.

for females recorded by [Oeppen and Vaupel \(2002\)](#) since 1840. Our forecast decreasing trend is a consequence of the use of the damped-trend models. However, it should be noted that both sexes show 80% prediction intervals wide enough to allow for continuing linear increases. The greater increase in forecasts for males than females between 2004 and 2024 does not therefore imply that male life expectancy will exceed female life expectancy in the future.

As noted above, the estimation of the uncertainty of forecast demographic processes presents considerable difficulties because estimates vary depending on the method used. In this paper, we have adjusted model-based estimates of the variance to ensure that the model one-step forecast variance is equal to the historical empirical one-step forecast variance. In the Australian case, this calibration increases the model variance for some ages and decreases it for others. For mortality rates, the adjustment reduces variances except at very old ages; for fertility rates, the adjustment tends to increase variances between ages 25 and 35 and decrease variances at ages 35 and above. The adjustment has little effect on net migration variances.

In population forecasting, several sources of uncertainty need to be taken into account. Our proposed method has accounted for all of the major sources of variation: the observational variation in the (Poisson) generation of births and deaths, the binomial variation in the sex ratio at birth, the observational variation in net migration, and the dynamic variation as the rates of mortality, fertility and migration change over time. This constitutes a more complete representation of uncertainty than most, if not all, other approaches.

In population forecasts, correlations among forecast errors in the demographic components affect the width of the prediction intervals ([Lee, 1999](#)). In our approach, correlations across age are taken into account by smoothing across age, and temporal correlations by the time series models. These correlations are then naturally incorporated into the simulated sample paths of each demographic component. However, we have not taken into account any correlation *between* the demographic components, including correlations between the sexes. The assumed independence among the three demographic components is a reasonable first approximation for developed countries (e.g., [Alho, 1992](#)). However, there is obvious dependence

between the sexes which we have not accounted for, which is why we have not produced forecasts for the two sexes combined. We leave this methodological development to another paper.

It is useful to compare our prediction intervals with those from previous research, although it is difficult to compare our forecasts with those for the two sexes combined because we only provide sex-specific forecasts. For a horizon of 20 years, [Lee and Carter \(1992\)](#) give a 95% life-expectancy prediction interval of 6.5 years for the two sexes combined, compared with 7.3 for males and 8.0 for females in our analysis. The [Wilson and Bell \(2004\)](#) forecasts of Australian life expectancy, which are based on ad hoc methods, have narrower prediction intervals than ours at a 20-year horizon, but are wider at forecast horizons of around 5–10 years. For fertility, the Wilson and Bell TFR forecast, based on a constrained ARIMA model, has a 95% prediction interval with a width of 1.4, which is about the same width as our 80% prediction interval, possibly because they use a shorter time series without the large pre-war fluctuations. Their model-based net migration pattern for the two sexes combined is similar to our sex-specific patterns. For forecast total population, the Wilson and Bell prediction interval at 20 years is 4.5 million people wide, compared with our sex-specific intervals of 2.0 million people for both males and females.

Further comparison is possible with the [Bongaarts and Bulatao \(2000\)](#) study, which reported (p. 339) an upper 95% relative error of 0.043 for the total Australian population at a ten-year forecast horizon; our corresponding relative error is 0.032 for males and 0.031 for females, indicating narrower prediction intervals (by roughly 25%) for each sex.

The main difference between our approach and those of most other authors is that our forecasts are obtained from a fully specified coherent stochastic model that is estimated from the historical data. There is much less room for subjective assumptions and ad hoc adjustments. Instead, the forecasts and prediction intervals are much more objective because they are dependent to a much greater extent on the available empirical information, although some judgements about the models used need to be made.

Time series methods such as we have used here are most useful for short- to medium-term forecasting. The width of the prediction intervals shows that the

methods become increasingly less informative over time, and as a result we have chosen to restrict our forecasts to 20 years.

This application of functional data models to fully stochastic population forecasting is a first exercise using data that are of good quality and well-behaved. Possible extensions of this research include extensions to other countries with different population histories, especially with respect to migration. The dependencies between male and female mortality and between male and female migration also need to be addressed, through, for example, restriction to non-divergent futures (see, for example, Li and Lee, 2005). Research into the demographic interpretation of the main principal components with a view to constraining forecast coefficients may also improve forecasting accuracy. Other possible extensions include taking into account cohort effects, modelling fertility by parity, and modelling interactions

between components, such as (past) fertility and (current) immigration.

Although the methods might seem relatively complicated, their implementation is relatively easy using the demography package for R (Hyndman, 2006), provided the historical births, deaths and population data are available in the same form as is used in Human Mortality Database (2006). Instructions on the use of the package are provided at www.robhyndman.info/Rlibrary/demography.

Acknowledgement

This paper was part of the International Institute of Forecasters' workshop on "Stochastic Demographic Forecasting" held at Universidad de Salamanca, July 2005. The authors would like to thank José Antonio Ortega Osona for organizing the workshop, and the contributors to the workshop for comments on this paper.

Appendix A. Population simulation equations

A.1. Notation

$B_t(x)$ = Births in calendar year t to females of age x

$D_t(x)$ = Deaths in calendar year t of persons of age x

$P_t(x)$ = Population of age x at 1 January, year t

$E_t(x)$ = Population of age x exposed to risk (30 June), year t

$D_t(x, x+1)$ = Cohort deaths in calendar year t of persons aged x at the beginning of year t

$D_t(B, 0)$ = Cohort deaths in calendar year t of births during year t

$G_t(x, x+1)$ = Net migration in calendar year t of persons aged x at the beginning of year t

$G_t(B, 0)$ = Net migration in calendar year t of births during year t

$m_t(x) = D_t(x)/E_t(x)$ = Central death rate for age x in year t

$f_t(x) = B_t(x)/E_t^F(x)$ = Fertility rate for females of age x in year t

$r_t(x)$ = Life table survivorship ratios for $x=0, 1, 2, \dots, p-2$

$r_t(B)$ = Life table survivorship ratio from birth to age 0

$r_t(p-1^+) =$ Life table survivorship ratio from $p-1^+$ to p^+

$R_t(x)$ = Population adjusted for the first half of migration

where $t=1, \dots, n$, $x=0, 1, 2, \dots, p-1, p^+$, and p^+ denotes the open-ended age group p and above. The superscripts M and F denote males and females respectively. Life table survivorship ratios are computed from $m_t(x)$ using standard methods.

ρ = male : female sex-ratio at birth; $B_t^F = \frac{1}{1+\rho} B_t$ and $B_t^M = \frac{\rho}{1+\rho} B_t$. \bar{A} indicates an estimate of A .

A.2. Algorithm for obtaining $P_{t+1}(x)$ given $P_t(x)$, $m_t(x)$, $f_t(x)$ and $G_t(x, x+1)$

$$\begin{aligned}
 R_t(x) &= P_t(x) + G_t(x, x+1)/2 & x = 0, \dots, p-2 \\
 R_t(p-1) &= P_t(p-1) + G_t(p-1^+, p^+)/4 \\
 R_t(p^+) &= P_t(p^+) + G_t(p-1^+, p^+)/4 \\
 \bar{D}_t(x, x+1) &= r_t(x)R_t(x) & x = 0, \dots, p-2 \\
 \bar{D}_t(p-1^+, p^+) &= r_t(p-1^+)[R_t(p-1) + R_t(p^+)] \\
 R_{t+1}(x+1) &= R_t(x) - \bar{D}_t(x, x+1) & x = 0, \dots, p-2 \\
 R_{t+1}(p^+) &= R_t(p-1) + R_t(p^+) - \bar{D}_t(p-1^+, p^+) \\
 \bar{E}_t(x) &= [R_t(x) + \bar{R}_{t+1}(x)]/2 & x = 1, \dots, p-1, p^+ \\
 \bar{D}_t(x) &= m_t(x)\bar{E}_t(x) & x = 1, \dots, p-1, p^+ \\
 D_t(x) &\sim \text{Poisson}(\bar{D}_t(x)) & x = 1, \dots, p-1, p^+ \\
 D_t(x, x+1) &= [D_t(x) + D_t(x+1)]/2 & x = 1, \dots, p-2 \\
 D_t(p-1^+, p^+) &= [D_t(p-1)/2 + D_t(p^+)] \\
 R_{t+1}(x+1) &= R_t(x) - D_t(x, x+1) & x = 1, \dots, p-2 \\
 R_{t+1}(p^+) &= R_t(p-1) + R_t(p^+) - D_t(p-1^+, p^+) \\
 B_t(x) &\sim \text{Poisson}(f_t(x)[R_t^F(x) + R_{t+1}^F(x)]/2) & x = 15, \dots, 49 \\
 B_t &= \sum_{x=15}^{49} B_t(x) \\
 B_t^M &\sim \text{Binomial}(B_t, \rho/(\rho+1)) \\
 B_t^F &= B_t - B_t^M \\
 R_t(B) &= B_t + G_t(B, 0)/2 \\
 \bar{D}_t(B, 0) &= r_t(B)R_t(B) \\
 \bar{R}_{t+1}(0) &= R_t(B) - \bar{D}_t(B, 0) \\
 \bar{E}_t(0) &= [R_t(0) + \bar{R}_{t+1}(0)]/2 \\
 \bar{D}_t(0) &= m_t(0)\bar{E}_t(0) \\
 D_t(0) &\sim \text{Poisson}(\bar{D}_t(0)) \\
 f_0 &= \bar{D}_t(B, 0)/\bar{D}_t(0) \\
 D_t(B, 0) &= f_0 D_t(0) \\
 D_t(0, 1) &= (1 - f_0)D_t(0) + D_t(1)/2 \\
 R_{t+1}(0) &= R_t(B) - D_t(B, 0) \\
 R_{t+1}(1) &= R_t(0) - D_t(0, 1) \\
 P_{t+1}(0) &= R_{t+1}(0) + G_t(B, 0)/2 \\
 P_{t+1}(x+1) &= R_{t+1}(x+1) + G_t(x, x+1)/2 & x = 0, \dots, p-1, p^+
 \end{aligned}$$

References

- Alho, J. M. (1992). Population forecasting theory, methods and assessments of accuracy: the magnitude of error due to different vital processes in population forecasts. *International Journal of Forecasting*, 8, 301–314.
- Alho, J. M., & Spencer, B. D. (1985). Uncertain population forecasting. *Journal of the American Statistical Association*, 80, 306–314.
- Australian Bureau of Statistics (2003). *Population projections, Australia: 2002–2101 Catalogue, Vol. 3222.0* Canberra: Australian Bureau of Statistics.
- Australian Bureau of Statistics (2006). *Population projections, Australia: 2004–2101 reissue Catalogue, Vol. 3222.0* Canberra: Australian Bureau of Statistics.
- Betz, F., & Lipps, O. (2004). Stochastic population projection for Germany based on the QS-approach to modelling age-specific fertility rates. *Tech. Rep. 59-2004, Mannheim Research Institute for the Economics of Aging*.
- Bongaarts, J., & Bulatao, R. A. (Eds.) (2000). *Beyond six billion: forecasting the world's population*. Washington, D.C.: National Academy Press.

- Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22(3), 547–581.
- Booth, H., Hyndman, R. J., Tickle, L., & de Jong, P. (2006). Lee–Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, 15(9), 289–310.
- Booth, H., Maindonald, J., & Smith, L. (2002). Applying Lee–Carter under conditions of variable mortality decline. *Population Studies*, 56(3), 325–336.
- Booth, H., Tickle, L., & Smith, L. (2005). Evaluation of the variants of the Lee–Carter method of forecasting mortality: a multi-country comparison. *New Zealand Population Review*, 31(1), 13–34.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B*, 26(2), 211–252.
- Bozik, J. E., & Bell, W. R. (1987). Forecasting age-specific fertility using principal components. *Proceedings of the American Statistical Association, Social Statistics Section* (pp. 396–401). Alexandria, Virginia: American Statistical Association.
- Brillinger, D. R. (1986). The natural variability of vital rates and associated statistics. *Biometrics*, 42, 693–734.
- Brouhns, N., Denuit, M., & Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics & Economics*, 31(3), 373–393.
- Chandola, T., Coleman, D. A., & Hiorns, R. W. (1999). Recent European fertility patterns: fitting curves to distorted distributions. *Population Studies*, 53, 317–329.
- De Beer, J. (1997). The effect of uncertainty of migration on national population forecasts: the case of the Netherlands. *Journal of Official Statistics*, 13, 227–243.
- De Jong, P., & Tickle, L. (2006). Extending Lee–Carter mortality forecasting. *Mathematical Population Studies*, 13(1), 1–18.
- Erbas, B., Hyndman, R. J., & Gertig, D. M. (2007). Forecasting age-specific breast cancer mortality using functional data models. *Statistics in Medicine*, 26(2), 458–470.
- George, M. V., & Perreault, J. (1992). Methods of external migration projections and forecasts. In N. Keilman, & H. Crujisen (Eds.), *National population forecasting in industrialized countries* (pp. 87–103). Amsterdam: Swets and Zeitlinger.
- Giroi, F., & King, G. (2006). *Demographic forecasting*. Cambridge: Cambridge University Press.
- Hilderink, H., Van der Gaag, N., Van Wissen, L., Jennissen, R., Román, A., Salt, J., Clarke, J., & Pinkerton, C. (2002). *Analysis and forecasting of international migration by major groups*. Part III, working papers and studies 3/2002/E/n17. The Hague: Eurostat.
- Human Mortality Database (2006). *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)*, downloaded on 1 May 2006 URL: www.mortality.org
- Hyndman, R. J. (2006). *demography: Forecasting mortality and fertility data, R package version 0.97* URL: www.robhyndman.info/Rlibrary/demography
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439–454.
- Hyndman, R. J., & Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51, 4942–4956.
- Keilman, N., & Pham, D. Q. (2004). Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European economic area. *Discussion Papers Research Department of Statistics Norway, Vol. 386*. (pp.) URL: <http://www.ssb.no/publikasjoner/Dp/pdf/dp386.pdf>
- Keilman, N., Pham, D. Q., & Hetland, A. (2002). Why population forecasts should be probabilistic — Illustrated by the case of Norway. *Demographic Research*, 6, 409–453.
- Lee, R. D. (1993). Modeling and forecasting the time series of U.S. fertility: age distribution, range, and ultimate level. *International Journal of Forecasting*, 9, 187–202.
- Lee, R. D. (1999). Probabilistic approaches to population forecasting. In W. Lutz, J. W. Vaupel, & D. A. Ahlburg (Eds.), *Frontiers of population forecasting*. Supplement to Population and Development Review, Vol. 24. (pp. 156–190) New York: Population Council.
- Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87, 659–675.
- Lee, R. D., & Miller, T. (2001). Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography*, 38(4), 537–549.
- Lee, R. D., & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association*, 89, 1175–1189.
- Li, N., & Lee, R. D. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography*, 42(3), 575–594.
- Miller, T. (2003) California's uncertain population future. Technical appendix. In R.D. Lee, T. Miller & R. Edwards, Special Report: the growth and aging of California's population: demographic and fiscal projections, characteristics and service needs. Center for the Economics and Demography of Aging. CEDA Papers: Paper 2003-0002CL.
- Miller, T., & Lee, R. D. (2004). A probabilistic forecast of net migration to the United States. *Report I in Stochastic infinite horizon forecasts for social security and related studies*. Working paper, Vol. 10917. National Bureau of Economic Research URL: <http://www.nber.org/papers/w10917>
- Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570), 1029–1031.
- Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: measuring and modelling population processes*. Oxford: Blackwell.
- Renshaw, A. E., & Haberman, S. (2003a). On the forecasting of mortality reduction factors. *Insurance: Mathematics & Economics*, 32(3), 379–401.
- Renshaw, A. E., & Haberman, S. (2003b). Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics & Economics*, 33(2), 255–272.
- Renshaw, A. E., & Haberman, S. (2003c). Lee–Carter mortality forecasting: a parallel generalized linear modelling approach for

- England and Wales mortality projections. *Applied Statistics*, 52 (1), 119–137.
- Rogers, A. (1990). Requiem for the net migrant. *Geographical Analysis*, 22, 283–300.
- Rogers, A., & Castro, L. (1981). *Model migration schedules. Research Report 81–30* Laxenburg, Austria: International Institute for Applied Systems Analysis.
- Rogers, A., Castro, L. J., & Lea, M. (2005). Model migration schedules: Three alternative linear parameter estimation methods. *Mathematical Population Studies*, 12, 17–38.
- Sykes, Z. M. (1969). Some stochastic versions of the matrix model for population dynamics. *Journal of the American Statistical Association*, 44, 111–130.
- United Nations (2001). Replacement migration: is it a solution to declining and aging populations? No. *ST/ESA/SER.A/206*, United Nations, New York.
- Wilmoth, J. R. (1993). Computational methods for fitting and extrapolating the Lee–Carter model of mortality change. *Technical report, Department of Demography, University of California, Berkeley* URL: <http://www.demog.berkeley.edu/jrw/Papers/LCtech.pdf>
- Wilson, T., & Bell, M. (2004). Australia's uncertain demographic future. *Demographic Research*, 11(8), 195–234.