Journal of Official Statistics, Vol. 13, No. 3, 1997, pp. 279–303

Comparing and Assessing Time Series Methods for Forecasting Age-Specific Fertility and Mortality Rates

William R. Bell[1]

Fertility and mortality rates exhibit strong age patterns, and various researchers have developed methods to capture this structure and use it in forecasting. Two general approaches have been developed. The *curve fitting* approach involves fitting parametric curves to the age-specific rates. The *principal components* approach involves computing principal components to obtain a linear transformation of the data with simplified structure. This article reviews and compares proposed alternative versions to these two approaches, and then evaluates the out-of-sample performance of the various alternatives in forecasting age-specific U.S. white male and female mortality rates. None of the approaches tried produced short-term forecasts more accurate than those obtained from a simple random walk with drift model applied to the rates for each age separately. Also, the curve fitting approach and low-dimensional principal components approaches clearly require a bias adjustment to avoid having approximation errors compromise short-term forecast accuracy.

Key words: Population projection; demographic forecasting; principal components.

1. Introduction

Population projections often involve the forecasting of age-specific fertility and mortality rates, which is a forecasting problem of large dimension. Age detail may be used for fertility rates to take advantage of the known age structure of the existing female population when forecasting births, or because age-specific fertility forecasts are of inherent interest. Age-specific mortality forecasts are needed to achieve age detail in the population projections. A characteristic feature of age-specific fertility and mortality rates is the smooth shape over age of the data each year. It seems desirable to use modelling and forecasting methods that capture this smooth shape over age, both to produce forecasts that appear internally consistent, and in the hope that this will improve the accuracy of short-term forecasts.

Several researchers have developed methods for capturing the smoothness of fertility and mortality rates over age and using this in forecasting. This article reviews two general approaches that have been explored. The *curve fitting* approach fits a parametric curve (a

[1] Statistical Research Division, U.S. Bureau of the Census, Washington, DC 20233-9100, U.S.A. e-mail: wbell@census.gov.
Acknowledgments: This article reports the general results of research undertaken by U.S. Census Bureau staff. The views expressed are attributed to the author and do not necessarily reflect those of the U.S. Census Bureau. I gratefully acknowledge the assistance of Brian Monsell with production of the principal components method forecasts, of Mark Otto with the forecast evaluations, and of Matthew Kramer with translating/summarizing for me two cited references in French. Comments by Stuart Scott and two referees materially improved the article. Any errors remain my own responsibility.

© Statistics Sweden

function of age) to approximate each year's fertility or mortality rates in terms of a set of curve "parameters." One then forecasts these parameters, thus producing forecasted curves, which are then taken (possibly with some modification) as forecasts of the rates themselves. The second approach uses the dimensionality reduction technique of *principal components* from multivariate analysis to linearly transform the mortality or (relative) fertility rates (generally after taking logarithms). The resulting principal component time series provide approximations to each year's fertility or mortality rates, with the accuracy of the approximations depending on the number (J) of principal components that are used. Having selected J, the first J principal component time series can be forecast, and these forecasts transformed back to forecasts of the fertility or mortality rates. Alternatively, aided by the simplified structure of the principal components time series, it can be feasible to forecast *all* the components and transform these back to forecasts of the rates without the approximation error. For the forecasting parts of both the curve fitting and principal components approaches, much of the recent literature has used univariate or multivariate ARIMA (autoregressive-integrated-moving average) time series models (Box and Jenkins 1976; Tiao and Box 1981).

This article focuses on comparing and assessing specific variants of the curve fitting and principal components approaches to forecasting age-specific fertility and mortality rates. Section 2 gives as background a brief review of some specific problems that motivate the search for such techniques. Section 3 discusses the curve fitting approach, focusing on the methods of Thompson et al. (1989); McNown and Rogers (1989); and Knudsen, McNown, and Rogers (1993). Section 4 discusses the approaches of Bozik and Bell (1987), Bell and Monsell (1991), Lee and Carter (1992), and Lee (1993) that make use of principal components. Section 5 then discusses the general advantages and disadvantages of the curve fitting and principal components approaches. Section 6 presents results of an empirical study comparing out-of-sample forecast performance for alternative variants of the curve fitting and principal components approaches to forecasting age-specific mortality rates. The data used in this study are central death rates for (mostly) 5-year age groups for U.S. white males and white females from 1940–1991. Section 7 provides conclusions and some suggestions for future research.

2. Problems in Forecasting Age-Specific Fertility and Mortality Rates

Figures 1 and 2 illustrate the character of the U.S. white age-specific fertility and mortality rate data. (The latter are analyzed in Section 6.) Figure 1.a shows white fertility rates for three years – 1927, 1957, and 1977. The three curves show similar shapes but quite different levels. The corresponding total fertility rates (TFRs), defined as the sum of all the age-specific fertility rates for a given year, were 2.78, 3.58, and 1.71. Figure 1.b shows the relative fertility rates obtained by dividing the fertility rates in Figure 1.a by the corresponding TFRs. By removing the effect of different levels, this more clearly shows the basic similarity in shape of the three curves, while also showing those differences in shape that evolved over the years shown. The data for 1927 look the most different, but some of the irregularities in the rates at the higher ages are probably due to errors in the age-specific female population estimates (denominators of the rates) for 1927.

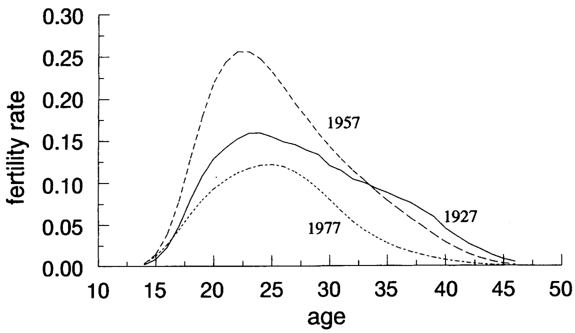

Fig. 1.a. U.S. white fertility rates

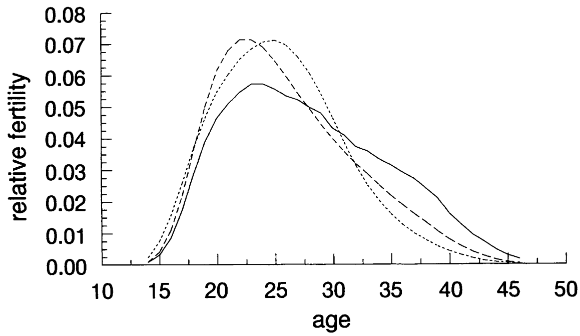

Fig. 1.b. U.S. white relative fertility rates

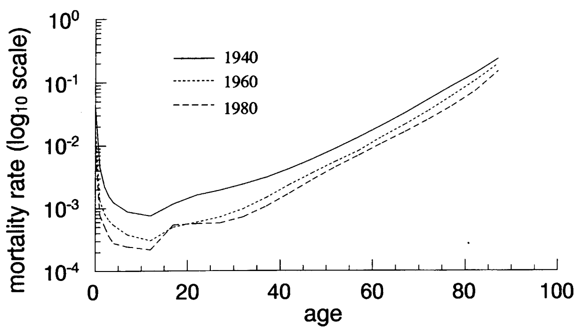

Fig. 2.a. U.S. white female mortality rates

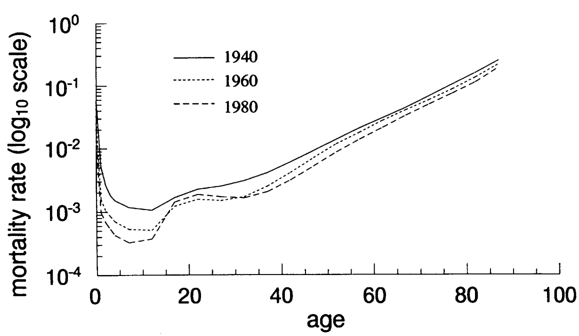

Fig. 2.b. U.S. white male mortality rates

Figures 2.a and 2.b show base 10 logarithms of the mortality rates for U.S. white females and males, respectively, for three years (1940, 1960, and 1980). The similarity in shape over the years of both the female and male mortality curves is evident, as is the general downward trend of mortality. Also, some changes over time in the shapes of the curves can be detected, particularly the increasing importance of the ''accident hump,'' the sharp increase in mortality at ages 15–19, in both the female and male mortality rates. (Note: Figure 2 actually shows central death rates rather than mortality rates, the distinction between which is discussed in Section 6. As this distinction is mostly unimportant for this article, the term ''mortality rate'' is used for either of the two.)

The highly structured nature of these data almost begs for a methodology that can capture this structure and use it in forecasting. Methods that have been proposed for doing so are discussed in Sections 3 and 4. Other forecasting approaches could be and have been pursued, however. Bell (1992) discusses possible problems with three such alternatives.

The first such alternative is analysis of the data on a cohort basis (data indexed by year of birth), rather than on a period basis (data indexed by calendar year, as in Figures 1 and 2). Since completion of cohort fertility requires about 30 years and completion of cohort mortality about 100 years, analysis of rates on a cohort basis presents massive missing data problems, which leads to formidable statistical problems. It also leads to practical problems for forecasting, as noted by Brass (1974). (See also De Beer's (1989) investigation of an age-period-cohort model suggested by Willekens and Baydar.) Further, Bell (1992) notes that U.S. cohort fertility rates do not exhibit the same degree of smoothness over age as the period rates shown in Figure 1.

The second such alternative is to directly develop a multivariate time series model for the complete set of age-specific fertility or mortality rates. This is generally infeasible due to the high dimensionality involved (unless very broad age groups are used) and the high cross correlations between the rates for different ages (a consequence of the smoothness of the rates over ages). Bell (1992) illustrates that this high cross correlation can have apparently unreasonable effects on model parameter estimates.

The third alternative is simply to forecast the fertility or mortality rates for each age separately. A possible drawback is that resulting forecasts may not show the same smooth shape over age as the historical data, particularly in the long-run. The extent to which this is a problem depends on how the rates at different ages are forecast and on the objective of the analysis. For illustration, Bell (1992) forecast U.S. white age-specific relative fertility rates separately at each age using ARIMA (1,1,0) models (first order autoregressive models for the first differenced series) applied to logged data. The resulting long-run forecasted fertility curve showed mild irregularities in shape over age. Using different models at different ages can easily produce much larger shape irregularities. De Beer (1992) obtained significant irregularities in shape from just a three-year-ahead forecast of the Netherlands fertility rates using different spline functions for each age. Irregularities in the shape of the forecast curve do not necessarily mean that the forecasts, taken individually, are bad, even though their joint appearance may seem unreasonable.

A second problem with the approach of separate forecasting at each age is its limitations in regard to measuring expected forecast uncertainty, since it does not directly provide estimates of the correlations of forecast errors at different ages. Given the smooth shapes over age of the fertility and mortality rates, these correlations would be expected to be quite high. Information on cross correlations over age of forecast errors is needed for computing forecast error variances for quantities, such as population, that depend on fertility or mortality at more than one age. (See, e.g., Alho 1992.) This problem might be addressed by simply estimating the correlations of forecast errors at different ages from the univariate model residuals, though the high dimensionality of the data means that there are a large number of correlations to be estimated.

Two other approaches are the use of mortality rates by cause of death, or use of fertility rates for births of first children, second children, etc. (parity-specific fertility rates). De Beer (1989) found mild accuracy gains from use of parity-specific data when forecasting fertility for the Netherlands both with his cohort ARIMA model, and also with ordinary ARIMA models. McNown and Rogers (1992) found little or no gain in forecast accuracy from using cause-specific versus aggregate U.S. mortality data in a curve-fitting approach. Wilmoth (1993) noted several problems with forecasting cause-specific mortality for Japanese data, one being that the behavior of the series that must be forecast (using the Lee-Carter version of the principal components approach discussed in Section 4.2.) was less regular by cause than for all causes combined. Alho (1991) discussed conditions under which aggregate and cause-specific mortality forecasts will yield similar results, observed that these conditions seemed to hold for mortality forecasts from the U.S. Social Security Administration, and also noted conditions under which forecasts based on aggregate mortality data can yield more accurate results than those based on cause-specific data. Finally, note that using cause-specific mortality rates or parity-specific fertility rates greatly increases the dimensionality of the forecasting problem, magnifying the problems the methods discussed in Sections 3 and 4 seek to address.

3. Curve Fitting Approaches

The fitting of curves to fertility and, particularly, mortality rates has a long history in demography and actuarial science, going back to the efforts of De Moivre and Gompertz. The objectives have usually been to estimate fertility or mortality curves with limited data, or to graduate (smooth) irregular curves of directly estimated rates. Hoem et al. (1981) fit a number of curves to Danish fertility data, and reviewed other work in this area. Hartmann (1987) reviews work on fitting mortality curves. Cramér and Wold (1935), as discussed in Section 3.2., developed an early curve fitting approach to forecasting mortality, and reviewed other early work on mortality forecasting. Rogers (1986) proposed an approach to population projection based on fitting curves to age-specific rates of fertility, mortality, marriage, divorce, and remarriage, and then projecting the curve parameters. His parameter projections relied heavily on judgement. More recently, various authors (discussed below) have combined curve fitting with forecasting of curve parameters from time series models.

In more detail, the basic curve fitting approach to forecasting is as follows. Let r_{it} be a set of fertility or mortality rates, or transformations of the same, observed for ages $i = 1, \ldots, K$ and years $t = 1, \ldots, n$. A parametric curve, $g(i, \eta_t)$, where η_t is an $m \times 1$ vector of curve parameters, is postulated as an approximation to the r_{it}. This approximating curve is intended to be used for every year, but with different values of η_t each year.

The curve is fitted to the data separately for each year $t = 1, \ldots, n$, possibly by least squares:

$$\text{find } \eta_t \text{ to min } \sum_i [r_{it} - g(i, \eta_t)]^2 \tag{1}$$

A weighted least squares fitting criterion can also be used. Since $g(i,\eta_t)$ is invariably a nonlinear function of i, such curve fitting requires nonlinear least squares software. Having determined η_t from (1) for $t = 1, \ldots, n$, one treats these as observations on an m-dimensional multivariate time series, and forecasts η_t, possibly using (univariate or multivariate) ARIMA models. The forecasts $\hat{\eta}_{n+\ell}$ of $\eta_{n+\ell}$ for $\ell > 0$ yield forecasted curves $g(i, \hat{\eta}_{n+\ell})$, which are taken as forecasts of the $r_{i,n+\ell}$.

3.1. *Curve fitting and forecasting fertility*

Thompson et al. (1989), and Knudsen, McNown, and Rogers (1993) applied the curve fitting approach to fertility forecasting. Knudsen et al. used a ''double exponential'' curve, which Rogers (1986) had suggested as a simpler alternative to an earlier model of Coale and Trussel (1974). Knudsen et al. fit a reparameterized version of the curve by nonlinear least squares, and forecasted the resulting four curve parameters (after taking logarithms) using univariate ARIMA models. They noted (p. 21) however, that, ''The task of forecasting the model schedule is made difficult by the strong interactions among the four parameters,'' and suggested using multiple time series methods instead.

Thompson et al. (1989) converted the fertility rates f_{it} to the total fertility rate and relative fertility rates: $\text{TFR}_t = \sum_i f_{it}$, and $r_{it} = f_{it}/\text{TFR}_t$. They then fit the following shifted gamma density to the r_{it} (which sum to one over the ages i) by weighted nonlinear least squares:

$$g(i, \alpha, \beta, A_0) = [\Gamma(\alpha)\beta^{\alpha}]^{-1}(i - A_0)^{\alpha - 1} \exp\{-(i - A_0)/\beta\} \tag{2}$$

Weights of four were given to ages 18–32, and one to the other ages, to give more weight in the fit to ages with higher fertility. Thompson et al. used a multivariate ARIMA model to forecast the (logarithms of the) TFR and the mean ($\alpha\beta$) and standard deviation ($\beta\sqrt{\alpha}$) of the gamma curve. The end point of the curve, A_0, was held fixed at 0 in the forecast period.

Thompson et al. show gamma curve fits to the U.S. white fertility rates for 1957 and 1977. While the gamma curves fit well and successfully capture the shape of the curves in both years, there are nevertheless significant deviations of the actual fertility rates from the fitted curves. This is perhaps to be expected when approximating 33 fertility rates by a curve involving only four parameters. Bozik and Bell (1987) show comparable fits for the gamma and the double exponential curve (which also uses four parameters) for a year of U.S. data. Hoem et al. (1981) found the gamma curve and Coale-Trussell models fit Danish fertility data about equally well, and better than the alternatives they tried except for spline functions (discussed below). Bell (1992) noted polynomial functions of age and, particularly, Fourier series, were not as successful at approximating fertility rates as the gamma curve.

Thompson et al. found it desirable to make a ''bias adjustment'' to their gamma curve forecasts to deal with the curve approximation error. This involved taking the deviations of the r_{it}s from the fitted gamma curve in the last year of data, and extrapolating these forward as constant forecasts of the deviations of future r_{it}s from the forecasted gamma curves (effectively assuming random walk models for these deviations). An empirical illustration showed improvements in accuracy from bias adjustment for the first two forecasts, but apparently diminishing effects as the forecast horizon increased. The bias adjustment was motivated by the observations that (i) the discrepancies between the fitted curves and the actual data were significant for some ages, and (ii) the discrepancies tended to persist and change slowly over time, so that they could be forecast in the short run. In the long run (say, beyond ten years) the errors in forecasting the curve parameters are likely to dominate, so that bias adjustment becomes less important. Knudsen, McNown, and Rogers (1993) noted that errors in their curve fits persisted into the forecast period, and that, in fact, ''Errors in approximating the true fertility age profile with the double-exponential model account for most of the forecast errors over the shorter intervals'' (p. 17). Despite this, they made no adjustment to their forecasts to address this problem.

An obvious generalization to the bias adjustment of Thompson et al. would be to develop and use time series models (not necessarily random walks) to forecast the approximation errors. The potential benefits of this generalization are unknown. It also goes against the objective of dimensionality reduction (raising the question, why not just model and forecast the rates directly?), particularly if a multivariate model is sought for the approximation errors.

In the U.S. Census Bureau implementation of the approach of Thompson et al., data on total births were available for the two most recent years, but without age-specific detail. These total birth figures were used to make a proportional adjustment to the TFR forecasts (described in Bell et al. (1988)) that constrained predicted total births to equal actual total births for these years.

3.2. *Curve fitting and forecasting mortality*

Cramér and Wold (1935) fit Makeham curves to Swedish mortality data for ages 30 and above, and then extrapolated (most of) the curve parameters using fitted logistic functions (in one case, a linear function was used). They did this on both a period and cohort basis. More recent authors have fitted curves to mortality rates for all ages. McNown and Rogers (1989) used the Heligman and Pollard (1980) curve for mortality rates q_i:

$$\begin{aligned} q_i &= Q_i/(1+Q_i) \\ Q_i &\approx A^{(i+B)^C} + D\exp\{-E[\log(i)-\log(F)]^2\} + GH^i \end{aligned} \tag{3}$$

where the t subscript is omitted in (3) for simplicity. Rogers (1986) used a modified version of (3). The three terms in the expression for Q_i in (3) model childhood mortality, the accident hump, and adult through old age mortality. The middle term is dropped for $i = 0$. Figure 3 (see Section 4.2.) shows the Heligman-Pollard mortality curve fit to 1980 U.S. white female mortality rates. While the fit appears good, this should be expected from an eight-parameter curve fitted to 22 data points. Note that this achieves much less dimensionality reduction than the 4-parameter gamma or double exponential curves achieve for 33 fertility rates.

McNown and Rogers (1989) fitted curves (3) to U.S. male and female mortality data, and forecast the eight curve parameters for each sex (after taking logs) using univariate ARIMA models. McNown and Rogers (1992) used a nine parameter ''multiexponential model'' to fit U.S. male and female cause-specific mortality rates. To deal with the high dimensionality, for each cause they held six shape parameters constant at historical average values, and explicitly forecast only (logs of) three level parameters using univariate ARIMA models.

3.3. *Related approaches*

There are other methods of analyzing fertility and mortality curves over age that will not be considered in detail here, though they are related to the curve fitting approach just described. These would include the fertility model of Coale and Trussell (1974) and the relational models of Brass (1974), both of which make use of ''standard'' age schedules. These models were not necessarily designed for forecasting, but for other purposes such as estimating fertility or mortality curves from limited data, or smoothing irregular curve estimates. In analysis with Danish fertility data, Hoem et al. (1981) found the gamma curve to give comparable fits to the Coale-Trussell model (as noted earlier), and better fits than the Brass models. Another approach closely related to the curve fitting approach is the fitting of spline functions over age. These were found by Hoem et al. (1981) to yield the best fits to the Danish fertility data. Splines, however, have the disadvantage that they generally involve a large number of parameters with no interpretation.

4. Principal Components Approaches

Bozik and Bell (1987) developed a principal components (PC) approach for time series forecasting of age-specific fertility rates. This approach was extended by Bell and Monsell (1991), where it was used in forecasting age-specific mortality rates. Lee and Carter (1992) (see also Carter and Lee 1992) explored a modified version of this approach for forecasting mortality rates, which Lee (1992, 1993) also applied to fertility rates. Actually, application of principal components analysis to mortality rates goes back at least to Ledermann and Breas (1959), who did their analysis not with time series data, but with life table data from many countries, both developed and developing. Sivamurthy (1987) pursued a similiar analysis with fertility rate data from different countries. Le Bras and Tapinos (1979) used the three principal component approximation suggested by Ledermann and Breas (1959) in forecasting mortality rates for France, although details of their forecasting methodology were not specified.

4.1. *The approaches of Bozik and Bell (1987) and Bell and Monsell (1991)*

In outline, the general approach is as follows. Let the r_{it} be as in Section 3, and let $r_t = (r_{1t}, \dots, r_{Kt})'$. Suppose a linear approximation of dimension $J \leq K$ is to be used to approximate the r_t. Such an approximation is defined by a $K \times J$ matrix Λ whose columns span the J-dimensional approximating space. Given Λ, the approximation to r_t each year is $\Lambda\hat{\beta}_t$, where $\hat{\beta}_t$ is obtained by least squares regression of r_t on Λ. The $\hat{\beta}_t$ for $t = 1, \dots, n$ are considered as a J-variate time series, which can be forecast using a time series model or other techniques. Having obtained these forecasts $\hat{\beta}_{n+\ell}$, the corresponding forecasts of $r_{n+\ell}$ are $\Lambda\hat{\beta}_{n+\ell}$.

It remains to determine Λ for given J. First, the columns of Λ are restricted to be orthonormal. This implies that $\hat{\beta}_t = (\Lambda'\Lambda)^{-1}\Lambda' r_t = \Lambda' r_t$, and the approximation to r_t is $\hat{r}_t = \Lambda\Lambda' r_t$. If $J = K$ then Λ is a $K \times K$ orthogonal matrix, so $\Lambda\Lambda' = I$ and $\hat{r}_t = r_t$ so there is no approximation error. Next, we ask what set of J vectors (columns of Λ) provide the approximation that minimizes the aggregate error, $\sum_t ||r_t - \hat{r}_t||^2$? The answer is that the columns of Λ are the first J principal component vectors (the J eigenvectors corresponding to the J largest eigenvalues) of the sum of squares and cross products matrix of the data, $\sum_t r_t r_t'$. Hence the term "principal components approach." One can also define weighted PCs corresponding to a weighted least-squares criterion that gives more weight to certain ages in determining the approximation (Bozik and Bell 1987).

We should distinguish the PC approach with $J < K$ from that with $J = K$. When $J < K$, the J PC series are taken to provide a reduced dimension approximation to the original K series, and the approximation error is typically ignored. Thus, Bozik and Bell (1987) used a four PC approximation for the logarithms of the single-year-of-age U.S. white relative fertility rates, and developed a five-variate times series model to forecast TFR along with the four PC series. One goal of the analysis reported there was to determine how many PCs were needed to provide a "sufficiently accurate" approximation, though no firm conclusion was reached. In contrast, when $J = K$ there is no dimensionality reduction, but the PC approach yields K time series with simplified structure. In particular, the PC series are much less cross-correlated than the original series. (Complete removal of cross-correlation via PCs is not achieved when the data are autocorrelated, as it is in the i.i.d. case.) This makes development of a multivariate model for *all* K PC series feasible. Thus, Bell and Monsell (1991) developed a time series model for all 22 PC time series obtained from the data on (log) U.S. white female mortality rates described in Section 6. Inverting the PC transformation (multiplying by Λ) then yields a multivariate time series model for the complete r_t vector. Contrast this with the discussion in Section 2 about the infeasibility of directly developing a multivariate model for r_t.

Figure 3 (see Section 4.2.) illustrates the one, two, and four PC approximations to the 1980 U.S. white female mortality rates. Comparisons of the accuracy of PC and HP approximations to the U.S. mortality data are made in Section 5.

Since the PC time series β_t are a linear transformation of the original series r_t, it is easy to translate their prediction error variances and covariances into prediction error variances for the rates themselves. Normal theory prediction intervals for the elements of r_t can thus be obtained. This is illustrated in Bozik and Bell (1987) and Bell and Monsell (1991). Alho (1992) used simplified versions of these models to produce prediction standard errors for U.S. population forecasts. Lee and Carter (1992) also discussed use of their approach (discussed next) for producing prediction intervals for mortality rates. Except with the approach of Bell and Monsell (1991) that uses all the PCs, such intervals ignore PC approximation error, which contributes to forecast error. With their approach Lee and Carter (1992) found the contribution of approximation error to be important for short term (less than ten years) prediction intervals of mortality rates for specific age groups, less important for prediction intervals for life expectancy, and unimportant for long-term prediction intervals.

4.2. The approach of Lee and Carter (1992)

Lee and Carter (1992) used principal components in modeling and forecasting (log) mortality rates for mostly 5-year age groups for the entire U.S. population (not for subgroups defined by race and sex). The approach was applied to corresponding data for males and females in Carter and Lee (1992), and to fertility rates for 5-year age groups in Lee (1992, 1993). The Lee-Carter approach differs from those of Bozik and Bell (1987) and Bell and Monsell (1991) in the following ways:

1. They first subtracted out age-specific means – the average (log) mortality curve over the history of the data.
2. They used a one PC approximation.
3. Given Λ, now a $K \times 1$ vector, they obtained the univariate time series β_t (k_t in their notation) by requiring the number of deaths in year t implied by the approximation to equal the actual number of deaths (or births, in the case of fertility).

They also suggested use of the singular value decomposition to perform the computations; this is merely a different way of computing the same quantities as PCs.

Given that a one-PC approximation is to be used, subtracting out the age-specific means is definitely recommended. This improves the approximation without requiring the forecasting of additional parameters, since the age-specific means are treated as constant over time, and so are merely added back to the forecasts of the mean-corrected data. Since use of two PCs provides, by definition, the best two-dimensional linear approximation, the Lee and Carter (1992) approximation has an accuracy that lies between the accuracy provided by one and by two PCs obtained without removing means. This is illustrated for the 1980 U.S. white female mortality rates in Figure 3, which shows all these approximations. (See also the comparison in Table 1 of Section 5.) Subtracting out means is useful with low-dimensional PC approximations, but it becomes less important the more PCs are used, and is unnecessary if all the PCs are used as in Bell and Monsell (1991).

Lee (1992) noted that determining k_t as in item 3 above has a couple of advantages, including permitting the determination of k_t for years for which only deaths (or births, in the case of fertility), and not age-specific rates, are available. Lee and Carter (1992) used this feature to extend their U.S. mortality data both forward and backward before modelling and forecasting k_t, and Lee (1993) did the same for fertility rates. This is similar to the proportional adjustment of TFR mentioned in Section 3.1. that was used by Thompson et al. (1989) (as detailed in Bell et al. 1988). This idea could also be applied to the PC approaches of Bozik and Bell (1987) and Bell and Monsell (1991).

One additional aspect to the approach used by Lee (1993) for fertility was to transform his fertility index (essentially, TFR) by the logistic transformation over a specified interval (he used (0,4)) to constrain point forecasts and prediction intervals. This technique was also suggested by Thompson (1989) and Alho (1990). In the U.S. Census Bureau implementation of Thompson et al. (1989), the transformation log(TFR−1) was used to constrain TFR forecasts and prediction intervals to lie above one.

The Lee-Carter approach has the advantage of requiring the forecasting of only a single time series (the first PC after the means are subtracted), and the corresponding disadvantage of more approximation error than if more PCs are taken as in Bozik and Bell (1987) or Bell and Monsell (1991). Two alternative simple modifications to the Lee-Carter approach

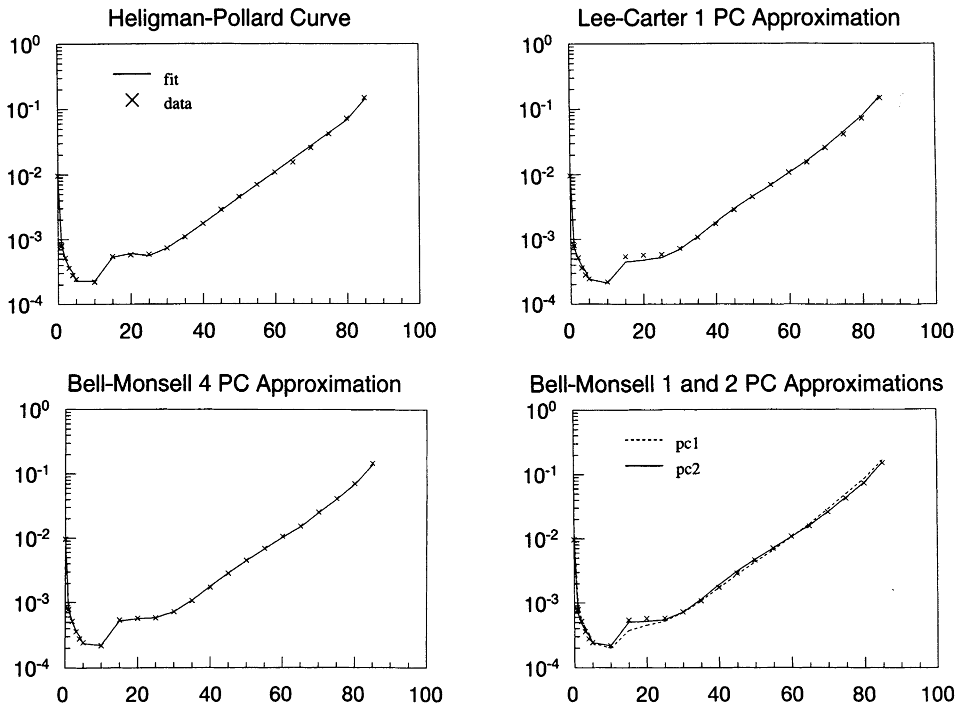

Fig. 3. Approximations to 1980 U.S. white female mortality rates

could avoid this drawback. One is to apply the "bias adjustment" procedure of Thompson et al. (1989). This is equivalent to forecasting all PCs after the first using random walk models (forecasts constant at the last observed values). The other possibility is to modify step 1 above by subtracting out r_n (the logged rates for the last year of data) rather than the mean of r_t over the years $1, \ldots, n$, a possibility noted by Lee and Carter (1992, pp. 665–666). This latter approach produces no approximation error in the last year of data, and hence no need for bias adjustment. (The value in year n of all the PC series is then zero, so that, in this case, ignoring all PCs other than the first is equivalent to forecasting them using random walk models. These results would not appear to be the same as when applying the bias adjustment to mean corrected data, however, since the computed PCs will be different.)

It is instructive to examine the form of the Lee-Carter forecasts with and without bias adjustment. Their one PC approximation to central death rates m_{it} (defined in Section 6) is $\log(m_{it}) = r_{it} \approx a_i + b_i k_t$, where the a_i are the sample means over $t = 1, \ldots, n$ of the r_{it}, and the b_i are the elements of the first PC vector computed using the data $r_{it} - a_i$. The forecast in year $n + \ell$ of the mortality index k_t is obtained from a random walk model with drift, so $\hat{k}_{n+\ell} = k_n + \hat{c}\ell$, where $\hat{c}$ is the estimated drift parameter. Thus, the forecast of $m_{i,n+\ell}$ is

$$\hat{m}_{i,n+\ell} = \exp\{a_i + b_i(k_n + \hat{c}\ell)\} = \hat{m}_{in}\rho_i^{\ell} \qquad (4)$$

where $\hat{m}_{in} = \exp(a_i + b_i k_n)$ is the one PC approximation to the death rates in the last year

(n) of data, and $\rho_i = \exp(b_i\hat{c})$. It turns out that $\hat{c} < 0$ and $b_i > 0$ for all i, so $\rho_i < 1$, which means that (4) yields forecasts $\hat{m}_{i,n+\ell}$ that decay geometrically from the $\hat{m}_{in}$. The approximation errors (biases) in the last year are

$$\text{bias}_i = \log(m_{in}) - \log(\hat{m}_{in}) = \log(m_{in}) - (a_i + b_i k_n)$$

The bias adjusted forecasts in the log-scale are $\log(\hat{m}_{i,n+\ell}) + \text{bias}_i$, for all $\ell > 0$, and the resulting bias adjusted forecasts ($\tilde{m}_{i,n+\ell}$) of the death rates themselves can be shown to be

$$\tilde{m}_{i,n+\ell} = m_{in}\rho_i^{\ell}$$

Thus, the bias-adjusted forecasts decay at the same rate (ρ_i) as the original forecasts, but from the death rates in the last year (m_{in}) rather than from the one PC approximation to these rates ($\hat{m}_{in}$). Similar expressions could be obtained for other models or procedures that might be used to forecast k_t. To the extent that error in the one PC approximation is important and persists over time, the bias adjustment should thus improve short-term forecasts.

5. Relative Advantages and Disadvantages of the Curve Fitting and Principal Components Approaches

The primary advantages to the curve fitting approach, relative to the principal components approach, are its familiarity and interpretability. Rogers (1986) and McNown and Rogers (1989) discuss how fitting curves to age-specific fertility or mortality rates can summarize the information in the data in terms of a few demographically meaningful parameters. As noted earlier, this advantage is better realized in fitting a four parameter curve to fertility rates for 30 plus ages than in fitting an eight or nine parameter curve to mortality rates for 20 plus age groups. Also, the high correlations between estimates of the Heligman-Pollard curve parameters (Hartman 1987) compromises their interpretability. The corresponding disadvantage to the principal component approach is that the resulting time series would seem to have no natural interpretation beyond their ability to provide approximations to the underlying rates. This is not entirely true, however, depending on how the approach is applied. For fertility rates Bozik and Bell (1987) model TFR plus principal component series for the relative fertility rates, and Lee (1993) defines his one principal component series so that it effectively approximates the TFR. Also, Lee and Carter (1992) discuss the demographic interpretation of their approach to forecasting mortality rates.

A minor advantage to the curve fitting approach is that the curves guarantee positive rate forecasts (perhaps with some restrictions on the curve parameters). This can be achieved in the principal components approach by taking logarithms of the data. It would also be necessary to apply any bias adjustment in either approach to the logged data (a proportional bias adjustment) to guarantee positivity.

The major disadvantage to the curve fitting approach is the approximation error involved. Table 1 gives aggregate mean absolute per cent errors (MAPEs, with averages taken over both the 52 years and 22 age groups of the data) for various approximations to U.S. white male and female mortality rates. (Outlier adjustments were made to the data as discussed in Section 6.) Similar results, though with larger numbers, were obtained for root mean squared errors (RMSEs). Notice that the Heligman-Pollard curve (HP) is overall more accurate than the Lee-Carter approximation (LC), whose accuracy lies between that of the Bell-Monsell one and two principal component approximations (PC1 and

Table 1. Aggregate mean absolute per cent error in approximations to U.S. white mortality rates

	HP	LC	PC1	PC2	PC3	PC4	PC5
Females	4.1	4.6	8.3	3.5	2.5	2.0	1.7
Males	3.8	5.1	10.3	4.4	2.3	1.9	1.6

Explanatory note: For each approximation, the aggregate MAPE (mean absolute per cent error) is defined as MAPE $= 100/(52 \times 22) \sum_{t=1941}^{1991} \sum_{i=1}^{22} |1 - \hat{m}_{it}/m_{it}|$. The notation of the column headings is as follows: HP = Heligman-Pollard curve, LC = Lee-Carter one principal component approximation (with means removed), PC1–PC5 = Bell-Monsell one to five principal component approximations (without means removed).

PC2), as noted in Section 4.2. Interestingly, the HP and three or more PC approximations are slightly more accurate for males than females; while the LC, PC1, and PC2 approximations are more accurate for females than for males. MAPEs (and RMSEs) for individual age groups vary over age by factors of usually two or less, though this variation is somewhat larger for PC1, PC2, and LC. An extreme case is age zero for HP, for which the approximation error is near zero. Heligman and Pollard (1980, p. 50) report that the B parameter of (3) has little effect on the fitted curve at ages above zero, implying that this parameter can be adjusted each year so the curve accurately reproduces the observed q_{0t}, which would explain the very low MAPE at age zero. Hartmann (1987, p. 34) relegates absence of effects of B to ages beyond one, but notes C in (3) is also closely related to q_{0t}.

Note that using only three PCs yields substantially more accurate approximations to the U.S. mortality rates than the eight parameter HP curve. (Also note that fitting the HP curve to all 52 years of data determines 416 parameters, while computing three orthonormal PC vectors, and their associated coefficients for each year, determines 216 total unknowns.) Bozik and Bell (1987) similarly noted the superiority of low-dimensional PC approximations to fertility rates relative to the gamma curve or double exponential fits. Furthermore, the accuracy of PC approximations is controllable by using more PCs, and no approximation error is involved if all PCs are forecast as in Bell and Monsell (1991).

In Section 6 we will see that the approximation error in the curve-fitting and low-dimensional PC approximations compromises the accuracy of short-term forecasts. This problem can be at least partly addressed through the bias adjustment procedure of Thompson et al. (1989). Though such a bias adjustment seems to have been used only in the gamma curve approach of Thompson et al. for forecasting fertility, there is no reason it cannot be used with other approximations to fertility or mortality rates, and this is explored in Section 6.

Along with compromising the accuracy of point forecasts, approximation error also causes problems for producing prediction intervals with the curve fitting and low-dimensional PC approaches. Ignoring the approximation error is inappropriate, and bias adjustment does not solve the problem, since it only affects the point forecasts. For low-dimensional PC approaches with bias adjustment, this problem might be addressed by formally recognizing the use of random walk models for any PCs involved in the bias adjustment. How successful this approach is would depend on the appropriateness of random walk models for these PCs. Apart from approximation error, obtaining prediction intervals for rates under a curve fitting approach would require either some sort of simulation method or asymptotic approximation, since the curves employed are highly nonlinear functions of their parameters. This contrasts with the ease of obtaining prediction intervals for the principal components approach since, as noted in Section 4.1., the (log) rates r_t are linear functions of the PCs.

Another possible disadvantage to the curve fitting approach is that some of the fitted curve parameters may show erratic behavior as time series, making them ill-suited to modelling and forecasting. Keyfitz (1982) raised this question in relation to a range of mortality curves and related approaches, e.g., the relational models of Brass (1974). Reparameterization of curves may help. In fact, Thompson et al. (1989) found the original gamma curve parameters α and β in (2) showed erratic time series behavior, whereas the alternative parameters $\alpha\beta$ and $\beta\sqrt{\alpha}$ – the gamma curve mean and standard deviation – showed much more stable behavior. The end point parameter A_0 remained problematic, and so was projected at its most recent value of zero. Knudsen, McNown, and Rogers (1993) observed the four original parameters of the double exponential function to be ''somewhat difficult to interpret,'' and so reparameterized the curve. One of the resulting parameters was closely related to TFR. Of the remaining three parameters, graphs of two of them still showed unstable time series behavior. For the Heligman-Pollard curve parameters in (3), several of the graphs in McNown and Rogers (1989) show unstable behavior. I found similar problems with some of the Heligman-Pollard curve parameters in the empirical study of Section 6. There may be value in investigating possible reparameterizations of the Heligman-Pollard or other mortality curves, to improve the time series behavior of the fitted curve parameters.

A potential disadvantage to the PC approach is that its transformation of the data is itself data dependent. That is, given the PCs, i.e., given Λ, the transformation is linear, but the determination of Λ depends on the data. The consequences of this for forecasting are unclear. If the structure of the data evolved quickly enough over time so that PCs computed now would soon provide a poor approximation in the future, there would be a clear problem. Given the quality of the approximation provided by a small number of PCs, however, this seems unlikely to be a significant practical problem. As it is, the complicated nature of the transformation from data to PCs makes their analysis not as clean as may be desired.

In particular, the data dependence of the PC transformation has some implications for time series modelling. For example, Bell and Monsell's (1991) model for white female mortality rates involves differencing only the first three β_{it}s, a decision based on the usual examination of sample autocorrelations for these time series. But the decision about how many in the set of β_{it}s to difference is intimately related to the issue of unit root testing with potentially cointegrated time series, as investigated by Stock and Watson (1988), among others. Without going into details, an important implication of their results is that even if *all* β_{it}s need to be differenced (no cointegration), this may not be apparent from the usual unit root tests applied to these series separately. There is a bias that makes the first few β_{it}s look the most nonstationary, and subsequent β_{it}s less and less so. This bias should be accounted for in unit root inference. Unfortunately, the results of Stock and Watson (1988) and others on this problem are (1) asymptotic, and (2) developed for a relatively small number of series (generally no more than five or six). The implications of these results for a large number of short to moderate length time series of age-specific fertility or mortality rates are unclear. They do suggest that perhaps more β_{it}s should be differenced than is readily apparent, and that perhaps other biases may be present in time series analysis of the PC series. On the other hand, the excellence of the PC approximations may render these issues practically unimportant.

6. Empirical Comparisons of Mortality Forecasts from the Curve Fitting and Principal Components Approaches

We now evaluate empirically the performance of variants of the curve fitting and PC approaches to forecasting mortality rates. The data used are central death rates for U.S. white males and white females from 1940–1991 for ages 0, 1, 2, 3, and 4; five-year age groups from 5–9 through 80–84; and 85 and over. Data with this level of age detail were not available beyond 1991 at the time this study was done. Central death rates are defined as $m_{it} = d_{it}/P_{it}$, where d_{it} is the number of deaths in year t to persons age i, and P_{it} is the corresponding midyear (July 1) estimate of population age i. For simplicity, the data will be referred to as "mortality rates," though actual mortality rates are defined somewhat differently. (Shryock and Siegel (1977, Chapters 14 and 15) note mortality rates q_{it} are defined by dividing d_{it} by a beginning-of-year population at exact age i. The differences between m_{it} and q_{it} are small for all but the lowest and highest ages. In any case, the distinction is mostly immaterial for the present study, since the methods considered could be applied to approximate and forecast either m_{it} or q_{it}, though a possible consideration for the Heligman-Pollard curve is noted shortly.)

Previous authors have done some limited empirical evaluations of their forecasting models. McNown and Rogers (1989) assessed the accuracy of their forecasts of male and female age-specific mortality rates over a ten-year forecast horizon. They found their forecasts to be superior to those of a random walk (no change) forecast. Carter and Lee (1992) examined forecasts of male and female life expectancy over a twelve-year horizon from two versions of their approach. They found their forecasts superior to those obtained by directly modelling and forecasting life expectancy. Both of these assessments are narrow in scope and are referenced to standards that might be expected to do poorly. Both also contain an important technical deficiency: they evaluated forecast accuracy from a single origin year, aggregating accuracy measures over different forecast leads. This is not advisable because of the heteroscedasticity of forecast errors for different leads, and the strong autocorrelation of forecast errors from a common origin (see, e.g., Box and Jenkins 1976, Chapter 5). A better strategy is to compute measures of forecast accuracy for fixed lead times using different forecast origins. This strategy is implemented here by first fitting models to data through 1980 and forecasting 1981 through 1991, then refitting to data through 1981 and forecasting 1982–1991, and so on. This provides eleven one-step-ahead out-of-sample forecast errors, ten at two-steps-ahead, and so on. I will focus on accuracy at leads one through five, because of the very small number of observations available for the more advanced forecast lead times.

As a first step, detection of additive outliers and level shifts as discussed in Chang, Tiao, and Chen (1988) and implemented in the REGARIMA time series program (U.S. Bureau of the Census 1995) was applied to the individual series $r_{it} = \log(m_{it})$ using ARIMA (1,1,0) models with trend constants and an outlier critical value of 3.0. The number of outliers found was not large, but level shifts were found in the 1980s for four series (females age 65, and males ages 5, 20, and 30). These series were adjusted for the level shifts, and the adjusted series were used in the modelling and forecast evaluations. The goal of this adjustment was to robustify the forecast evaluations against outliers. If outliers in the forecast period are not modified, they may distort forecast comparisons so that whichever approach happens by chance to forecast the outliers best comes out best in the comparisons. Since the Gaussian models used here can make no claim to any inherent ability to forecast outliers, it seems best to adjust for outliers, to avoid letting them drive the forecast comparisons. Given the small number of outliers found, the results would probably be essentially the same without this adjustment.

The models evaluated were as follows:

1. random walk with drift models (RWD) applied separately to each $\log(m_{it})$,
2. the curve fitting approach using the Heligman-Pollard curve (HP),
3. same as 2 but with bias adjustment,
4. the one PC approximation of Lee and Carter (1992), denoted LC,
5. same as 4 but with bias adjustment, and
6. the approach of Bell and Monsell (1991) using all the PCs.

For each model forecasts of all the m_{it}s were generated through 1991 from each forecast origin year 1980 to 1990. Then RMSEs for the logged data, and MAPEs for the original data, were computed for each age group and each forecast lead, averaging the results over the available forecast origin years. For example, if $\hat{m}_{i,n+\ell}$ denotes the forecast of $m_{i,n+\ell}$ from forecast origin year n for any of the above models, then the lead one forecast MAPE for that model at age i is

$$MAPE(i,1) = (100/11) \sum_{n=1980}^{1990} |1 - \hat{m}_{i,n+1}/m_{i,n+1}| \tag{5}$$

Summary measures were also computed for each forecast lead by averaging the MAPEs and RMSEs over all ages. This is not necessarily estimating a well-defined quantity, but was computed to give an overall indication of the quality of the forecasts for a given lead.

Results for RMSEs tended to be in the same direction as the MAPEs, but showed somewhat more extreme differences. Thus, results will be presented only for the MAPEs. Before doing so a brief summary of the time series analysis performed for each model will be given.

The random walk with drift model for each age i is $\log(m_{it}) - \log(m_{i,t-1}) = c_i + \varepsilon_{it}$, where ε_{it} is a series of random errors (i.i.d. $N(0, \sigma_i^2)$). Forecasts of $\log(m_{it})$ from this model follow a straight line connecting the first and last data points (the "ruler method of forecasting.") When transformed back to the original scale, these forecasts decay exponentially from the last data point (n) at the rate $\exp(\hat{c}_i)$ (where $\hat{c}_i = [\log(m_{in}/m_{i1})]/(n-1)$). Outlier detection was performed as part of using this model, but this affects the results only if an outlier is detected at the first or last data point. (Note that as the forecast origin is advanced the last data point (n) shifts from 1980 through 1990.) Actually, for a few series a number of outliers were found at the beginning of the series, including additive outliers in 1940 and level shifts early in the 1940s. In such cases the first observations of the series were dropped so that the series used started past these outliers. In no case was a series started later than 1950.

The Heligman-Pollard curves were fitted by minimizing $\sum_i (m_{it}/\hat{m}_{it} - 1)^2$, where $\hat{m}_{it}$ is the fitted HP curve for year t. The fitting was done with the Minpack software (More, Garbow, and Hillstrom 1980). Time series modelling of the curve parameters led to random walk models for all but one parameter. The one exception was the B parameter for males, for which an ARIMA (0,1,1) model was used. Drift parameters were significant and included in the models for the level parameters (A, D, and G), except for the D parameter for females. Only one of the other parameters was found to have a possibly significant drift – the H parameter for males (its t-statistic was marginal at 1.95). These determinations of series with drift corresponded well to the visual appearance of time series plots of the HP parameters. Outlier detection was performed as part of the modelling. Most of the series showed no outliers, but a few did, and the F parameter for females showed many outliers, depending on the outlier critical value used. A few of the series showed several outliers in the 1940s, so these series were redefined to start in 1950.

(Through my own inadvertent errors, the above fitting criterion that I used differs from that originally recommended by Heligman and Pollard (1980). They minimized $\sum_i(\hat{q}_{it}/q_{it} - 1)^2$, with q_{it} obtained from m_{it} via the approximation $q_{it} \approx m_{it}/(1 + .5m_{it})$. Hartmann (1987) used the same fitting criterion, but with q_{it} obtained directly from Swedish life tables. Relative to these references, I substituted m_{it} for q_{it}, and inverted the ratio in the criterion as well. (If consulting either reference, note carefully how their notation differs from that used here.) I have since fitted the HP curve as originally recommended by Heligman and Pollard, inverting the fitted results via $\hat{m}_{it} \approx \hat{q}_{it}/(1 - .5\hat{q}_{it})$. The overall quality of the resulting fits for males and females was quite similar to that for my original fits (cf. Table 1), though with differences for some individual ages. It may have been preferable to follow Heligman and Pollard's original fitting recommendation, which would have generated different time series of curve parameters to be forecast. Nevertheless, since my inadvertent modifications at least yielded fits to the data that were as reasonable as would have been otherwise obtained, it seems unlikely that this had a major effect on the forecasting results.)

The Lee and Carter (1992) RWD model for the first PC series was used without modification. However, the first PC was used rather than their modified "second stage estimate" obtained by determining the mortality index so that the approximation in each year reproduces actual total deaths. Outlier detection was performed as part of the modelling, but no outliers were found in the first PC series for either males or females.

The bias adjustment of Thompson et al. (1989) was applied to the HP and LC forecasts as described in Sections 3 and 4.

The multivariate time series model for the female PC series in Bell and Monsell (1991) was used, although it was of course reestimated with the data for each time frame used. Even for the same time frame used in Bell and Monsell (1991), the results would not have been exactly the same, because the mortality rate data were revised using population estimates derived from results of the 1990 census. Otherwise, the model was applied as discussed in Bell and Monsell (1991). For the male PC series, a multivariate time series model of similar form to that for females was used. Details are omitted.

The RWD model applied separately for each age is presented here as a benchmark ("naive" model) for measuring the performance of the other models. Thus, the MAPEs will be presented for the RWD, and the ratio of the MAPEs for the other models to those for the RWD will be presented. Note that use of the ordinary random walk model with no drift parameter (a no change forecast) would not be an appropriate benchmark, due to the obvious downward trend of mortality rates.

Table 2. Average forecast MAPEs for the RWD model, and ratios of average forecast MAPEs for the other models to those of the RWD model

Forecast lead	MAPE ratios					
	RWD MAPE	HP	HP + bias	LC	LC + bias	PC
Females						
1	2.6	1.93	1.34	2.34	0.98	1.22
2	3.4	1.52	1.13	1.89	1.00	1.18
3	4.2	1.26	1.03	1.66	1.02	1.17
4	5.2	1.04	0.94	1.48	1.04	1.11
5	6.0	0.99	0.87	1.37	1.05	1.10
Males						
1	2.7	2.45	2.01	3.03	0.99	1.14
2	3.8	1.90	1.66	2.43	0.99	1.06
3	4.8	1.61	1.49	2.05	1.01	1.03
4	5.4	1.51	1.48	1.97	1.04	1.03
5	6.1	1.39	1.41	1.86	1.07	1.06

Explanatory note: The average forecast MAPE (mean absolute per cent error) for each model is the average over all ages i of its age-specific MAPEs defined in (5). The notation of the column headings is as follows: RWD = random walk with drift model applied to each age separately, HP = curve fitting approach using Heligman-Pollard mortality curve, HP + bias = HP with bias adjustment of forecasts, LC = Lee-Carter one-PC approximation approach, LC + bias = LC with bias adjustment of forecasts, and PC = Bell-Monsell approach using all PCs. The first column following forecast lead shows MAPEs for the RWD model; the remaining columns show the ratios of MAPEs for the other models to those for the RWD.

Table 2 presents the summary measures obtained from the averages of the MAPEs over all ages, for forecast leads one through five. The average MAPEs for the RWD model show that mortality rates for males were a little harder to forecast than those for females. They also show the expected behavior of increasing forecast errors with increasing forecast lead. The MAPE ratios show that the only model that appears competitive with the RWD model is the Lee-Carter model with bias adjustment, which seems roughly on par with the RWD. The next best choice is the Bell-Monsell PC approach, which is moderately worse than the RWD for females, and just slightly worse for males. Particularly worth noting are the results for the HP and LC forecasts without bias adjustment. These results are very poor, clearly showing the importance of the bias adjustment for short-term forecasting. This is not surprising – comparing Tables 1 and 2 reveals that the approximation error for the HP and LC approaches has about the same magnitude as the short-term forecast error of the RWD model. Bias adjustment significantly improves the HP forecasts, making them more competitive with those of the RWD model for females, though it does not completely solve their problems for males. Notice that the MAPE ratios for the HP and LC forecasts without bias adjustment decline with increasing forecast lead. This shows that approximation error diminishes in importance as the forecast lead advances and more inherent forecast error is accumulated. The bias adjusted HP forecast MAPE ratios also decline with increasing lead, for unknown reasons.

Figures 4 through 7 show MAPE results graphically by age. The first graph in each figure shows the MAPEs by age for the RWD model (for males or females for forecast lead one or two, as noted). Most of the MAPEs are below seven per cent, and MAPEs for the ages above 40 tend to be even lower. The other graphs in Figures 4 through 7 show the logarithms of the ratios of the MAPEs for the other models to those of the RWD model. Values of the log ratios

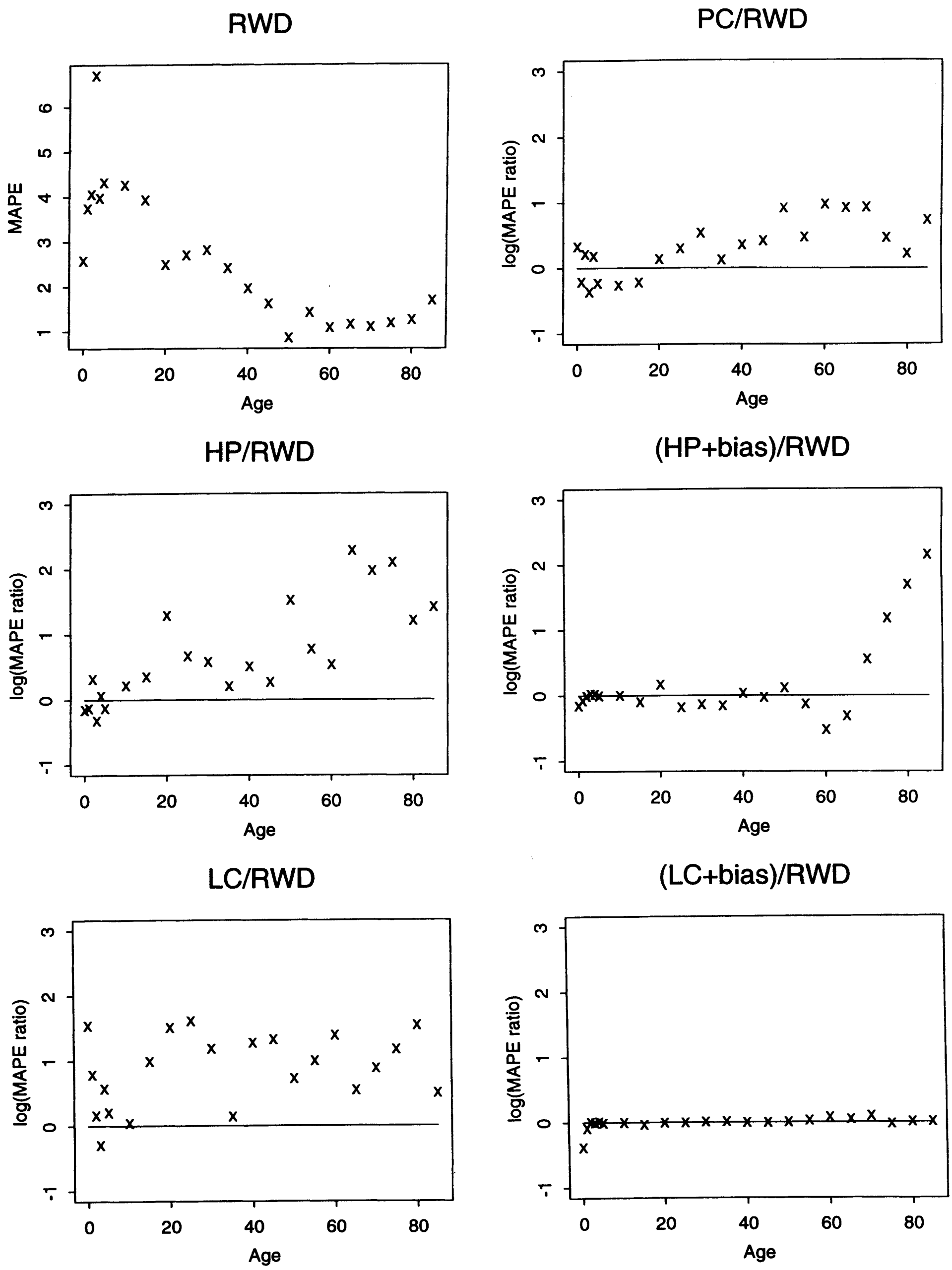

Fig. 4. U.S. white female mortality rates–lead 1 forecast accuracy

Explanatory note: The first graph shows mean absolute percentage errors (MAPEs) by age at lead 1 for the random walk with drift model applied to each age separately (RWD). The remaining graphs show logs of ratios of the corresponding MAPEs for other models, to those for the RWD. The notation is as follows: PC = Bell-Monsell approach using all 22 principal components; HP = Heligman-Pollard curve fitting approach; HP + bias = HP with bias adjustment of forecasts; LC = Lee-Carter one principal component approach; LC + bias = LC with bias adjustment of forecasts.

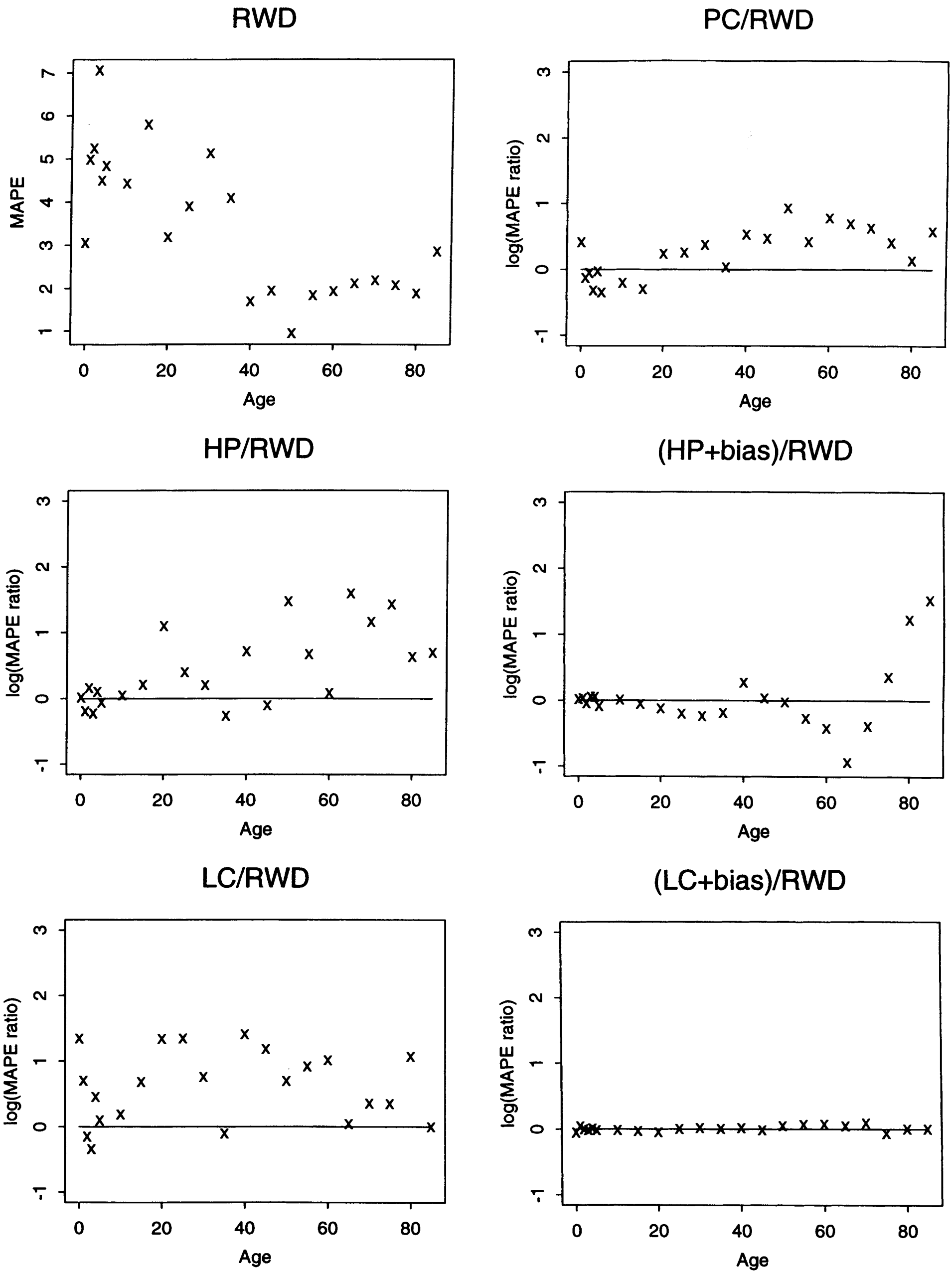

Fig. 5. U.S. white female mortality rates–lead 2 forecast accuracy

Explanatory note: Notation and arrangement of graphs is as for Figure 4

below 0 indicate better performance than the RWD model; values above 0 indicate worse performance. As with Table 2, the graphs reveal that among the more "sophisticated models" only the bias adjusted LC forecasts appear competitive with those of the RWD model. It is perhaps surprising how poorly some of the alternative models do. Note, for example, the very poor performance at advanced ages for the bias adjusted HP forecasts.

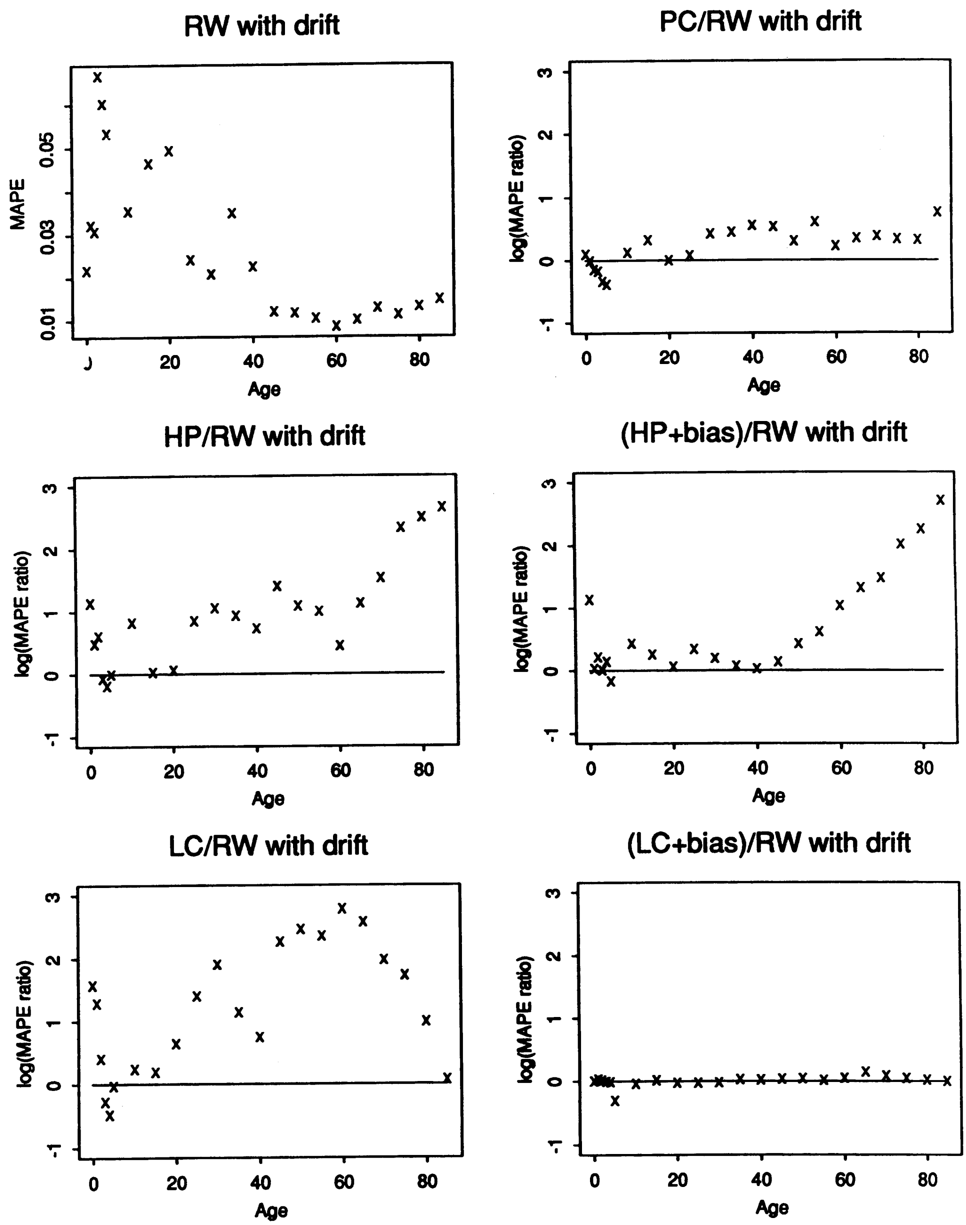

Fig. 6. U.S. white male mortality rates–lead 1 forecast accuracy

Explanatory note: Notation and arrangement of graphs is as for Figure 4

The reason that the performance of the bias adjusted LC forecasts is so close to that of the RWD model is not entirely clear. Note that both models produce forecasts that decline exponentially from the mortality rates in the last year, though they do so at different rates. The estimated drift parameters $\hat{c}_i$ in the RWD model were compared to the corresponding quantities $b_i\hat{c}$ in the Lee-Carter model (note Equation (4)), but the two sets of values did not appear very similar. Perhaps the fact that the forecast functions of the two models are

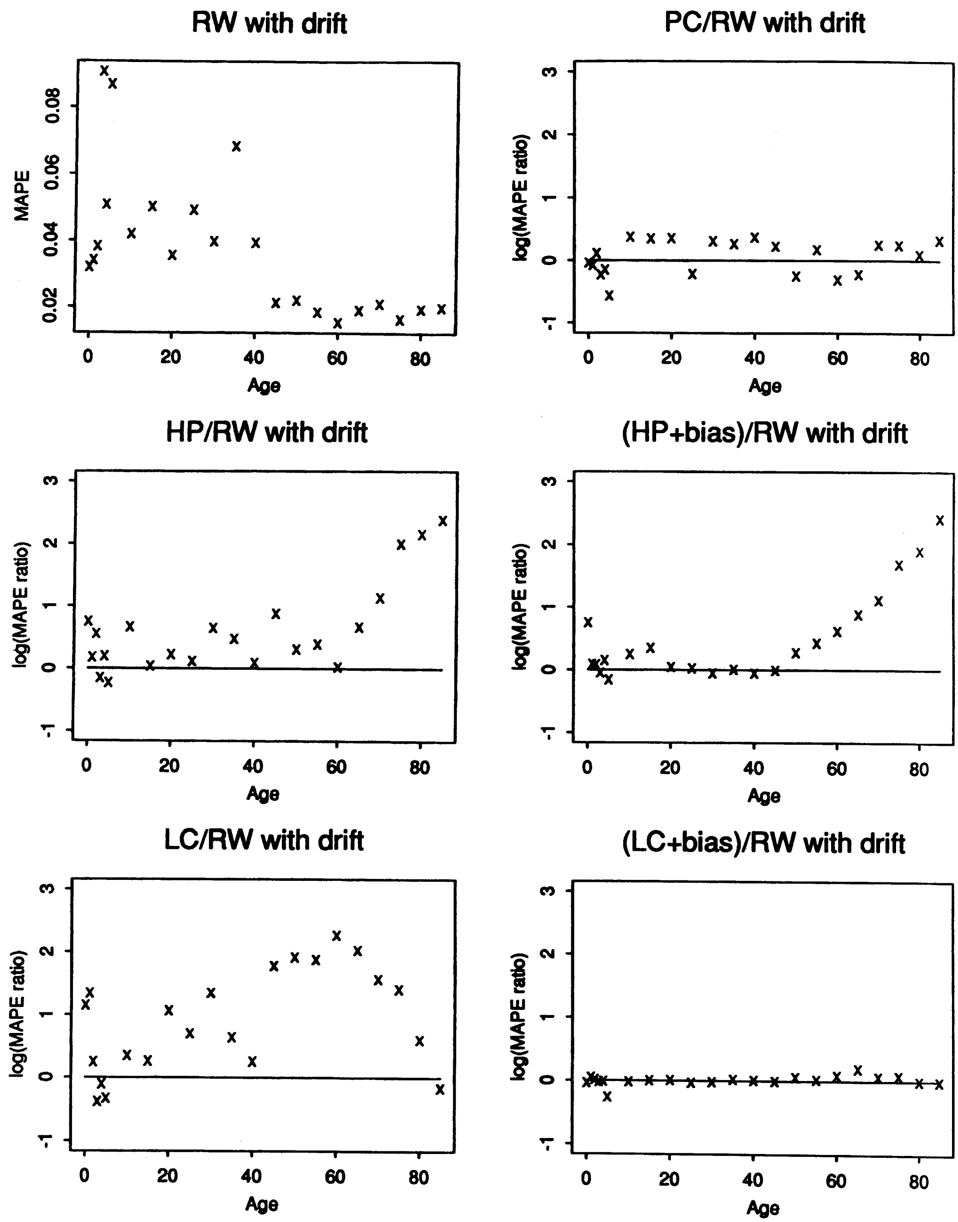

Fig. 7. U.S. white male mortality rates–lead 2 forecast accuracy

Explanatory note: Notation and arrangement of graphs is as for Figure 4

of this same form is more important to forecast performance than the exact values of the estimated rates of decline in mortality at each age.

7. Conclusions

While a single study using data from a single country and evaluating forecasts over only an eleven-year period cannot be taken as definitive, two conclusions still emerge from the results. The first is tentative, due to limitations of the data. The second is rather firm.

The tentative conclusion is that the curve fitting and principal components approaches should not be expected to greatly improve the accuracy of short-term mortality forecasts relative to using a simple random walk with drift model for each age separately. Similar studies of this question with data from other countries would be instructive. Also, future research should perform similar assessments of the curve fitting and principal components approaches to forecasting fertility. In the case of fertility the simple random walk model *without* drift is a suitable choice of a ''naive'' model for comparisons, as there has been no long-term monotonic trend in fertility rates in developed countries. Thus, the benchmark forecast is no change in fertility. The question is then whether the curve fitting and principal components approaches can successfully capture and forecast variations over time in the level (TFR) and shape of the fertility curve, and thus produce better forecasts than the random walk model?

The rather firm conclusion from the results here is that when using a curve fitting or low-dimensional principal components approach to forecasting (such as Lee-Carter), the forecasts should be bias adjusted in some way (e.g., as in Thompson et al. 1989), to avoid unnecessarily large short-term forecast errors due to persistent approximation errors. There is no reason to expect that approximation error would not compromise the accuracy of short-term forecasts from the curve fitting or Lee-Carter one PC approaches when they are applied to other data sets of mortality or fertility rates. (Note, e.g., the results of Thomson et al. 1989; and Knudsen, McNown, and Rogers 1993.) As demonstrated in Section 4.2., without bias adjustment the forecasts emanate not from the rates in the last year of data, but from the fitted curve or PC approximation to these rates. Thus, it should be expected that problems from approximation error will be avoided only under the unlikely and extremely fortunate circumstance that the approximation is essentially error free in the last year of data. In this case the bias adjustment will be negligible anyway. In the common case of significant approximation error, bias adjustment is easy and can be expected to lead to significant improvements in short-term forecasts. Since there is no reason to expect bias adjustment to be harmful to long-term forecasts, there is no reason not to use it. Finally, as noted in Section 4.2., the bias adjustment integrates well with the PC approach, emphasizing a point implicit in Bell and Monsell (1991), that the central issue is not how many PCs to use in the approximation, but how to forecast all the PCs.

It is possible that the curve fitting and PC approaches may improve longer-term forecasts, but empirical evaluation of this requires a lot of data, more than was available for the study of Section 6. In any case, these approaches can at least facilitate construction of long-term forecasts by reducing the dimension of the forecasting problem, thus allowing the forecaster to concentrate attention on forecasting a few quantities rather than a large number of age-specific mortality or fertility rates. Also, the curve fitting and PC approaches may have value for describing forecast error variation in a way that allows for the strong correlation over both time and age. The principal components approach is more convenient than the curve fitting approach in this respect, and, if all PCs are used, it avoids problems with approximation error compromising estimates of forecast error variance.

8. References

Alho, J.M. (1990). Stochastic Methods in Population Forecasting. International Journal of Forecasting, 6, 521–530.

Alho, J.M. (1991). Effect of Aggregation on the Estimation of Trend in Mortality. Mathematical Population Studies, 2, 53–67.

Alho, J.M. (1992). The Magnitude of Error Due to Different Vital Processes in Population Forecasts. International Journal of Forecasting, 8, 301–314.

Bell, W.R., Long, J.F., Miller, R.B., and Thompson, P.A. (1988). Multivariate Time Series Projections of Parameterized Age-Specific Fertility Rates. Research Report 88/16, Statistical Research Division, U.S. Bureau of the Census.

Bell,W.R. (1992). ARIMA and Principal Component Models in Forecasting Age-Specific Fertility. National Population Forecasting in Industrialized Countries, Nico Keilman and Harri Cruijsen (eds.). Amsterdam: Swets and Zeitlinger, 177–200.

Bell, W.R. and Monsell, B.C. (1991). Using Principal Components in Time Series Modeling and Forecasting of Age-Specific Mortality Rates. Proceedings of the American Statistical Association, Social Statistics Section, 154–159.

Box, G.E.P. and Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control. San Francisco: Holden Day.

Bozik, J.E. and Bell, W.R. (1987). Forecasting Age Specific Fertility Using Principle Components. Proceedings of the American Statistical Association, Social Statistics Section, 396–401.

Brass, W. (1974). Perspectives in Population Prediction: Illustrated by the Statistics of England and Wales (with Discussion). Journal of the Royal Statistical Society, Series A, 137, 532–583.

Carter, L.R. and Lee, R.D. (1992). Modeling and Forecasting U.S. Sex Differentials in Mortality. International Journal of Forecasting, 8, 393–411.

Chang, I., Tiao, G.C., and Chen, C. (1988). Estimation of Time Series Parameters in the Presence of Outliers. Technometrics, 30, 193–204.

Coale, A.J. and Trussell, T.J. (1974). Model Fertility Schedules: Variations in the Age Structure of Childbearing in Human Populations. Population Index, 40, 185–258.

Cramér, H. and Wold, H. (1935). Mortality Variations in Sweden: A Study in Graduation and Forecasting. Skandinavisk Aktuarietidskrift, 161–241.

De Beer, J. (1989). Projecting Age-Specific Fertility Rates by Using Time Series Methods. European Journal of Population, 5, 315–346.

De Beer, J. (1992). General Time-Series Models for Forecasting Fertility. National Population Forecasting in Industrialized Countries, Nico Keilman and Harri Cruijsen (eds.). Amsterdam: Swets and Zeitlinger, 147–176.

Hartmann, M. (1987). Past and Recent Attempts to Model Mortality at All Ages. Journal of Official Statistics, 3, 19–36.

Heligman, L. and Pollard, J.H. (1980). The Age Pattern of Mortality. Journal of the Institute of Actuaries, 107, part I, 49–80.

Hoem, J.M., Madsen, D., Nielsen, J.L., Ohlsen, E.M., Hansen, H.O., and Rennermalm, B. (1981). Experiments in Modelling Recent Danish Fertility Curves. Demography, 18, 231–244.

Keyfitz, N. (1982). Choice of Function for Mortality Analysis: Effective Forecasting Depends on a Minimum Parameter Representation. Theoretical Population Biology, 21, 329–352.

Knudsen, C., McNown, R., and Rogers, A. (1983). Forecasting Fertility: An Application of Time Series Methods of Parameterized Model Schedules. Social Science Research, 22, 1–23.

Le Bras, H. and Tapinos, G. (1979). Perspectives a Long Terme de la Population Francaise et Leurs Implications Économiques. Population, 34 (Special Edition), 1391–1450. [In French].

Ledermann, S. and Breas, J. (1959). Les Dimensions de la Mortalité. Population, 14, 637–682. [In French].

Lee. R.D. (1992). Stochastic Demographic Forecasting. International Journal of Forecasting, 8, 315–327.

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. and Carter, L.R. (1992). Modeling and Forecasting the Time Series of U.S. Mortality. Journal of the American Statistical Association, 87, 659–671.

McNown, R. and Rogers, A. (1989). Forecasting Mortality: A Parameterized Time Series Approach. Demography, 26, 645–660.

McNown, R. and Rogers, A. (1992). Forecasting Cause-Specific Mortality Using Time Series Methods. International Journal of Forecasting, 8, 413–432.

More, J.J., Garbow, B.S., and Hillstrom, K.E. (1980). User Guide for MINPACK-1. Report ANL-80-74. Argonne National Laboratory, Argonne, IL.

Rogers, A. (1986). Parameterized Multistate Population Dynamics and Projections. Journal of the American Statistical Association, 81, 48–61.

Shryock, H.S., Siegel, J.S., and associates (1971). The Methods and Materials of Demography, Vol. 2. U.S. Department of Commerce, U.S. Bureau of the Census, Washington, DC: U.S. Government Printing Office.

Sivamurthy, M. (1987). Principal Components Representation of ASFR: Model of Fertility Estimation and Projection. CDC Research Monograph Number 16. Cairo Demographic Center, 655–693.

Stock, J.H. and Watson, M.W. (1988). Testing for Common Trends. Journal of the American Statistical Association, 83, 1097–1107.

Thompson, P.A. (1989). A Transformation Useful for Bounding a Forecast. Statistics and Probability Letters, 8, 469–475.

Thompson, P.A., Bell, W.R., Long, J.F., and Miller, R.B. (1989). Multivariate Time Series Projections of Parameterized Age-Specific Fertility Rates. Journal of the American Statistical Association, 84, 689–699.

Tiao, G.C. and Box, G.E.P. (1981). Modeling Multiple Time Series With Applications. Journal of the American Statistical Association, 76, 802–816.

U.S. Bureau of the Census (1995). REGARIMA Reference Manual, Version 1.0. Statistical Research Division, U.S. Bureau of the Census, Washington, DC.

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, CA.

Received June 1995
Revised June 1997