BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-88/16

MULTIVARIATE TIME SERIES PROJECTIONS OF PARAMETERIZED AGE-SPECIFIC FERTILITY RATES

by

William Bell and John Long, U.S. Bureau of the Census

Robert B. Miller, University of Wisconsin-Madison

Patrick A. Thompson, The Ohio State University

Recommended by: Nash J. Monsour

Report completed: October 19, 1988

Report issued: October 19, 1988

MULTIVARIATE TIME SERIES PROJECTIONS OF PARAMETERIZED AGE-SPECIFIC FERTILITY RATES

William Bell and John Long, U.S. Bureau of the Census

Robert B. Miller, University of Wisconsin-Madison

Patrick A. Thompson, The Ohio State University

ABSTRACT

Projection of individual age-specific fertility rates is a forecasting problem of high dimension. We solve this dimensionality problem by using a scaled and shifted gamma curve to approximate the age-specific rates and forecasting the curve parameters using a multivariate time series model. The resulting time series forecasts of parameters are then used to project fertility curves, and hence individual age-specific fertility rates. This reduces the dimensionality of the forecasting problem and also guarantees that long run projections of age-specific fertility rates will exhibit a smooth shape across age similar to historical data. For short-term projections it is also important to forecast the age-specific deviations from the fitted curves, which can be done by simple methods.

The paper applies this approach to age-specific fertility data for U.S. white women from 1921-1984. The resulting forecasts are examined, and the multivariate time series model is used to investigate possible relations between the fitted curve's parameters expressed as the total fertility rate, the mean age of childbearing, and the standard deviation of age at childbearing. Also, the application of the general approach in a recent set of Census Bureau population projections is discussed.

Key Words: Fertility Projection, Multivariate Time Series, Non-linear Least Squares

AUTHORS' FOOTNOTE

William Bell is Mathematical Statistician in the Statistical Research Division at the Bureau of the Census, U.S. Department of Commerce. John Long is Chief, Population Projections Branch, at Census. Robert Miller is Professor of Business and Statistics at the University of Wisconsin-Madison. Patrick Thompson is Professor of Management Sciences in the College of Business at The Ohio State University.

This paper is based on work supported by the National Science Foundation under grant SES 81-22051, "Research to Improve the Government-generated Social Science Data Base". The research was partially conducted at the Bureau of the Census while two of the authors (Thompson and Miller) were participants in the American Statistical Association/Census Bureau Research Program which is supported by the Census Bureau and through the NSF grant. Additional funding was provided by Ohio State University Seed Money Grant 221772 in 1985 and by Census through Ohio State University Research Foundation grant RF-717887 in 1985-86. Any opinions, findings, conclusions and recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Census Bureau.

The authors wish to thank Gregory Spencer (Population Division, U.S. Bureau of the Census) for supplying data and for assistance in the analysis.

This work was previously reported on in Thompson et. al. (1987). There are some differences in the material covered in this and the previous report: the present paper includes some new analysis with the time series model, and discusses the use of the approach in a recent set of Census Bureau population projections; the previous report performed some analyses on a cohort (as opposed to period) basis. Also, there are some slight numerical differences between the results here and those obtained for the previous report, because of the availability of one additional year of data (1984) for the present paper, and revision of data for the years 1980-1983.

MULTIVARIATE TIME SERIES PROJECTIONS OF PARAMETERIZED AGE-SPECIFIC FERTILITY RATES

1. Introduction

Fertility projections are a vital part of any system for predicting the size and structure of a population. Demographers traditionally use the cohort-component method of population projections (Shryock and Siegel, 1976). This method projects fertility by multiplying age-specific birth rates by the population of females in each age group, yielding a forecast of the total number of births. In this approach, the size and age composition of the female population of childbearing ages has a major impact on the projected number of births. Since most of the mothers for the first twenty-five years of the projection period are already alive at the time the projection is made, the size and age composition of the female population is the most predictable element in short-term fertility projections.

Initially, demographers were satisfied simply to hold the age-specific fertility rates constant and let the number of future births vary only with the changes in the size and age composition of the population of women of child-bearing ages (Whelpton, 1947). While this method worked fairly well for mortality rates, age-specific fertility rates were often too volatile for such a method.

Over time, the projection method employed by the U.S. Bureau of the Census evolved into a three part process. First, assumptions are made as to the ultimate age-specific fertility rates at the end of the projection period. Second, an interpolation method is devised to get from the last year of actual data to the ultimate year of projections. Third, the projections are adjusted to allow for short-term variations in period fertility. Alternative (high and low) fertility levels are set judgmentally at the beginning and end (ultimate) of the projection period, and interpolations are made between these to produce alternative fertility projections.

This demographic method relies heavily on substantive judgment in setting realistic assumptions of future fertility schedules. Those assumptions and the methods employed have varied from projection to projection. In Figure 1, we show the approach used in a previous Census Bureau projection (U.S. Bureau of the Census, 1984). This method emphasized a cohort approach to fertility in the ultimate and intermediate stages but a period approach in the short term, third stage.

There has been considerable controversy over whether cohort or period analysis of fertility is best. Period-based methods of forecasting fertility were largely replaced by cohort methods in the 1960's when projections were heavily based on birth expectations surveys of women in various stages of their childbearing years. Recent emphasis on the effects of economic conditions that simultaneously affect the fertility of women at all childbearing ages argues for a period approach. Promising efforts have also been made that are aimed at combining both period and cohort analysis, eg. Willekens' (1984) age-period-cohort model.

Without entering the period versus cohort controversy directly, we have chosen to perform our analysis on a period basis for essentially practical reasons. The use of age-specific data on a cohort basis creates massive incomplete data problems for time series analysis, since a cohort fertility record is not complete for some 30 years after the first births to the cohort are observed. Moreover, while fertility rates on a period basis follow a smooth curve over age, this is not true of rates on a cohort basis (See Figure 2). Our goal in this analysis is to improve the short-term fertility forecasts of the U.S. Bureau of the Census national population projections. Period fertility analysis emphasizes the short-term period trends in fertility that affect all ages simultaneously. Cohort fertility appears more important in the intermediate and long term. Thus, in this paper our main emphasis will be on projecting age-specific fertility on a period basis.

1.1 Time Series Forecasts of Fertility

The period approach fits well with the statistical time series forecasting tradition. Time series analysis has developed its own approaches to forecasting fertility independently. These methods are usually applied on a period basis and are particularly suited for analyzing short-term variations and providing information about future variability. These are exactly the characteristics needed in projecting several series of fertility rates for the first five to ten years of a projection.

In this paper, we shall follow the time series tradition in developing a method to forecast age-specific fertility rates. Multiplying these forecasts by forecasts of the size of the age-specific female population would then yield fertility forecasts deriving from both the time series and demographic cohort-component traditions. In this way, the advantages of the demographic tradition in taking account of the predictability of the size and age composition of the female population can be combined with the more statistically rigorous time series techniques of modeling the short-term variability of the age-specific fertility rates. The advantages of combining the major traditions of population projections have been outlined by Long (1984) and Land (1986). In actual projections, we have also combined short term time series forecasts of period age-specific fertility rates with long term demographic projections of completed cohort fertility rates determined through judgmental analysis of birth expectation surveys and other information.

Several authors have applied time series methods by themselves, using autoregressive integrated moving average (ARIMA) methods to forecast total births (Dodd, 1980; McDonald, 1981; and Saboia, 1977). While these efforts yielded some insights into the use of time series methods on fertility, the forecasts ignored the advantage of using cohort-component methods (Long, 1981). This omission was partially remedied by Lee (1974,1975), who applied time series methods to the total fertility rate (TFR), the sum of all age-specific rates that occur in a given year. TFR measures the overall level of fertility within a standardized age structure. Carter and Lee (1986) also used a standardized age structure in analyzing an "index of fertility" constructed from time series of births and marriages. If, however, the age pattern of fertility changes, it can only be captured by a method that accounts for fertility by single years of age.

The main difficulty faced in using time series methods to forecast age-specific fertility is the dimensionality of the problem. We shall use data for women at ages 14 and under, single years of age from 15 through 44, and 45 and over, yielding 32 time series to be modeled and forecast. This is a large number of series even for univariate modeling - if this were done, quite likely one would consider only a few simple models. The system is far too large for general multivariate models - a general AR(1) model would have 32x32=1024 autoregressive parameters. One could attempt to use a multivariate model with a very simplified structure. The CARIMA model of deBeer (1985) can be viewed as one example. It essentially lets the fertility rate at a given age for a given cohort depend on the rates at prior ages for the same cohort and on the rates at the same age for prior cohorts. The model is restricted to be the same for all cohort-age combinations. This restriction was necessary in deBeer's application with limited time series data, but using the same model at all ages seems overly restrictive with age-specific time series of reasonable length such as we have.

Another problem we must face is possible inconsistency of the 32 forecasted age-specific fertility series. One would expect that future fertility rates would be smooth functions of age with a shape similar to that observed historically. One potential problem with using separate univariate time series models to forecast the 32 age-specific rates is that the forecasts at various ages may follow inconsistent trends so that the resulting distribution of fertility across age in the forecast years may not make sense. Some constraints on the univariate models may be necessary to avoid this.

The approach we shall adopt emphasizes the advantages of the ARIMA models for short-term fertility forecasting while at the same time addressing the dimensionality and consistency problems. We divide the task into three parts. First, each of the historical annual sets of 32 age-specific fertility rates is fit with a curve that can be represented by a few (eg., four) parameters. Second, the resulting time series of parameters are modeled using multivariate time series models and the models are used to forecast the parameters for future years. Third, for each of the future years the forecasted parameters are used to generate a model curve of 32 forecasted age-specific fertility rates. The forecasted age-specific rates would then be incorporated into the cohort-component model by multiplying them by the forecasted female population by age for the corresponding year. Rogers (1986) suggests much the same approach, although for his application to Swedish fertility he does not perform time series analysis of the curve parameters.

The advantages of this method are that it reduces the dimension of the forecasting problem from 32 to the number of curve parameters (in our case 4), and the use of the curves forces even long-term fertility rate projections to exhibit the same smooth distribution across age as historical data. There is one drawback to the method. In its basic form it forecasts not the future rates themselves, but the curves that will be fit to these rates, thus assuming that the error in using the curve to approximate the rates is negligible for forecast purposes. While this is certainly true for long-term projections, it is not true in the short-term (1-5 years ahead). We deal with this by using simple methods to forecast the deviations of the curve from the rates at each age (the age-specific biases). This greatly improves the accuracy of short-term forecasts, while still accomplishing the goals of dimensionality reduction (since most of the attention can be focused on modeling and forecasting the curve parameters) and producing consistent long-term projections.

1.2 Fertility Curves

Hoem, et. al (1981) investigated the fit of a variety of curves to Danish age-specific fertility rates. They obtained the best fit with the Coale-Trussell (1973) function and Gamma density, both of which fit about equally well. (They actually got still better fits with a cubic spline, but this involved 10 parameters, too many for our purposes). The Coale-Trussell function is somewhat complicated, involving tables of two age-specific functions and the integral of what Rogers (1986) calls a double exponential function. Rogers (1986) suggests direct use of the double exponential function as a simpler alternative. We shall use the Gamma curve, scaled and shifted, because of its simplicity and good fit relative to other curves that have been considered. (We found the double exponential curve to give comparable fits.) The gamma curve parameters can be easily transformed into the TFR, mean age of childbearing, and standard deviation of the age of childbearing - terms readily recognizable by fertility researchers. This interpretability in demographic terms becomes useful later when we combine the short term ARIMA forecasts with long-term (ultimate) judgmental forecasts that are defined in similar terms.

The basic approach of fitting a parametric curve to fertility rates and forecasting the curve parameters can certainly be used with curves other than the Gamma. The choice among curves with a reasonable fit is not critical for forecasting purposes, since no parametric curve is likely to be immune from the bias problem in short-term forecasting mentioned earlier. Note that we are not proposing the Gamma curve here as a "model" for fertility in the sense of using it to provide statistical estimates of age-specific fertility rates for populations in which only summary fertility data are available. Our data contain complete age-specific detail and are obtained from the essentially complete reporting of births in the Vital Statistics Registration System so there is no sampling error. We use the Gamma curve merely as a device for approximating the age-specific fertility rates to reduce the dimensionality of the forecasting problem.

2. Approximating Period Fertility Rates with Gamma Curves

The historical fertility data base we use here consists of age-specific fertility rates to United States white women from 1921 through 1984 (Heuser, 1976 and more recent unpublished data). These are recorded for women of ages 14 and under, single ages of 15 through 44, and 45 and over, yielding an initial data matrix of 64 observations on 32 concurrent time series. Figure 2a shows the period fertility rates for three different years (1927, 1957, and 1977). Even though total fertility was markedly different in these years, the three sets of rates have a similar smooth shape over age. Figure 2b shows age-specific fertility rates for the 1902, 1932, and 1952 birth cohorts, which reach age 25 (one of the peak childbearing ages) in the years 1927, 1957, and 1977 respectively. In addition to illustrating the incomplete data problems when analyzing data on a cohort basis, Figure 2b shows fertility rates for different cohorts do not follow the same smooth shape across age. The problems are most pronounced in recent cohorts, of which the 1952 cohort provides an example.

To approximate the fertility rates each year, we fit a shifted Gamma probability density function to the scaled age-specific rates, using the parametric form:

$$\gamma_i = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}(i-A_0)^{\alpha-1}\exp(-(i-A_0)/\beta) \quad i \geq A_0 \tag{2.1}$$

where

$$\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1}\exp(-u)du$$

This density has probability starting at the point A_0, expected value $A_0 + \alpha\beta$, and variance $\alpha\beta^2$. Since the Gamma curve integrates to one, in fitting it we rescale the age-specific fertility rates so that they sum to unity in each year. Let F_{it} denote the fertility rate per woman of age i in year t, and define the total fertility rate (TFR) in year t as:

$$TFR_t = \sum_{i=14}^{45} F_{it} \tag{2.2}$$

and compute the "relative" fertility rate to women age i in year t as:

$$R_{it} = F_{it}/TFR_t, \quad i = 14, 15, \ldots, 45; \; t = 1921, \ldots, 1984.$$

R_{it} is the proportion of births in year t that occur to mothers of age i.

Now let γ_{it} denote the height, at age $i > A_{0t}$, of a shifted Gamma curve with parameters A_{0t}, α_t and β_t. We fit the curve to the R_{it} in each year t by minimizing the sum of weighted squared errors:

$$WSSE_t = \sum_{i=14}^{45} [w_i(R_{it}-\gamma_{it})]^2 \tag{2.3}$$

where w_i is the weight for age i. The parameter values for α and β are determined through a derivative-based nonlinear least squares procedure operating conditional on values of A_0. Values for the starting point of the curve, A_0, are obtained from a one-dimensional search constrained to the region $0 \leq A_0 \leq 14$. In practice, the values γ_{it} are computed at ages 14.5, 15.5, ..., 45.5 since fertility is recorded at the mother's last birthday.

Since a main reason for forecasting fertility rates is to forecast total births, we use the weights w_i to give more emphasis in (2.3) to ages with high fertility. In recent years most births (80-85 percent) are to women aged 18 through 32. We thus used the simple weighting scheme $w_i = 4$ for i = 18 through 32, $w_i = 1$ otherwise. Sensitivity analysis, examining the response of the Gamma curve parameters to changes in the w_i structure, showed that the exact weight used for ages 18 through 32 had much less impact than the fact that ages 18 through 32 were singled out for special treatment. For example, both α_{1982} and β_{1982} change only one unit in their third significant digits when weights of 2 through 6 are used.

Our weighting scheme disagrees with the one used by Rogers (1986), who suggests a Chi-squared goodness of fit criterion. Although not stated explicitly, Rogers apparently minimized $\Sigma(O_i-E_i)^2/O_i$, where O_i and E_i denote, respectively, the observed rate and the curve value at age i; in our case these are our R_{it} and γ_{it}. This approach gives <u>more weight</u> to the ages with the <u>lowest fertility rates</u> (smallest O_i). Different individuals are, of course, free to choose different weights reflecting their own "loss functions". We feel the weighting scheme we used accomplished our goal of giving enough weight to ages with high fertility to help the curves yield reasonable approximations to total births, while not ignoring the other ages to the extent that the fit at these ages was completely unreasonable.

The results of the curve fits are summarized in Figure 3, which traces the parameter values over time. Apart from the years affected by World War II, examination shows remarkable stability in the curve fits through 1968. Although total fertility varies considerably over this period, the Gamma parameters remain relatively stable. The 1970s, however, show a marked change in this pattern; all three parameters shift rapidly to new levels. This shift in the pattern of relative fertility is evident in Figure 4, where we show the Gamma curve fits for two years. In the early part of our data base, there is more fertility to older women; in later years, fertility is concentrated to women in their twenties and the overall curve is more symmetric in shape. Note also that the Gamma curve fits much better in the more recent year, an important consideration since forecasts reflect behavior nearest to the forecast origin.

Some comments on the behavior of A_{0t} in Figure 3 are in order. Through 1968, the weighted sum of squares (2.3) was minimized by setting A_0 to its upper limit of 14 (with a few exceptions including years affected by WWII). This upper limit was imposed so that the Gamma curve would yield an ordinate at age 14. In more recent years, as α_t and β_t shifted to make the curve more nearly symmetric, A_{0t} declined and eventually (2.3) became rather insensitive to values of A_0 from 0 through negative numbers large in magnitude. To simplify the least squares problem, a lower limit of 0 was used for A_0 with little sacrifice in the quality of the fit.

In modeling the Gamma system over time, we first transform the α and β parameters. Since the Gamma density (2.1) has expected value $A_0 + \alpha\beta$ and variance $\alpha\beta^2$, we form the new series:

$$MACB_t = A_{0t} + \alpha_t\beta_t \quad \text{and} \quad SDACB_t = \beta_t \cdot (\alpha_t)^{1/2} \tag{2.4}$$

These are, respectively, the mean age of childbearing and the standard deviation of age at childbearing, computed from the Gamma density. These time series, in contrast to the α and β parameters, have more stable traces over time: as indicated by Figures 5 and 6 the abrupt shifts in level displayed in Figure 3 are not evident in the traces of MACB and SDACB. These series also have meaningful demographic interpretations.

3. Time Series Modeling of Curve Parameters

For modeling purposes we transform TFR_t, $MACB_t$, and $SDACB_t$ by taking logarithms to get $T_t = \ell n(TFR_t)$, $M_t = \ell n(MACB_t)$, and $S_t = \ell n(SDACB_t)$. We model and forecast T_t, M_t, and S_t, and exponentiate the forecasts to get forecasts of TFR_t, $MACB_t$, and $SDACB_t$. A major reason for transforming TFR is to prevent forecasts, or forecast limits, from going below 0, which is a lower limit TFR cannot reach. (A stricter lower limit than 0 can be used if it is believed TFR will never fall below some other level.) Taking logarithms does not materially affect the modeling of MACB and SDACB since the transformation is nearly linear over the range of those series. However, transforming MACB and SDACB along with TFR helps interpretation of the models used, since any additive relations between T, M, and S can be interpreted as multiplicative relations between TFR, MACB, and SDACB.

Note that the endpoint, A_{0t}, has almost a "two-point" (either 0 or 14) distribution (Figure 3), and these are the limits imposed on the parameter during the least squares fit. We treat A_{0t} as deterministic since its graph shows it would be ill-suited to modeling, and forecast A_{0t} simply by projecting it at its most recent level of 0.

3.1 Univariate Models

The plots of TFR, MACB, and SDACB (Figures 5-7) all show unusual behavior for the years 1942 through 1947 due to the effect on fertility of World War II. To account for these effects we use indicator variables for the years 1942 through 1947. This leads us to consider univariate time series models for all three series of the form:

$$y_t = \beta_1 I42_t + \cdots + \beta_6 I47_t + [\theta(B)/\phi(B)\delta(B)]a_t$$

where $I42_t$, ..., $I47_t$ are the war-year indicator variables; β_1, ..., β_6 are parameters; $\theta(B)$, $\phi(B)$, and $\delta(B)$ are the moving-average, autoregressive, and differencing operators in the backshift operator B; y_t is the particular series being modeled, and a_t is a sequence of i.i.d. $N(0,\sigma^2)$ random variables (white noise).

The ARIMA structure is identified following the scheme suggested by Bell and Hillmer (1983) for models of this type. The differencing order is tentatively identified as one for all three series by noting that the sample ACF (autocorrelation function) of the original data fails to damp out, while that of the differenced data does. The war-year effects are removed by regressing each differenced series on the differenced indicator variables. Examining sample ACFs and PACFs (partial ACF) for the three series suggests an AR(3) model for ∇T_t, and AR(1) or AR(2) models for ∇M_t and ∇S_t (∇ = 1-B denotes first difference). Conditional least squares estimates for the AR model parameters are given in Table 1. T-statistics for the last AR parameter in all 3 models are about 2; so that these might be dropped. The AR(2) parameter for T_t is estimated near 0, but we keep it in the model

Table 1

Parameter Estimates from Univariate Models

Series	β_1	β_2	β_3	β_4	β_5	β_6	ϕ_1	ϕ_2	ϕ_3	$\sigma^2 \times 10^4$
∇T_t	.066	.061	-.047	-.120	.025	.098	.54	-.01	.24	8.603
	(2.6)	(1.6)	(-1.1)	(-2.9)	(.7)	(4.0)	(4.3)	(-.1)	(1.9)	
∇M_t	-.0042	.010	.029	.037	.020	.0032	.49	.22		.100
	(-1.5)	(2.5)	(6.0)	(7.7)	(5.0)	(1.2)	(3.9)	(1.7)		
∇S_t	-.027	.003	.046	.055	-.004	-.010	.55	.24		.514
	(-4.3)	(.3)	(4.2)	(5.0)	(-.5)	(-1.6)	(4.3)	(1.8)		

Note: The numbers in parentheses below the parameter estimates are T-ratios.

Table 2

Multivariate Model Identification Exhibits

Simplified Cross-Correlation Matrices, Lags 1-8

LAG 1	LAG 2	LAG 3	LAG 4
$\begin{bmatrix} + & \cdot & \cdot \\ \cdot & + & + \\ \cdot & + & + \end{bmatrix}$	$\begin{bmatrix} + & \cdot & \cdot \\ \cdot & + & + \\ \cdot & + & + \end{bmatrix}$	$\begin{bmatrix} + & \cdot & \cdot \\ \cdot & \cdot & + \\ \cdot & \cdot & + \end{bmatrix}$	$\begin{bmatrix} + & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$

LAG 5	LAG 6	LAG 7	LAG 8
$\begin{bmatrix} + & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} + & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} + & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & - & - \\ \cdot & \cdot & - \\ \cdot & \cdot & \cdot \end{bmatrix}$

NOTE: + indicates a T-ratio ≥ 2, - indicates a T-ratio ≤ -2, and $\cdot$ indicates $-2 < T < 2$.

Stepwise Autoregressive Fits

AR Lag:	1	2	3	4	5	6
$\lvert\Sigma_j\rvert \times 10^{13}$:	2.61	2.01	1.71	1.32	1.22	1.08
χ^2 test:	59.9	11.4	6.6	9.6	2.7	3.8
AIC:	-1633.5	-1630.4	-1621.7	-1618.4	-1604.9	-1593.8

NOTES: The asymptotic χ^2 test statistic of $\Phi_j = 0$ in an AR(j) fit is $-(49.5 - 3j)\ell n(|\Sigma_j|/|\Sigma_{j-1}|)$, starting from $|\Sigma_0| = 9.47 \times 10^{-13}$. With 9 degrees of freedom, the 5% critical value is 16.9.

AIC denotes the Akaike Information Criterion, computed here as $57\ \ell n|\Sigma| + 2(9j)$ for the lag j fit. Akaike (1973) suggests picking the AR model for which this criterion is minimized.

since we keep the AR(3) parameter. While we could use these models to forecast T_t, M_t, and S_t individually, we instead use them to guide identification of a multivariate model that allows for the possibility of relationships between the variables.

3.2 Multivariate Models

To begin, the war-year effects estimated via the univariate models are subtracted from the series (taking $T_t - \hat{\beta}_1 I42_t - \cdots - \hat{\beta}_6 I47_t$, etc.), and the resulting series are first differenced. Henceforth, let T_t, M_t, and S_t denote the war-year adjusted series, let $Y_t = (T_t, M_t, S_t)'$, and let the differenced vector series be ∇Y_t. Sample cross-correlation matrices for ∇Y_t and results from stepwise autoregressive fits (Tiao and Box, 1981) are given in Table 2. An AR model appears more appropriate than an MA model, since the stepwise AR tests truncate while the cross-correlation matrices do not.

While the overall statistics suggest an AR(1) model, we found in the AR(3) fit that the (1,1) element (corresponding to T_t) of the third AR parameter matrix is significant. This is not surprising given that the univariate model for ∇T_t is AR(3). No other parameter estimates in the second- and third-order matrices of the AR(3) fit exceed twice their standard errors, including the (2,2) and (3,3) elements of the second matrix corresponding to parameters in the univariate AR(2) models for ∇M_t and ∇S_t. Rather than use a full AR(3) model with its 27 parameters (not counting variances and covariances of the residuals) we consider a model of the following form:

$$\nabla Y_t = \Phi_1 \nabla Y_{t-1} + \Phi_2 \nabla Y_{t-2} + \Phi_3 \nabla Y_{t-3} + a_t$$

with

$$\Phi_1 = \begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \end{bmatrix} \qquad \text{and} \qquad \Phi_2, \Phi_3 = \begin{bmatrix} x & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \tag{3.1}$$

where x indicates a parameter to be estimated, • an element fixed at 0 (and hence not a parameter to be estimated), and $a_t = (a_{1t}, a_{2t}, a_{3t})'$ is a vector white noise series with covariance matrix Σ. This model was estimated by conditional likelihood with the following results:

$$\hat{\Phi}_1 = \begin{bmatrix} .46 & 2.8 & -1.0 \\ .014 & .34 & .14 \\ .045 & .26 & .63 \end{bmatrix} \quad \text{with standard errors} \quad \begin{bmatrix} .12 & 1.5 & .60 \\ .010 & .15 & .06 \\ .023 & .35 & .14 \end{bmatrix}$$

$\hat{\Phi}_2(1,1) = .03$ with standard error .13, $\hat{\Phi}_3(1,1) = .32$ with standard error .12, and

$$\hat{\Sigma} = 10^{-4} \times \begin{bmatrix} 8.190 & & \\ -.040 & .094 & \\ -.639 & .133 & .509 \end{bmatrix}$$

Diagnostic checks reveal no obvious inadequacies with the above multivariate model, with the exception of some outliers in the residuals of T_t and M_t (which will not materially affect the forecasts). A discussion of this and some alternative models considered is contained in the Appendix. Thus, we shall use the above model for forecasting the gamma curve parameters, and in so doing do not need to explicitly account for the war-year variables since they are all zero over the forecast horizon.

3.3 Relations Among the Curve Parameters

An important feature of this model is that it allows us to examine relations between the total fertility rate and the mean and standard deviation of age at childbearing. One of the primary questions addressed in setting ultimate fertility assumptions is whether the level of fertility affects the distribution by age. In previous Census Bureau projections these assumptions were made independently, albeit with some speculation about the possible relations. In the current set of projections, we wanted to be able to examine empirical evidence on the relations, and use this in our long-term as well as short-term fertility projections.

The model (3.1) allows for effects of the variables T_t, M_t, and S_t on each other through the off-diagonal elements of Φ_1 (dynamic effects) and Σ. The t-statistics corresponding to $\hat{\Phi}_1$, and the residual correlation matrix, $\hat{C}$, corresponding to $\hat{\Sigma}$, are the following:

$$\hat{\Phi}_1: \begin{bmatrix} 3.8 & 1.9 & -1.7 \\ 1.4 & 2.3 & 2.3 \\ 2.0 & .7 & 4.5 \end{bmatrix} \qquad \hat{C} = \begin{bmatrix} 1.00 & & \\ -.05 & 1.00 & \\ -.31 & .61 & 1.00 \end{bmatrix} \qquad (3.2)$$

The off-diagonal elements of Φ_1 are at most marginally significant, suggesting there are no strong dynamic relationships involved. As a more formal check on this we fit a model with Φ_1 restricted to be diagonal; this amounts to fitting three univariate models jointly. The residual covariance matrix for this model was

$$\hat{\Sigma}_2 = 10^{-4} \times \begin{bmatrix} 8.659 & & \\ -.115 & .107 & \\ -.656 & .151 & .566 \end{bmatrix}.$$

An asymptotic χ^2_6 - test of the restrictions on Φ_1 (6 degrees of freedom for the 6 off-diagonal elements) is obtained using $-m\ \ell n|\hat{\Sigma}|/|\hat{\Sigma}_2|$ (see Hannan 1970, pp. 339-341). The choice of the 0(n) multiplier m is not obvious; we followed Rao (1965, chapter 8) and used $m = (51+2) - .5(3+2+1) = 50$, where 3 is the dimension (number of equations), 2 is the number of parameters constrained in each equation, and 51 is the residual degrees of freedom in each equation of the full model (64 observations, less 4 lost to differencing and AR order, less 6 war-year and 3 AR(1) parameters estimated, but ignoring the estimation of the lag 2 and 3 parameters in the T_t equation). This yields $-m\ \ell n|\hat{\Sigma}|/|\hat{\Sigma}_2| =$ 15.8, which is about the 98.5 percentile of the χ^2_6 distribution, also suggesting that dynamic relationships are marginally significant but not strong.

We also could consider using models of transfer function form for T_t, M_t, and S_t. This is appropriate if the Φ_1 matrix is lower triangular, possibly after reordering of the variables in $Y_t = (T_t, M_t, S_t)'$ (Tiao and Box 1981, pp. 802-803). The t-statistics for Φ_1 in (3.2) do not suggest any obvious transfer function structure, though they do not argue strongly against any such structures either. We experimented with some transfer function models and, not surprisingly, found dynamic effects at most marginally significant. If we accept the hypothesis of no dynamic effects then we can use a simple transfer function model where variables depend on each other only at the current time, and the variables can be entered in any order. Modeling variables in the order $T_t \rightarrow M_t \rightarrow S_t$ yields a univariate ARIMA(3,1,0) model for T_t (as given previously) and the following models for M_t and S_t (standard errors of parameter estimates in parentheses):

$$M_t = -.0140\ T_t + a_{2t}/(1-.53B-.20B^2)(1-B)$$
$$\quad (.0134) \qquad\qquad (.13)\ (.13)$$

$$S_t = -.061\ T_t + 1.18\ M_t + a_{3t}/(1-.44B-.29B^2)(1-B) \tag{3.3}$$
$$\quad (.025) \qquad (.24) \qquad\qquad (.13)\ (.12)$$

We see the only strong contemporaneous relationship is that between M_t and S_t, with that between T_t and S_t being marginally significant. This provides a significance check on what we would also infer from the correlation matrix $\hat{C}$ in (3.2).

Rogers (1986), in the absence of time series data on his curve parameters, suggests a regression approach to projecting the curve parameters. In our case this approach would develop projections of T_t independently (either by judgment or using models), regress M_t and S_t on T_t, and use the regression relations to develop M_t and S_t projections from the T_t projections. The weak relations of T_t to M_t and S_t in (3.3) show this would be a poor approach in the case of the gamma curve. Since TFR refers to the overall level of fertility and MACB and SDACB to the shape of the curve, the weak relations in (3.3) suggest the shape of the fertility curve depends little on the level. Thus, we would expect the approach Rogers (1986) suggests would not work well with other curves applied to our data, and would recommend caution in applying it to other fertility data sets.

4. Forecast Results

We use the multivariate model (3.1) to forecast T_t, M_t, and S_t because the dynamic relationships in the model are (marginally) statistically significant. Since these relationships are not strong, use of univariate (diagonal Φ_1) or transfer function models to forecast T_t, M_t, and S_t would not produce greatly different results. To use (3.1) in forecasting we incorporate the differencing into the AR operator and rewrite it as

$$Y_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \Pi_3 Y_{t-3} + \Pi_4 Y_{t-4} + a_t \tag{4.1}$$

$$\Pi_1 = I + \Phi_1 \qquad \Pi_2 = \Phi_2 - \Phi_1 \qquad \Pi_3 = \Phi_3 - \Phi_2 \qquad \Pi_4 = -\Phi_3 .$$

Forecasts $\hat{Y}_n(\ell)$ of $Y_{n+\ell} = (T_{n+\ell}, M_{n+\ell}, S_{n+\ell})'$ from any origin n are computed recursively for $\ell = 1,2,\ldots$ from

$$\hat{Y}_n(\ell) = \Pi_1 \hat{Y}_n(\ell-1) + \Pi_2 \hat{Y}_n(\ell-2) + \Pi_3 \hat{Y}_n(\ell-3) + \Pi_4 \hat{Y}_n(\ell-4) \tag{4.2}$$

where $\hat{Y}_n(j) = Y_{n+j}$ for $j \leq 0$. The variance matrix of the ℓ-step ahead forecast error, $Y_{n+\ell} - \hat{Y}_n(\ell)$, is given by

$$V(\ell) = I + \psi_1 \Sigma \psi_1' + \cdots + \psi_{\ell-1} \Sigma \psi_{\ell-1}' \tag{4.3}$$

where the ψ_j matrices are obtained by equating coefficients of B^j in $(I+\psi_1 B+\psi_2 B^2+\cdots)(I-\Pi_1 B-\Pi_2 B^2-\Pi_3 B^3-\Pi_4 B^4) = I$. The diagonal elements, $v_{ii}(\ell)$ $i = 1,2,3$, of $V(\ell)$ are the variances of the forecast errors for the series T_t, M_t, and S_t, and can be used to produce forecast intervals for these series assuming normality. For example, $\hat{T}_n(\ell) \pm k(v_{11}(\ell))^{1/2}$ provides a 67% (k=1), or 95% (k=2), etc. forecast interval for $T_{n+\ell}$. Exponentiating the point forecasts and interval limits for T_t, M_t, and S_t yields point and interval forecasts for TFR_t, $MACB_t$, and $SDACB_t$. In applying these results we substitute the estimates of Φ_1, Φ_2, Φ_3, and Σ given following (3.1) into (4.2) and (4.3).

Figures 5-7 show point and 67% interval forecasts from origin 1984 for TFR, MACB, and SDACB, and point forecasts from 1980 for MACB and SDACB. For the latter, model (3.1) was refit using only data through 1980. (The point forecasts from 1980 for TFR are not shown because they would almost coincide on the graph with the 1981, ..., 1984 data and 1984 forecasts.) These forecasts are determined primarily by the last two years of data; (3.1) is almost an AR(1) model in the first difference, so only the (1,1) elements of Π_3 and Π_4 in (4.2) are nonzero. This results in very flat point forecasts from 1980 since the change from 1979 to 1980 is small for all three series. The sharp increase in MACB and SDACB from 1980 through 1984 produces MACB and SDACB forecasts from 1984 that increase before leveling off. The long-term point forecasts all level off due to the first differencing and absence of a constant term in (3.1). This behavior is convenient for long-term forecasting, since it facilitates comparison with traditional long-term judgmental forecasts made by specifying an ultimate level for the series being forecast, as discussed in section 1. Models with a constant term or other degrees of differencing would exhibit different long-run forecast behavior.

The forecast intervals in Figures 5-7 show there is considerable uncertainty associated with long-term forecasts for all three series. We shall focus on TFR. Figure 7 for TFR indicates what demographers refer to as the replacement level of TFR, which is currently approximately 2.1 in the United States. If TFR remains constant at this level, with no change to mortality and relative fertility rates and zero net migration, then population size remains approximately constant (fluctuates above and below a constant level). Figure 7 shows the TFR for white women in the U.S. dropped below replacement level for the first time in 1972, and has remained well below replacement level since. While the point forecasts for TFR in Figure 7 stay below 2.1, however, the forecast limits show considerable uncertainty in how long TFR will remain below replacement level. The upper 67% limit exceeds 2.1 as early as 1992; an upper 95% limit reaches 2.1 in 1988. (Keep in mind we are forecasting from 1984 here.)

The forecasts of TFR, MACB, and SDACB, along with the forecast of A_0 as 0, can be used to produce forecasted scaled and shifted gamma curves. These are forecasts of the curves that will be fit to the future data when it becomes available. To forecast the actual age-specific fertility rates we also want to forecast the deviations of the rates from the curve. We do so as follows. The relative fertility rates, R_{it}, are defined following (2.2). Let γ_{it} be the ordinate of a gamma curve approximation to R_{it}, and define the "bias" for age i, year t as

$$b_{it} = R_{it} - \gamma_{it} \qquad i = 14, \ldots, 45 \tag{4.4}$$

We use the relation

$$F_{it} = TFR_t(\gamma_{it}+b_{it}) \tag{4.5}$$

to produce "bias adjusted" forecasts.

We briefly examined ACF's for the 32 bias time series (one for each age) and found all series appeared to need first differencing. For most of the bias series autocorrelation remaining after a first difference is applied is not strong. The minor exceptions tend to be at the very low or very high ages, which are the least important in determining births. For simplicity, we use the random walk model $\nabla b_{it} = \epsilon_{it}$ with ϵ_{it} white noise for all the bias series. The forecast of $b_{i,n+\ell}$ is then just

$$\hat{b}_{in}(\ell) = b_{in} \tag{4.6}$$

that is, the biases are all forecast to remain constant at their values in the forecast origin year. Forecasts of $A_{0,n+\ell}$, $MACB_{n+\ell}$, and $SDACB_{n+\ell}$ yield forecasts of $\gamma_{i,n+\ell}$; these are then combined as in (4.5) with the forecast of $TFR_{n+\ell}$ and bias forecasts from (4.6) to yield forecasts of the age-specific rates, $\hat{F}_{i,n}(\ell)$. This was done from both the 1980 and 1984 forecast origins.

In Figure 8a,b,c,d we show fertility rate data and fitted gamma curves for the years 1981 through 1984 compared to forecasts of the data and gamma curves from 1980. We use these to illustrate the roles of the gamma curve and bias forecasts in forecasting the fertility rates, $F_{i,n+\ell}$. Consider first the fitted (solid line) and forecasted (dotted line) gamma curves in Figures 8a - 8d. The forecasted curves appear quite accurate for 1981 and 1982, less so for 1983 and 1984. From Figures 5 and 6 we see the MACB and SDACB forecasts undershoot the actual values by amounts that increase steadily from 1981 through 1984. This results in forecasted curves that peak too early and are too narrow by amounts that increase from 1981 through 1984, as can be seen in Figures 8a - 8d. In Table 3 we give TFR forecasts, actual values, forecast errors, and forecast standard errors from the model

Table 3

TFR Forecasts (from 1980), Data, and Errors
for 1981-1984

Year	Forecast[1]	Actual	% Error[2]	Forecast % Standard Error[3]
1981	1.715	1.723	-.46%	2.9%
1982	1.720	1.730	-.58%	5.5%
1983	1.725	1.701	+1.41%	8.0%
1984	1.720	1.710	+ .58%	10.8%

[1]Computed as $\exp(\hat{T}_{1980}(\ell))$ for $\ell = 1, \ldots, 4$ with $\hat{T}_{1980}(\ell)$ the forecasts computed from (4.2) with model parameters estimated using data through 1980.

[2]Computed as 100(Forecast - Actual)/Actual.

[3]Computed as $100(\exp(v_{11}(\ell)^{1/2})-1)$ for $\ell = 1, \ldots, 4$ with $v_{11}(\ell)$ the first diagonal element in (4.3), using the model parameters estimated with data through 1980.

(the latter two in percents) for 1981 - 1984. We see the TFR forecasts for 1981-1984 were quite accurate. The errors in the MACB and SDACB forecasts causing the forecasted curves to be too peaked are partially offset in 1981 and 1982 for ages around the peak (say, ages 20-30) by TFR forecasts that are slightly low, and these errors are accentuated some in 1983 and 1984 for ages around the peak by TFR forecasts that are slightly high. The result of this is that the forecasted curves show noticeably more error at the peak ages in 1983 and 1984 than in 1981 and 1982.

While the forecasted gamma curves for 1981 and 1982 are quite close to the fitted gamma curves, they are not satisfactory as forecasts of the actual rates, F_{it}. Forecasting the biases to remain constant at their 1980 values for all ages (n = 1980 in (4.6)), and adjusting the forecasted curves using (4.5) effects a noticeable improvement in forecasts of the age specific rates. The bias-adjusted forecasts are almost uniformly better across age than the raw gamma curve forecasts for 1981 and 1982. With the larger error in the curve forecasts for 1983 and 1984 the bias adjustment is less important. It still makes some improvement, particularly at the very young ages.

The accuracy of the gamma curve forecasts for 1981-1984 depended to a large extent on the quite accurate TFR forecasts for those years. From Table 3 we see that the historical data suggests much larger errors in TFR forecasts could be expected (unless one interprets the flatness of the TFR graph in recent years as an indication that the variability in this series has decreased). Considering the magnitude of the forecast errors that can occur for TFR, one sees that while the bias forecasts will be useful and important in the short-term (especially 1-2 years ahead), they will generally not be important for medium- to long-term forecasts (certainly beyond 10 years ahead), except possibly at a very few ages with low fertility.

5. Implementation in Census Bureau Population Projections

In its latest set of national population projections, the Census Bureau projected fertility by combining short-term results from time series methods similar to those of this paper with long-term results based on demographic judgments (age-specific fertility data through 1984 was used). The time series methods used differed from the presentation here in that transfer function instead of multivariate time series models were used (primarily to facilitate outlier detection and modification), ℓn(TFR-1) was modeled instead of ℓn(TFR) (to prevent TFR forecasts and forecast limits from going below 1), and an adjustment was made to the 1985 and 1986 TFR forecasts using preliminary information on total births for those years. The models actually used and the adjustment for 1985 and 1986 total births are described in the Appendix. The slightly different time series methods were presented here to simplify exposition. In what follows we describe how demographic judgment was used in conjunction with time series model results, and compare results for the Census Bureau projections with the results from the time series model (3.1) presented here.

In the official Census Bureau projections, traditional judgmental methods were used to set the ultimate fertility level and mean age of childbearing for women completing their fertility after the year 2020. These values were then compared with values for past cohorts to produce an acceptable pattern of age-distribution of fertility. These results became the ultimate distribution of cohort fertility shown in Figure 9. Ultimate fertility forecasts can be obtained from the time series model (3.1) used since the forecasts of all quantities needed (TFR, MACB, SDACB, and the age-specific biases) achieve ultimate levels. (See Figures 5-7 and recall that the biases are forecast at their 1984 values.) The results for the age-distribution of fertility (the relative fertility rates) are roughly comparable for the two approaches as shown in Figure 9, with the Census Bureau rates having a somewhat narrower and later peak (reflecting a somewhat lower SDACB and higher MACB).

The official projections used bias adjusted forecasts from time series transfer function models for the first six years of the forecast period for the age-specific fertility rates for ages 14-39. However, the time series models forecasted increases in MACB and SDACB that produced forecasted increases in fertility to women age 40 and over that, while small in absolute magnitude, were large in percentage terms. Census Bureau demographers felt they could not accept such rapid rises in fertility at these ages for the projections to have face validity, so the rates at these ages were forecast to remain constant at current levels. The results of this adjustment were inconsequential for the total birth and population projections since the fertility rates above 40 are so small. (The magnitude of the difference this makes at ages 40 and over is reflected in Figure 9.) Therefore differences between the Census projections and model (3.1) forecasts for the years 1985 through 1990 are due almost entirely to the differences in the use of time series models mentioned earlier, especially to the adjustment of the 1985 and 1986 TFR forecasts for preliminary data on total births. The effects of this adjustment can be seen in the 1984-1986 TFR numbers in Table 4. The Census Bureau projections then interpolated

Table 4

Projected Total Fertility Rates and Empirical Mean and Standard Deviation of Age at Childbearing of White Women from Census Bureau Population Projections and Forecasts from Model (3.1): 1984 to 2020

			AGE OF CHILDBEARING			
	TOTAL FERTILITY RATE		MEAN		STANDARD DEVIATION	
Year	Census	Model (3.1)	Census	Model (3.1)	Census	Model (3.1)
1984	1.710	1.710	26.51	26.51	5.52	5.52
1985	1.746	1.719	26.60	26.58	5.56	5.58
1986	1.747	1.714	26.66	26.63	5.58	5.63
1987	1.749	1.715	26.70	26.66	5.58	5.67
1988	1.769	1.717	26.73	26.69	5.58	5.69
1989	1.778	1.717	26.75	26.71	5.59	5.71
1990	1.781	1.717	26.78	26.72	5.60	5.73
1991	1.778	1.718	26.75	26.73	5.56	5.74
1992	1.778	1.718	26.76	26.74	5.56	5.74
1993	1.778	1.718	26.77	26.74	5.55	5.75
1994	1.779	1.718	26.78	26.74	5.55	5.75
1995	1.779	1.718	26.79	26.75	5.55	5.76
2000	1.780	1.718	26.83	26.75	5.54	5.76
2010	1.791	1.718	26.89	26.76	5.53	5.76
2020	1.800	1.718	26.92	26.76	5.54	5.76

Notes: The figures for 1984, the last year of complete data, are shown for comparison. The Census TFR projections for 1985 and 1986 were adjusted for preliminary data on total births for those years (see Appendix). The mean and standard deviation of the age of childbearing are computed empirically from the projected fertility distribution by age in each year. The forecasts from model (3.1) are "bias adjusted" as described in the text.

Sources: U.S. Bureau of the Census, Projections of the Population for the United States by Age, Sex, and Race: 1987-2080 (P-25, No. 1018) and unpublished data.

Table 5

High, Middle, and Low Series of Census Total Fertility Rate Projections Compared to Model (3.1) Forecasts with 67% Forecast Intervals

	Census			Model (3.1)		
Year	Low Series	Middle Series	High Series	Lower Limit	Forecast	Upper Limit
1984		1.710			1.710	
1985		1.746		1.670	1.719	1.769
1986		1.747		1.626	1.714	1.806
1987	1.718	1.749	1.780	1.591	1.715	1.847
1988	1.716	1.769	1.826	1.554	1.717	1.898
1989	1.708	1.778	1.855	1.513	1.717	1.949
1990	1.689	1.781	1.885	1.473	1.717	2.001
1991	1.671	1.778	1.901	1.434	1.718	2.056
1992	1.662	1.778	1.916	1.396	1.718	2.113
1993	1.652	1.778	1.931	1.359	1.718	2.170
1994	1.642	1.779	1.946	1.324	1.718	2.229
1995	1.632	1.779	1.961	1.290	1.718	2.288
2000	1.583	1.780	2.035	1.140	1.718	2.588
2010	1.514	1.791	2.159	0.924	1.718	3.192
2020	1.500	1.800	2.200	0.780	1.718	3.784

Notes: The figures for 1984, the last year of complete data, are shown for comparison. The Census TFR projections for 1985 and 1986 were adjusted for preliminary data on total births for those years (see Appendix). These are treated like actual data, which is why there are no low or high series Census projections for those years. The model (3.1) forecasts and limits are "bias adjusted" as described in the text.

Sources: U.S. Bureau of the Census, Projections of the Population of the United States by Age, Sex, and Race: 1988-2080, Current Population Reports, Series P-25, No. 1018 and unpublished data.

between the 1990 projections and the judgmentally determined ultimate values to get age-specific rates for the intermediate years. The total fertility rate and mean age of childbearing of the resulting forecasts are shown in columns 2 and 4 of Table 4. The TFR forecasts are within one-tenth of a child of the forecasts using model (3.1) shown in column 3 of Table 4, and the mean age and standard deviation projections in columns 4 and 6 do not differ greatly from the corresponding forecasts from model 3.1 in columns 5 and 7. (The mean age and standard deviation in Table 4 are computed empirically from the projected age-specific relative fertility rates ($\hat{g}_{kt}$) according to EMACB = $\sum_{k=14}^{45} \hat{g}_{kt}(k+.5)$ and ESDACB = $[\sum_{k=14}^{45} \hat{g}_{kt}(k+.5-\text{EMACB})^2]^{.5}$. The results in columns 5 and 7 differ from the forecasted gamma curve MACB and SDACB because of the "bias adjustment", and also because the gamma curve is continuous whereas the $\hat{g}_{kt}$ are discrete.)

Time series methods were also used to aid in developing high and low fertility variants of the projections. This was done by setting TFR projections at upper and lower 67% limits through 1990 from a model for ℓn(TFR-1) analogous to that for ℓn(TFR) in Table 1 for the high and low fertility projections respectively. By 2020, the lower forecast limit for the total fertility rate in model (3.1) reaches .78 while the upper limit reaches 3.78 (see Table 5). This range of three children is far wider than the traditional one child range between high and low projections in judgmental demographic projections of U.S. fertility. This large range is indicative of the large amount of uncertainty inherent in forecasts that are based only on the historical time series and do not take into account the substantive knowledge available to demographers. Even though the forecasts in this instance do not differ greatly from the middle fertility series projected by demographic methods, demographers would be unlikely to accept the wide range between high and low fertility produced by time series models in the middle to long-run. This is primarily because demographers contend that the large swings in fertility during the baby boom and bust of 1945 through 1974 were generated by a fertility process that is not likely to be repeated. The recent stability in fertility rates since 1974 has been cited to support this argument. Consequently, in the Census Bureau projections the high and low fertility variants from the time series model through 1990 were interpolated to values in 2020 of 2.2 and 1.5 respectively (Table 5).

6. Conclusions

The approach to fertility forecasting set out here accomplishes the goals of reducing the dimensionality of the forecasting problem, and providing short-term forecasts of reasonable accuracy and long-term forecasts that capture the smooth shape across age of historical data. It allows the forecaster to concentrate efforts on modeling and predicting demographically meaningful quantities (TFR, and MACB and SDACB for the gamma curve), since the gamma curves capture much of what is going on in the data. The deviations of the fitted curves from the age-specific rates can be forecast by simple means and the forecasted curves adjusted accordingly. These "bias adjustments" are important for short-term forecast accuracy, much less important for medium to long-term forecasting.

In recent Census Bureau projections (U.S. Bureau of the Census, 1988) short-term forecasts (through 1990) from a time series model similar to that presented here were combined with long-term projections from the judgmental approach described at the beginning of this paper. Long-term forecasts by any method are likely to contain a substantial amount of error, and so are perhaps best guided by judgment and substantive theory rather than by atheoretical time series forecasts. Hence, the Census Bureau projections used the forecasted, bias-adjusted gamma curves for the first few years of the projection period and then interpolated between these and assumed ultimate levels determined by demographic theory.

In future research, an alternative to simply interpolating between short-term time series and long-term judgmental forecasts may be explored. Thompson and Miller (1986) present a method for producing forecasts that reach a specified future "target" by adding an intervention component into the forecast function and solving for the forecast intervals using Bayesian methods. In the meantime, the approach used in this paper allows both middle and alternative (high and low) fertility projections to be based on both historical time series data and demographic judgment.

APPENDIX: ADJUSTMENTS FOR PRELIMINARY DATA AND ALTERNATIVE MODELS

Adjustment for 1985 and 1986 Preliminary Data

There is about a two year lag between the times when age-specific birth statistics are collected and released by the National Center for Health Statistics. Our model used fertility data through 1984, but we performed our analysis in mid 1987. Certain 1985-86 information (the total number of white births and age-specific population counts) was known when we produced our 1984-origin forecasts. This was taken advantage of in the Census Bureau projections presented, but was not done with the multivariate time series model (3.1) in this paper to simplify the exposition. Here we discuss how this adjustment of the Census Bureau projections was done.

First, we forecasted TFR_{1985}, $MACB_{1985}$ and $SDACB_{1985}$ to obtain α_{1985} and β_{1985} forecasts. From these we computed the relative fertility rate forecasts $\{\hat{\gamma}_{i,1985};\ i = 14 \text{ to } 45\}$. Multiplying these by $\hat{TFR}_{1985}$ yields the set $\{\hat{F}_{i,1985};\ i = 14 \text{ to } 45\}$. These age-specific fertility rates were then multiplied by the known age-specific population counts, and summed to arrive at a forecast of total white births for 1985. Our forecast was slightly lower than the actual, so we adjusted $\hat{TFR}_{1985}$ upwards, by the ratio of actual to forecasted total births.

Repeating the process, we generated forecasts for 1986 (using the adjusted TFR forecast for 1985) and in similar fashion adjusted $\hat{TFR}_{1986}$. The forecasts of TFR extend from origin 1984, but the forecast limits extend from origin 1986, since we behaved as if the adjusted forecasts of TFR for 1985 and 1986 were the actual TFR values. The original TFR forecasts for

1985 and 1986 were 1.720 and 1.708; the adjustments brought these to 1.746 and 1.747 (note Tables 4 and 5).

Alternative Models

We experimented with several variations on our model (3.1) for forecasting. There is little evidence, such as clearly significant or insignificant test statistics, for discriminating among most of these alternatives, and most have only minor effects on the forecasts. Two changes that had a major impact were including constant terms in the models and taking second differences of the time series. Most of the discussion here is taken from Thompson, et. al. (1987), and thus uses data only through 1983, and without the revisions to the 1980-83 data. While we did not redo all the analyses discussed here with the new data, the results would be unlikely to differ appreciably. The differences in results for those models that were fit both to the earlier and revised data were very slight.

Letting the (2,2) and (3,3) elements of Φ_2 be nonzero produced an improvement in the residual autocorrelations for M_t and S_t, but t-statistics for these parameters were only 1.2 and 1.7, respectively, and the impact on the forecasts was slight.

Outliers were investigated with both the univariate and transfer function models using a program by Bell (1983), which implements a modification of a procedure suggested by Chang (1982). The procedure checks for both AO's (additive outliers -- affecting only one time point, such as the war-year variables) and LS's (level shifts -- where the series moves to and stays at a new level). Outliers were found in the early 1970's for all three series, and the mid 1960's for T_t and M_t. The reasons for this are easy to see from Figures 5-7. TFR drops rapidly, resulting in a LS for 1965, and AO's for 1971 and 1972 followed by a LS in 1973 for T_t. MACB moves less dramatically, but for M_t we still found a LS in 1966, and an AO in 1971 followed by a LS in 1972. SDACB is better-behaved, with only an AO in 1971 for S_t. Other outliers found were an AO in 1933 and a LS in 1979 for T_t, and an AO in 1950 for M_t.

Despite the statistical significance of the outlier terms their impact on the forecasts was very slight. They do inflate the residual variance estimates, and would thus affect forecast intervals. We proceeded here with the multivariate model (3.1) without outlier terms since we presented forecast intervals only for TFR, and only as an exercise in portraying uncertainty in projections.

The model (3.1) and those above do not include constant terms. Since all these models contain a first difference, a constant term in such a model would be a slope, with the forecasts eventually following a straight line with this slope as the forecast lead increases. Estimates of constant terms can be thought of as average slopes over the length of the series. Constant terms were tried in transfer function models for T_t, M_t, and S_t (similar to (3.3), but with outlier terms), with resulting t-statistics of -.3, -1.3, and -2.1. While only S_t offers much evidence for a constant term, it is worth at least considering constant terms seriously since they have a large impact on the forecasts, causing TFR forecasts to drop slightly, and causing long-term MACB and SDACB forecasts to drop dramatically. Models with constant terms were not chosen partly because the statistical evidence for them was not strong, and partly because the resulting forecasts were thought to be less reasonable than those from models without constant terms.

Using a second difference, ∇^2, also has a dramatic effect on the forecasts. These models are relevant since the AR operators for all three series have a root not far from one. For example, the model for ∇T_t given in Table 1 can be factored as

$$(1 - .55B + .03B^2 - .25B^3) = (1 - .86B)(1 + .32B + .28B^2)$$

and the first factor is not far from $\nabla = 1\text{-}B$. A unit root test (Dickey, Bell, and Miller 1984) for using $\nabla^2 T_t$ was insignificant at the 10% level. The autoregressive operators in Table 1 for ∇M_t and ∇S_t factor into $(1-.77B)(1+.28B)$ and $(1-.84B)(1+.29B)$, respectively. If the transfer function models for T_t, M_t, and S_t were modified to include second differences, this effectively sets to one the AR roots near unity. When this was done, examining residual autocorrelations suggested adding first-order moving average terms to the models for M_t and S_t to compensate for changing the AR roots to one, leading to ARIMA (0,2,1) models.

Forecasts from the second difference models also follow a straight line as the forecast lead increases; in this case the slope of the line is determined primarily from the most recent data. The results were TFR forecasts that declined faster than those from the model with one difference and a constant term (there are no constant terms in the second difference models), and MACB and SDACB forecasts that rose rapidly, at about the rate of the last three or four data points for each series. Though the statistical evidence against the second difference models is not strong, the longer-term forecasts from these models, particularly for M_t and S_t, were judged unreasonable. Thus, the second difference models were not pursued further.

Actual Models Used in Census Bureau Projections

The actual models used for the projections described in U.S. Bureau of the Census (1988) were of transfer function form, similar to (3.3) except for the following. First, $\ell n(TFR_t\text{-}1)$ instead of $T_t = \ell n(TFR_t)$ was modeled with a univariate ARIMA (3,1,0) model, though T_t was used as an input variable in the models for M_t and S_t as in (3.3). Second, the outlier terms described above were included in the model (as well as the war-year indicator variables). Third, the ARIMA structure for M_t and S_t was (1,1,0), since the AR(2) terms did not appear necessary when outliers were allowed for. The resulting fitted models, omitting the outlier and war year effects for simplicity, were as follows:

$$\ell n(TFR_t\text{-}1) = a_{1t}/(1-\underset{(.11)}{.42}B + \underset{(.13)}{.09}B^2 - \underset{(.12)}{.51}B^3)(1-B) \qquad \hat{\sigma}_1^2 = .03915$$

$$M_t = -\underset{(.011)}{.0270}T_t + a_{2t}/(1-\underset{(.08)}{.78}B)(1-B) \qquad \hat{\sigma}_2^2 = .0630 \times 10^{-4}$$

$$S_t = -\underset{(.023)}{.0273}T_t + \underset{(.23)}{1.571}M_t + a_{3t}/(1-\underset{(.12)}{.53}B)(1-B) \qquad \hat{\sigma}_3^2 = .3458 \times 10^{-4}$$

The inclusion of outlier terms results in changes in parameter estimates, though this does not greatly affect the forecasts. The variance estimates are not comparable to any given previously, since in the first equation $\ell n(TFR_t\text{-}1)$ is used instead of $T_t = \ell n(TFR_t)$, and in the second and third equations M_t and S_t depend, respectively, on T_t and on (T_t,M_t) at the current time. The above fitted models could be used directly in forecasting, ignoring the war-year and outlier variables, since the forecasts actually turn out to be the same whether or not the war-year and outlier variables are included. (This is because autoregressive models were used, and the war-year and outlier variables are all zero near the end of the series and over the forecast horizon.)

REFERENCES

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," 2nd International Symposium on Information Theory, ed. B.N. Petrov and F. Csaki, Budapest: Akademiai Kiado, 267-281.

Bell, W.R. (1983), "A Computer Program for Detecting Outliers in Time Series," in Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 634-639.

Bell, W.R. and Hillmer, S.C. (1983), "Modeling Time Series with Calendar Variations," Journal of the American Statistical Association, 78, 526-534.

Carter, L.R. and Lee, R.D. (1986), "Joint Forecasts of U.S. Marital Fertility, Nuptiality, Births and Marriages Using Time Series Models," Journal of the American Statistical Association, 81, 902-911.

Chang, I. (1982) "Outliers in Time Series," Ph.D. Thesis, University of Wisconsin-Madison, Department of Statistics.

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.

de Beer, J. (1985), "A Time Series Model for Cohort Data," Journal of the American Statistical Association, 80, 525-530.

Dickey, D.A., Bell, W.R., and Miller, R.B. (1986) "Unit Roots in Time Series Models: Tests and Implications," The American Statistician, 40, 12-26.

Dodd, D.A. (1980), "Short-Term Birth Forecasting: A Serial Correlation Process," presentation at the annual meeting of the Box-Jenkins Computer User Group, Princeton, NJ, July 21.

Heuser, R.L. (1976), "Fertility Tables for Birth Cohorts by Color: United States, 1917-1973," DHEW Publication No. HRA 76-1152.

Hannan, E.J. (1970) Multiple Time Series, New York: Wiley.

Hoem, J.M., Madsen, D., Nielsen, J.L. and Ohlsen, E. (1981), "Experiments in Modelling Recent Danish Fertility Curves," Demography, 18, 231-244.

Land, K.C. (1986), "Methods for National Population Forecasts: A Review," Journal of the American Statistical Association, 81, 888-901.

Lee, R.D. (1974), "Forecasting Births in Post-Transitional Populations: Stochastic Renewal with Serial Correlated Fertility," Journal of the American Statistical Association, 69, 607-617.

Lee, R.D. (1975), "Natural Fertility, Population Cycles, and the Spectral Analysis of Births and Marriages," Journal of the American Statistical Association, 70, 2 -304.

Long, J.F. (1981), comment on "Modeling Demographic Relationships: An Analysis of Forecast Functions for Australian Births," by J. McDonald, Journal of the American Statistical Association, 76: 796-798.

Long, J.F. (1984), "U.S. National Population Projection Methods: A View from Four Forecasting Traditions," Insurance: Mathematics and Economics, 3, 231-239.

McDonald, J. (1979), "A Time Series Approach to Forecast Australian Total Live Births," Demography, 16, 575-602.

McDonald, J. (1981), "Modeling Demographic Relationships: An Analysis of Forecast Functions for Australian Births," Journal of the American Statistical Association, 76: 782-792.

Rao, C.R. (1965) Linear Statistical Inference and Its Applications, New York: Wiley.

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

Saboia, J.L.M. (1977), "Autoregressive Integrated Moving Average (ARIMA) Models for Birth Forecasting," Journal of the American Statistical Association, 72: 264-270.

Shryock, H.S., and Siegel, J.S. (1976) The Methods and Materials of Demography, New York: Academic Press.

Thompson, P.A., Bell, W.R., Long, J.F., and Miller, R.B. (1987) "Multivariate Time Series Projections of Parameterized Age-Specific Fertility Rates," Working Paper 87-3, College of Business, The Ohio State University.

Thompson, P.A. and Miller, R.B. (1986), "Sampling the Future: A Bayesian Approach to Forecasting from Univariate Time Series Models," Journal of Business and Economic Statistics, 4: 427-436.

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 (1984), "Projections of the Population of the United States, by Age, Sex and Race: 1983-2080," Current Population Reports, Series P-25, No. 952, U.S. Government Printing Office, Washington.

U.S. Bureau of the Census (1988), "Population Projections of the United States, by Age, Sex, and Race: 1988 to 2080," Current Population Reports, Series P-25, No. 1018. Washington, D.C.: Government Printing Office.

Whelpton, P.K. (1947), Forecasts of the Population of the United States: 1945-1975 Washington, D.C.: U.S. Bureau of the Census.

Willekens, F. (1984) "An Age-Period-Cohort Model" Paper presented at the annual meeting of the Population Association of America, Minneapolis.

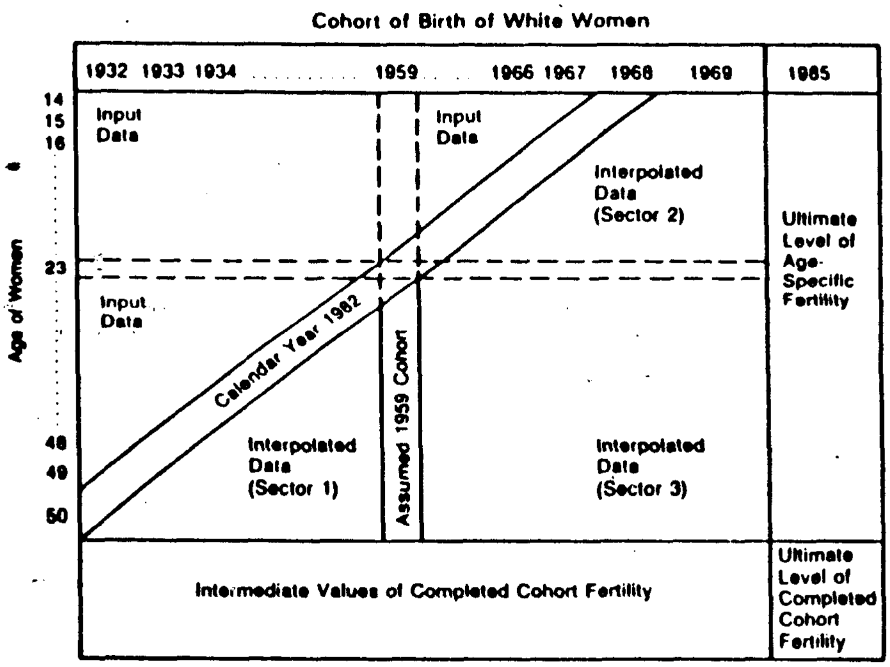

Figure 1: The fertility table used by Census. All the table entries below the "Calendar Year 1982" diagonal were unobserved. After supplying an assumed 1959 cohort, Sector 1 was completed by interpolation. An assumed 1985 cohort allowed completion of Sectors and 3 by interpolation. Constant rates were assumed throughout.

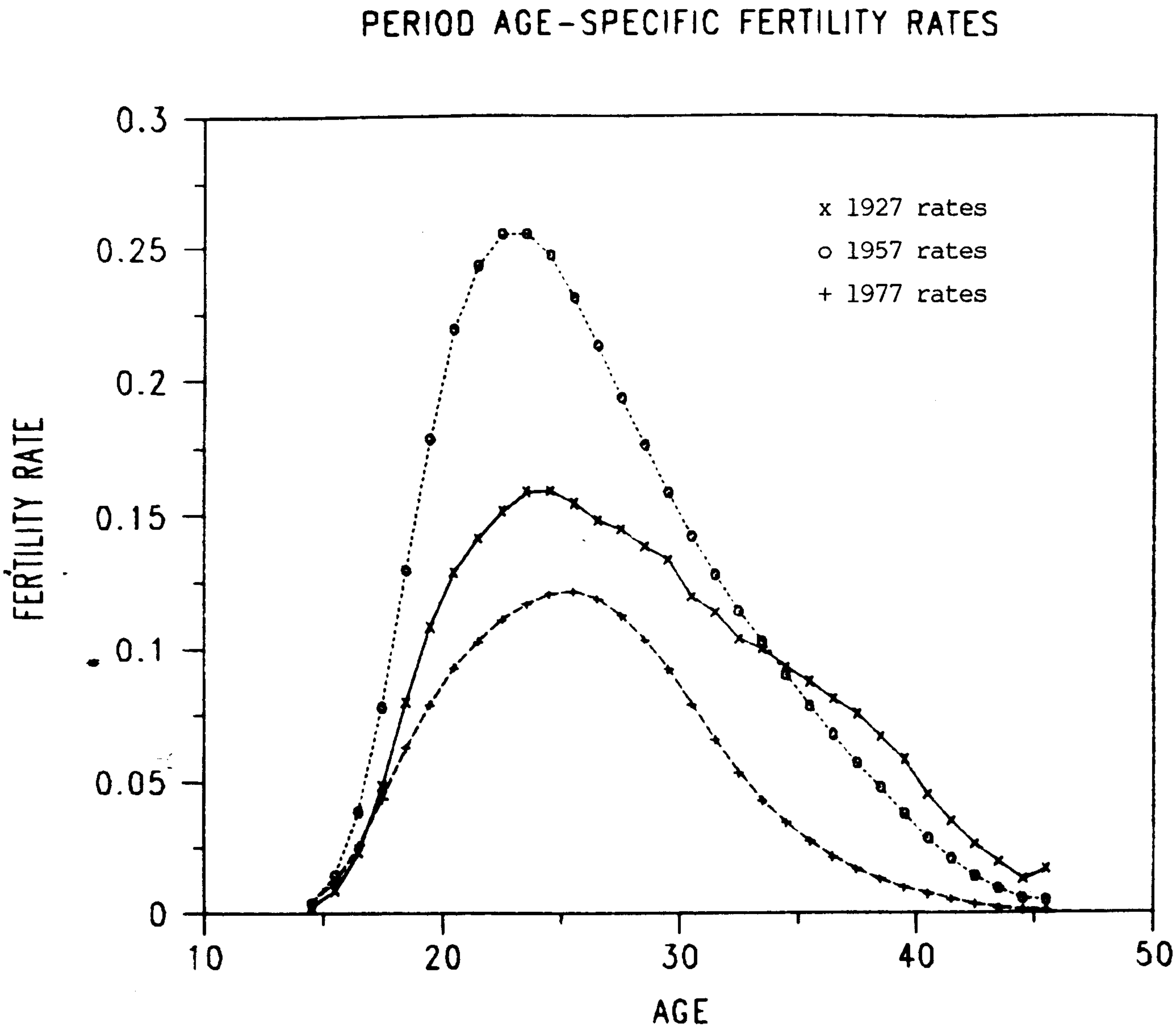

Figure 2a. Period age-specific fertility rates for three years, U.S. white women. Total fertility differs in these years, and the age-specific pattern shifts, but the rates for all three years have a similar smooth shape across age that is well-approximated by a scaled and shifted gamma density. The largest deviations from this shape occur in the early years of data (1927, for example), which are the least important for forecasting. The rates are plotted at the mother's age at last birthday plus .5.

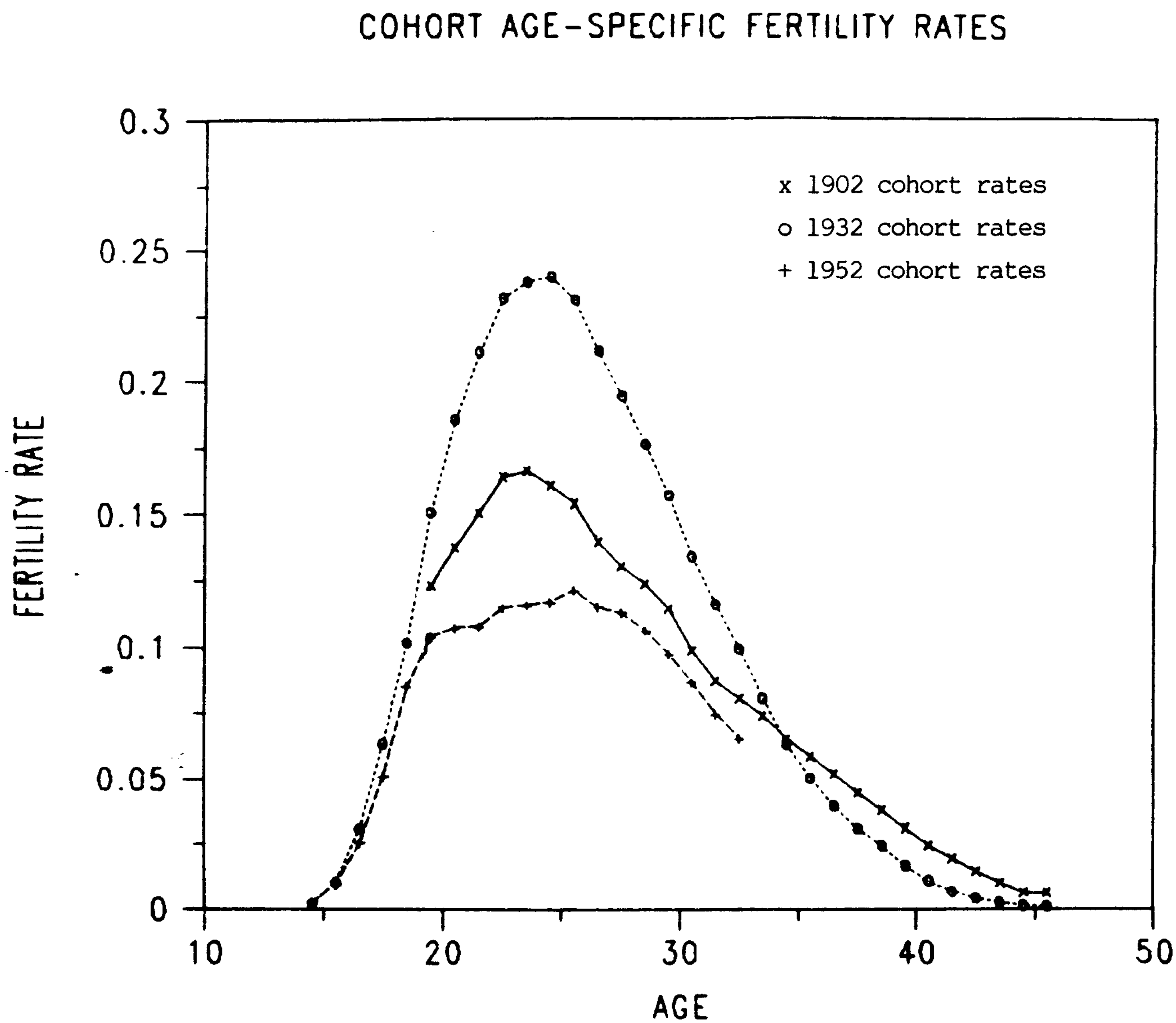

Figure 2b. Cohort age-specific fertility rates for three cohorts, U.S. white women. Since we are only using data for 1921-1984, the 1902 cohort is incomplete at ages 14-19, and the 1952 cohort is incomplete at ages 33-45. In contrast to period rates, cohort rates do not follow such similar smooth shapes across age. Large deviations from a common smooth shape occur in recent cohorts, which are the most important for forecasting. The fertility rates for recent cohorts are relatively flat from ages 19-30, as illustrated by the 1952 cohort.

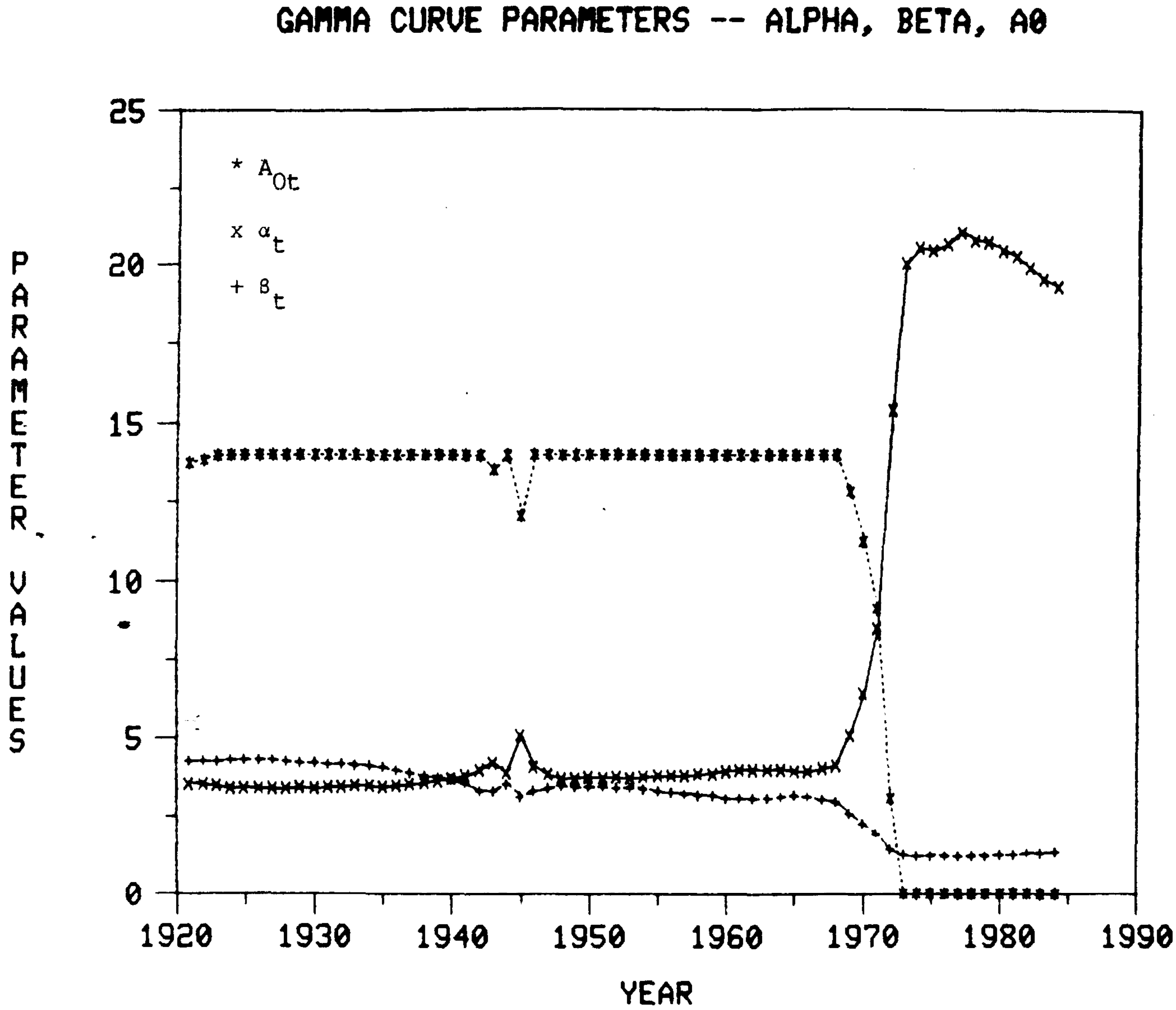

Figure 3. Parameters of gamma curves fitted to relative fertility rates of U.S. white women, 1921-1984. Except for years affected by World War II (1942-47), the parameters remain relatively stable through 1970, after which they shift rapidly to new levels. In fitting the gamma curves values of the endpoint parameter A_{0t} were constrained to the interval [0, 14] (see text).

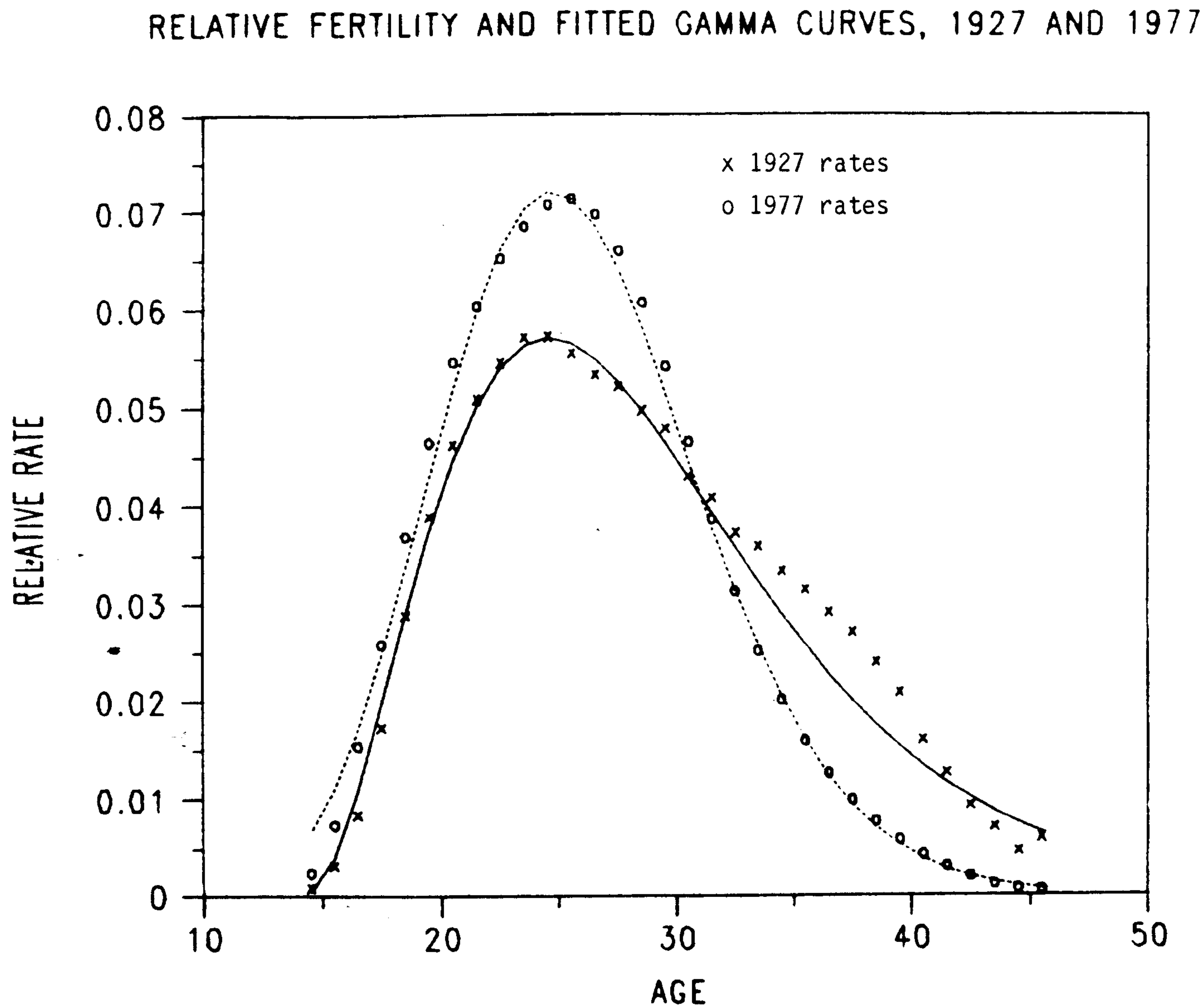

Figure 4. Two fitted relative fertility curves. We exhibit the fitted curves from 1927 and 1977, which have, respectively, the smallest and largest α values. The three adjustable parameters (A_0, α and β) allow the Gamma curves to summarize a variety of age-specific fertility patterns. The observed fertilities in these years are also shown; in general, the fitted curves provide good overall summaries of the data.

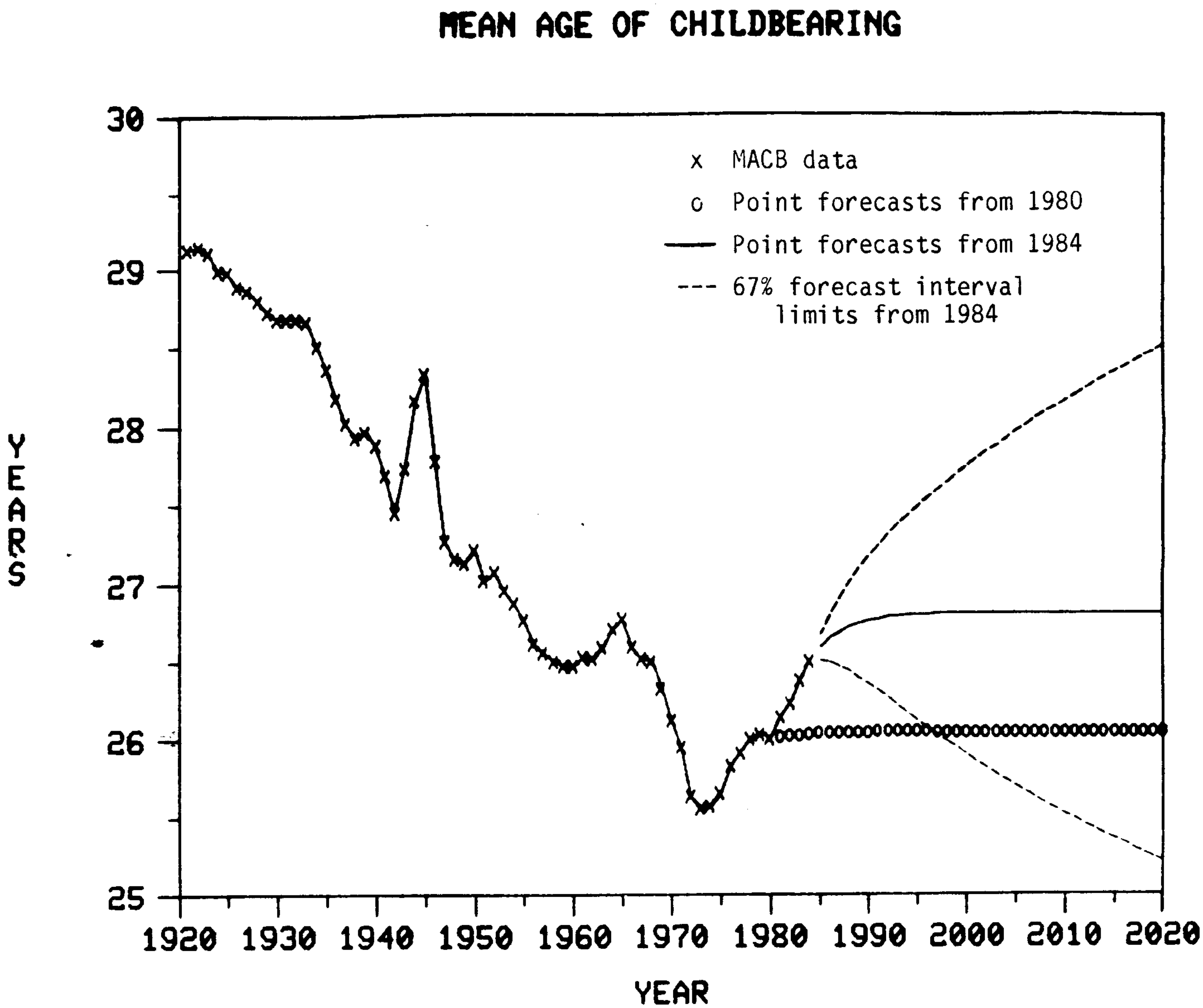

Figure 5. MACB data with point forecasts from 1980 and point and 67% interval forecasts from 1984. Forecasting for $M_t = \ln(MACB_t)$ uses model (3.1) with estimated parameters, and then exponentiates to provide point forecasts and interval limits for $MACB_t$. Model parameters are estimated using data up through the forecast origin years (1980 and 1984).

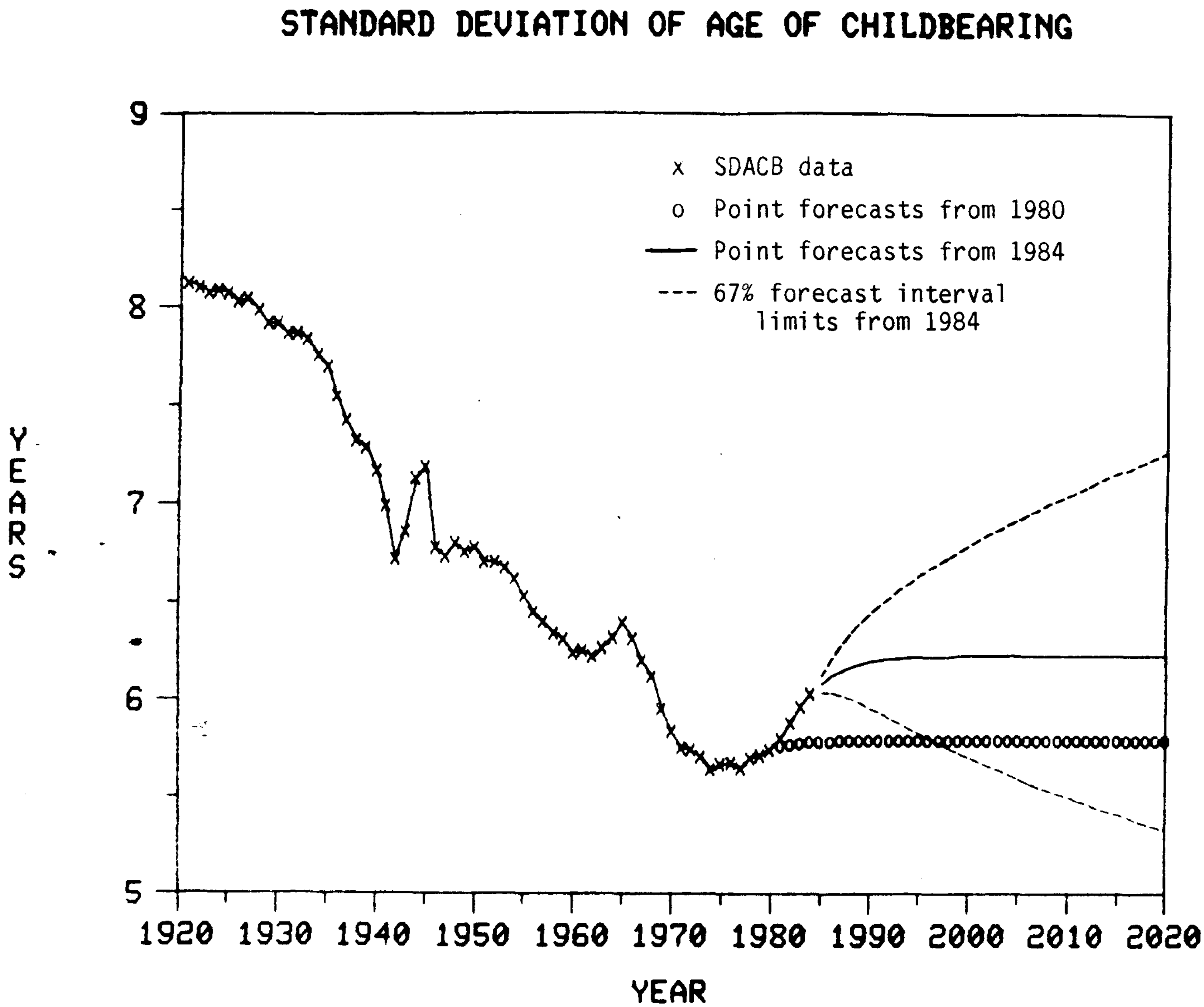

Figure 6. SDACB data with point forecasts from 1980 and point and 67% interval forecasts from 1984. Forecasting for $S_t = \ell n(SDACB_t)$ uses model (3.1) with estimated parameters, and then exponentiates to provide point forecasts and interval limits for $SDACB_t$. Model parameters are estimated using data up through the forecast origin years (1980 and 1984).

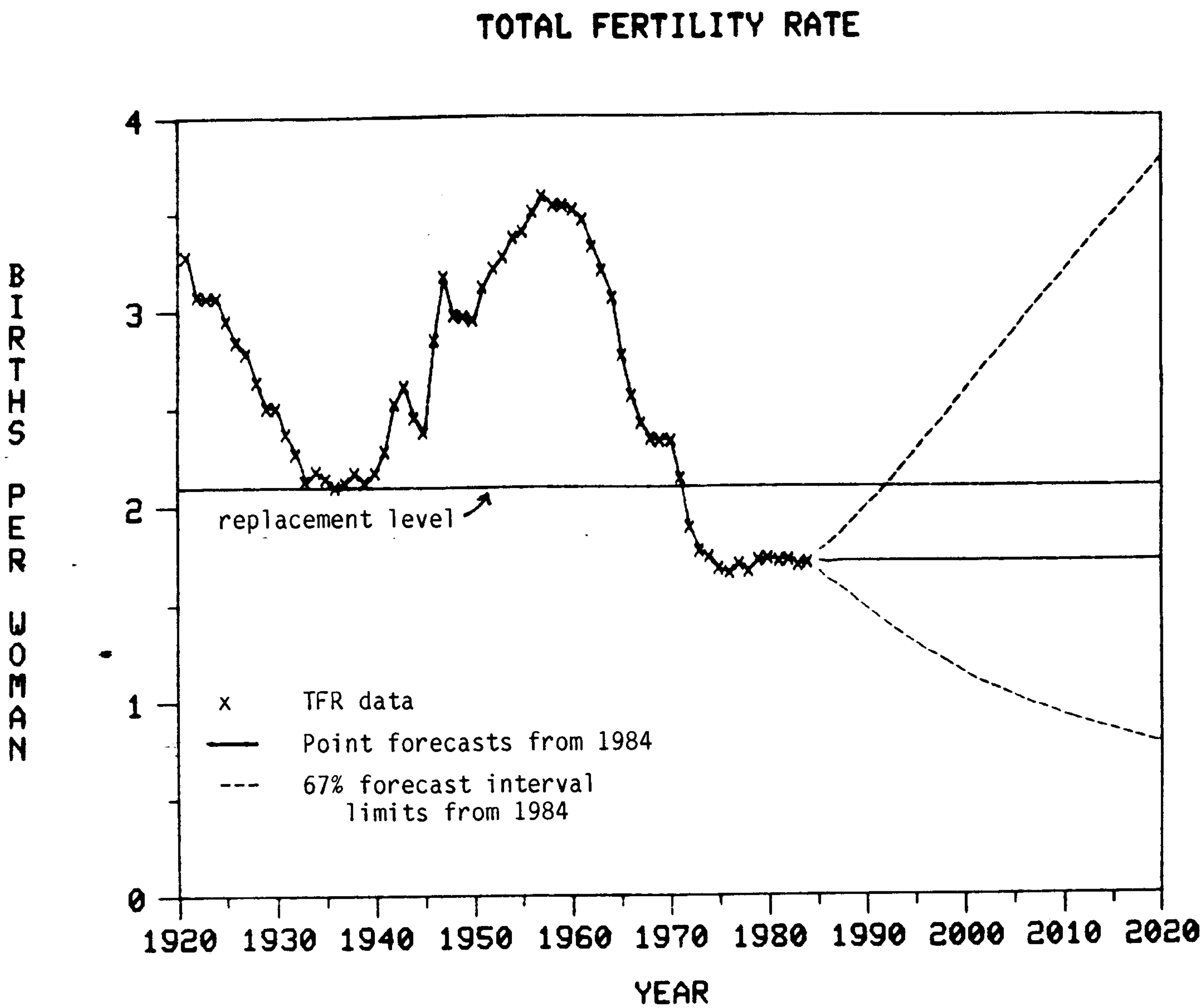

Figure 7. TFR data with point and 67% interval forecasts from 1984. Forecasting for $T_t = \ln(TFR_t)$ uses model (3.1) with estimated parameters, and then exponentiates to provide point forecasts and interval limits for TFR_t. Notice the asymmetry of the resulting forecast intervals. Forecasts from 1980 are not shown since they would almost coincide with 1981-1984 data and the 1984 point forecasts. Also shown for reference is a line at 2.1, the "replacement level" for TFR (see text).

Figures 8a - 8d. Actual (X) and forecasted (0) fertility rates, and fitted (—) and forecasted (--) gamma curves, 1981-1984. The curve parameters are forecasted from 1980 using model (3.1) estimated with data through 1980. The forecasted parameters produce forecasted (scaled and shifted) gamma curves (dotted lines) which may be compared to the fitted (scaled and shifted) gamma curves (solid line) which are obtained when the data for a given year become available. The forecasted curves are then "bias adjusted" (see text) to produce forecasts of age-specific fertility rates (0) which may be compared to the actual fertility rates (X). The graphs show the importance of bias adjusting the forecasted gamma curves in producing short-term, age-specific, fertility rate forecasts.

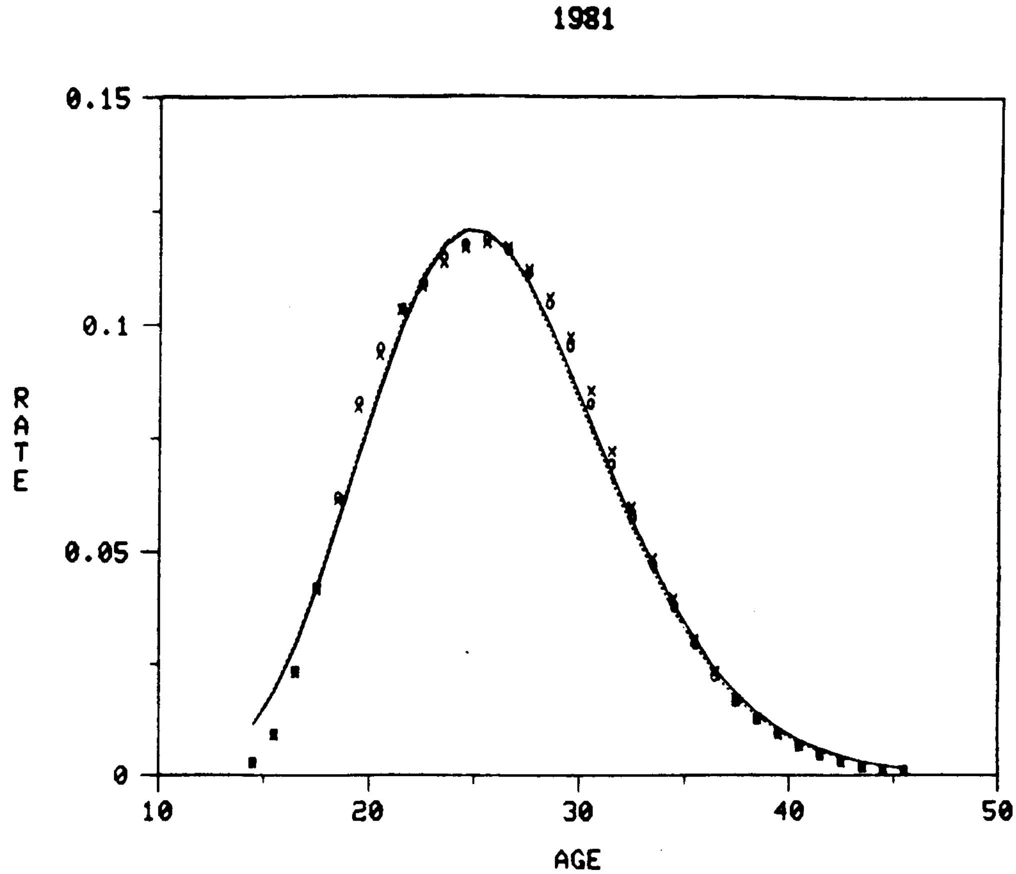

Figure 8a

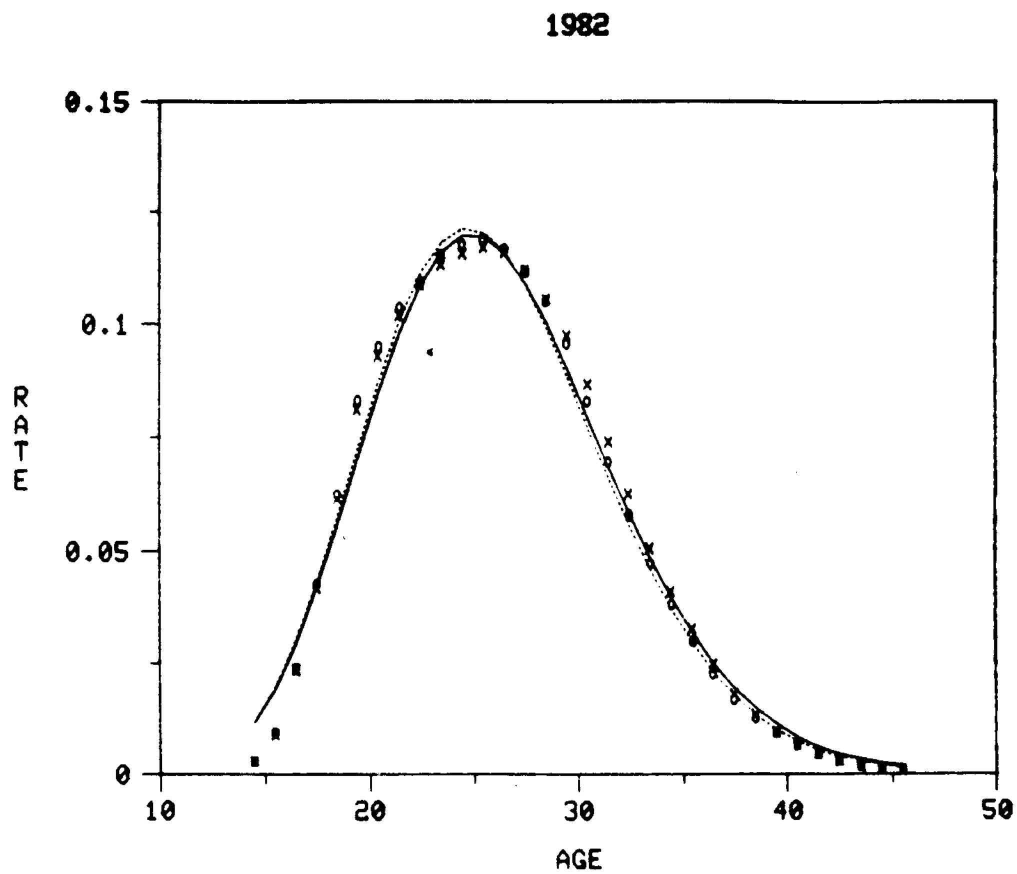

Figure 8b

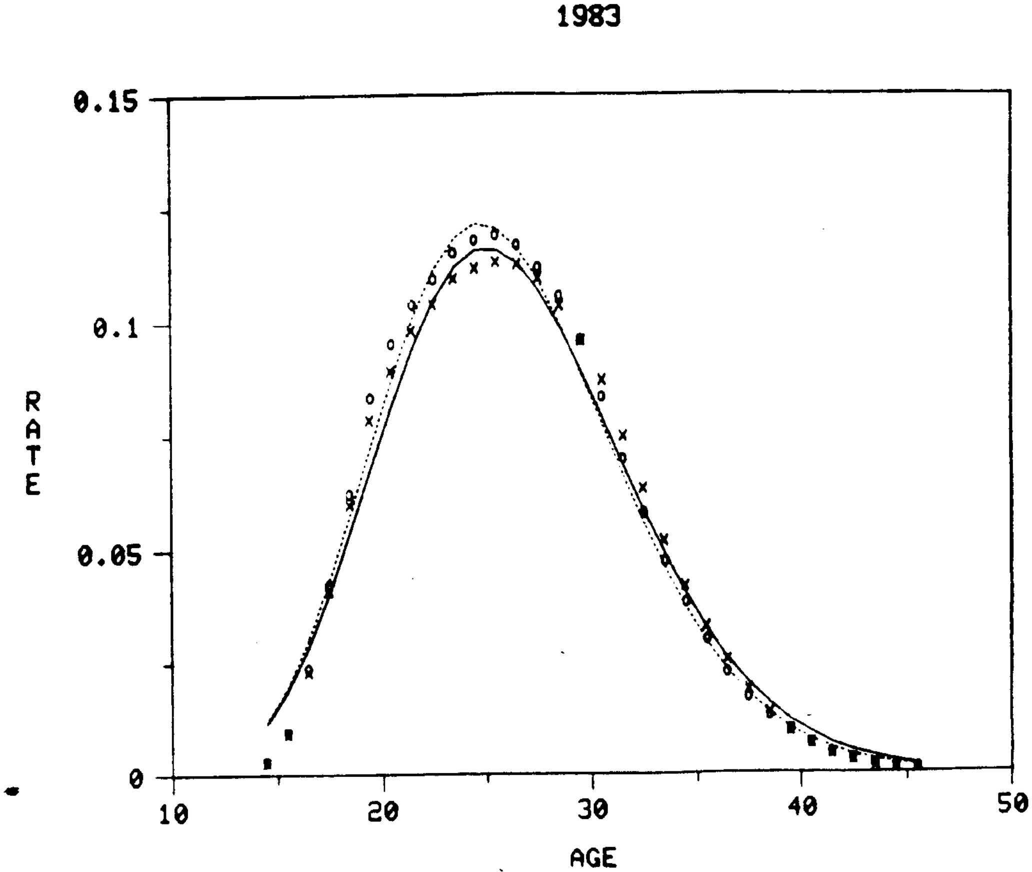

Figure 8c

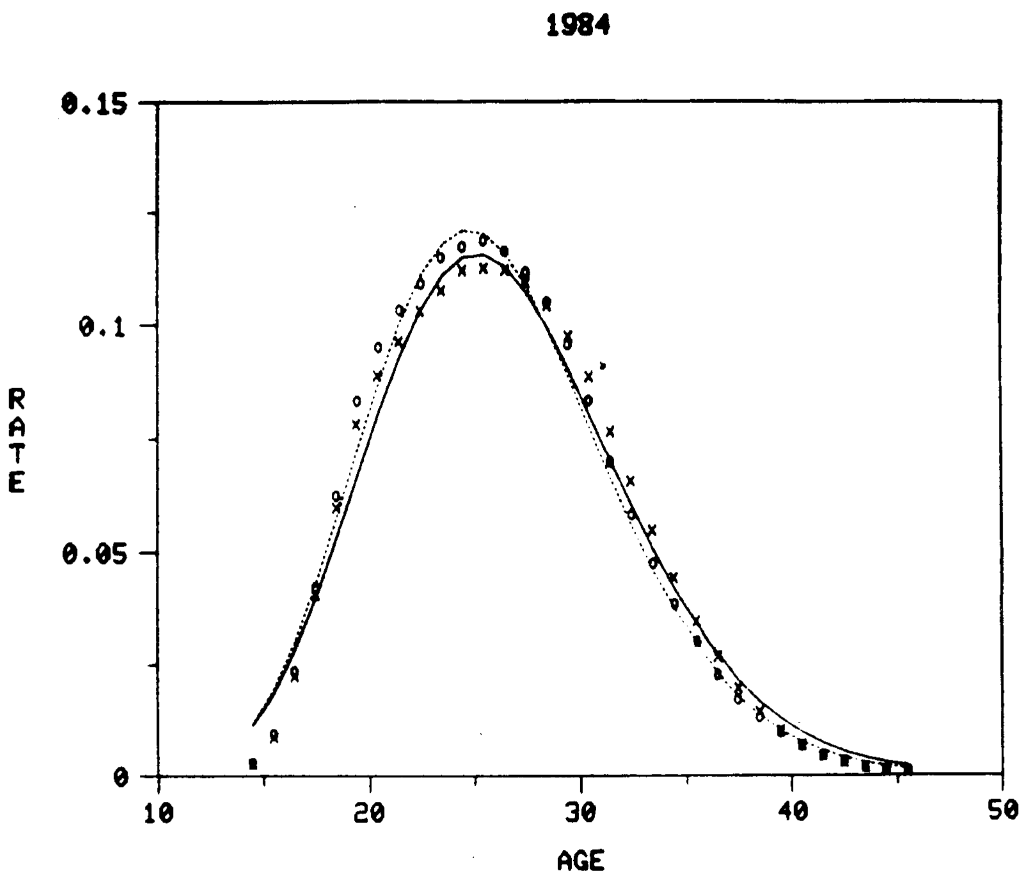

Figure 8d

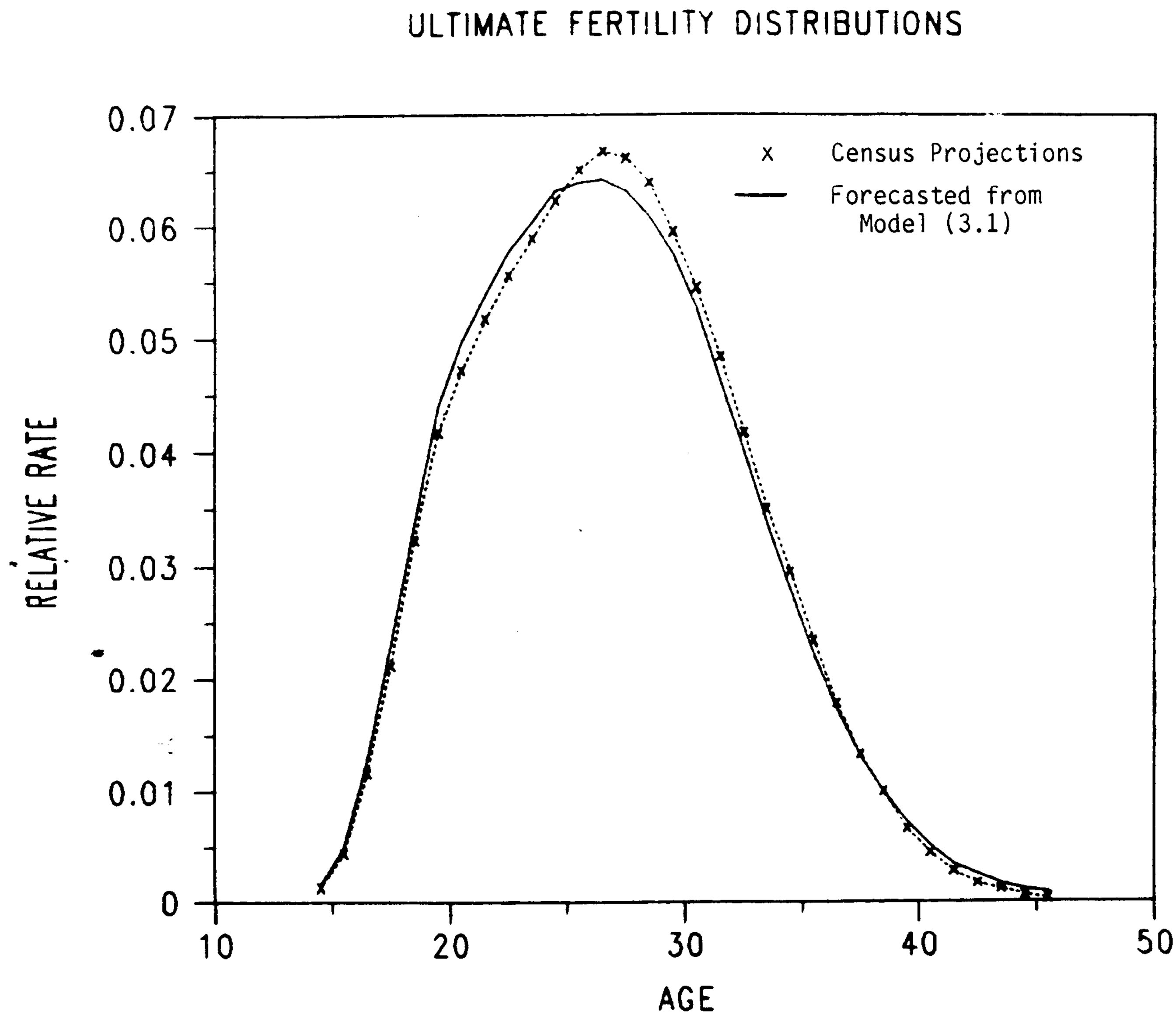

Figure 9. Ultimate fertility distribution -- Census projections and forecasts from model (3.1). The graph compares the ultimate fertility distribution (relative fertility rates) from the U.S. Bureau of the Census (1988) projections, which was determined judgmentally as described in the text, with the ultimate fertility distribution obtained from the bias adjusted gamma curve resulting from the ultimate forecasts of the curve parameters from model (3.1) (though TFR is not needed here). The Census projections show a slightly higher ultimate MACB and slightly lower ultimate SDACB than those resulting from the model (see Table 3), corresponding here to the Census projections having a slightly later and narrower peak. Notice also that the Census projections for ages 40 and over are lower than the model forecasts. The Census projections hold these relative rates at their 1984 values to avoid projecting (as the model does) an increase in fertility at these ages that is large in percentage terms though small in absolute magnitude.