BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-88/19

APPLYING TIME SERIES MODELS IN FORECASTING AGE-SPECIFIC FERTILITY RATES

by

William Bell
Statistical Research Division
Bureau of the Census
Washington, D.C. 20233

Recommended by: Nash J. Monsour

Report completed: September 9, 1988

Report issued: September 9, 1988

APPLYING TIME SERIES MODELS IN FORECASTING AGE-SPECIFIC FERTILITY RATES

by

William Bell

Statistical Research Division
U.S. Bureau of the Census
Washington, D.C. 20233

September 6, 1988

This paper will be presented at the International Workshop on National Population Projections in Industrialized Countries from October 25-28, 1988 at The Hague, The Netherlands.

Applying Time Series Models in Forecasting
Age-Specific Fertility Rates

William Bell
U.S. Bureau of the Census

1. Introduction

Fertility projections often involve the forecasting of age-specific fertility rates. This may be done to take advantage of the known age structure of the existing female population in using the cohort-component approach to fertility projection, or because age-specific fertility is itself of interest. If rates for single years of age are used, this creates a forecasting problem of large dimension, with fertility rates for 30 or more ages to forecast. If long-term projections are being made, care must also be taken to insure that projections for different ages are consistent in the sense of the long-term shape of fertility across age looking reasonable in comparison to historical data.

One approach to general forecasting problems involves the use of statistical time series models. In this approach one selects a particular time series model for the series to be forecast, fits the model to the data, and uses the fitted model to produce point and interval forecasts. The autoregressive-integrated-moving average (ARIMA) models discussed by Box and Jenkins (1970) comprise one popular class of models. These models have been applied to fertility projection and related problems by Lee (1974, 1975), Saboia (1977), McDonald (1979, 1981), Miller and Hickman (1981), Land and

Cantor (1983), Miller and McKenzie (1984), Carter and Lee (1986), Miller (1986), Bell et al. (1988), and Bozik and Bell (1988). The use of these models need not be exclusive: they can be used in combination with other models (e.g. econometric models) and with demographic judgmental projections. Bell et al. (1988) projected fertility by combining short-term forecasts from time series models with long-term demographic judgmental projections.

This paper is concerned with applying time series models in doing fertility projections. In section 2 we give a review of ARIMA time series models and their use in forecasting. Section 3 discusses general considerations in using time series models to forecast fertility. We suggest time series models are useful tools for short-term forecasting with 50 or more observations of age-specific fertility time series data analyzed on a period basis. Time series models can be used in a more judgmental fashion if less data are available, and in long term forecasting. Transformation to the TFR (total fertility rate) and relative fertility rates, and then to the logarithms of these, is likely to be helpful. Forecast intervals from time series models provide useful guides to the amount of uncertainty in the forecasts, and can be used to help develop alternative projections. Section 4 considers in more detail some approaches to applying time series methods in forecasting age-specific fertility, with particular attention to the dimensionality and consistency problems mentioned earlier. We present results of some recent research at the U.S. Census Bureau where these problems were addressed by fitting scaled and shifted gamma curves to age-specific fertility rates, and another approach based on principal components.

The analysis presented in sections 3 and 4 primarily uses fertility rate data for white women in the U.S. from 1921-84. Data for women of ages 14 and under, single years of age 15 through 48, and 49 and over are available. (The upper age limit actually used varies in different analyses.) Some analyses are also shown of fertility rate data for women in The Netherlands for 1950-86, covering ages 15 and under, single years of ages 16 through 48, and 49 and over.

It should be kept in mind that many of the general ideas discussed here could be applied to other problems, such as forecasting age-specific mortality rates, age-specific marriage rates, etc. We expect to investigate some of these applications (particularly mortality projection) at the Census Bureau in coming years.

2. ARIMA Time Series Models and Their Use In Forecasting

Detailed discussions of ARIMA time series modeling are given in the books by Box and Jenkins (1970) and Abraham and Ledolter (1983); the latter emphasizes the use of ARIMA models in forecasting. Brockwell and Davis (1987) is a more theoretical book on time series models that covers some recent developments. Tiao and Box (1981) discuss multivariate ARIMA modeling. Bell (1984) also gives a brief review of time series models and gives examples of forecasting of birth rate series. Several statistical packages are available for doing time series modeling and forecasting; most of the analysis for the examples presented later was done with the SCA statistical package (Liu et al 1986).

We shall only give a brief summary of time series models here, and refer the reader to the above references for thorough discussions of modeling. We shall assume the time series has already been modeled and we are proceeding with the fitted model as if it were the true model. We shall be primarily concerned in this paper with the application of these models in forecasting age-specific fertility.

2.1 ARIMA Time Series Models

Let Y_t be the (univariate) time series being modeled and forecast. A simple time series model is the autoregressive model (AR(p))

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + a_t \tag{2.1}$$

where $\phi_1, \ldots, \phi_p$ are parameters, p is the the order of the model, the a_t's are independent, identically distributed $N(0,\sigma^2)$, and we assume, for now, $E(Y_t) = 0$. (2.1) looks much like a regression model for Y_t in terms of its own past values - hence the name autoregressive model. Letting B be the backshift operator ($BY_t = Y_{t-1}$ and $B^jY_t = Y_{t-j}$) we can write (2.1) as

$$(1-\phi_1 B - \ldots - \phi_p B^p)Y_t = a_t \tag{2.2}$$

or $\phi(B)Y_t = a_t$ where $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$.

The moving average model (MA(q)) is

$$Y_t = a_t - \theta_1 a_{t-1} - \ldots \theta_q a_{t-q} \tag{2.3}$$

or $Y_t = (1-\theta_1 B - \ldots - \theta_q B^q)a_t$ or $Y_t = \theta(B)a_t$. For reasons we shall not go into here, we shall generally assume the zeroes of the polynomials

$1-\phi_1 x - \ldots - \phi_p x^p$ and $1-\theta_1 x - \ldots - \theta_q x^q$ are greater than 1 in absolute value. Additional flexibility results from combining (2.2) and (2.3) to get the autoregressive-moving average model (ARMA(p,q))

$$(1-\phi_1 B - \ldots - \phi_p B^p)Y_t = (1-\theta_1 B - \ldots - \theta_q B^q)a_t. \tag{2.4}$$

We can allow for a nonzero mean of Y_t in (2.2) - (2.4) by replacing Y_t by $Y_t-\mu$, where $\mu = E(Y_t)$ is the same for all t.

In the above we essentially assumed the time series Y_t to be stationary. This effectively means the statistical properties of any segment of the Y_t series are the same as those of any other segment, including the assumption that $E(Y_t) = \mu$ remains constant over time. Stationarity is an unrealistic assumption for most demographic and economic time series. One generalization of the models to deal with this, allowing $E(Y_t)$ to vary over time, will be discussed later. Another useful generalization assumes that not Y_t itself, but some difference of Y_t, is stationary. For example, we might use an ARMA model (2.4) for the first difference of Y_t:

$$Y_t - Y_{t-1} = (1-B)Y_t = \nabla Y_t \qquad (\nabla = 1-B)$$

or the second difference

$$Y_t - 2Y_{t-1} + Y_{t-2} = (1-B)[(1-B)Y_t] = (1-B)^2 Y_t = \nabla^2 Y_t \ .$$

In general, we may need to take the d^{th} difference, $\nabla^d Y_t$, though rarely is d larger than 2. Substituting $\nabla^d Y_t$ for Y_t in (2.4) yields the ARIMA (p,d,q) model

$$(1-\phi_1 B - \ldots - \phi_p B^p)\nabla^d Y_t = (1-\theta_1 B - \ldots - \theta_q B^q)a_t \qquad (2.5)$$

or $\phi(B)\nabla^d Y_t = \theta(B)a_t$. (The "I" in ARIMA stands for "integrated", the inverse of differencing). If $\nabla^d Y_t$ has a constant nonzero mean, this can be allowed for by adding a parameter θ_0 to the right hand side of (2.5):

$$\phi(B)\nabla^d Y_t = \theta_0 + \theta(B)a_t \qquad (2.6).$$

As will be discussed in section 3, another useful generalization is to let Y_t be some transformation (e.g. logarithm) of the original series of interest.

Two further generalizations of these models are worth mentioning. The first is seasonal models for sub-annual series (e.g. monthly) with an annual cycle. These models allow additional AR, MA, and differencing operators in the backshift operator B raised to the seasonal period (e.g. B^{12} for monthly data). This lets the models explain the strong relations between observations that are one or more full years apart. One useful model is the $\text{ARIMA}(0,1,1)\text{x}(0,1,1)_{12}$ model

$$(1-B)(1-B^{12})Y_t = (1-\theta_1 B)(1-\theta_{12}B^{12})a_t.$$

This model and others are discussed in detail by Box and Jenkins (1970). Land and Cantor (1983) used such models with monthly birth and death rate series.

We shall not consider seasonal models further since the applications we are concerned with here are to annual time series of fertility rates.

The second generalization is to multivariate, or vector, ARIMA models. If not one, but k time series $Y_{1t},\ldots,Y_{kt}$ are to be modeled, we generalize the previous models by replacing Y_t by $\underset{\sim}{Y}_t = (Y_{1t},\ldots,Y_{kt})'$, a_t by $\underset{\sim}{a}_t = (a_{1t},\ldots,a_{kt})'$ which is iid $N(0, \Sigma)$ where $\Sigma = Var(\underset{\sim}{a}_t)$ is kxk, and parameters ϕ_i and θ_j by kxk parameter matrices Φ_i and Θ_j. Thus, the multivariate ARIMA(p,d,q) model generalizing (2.5) is

$$(I-\Phi_1 B - \ldots - \Phi_p B^p)[\nabla^d \underset{\sim}{Y}_t] = (I - \Theta_1 B - \ldots - \Theta_q B^q)\underset{\sim}{a}_t \qquad (2.7)$$

where I is the kxk identity matrix. Notice in (2.7) that $\nabla^d = (1-B)^d$ remains a scalar operator, indicating each series Y_{it} is differenced the same number of times. This is common in practice, though not essential. One can also allow nonzero means by replacing $\underset{\sim}{Y}_t$ by $\underset{\sim}{Y}_t - \underset{\sim}{\mu}$ with $\underset{\sim}{\mu}$ a k×1 mean vector (if d=0), or putting a kx1 vector $\underset{\sim}{\theta}_0$ on the right hand side of (2.7), analogous to (2.6). The vector AR(p) model can be written as

$$\underset{\sim}{Y}_t = \Phi_1 \underset{\sim}{Y}_{t-1} + \ldots + \Phi_p \underset{\sim}{Y}_{t-p} + \underset{\sim}{a}_t \qquad (2.8)$$

Such models (possibly with $\nabla^d \underset{\sim}{Y}_t$, $\nabla^d \underset{\sim}{Y}_{t-1},\ldots$, instead of $\underset{\sim}{Y}_t$, $\underset{\sim}{Y}_{t-1}$, ...) are used extensively in econometrics. (2.8) yields an equation for each Y_{it} in terms of p past lags of itself and each of the other series Y_{jt} $j\neq i$, and a shock a_{it} which may be correlated with shocks a_{jt} from the other equations ($j\neq i$), implicitly allowing contemporaneous relations between Y_{it} and Y_{jt}.

The applications in section 4 exclusively use autoregressive models for differenced data, both univariate and multivariate. Evidence from the data for moving average terms in models was not strong. Models with MA terms tend to be more important with seasonal time series.

2.2 Regression Models with ARIMA Errors

A linear regression model for the time series Y_t is

$$Y_t = \beta_1 X_{1t} + \dots + \beta_m X_{mt} + e_t \tag{2.9}$$

where $X_{1t}, \dots, X_{mt}$ are explanatory (independent) variables, deterministic variables observed over time, and the error term e_t has mean 0. The expression $\beta_1 X_{1t} + \dots + \beta_m X_{mt}$ models a changing mean function, $E(Y_t)$. Standard regression analysis would assume the e_t are uncorrelated over time, but this is unrealistic for time series data. Instead, we can let e_t follow an ARIMA model. We write the combined model as (2.9) and (2.5) with e_t substituted for Y_t in (2.5), or we write the model in one equation as

$$\phi(B)\nabla^d[Y_t - \sum_1^m \beta_i X_{it}] = \theta(B)a_t \tag{2.10}$$

We can look at the model (2.10) as either generalizing the error structure of the regression model (2.9), or as generalizing the mean function of (2.5) (where we assumed $E(\nabla^d Y_t) = 0$).

Regression terms in time series models such as (2.10) have many uses. Bell and Hillmer (1983) give an approach to developing such models, with particular attention to modeling calendar "trading day" and holiday effects in monthly economic time series. Miller and McKenzie (1984) apply such models to monthly birth rate series, finding calendar effects in the data, and showing the use of regression variables to account for these produces much better models than pure ARIMA models such as were used by Land and Cantor (1983). Modeling effects of known interventions (Tiao and Box 1975) and unknown outliers (Bell 1983) are useful applications where the X_{it} are appropriate indicator variables; such applications will be illustrated later. Another useful application with seasonal data is the modeling of a deterministic seasonal component with a regular annual cycle.

Polynomial regression, i.e. use of $1, t, t^2, \ldots$ for $X_{1t}, X_{2t}, X_{3t}, \ldots$ in (2.9), has often been used in time series forecasting. However, if we use polynomial X_{it} in (2.10) with d>0, some of the terms are wiped out since $\nabla^d t^j = 0$ for $j=0,1,\ldots,d-1$; hence the coefficients $\beta_1, \ldots, \beta_{d-1}$ would not be estimable. Thus, the ARIMA(p,d,q) model (2.5) implicitly allows for a polynomial of degree d-1 in modeling and forecasting Y_t. This implicit polynomial is adaptive and need only apply locally, in the sense that the polynomial coefficients are effectively redetermined as each new data point is added. The model (2.6) implicitly allows for a polynomial of degree d which is non-adaptive in the sense that the coefficient of t^d is $\theta_0/d!$ at all time points. Because of these results, it is rarely necessary to explicitly include polynomial terms in ARIMA models beyond the θ_0 in (2.6).

The problem with using polynomial regression in modeling and forecasting time series is that the assumption that the regression error terms are uncorrelated over time is virtually always unrealistic. Bell (1984) points out that this has the following bad effects: the behavior of long run forecasts is unreasonable (tending to $+\infty$ or $-\infty$), if the fit of the curve at the end of the series is poor short run forecasts are likely to be bad, and variances of forecast errors from regression theory are usually highly unrealistic. ARIMA models tend not to suffer from these drawbacks. In fact, if a polynomial regression model of degree d is really appropriate, then the time series modeling process should lead approximately to the model

$$\nabla^d Y_t = (1-B)^d Y_t = \theta_0 + (1-B)^d a_t \ .$$

Solving this difference equation for Y_t leads back to the polynomial regression model. (See Abraham and Ledolter 1983.) Thus, ARIMA models allow for polynomial regression when appropriate.

If X_t in (2.10) (assuming only one X for simplicity of discussion) is not deterministic but is itself a stochastic time series, then (2.10) becomes a particular case of a transfer function model; these are discussed in detail in Box and Jenkins (1970). For forecasting purposes the distinction between a deterministic and stochastic X_t is important since if X_t is stochastic its future values are not known, and to forecast Y_t using (2.10) we must also forecast X_t. This shifts some of the responsibility for forecasting Y_t onto forecasting X_t, and the error in forecasting X_t leads to part of the error in forecasting Y_t (Box and Jenkins 1970). Transfer function models are useful in forecasting if (1) the relationship between Y_t and X_t is strong, so that X_t explains a significant amount of the variability in Y_t, and (2) X_t can be more accurately forecast than Y_t. One case where (2) occurs is for "leading indicators", where Y_t this period depends heavily on the value of some other variable last period, (X_{t-1}) or in some other preceeding period (X_{t-r}), in which case the explanatory variable is known exactly at least through r time periods ahead when forecasting Y_t. McDonald (1981) found marriages to be a useful leading indicator in transfer function models for forecasting first births in Australia. For the most part the preceeding comments apply also to jointly modeling and forecasting Y_t and X_t with a multivariate ARIMA model, of which transfer function models are a special case (Tiao and Box 1981).

2.3 Forecasting With ARIMA Models

Given an ARIMA model, forecasts having minimum mean squared error may be easily computed recursively as shown in Box and Jenkins (1970). (Brockwell and Davis (1987) consider some additional theoretical details.) Denote such a forecast of $Y_{n+\ell}$ as $\hat{Y}_n(\ell)$, where n is the forecast origin (time of the last data point), and ℓ the forecast lead. When using a model with regression terms, (2.10), one forecasts $\tilde{Y}_{n+\ell} = Y_{n+\ell} - \sum_i^m \beta_i X_{i,n+\ell}$, and then adds $\sum_i^m \beta_i X_{i,n+\ell}$ back in to get $\hat{Y}_n(\ell)$. Variances of forecast errors, $V(\ell) = \mathrm{Var}(Y_{n+\ell} - \hat{Y}_n(\ell))$, are also easily computed. One can then compute forecast intervals under normality for the future values $Y_{n+\ell}$ from

$$\hat{Y}_n(\ell) - k(V(\ell))^{1/2} < Y_{n+\ell} < \hat{Y}_n(\ell) + k(V(\ell))^{1/2} \qquad (2.11)$$

where, for example, k=1 would yield an approximate 67% interval, and k=2 an approximate 95% interval. Time series software packages such as SCA (Liu et al 1986) can perform the computations required.

Consider the "forecast function", $\hat{Y}_n(\ell)$ as a function of forecast lead ℓ. Box and Jenkins note that for an ARIMA(p,d,q) model $\hat{Y}_n(\ell)$ satisfies a difference equation for $\ell > q-p-d$ (for all ℓ if $q \le p+d$). As such, it is a linear combination of damped exponentials, damped sine waves, and a polynomial (if $d > 0$). For large ℓ the polynomial term will eventually dominate. One can also obtain results about the behavior of the forecast error variance function, $V(\ell)$. The behavior of $\hat{Y}_n(\ell)$ and $V(\ell)$ for large ℓ with an ARIMA(p,d,q) model can be summarized as follows:

(i) If $d = 0$ then $\hat{Y}_n(\ell) \to \mu$ and $V(\ell) \to Var(Y_t)$ as $\ell \to \infty$ where $\mu = 0$ for (2.5), $\mu = \theta_0/(1-\phi_1 - \dots - \phi_p)$ for (2.6), and μ is replaced by $E(Y_{n+\ell}) = \sum_1^m \beta_i X_{i,n+\ell}$ for (2.10).

(ii) If $d > 0$ then $\hat{Y}_n(\ell)$ is eventually dominated by a polynomial of degree d-1 for (2.5), and of degree d for (2.6). For (2.10) $\hat{Y}_n(\ell)$ is eventually the sum of a polynomial of degree d-1 and the regression function $\sum_1^m \beta_i X_{i,m+\ell}$. Also, if $d > 0$

$$V(\ell) \to \infty \text{ as } \ell \to \infty .$$

Table 1 summarizes the behavior of $\hat{Y}_n(\ell)$ and $V(\ell)$ for some popular models.

3. General Considerations in Using Time Series Models to Forecast Age-Specific Fertility

The general considerations we discuss here are not necessarily restricted to the use of ARIMA time series models in forecasting age-specific fertility. It should be clear that many apply to the use of other types of models, and even to judgmental forecasting. Some obviously apply to other forecasting problems, such as forecasting age-specific mortality.

3.1 Length of Series

One of the first questions that arises when considering the use of time series models for forecasting is how much time series data there is to work with. This question is important because the usual approach uses the available time series data to select and fit a model, and then uses the fitted model in forecasting as if it were the "true model". The more data we have (the longer the time series) the more likely it is that we will select a model that can approximate the structure of the data well, and the better our parameter estimates will be. Both of these make the assumption that we know the "true model" more reasonable. Recent research (Thompson and Miller 1986, Ansley and Kohn 1986) attempts to deal with the effect of uncertainty about model parameters in forecasting.

As to exactly how much data is needed for time series modeling, Box and Jenkins (1970, p. 18) suggest at least 50 observations, though this requirement obviously can depend on the particulars of the series involved. We generally prefer to use as much data as possible, subject to some qualifications mentioned later. Some things can be done using time series models with more limited data. Model selection might be based on properties of the resulting forecast function, since the data will be unlikely to supply definitive evidence about what model is appropriate. This approach was taken by Bell et al. (1986) in forecasting household headship proportions with 27 years of annual data. When forecasting many short time series it may be possible to make assumptions that effectively increase the sample size for estimating the parameters. For example, with few observations over time of age-specific fertility rates, we might consider assuming the same model for each age or across groups of ages. Thisted and Wecker (1981) discuss a generalization of the "same model" idea. In general, the more limited are the available data, the more assumptions one must make in using time series models, bringing this closer to doing judgmental forecasting. With very limited data one might pick both the model and its parameters judgmentally, according to the behavior of the resulting forecasts.

However much data is available, there may still be a question about how much of the data should be used in modeling. While generally more data are better, this assumes that the same model applies over all the data. However, there may be known events thought to affect the data, or other reasons to suspect the structure of the data may have changed over time. There are two basic ways of dealing with this: drop the affected data, or attempt to model the effects. Regression terms are useful for the latter as discussed in section 2. Here we shall mention some of the data considerations in the development of time series models to forecast white fertility in the U.S. (see Bell et al. 1988).

Age-specific fertility rate data were available for the years 1917-84, but ultimately data for 1917-20 were dropped because of concerns about effects of World War I and the 1919 flu epidemic on fertility. It was thought that retaining these four years of data was not worth the trouble of trying to model these effects. The data for 1942-47 appear to be affected by World War II (see Figures 3, 8, 9, and 12). While some approaches to fitting time series models will handle missing data (e.g. Jones 1980), so that data for these years could potentially have been dropped, our existing computer software did not allow this. Instead, indicator regression variables for the years 1942-47 were used as suggested in section 2. Other empirically determined outliers were also handled this way. (See Bell et al. 1988.) Miller and Hickman (1981) discuss empirical evidence that different models should be used for pre- and post- World War II U.S. fertility data. However, their analysis used the suspect 1917-20 data; reanalysis of TFR without this data suggested a possibility of model change but did not offer conclusive evidence. There was also some concern about changes in the quality of the data over time. While birth registration in the U.S. in recent years is essentially complete, this was not so in earlier years. However, our data were corrected for underregistration of births (Passel, Rives, and Robinson 1977). Perhaps a more serious consideration was the quality of population estimates in earlier years. In particular, population figures for 1921-29 were obtained by interpolating between the 1920 and 1930 censuses, rather than estimated by demographic analysis as were figures for later years. This might suggest dropping the 1920-29 data, though we did not try this.

Apart from modeling considerations, sometimes the relevance of data in the distant past for forecasting is questioned. However, forecasts from time series models depend most heavily on recent data, with diminishing weight given to data distant from the forecast origin. (See Table 1.) In fact, forecasts from an AR(p) or ARIMA(p,d,0) model are determined entirely from the last p or last p+d data points, respectively. Thus the question of whether all available data should be used is really a consideration for modeling, not directly for forecasting.

3.2 Period or Cohort Basis

A fundamental question when analyzing age-specific fertility time series is whether the data should be analyzed on a period (indexed by calendar year) or cohort (indexed by year of mother's birth) basis. Demographic theories about fertility may be formulated on a cohort basis (e.g. the Easterlin hypothesis), while economic theories may emphasize the simultaneous effects of economic conditions on fertility at all ages, suggesting a period basis. Data from birth expectations surveys may fit naturally into a cohort analysis, but might also be thought of as reflecting current attitudes that may change over time and thus be useful in a period-based analysis.

We have not attempted to directly resolve the period-cohort controversy, but have done our analysis on a period basis for two essentially practical reasons. First, 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 its first births are observed. Second, in our data cohort fertility rates do not follow smooth patterns across age as do period fertility rates.

Both of these problems are illustrated in Figure 1. Figure 1a shows the U.S. period age-specific fertility rates for 1927, 1957, and 1977. Although total fertility was markedly different in these years, the three sets of rates have a similar smooth shape over age. Figure 1b shows U.S. fertility rates for the 1902, 1932, and 1952 birth cohorts, which reach age 25 in 1927, 1957, and 1977 respectively. This shows the incomplete data problems when analyzing data on a cohort basis, and shows rates for different cohorts do not follow the same smooth shape across age. The irregular shape of the rates is most pronounced in recent cohorts (for example, the 1952 cohort), which are most important for forecasting. The deviations of period fertility rates from a common shape are most pronounced in early years (e.g. 1927) which are the least important for forecasting. These problems would make it difficult to use the approaches to dimensionality reduction discussed in section 4 with the data on a cohort basis. Figures 2a and 2b show the corresponding period and cohort fertility rates for The Netherlands. (We did not have data for the year 1927 or for the 1902 cohort.) The same basic conclusions hold, though to a lesser degree; in particular, the 1952 cohort rates are smoother than in the U.S. data.

3.3 Transformation of the Data; Use of the Total Fertility Rate (TFR) and Relative Fertility Rates

If F_{it} is the fertility rate for mothers of age i in year t, the total fertility rate (TFR) in year t is

$$TFR_t = \sum_i F_{it}$$

where the sum extends over the range of childbearing ages. Now define the relative fertility rate for age i, year t by

$$R_{it} = F_{it} / TFR_t \qquad (3.1)$$

which is the proportion of births in year t that occur to mothers of age i. TFR measures the overall level of fertility in year t, while the R_{it} describe the distribution across age of fertility in year t (the shape of the fertility curve). The transformation from the F_{it} to TFR_t and the R_{it} can be quite useful for modeling and forecasting.

An interesting question is what relation is there, if any, between the level of fertility and the shape of the fertility curve, i.e. between TFR_t and the R_{it}. The analyses of Miller (1986), Bell et al. (1988), and Bozik and Bell (1988) all indicate at most a weak relationship in the U.S. data, suggesting the shape of the fertility curve depends little on the level of fertility. All at least suggest TFR can be modeled and forecast separately first, then the R_i can be forecast, quite possibly without allowing any dependence of their forecasts on those for TFR. This evidence is in sharp contrast to the approach suggested by Rogers (1986) which involves regression on TFR of the parameters of a curve fit to the R_i, to develop forecasts of the R_i from those of TFR.

The above discussion suggests an easy first step into using time series models for projecting fertility is to model and forecast TFR, and then develop R_i forecasts separately. The R_i could be projected to remain constant at their most recent values, or other forecasts could be developed judgmentally. (Note that since $\sum_i R_{it} = 1$ we can determine forecasts for R_{49}, say, from those of the other R_i's.) Using time series models to forecast the R_i leads us to consider the dimensionality problem discussed later.

Along with transforming to TFR and the R_i, we should consider other transformation of these series. The logarithm is a particularly useful transformation for any time series that is always positive. Forecasts and forecast interval limits for a log-transformed series may be exponentiated to yield forecasts and interval limits for the original series. The use of the logarithmic transformation has several important benefits:

(1) Taking logarithms often makes the model assumptions of normality and stable variance much more tenable. If a series shows more variability when it is at high levels than when it is at low levels, then taking logarithms may correct this and also make model residuals look more nearly normally distributed. Figure 3 shows TFR and log(TFR) for the U.S., and Figure 4 for The Netherlands. Taking logarithms appears beneficial, since it pulls down the baby boom peak in the data and relatively enhances the variability when TFR is at lower levels, as in recent years. Further modeling analysis has confirmed the benefits of taking log(TFR), though it has also indicated outliers that need to be dealt with whether or not logs are taken (Bell et al. 1988).

(2) Taking logarithms prevents forecasts and forecast interval limits for the original series from going below zero. This is attractive for both TFR and the R_i since they can never be negative. Figure 3 shows the results for TFR in the U.S. using a multivariate model given in Bell et al. (1988) for log(TFR). Figure 4 shows results for The Netherlands using an ARIMA(1,1,0) model for log(TFR). The forecast intervals for log(TFR) are

symmetric about the point forecast; when exponentiated the resulting intervals are asymmetric, with the lower limit asymptotically approaching zero. If untransformed TFR were modeled and forecast, the lower forecast limit would quickly go below zero, which does not make sense. It is possible to enforce a different lower limit than zero if it is believed a series will never fall below some other level. In the most recent set of U.S. Census Bureau fertility projections the transformation log(TFR-1) was used to prevent TFR forecasts and limits from going below 1.

(3) Additive relationships between logged series are interpretable as multiplicative relationships between the original series. In particular, corresponding to (3.1) we have

$$\log(F_{it}) = \log(TFR_t) + \log(R_{it})$$

Thus forecasts of $\log(TFR_t)$ and the $\log(R_{it})$'s are easily translated into forecasts of the $\log(F_{it})$'s, an ey can then be exponentiated into forecasts of the F_{it}'s. Also, use a regression model with time series errors (section 2) where both the dependent and independent variables are logged, the resulting regression equation implies multiplicative relations between the original variables.

(4) The first difference of a logged series is approximately a growth rate. That is

$$\nabla \log(Y_t) = \log(Y_t) - \log(Y_{t-1}) = \log(Y_t/Y_{t-1})$$

$$= \log\left(1 + \frac{Y_t - Y_{t-1}}{Y_{t-1}}\right)$$

$$\approx \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

ignoring all but the first term in $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$, which is approximately correct as long as $(Y_t - Y_{t-1})/Y_{t-1}$ is small. Since first differencing seems advised for most series, if we have also taken logarithms we often will effectively be modeling growth rates. Of course we could actually use the growth rate transformation, $(Y_t - Y_{t-1})/Y_{t-1}$ if this is preferred. The results should be close to those for $\nabla\log(Y_t)$ in most cases.

Other transformations than the logarithm are possible. Since the R_{it}'s are bounded below by zero <u>and</u> above by 1, another possible transformation for them is the logistic, defined by

$$Z_{it} = \text{logit}(R_{it}) = \log \frac{R_{it}}{1-R_{it}} = \log(R_{it}) - \log(1-R_{it})$$

with inverse $R_{it} = \exp(Z_{it})/(1 + \exp(Z_{it}))$. This will constrain forecasts and intervals for R_{it} to the interval (0,1). Bell et al. (1986) used the logistic transformation in projecting household headship proportions. However, in projecting relative fertility rates the logistic should yield results quite close to the logarithm, since the R_{it} are all rather small and not near 1, so that the $\log(1-R_{it})$ term is approximately linear in R_{it} and does not affect the analysis. Even upper forecast interval limits for R_{it} using $\log(R_{it})$ would usually not approach 1 for many years into the forecast horizon.

Box and Cox (1964) discuss the general family of power transformations, of which the logarithm is a limiting case. Ansley, Spivey, and Wrobleski (1977) discuss their use with time series models. Miller and Hickman (1981) use a square-root transformation of birth rates for five year age groups. However, it is not clear how to handle forecasts or limits that go below zero with this transformation, since simply squaring them reflects them back away from zero, which makes little sense. Thompson (1987) considers use of the logistic transformation on an interval whose endpoints are estimated from the data. This can be used with any series (he considers U.S. TFR) to produce forecasts and limits bounded both above and below.

3.4 Forecast Intervals and Alternative Projection

The Census Bureau has traditionally provided alternative sets of population projections derived from alternative assumed future courses of fertility, mortality, and migration. The most recent set of projections (U.S. Bureau of the Census 1988) includes 30 such projections. The alternative fertility projections were developed from upper and lower 67% forecast interval limits for TFR from a time series model for log(TFR-1) used through 1990, and interpolated to assumed ultimate high and low TFR values of 1.5 and

2.2 in 2020. Figure 5 shows the results. The most striking thing is how wide the forecast intervals of Figure 3a are relative to the alternative projections of Figure 5 beyond about 1994.

Forecast intervals from time series models attempt to reflect the uncertainty about the future course of the time series in the context of the model and data used in forecasting. As noted in section 2, in models with differencing the forecast error variance tends to infinity as the forecast lead time increases, resulting in ever widening forecast intervals. This basically says that the past data on the time series has little to say about the distant future. In the case of TFR, the wide movements in the historical data over relatively short spans of time (e.g. the post-war baby boom) suggest this conclusion without use of a time series model.

There are two ways of viewing this difference between forecast intervals from time series models and alternative demographic projections. One viewpoint is that long term judgmental projections incorporate demographic knowledge that is not used in the time series model forecasts, i.e. the long term future is not as uncertain as an analysis based solely on historical time series data would lead us to believe. Another viewpoint is that alternative demographic projections are not meant to reflect uncertainty in the same sense that forecast intervals do (giving an interval in which the value of the series at some future time point can be expected to fall with a certain probability). Forecast interval limits reflect the limits of what can be expected (based on the model) with a given probability, whereas users of alternative projections may not wish to plan for such extremes, at least not in the long term. (Short-term intervals, up to 5 or so years ahead, are not so wide.) More discussion between statisticians and demographers is warranted on this point, and should consider how alternative projections are actually used. Otherwise, people may be led to adopt more cynical versions of these viewpoints, such as that time series forecast intervals are ridiculously wide or the limits described by alternative demographic projections are ridiculously narrow.

4. The Dimensionality and Consistency Problems and Time Series Methods

The dimensionality and consistency problems in forecasting age-specific fertility were alluded to earlier. The dimensionality problem is the large number of time series to be forecast - 30 or more if data for single years of age are being used. A common means of addressing this problem is simply to use data grouped into broader age intervals, such as 5-year intervals. However, if there are benefits to be gained from using the cohort-component approach over simply forecasting births directly, then use of broader age intervals should lessen these benefits. Also, we shall see in section 4.4 that use of broader age intervals can be viewed as a non-optimal choice among linear transformations to reduce dimensionality.

The consistency problem refers to the possibility that if age-specific fertility rates are forecast or projected by a model or procedure that does not take into account their strong relationships, then the long-term projections may show a distribution across age that does not make intuitive sense in terms of the fertility curve not having the same sort of smooth shape as historical data. This does not mean that such forecasts are necessarily bad. Any long-term forecasts are likely to be substantially in error, and errors in forecasting TFR long-term are likely to be more important than errors in forecasting the age-distribution of fertility (the R_i's). The concern may be more over the intuitive appeal of the forecasts when looked at from either a period or cohort perspective. Even though recent U.S. cohorts have not shown a smooth distribution of fertility across age, this may be desired in long-term forecasts.

The dimensionality and consistency problems are connected. If the dimension is low (as when broad age intervals are used) it is not difficult to achieve consistency, nor is consistency of much concern. If historical data did not show such strong consistency (smooth shape of fertility rates over age) then dimensionality would be easier to address; in the extreme case where the time series are all unrelated they can all be forecast separately and we need not worry about reducing the dimensionality of the problem.

In the first two sections of this chapter we consider how two direct attempts at applying time series methods run up against the consistency and dimensionality problems. In the second two sections we present two approaches that have been investigated recently at the U.S. Census Bureau for dealing with these problems. In what follows we shall use the U.S. data for 1921-84 transformed to log(TFR) and the log(R_i)'s as discussed in section 3.3 (except for direct use of the R_i's in section 4.3). The problems and approaches illustrated here apply as well to use of data without the logarithm or relative fertility rate transformations.

4.1 Univariate Time Series Models for Each Age Separately

Section 3.3 suggested a simple time series approach to forecasting age-specific fertility is to transform to TFR and the R_i, model and forecast log(TFR), and forecast the R_i to remain constant at their values in the last year of data. This corresponds to using separate univariate random walk models for the R_i (or $\log(R_i)$). This is not necessarily a bad approach. There is no consistency problem since the forecasted shape of the fertility curve each year is the same as in the last year of data, and no dimensionality problem since only TFR is being modeled. However, this approach will not forecast any change in the shape of the fertility curve. In recent years in the U.S. TFR has remained relatively stable (see Figure 3), while there have been changes in the distribution of fertility over age (reflected in Figures 8 and 9). We wished to consider methods that might forecast changes in the fertility distribution over age, at least in the short-term.

The next step up in generality is to use univariate time series models to forecast $\log(TFR_t)$ and each $\log(R_{it})$ series separately. However, this approach runs into the consistency problem since the changes forecast by the univariate models for the $\log(R_{it})$'s may result in long-term forecasted fertility distributions that lack intuitive appeal. Figure 6 shows the results of fitting and forecasting each $\log(R_{it})$ with a model with regression indicator variables for the years 1942-47 (to handle the effects of World War II on the data) and ARIMA(1,1,0) errors. Since the indicator variables are all zero after 1947 they affect only the model fitting, not directly the forecasting. The (1,1,0) model forecasts head towards an ultimate level and the forecast error variance grows unboundedly with the forecast lead. (See Table 1.) Figure 6 shows the forecasts and 67% intervals for the year 2020 using data through 1984. The forecasts at all ages in fact stabilized considerably before 2020, so the forecasts in Figure 6 can be viewed as ultimate values. (This will also be the ultimate forecast fertility distribution looked at on a cohort basis if we use TFR forecasts that stabilize.) Notice that this ultimate fertility distribution does not exhibit the same smooth shape across age observed in historical data (period basis). These forecasts are not necessarily "bad". The 67% intervals shown reflect considerably uncertainty in the future data (here for 2020), and it would be easy to draw a smooth curve within the band described by the interval limits over age. Also, the uncertainty in forecasting TFR_{2020} (see Figure 3a) seems more important. Still, one might prefer long-term forecasts that do capture the smooth shape of historical data.

• If different univariate models were used for different $\log(R_{it})$'s, instead of the common (1,1,0) model, then we would be using different types of forecast functions at each age. (See Table 1). The inconsistency of the resulting long-term forecasts would then likely be worse. Use of a different common model than the (1,1,0) might yield more consistent results, but we are merely trying to illustrate the problem here. Also, the (1,1,0) model is not an arbitrary choice - examination of autocorrelation and partial autocorrelation functions (as discussed in Box and Jenkins 1970) suggests it is a reasonable model choice for $\log(R_{it})$ for most if not all ages.

There is another possible drawback to the use of separate univariate models. While the forecast intervals from the univariate models such as those shown in Figure 6 do reflect uncertainty in the individual age specific rates, if one wishes to eventually produce forecast intervals for total births, then separate univariate models will not work - one needs a multivariate model that accounts for the strong relationships between the series. However, if only point forecasts are desired, and one is either not concerned with the consistency problem or is willing to forecast the R_{it} to remain constant at current values (random walk model), then the use of separate univariate forecasts for the $\log(R_{it})$ may not be a bad approach.

4.2 Multivariate Models for All Ages Jointly

An appropriate multivariate model for all the series, $\log(TFR_t)$, $\log(R_{14t}),\ldots,\log(R_{45t})$ say, would not have a consistency problem, since it would capture the relationships between the series and this would be reflected in the forecasts. Also, it could produce forecast intervals for functions of the series such as total births. (At least this could be done short-term, taking projections of the number of women at each childbearing age as given.) Unfortunately, the dimensionality problem posed in modeling this many time series is severe. For example, an unrestricted multivariate (1,1,0) model would have 33 parameters in each equation for a total of 33x33 = 1089 elements in its Φ_1 parameter matrix. This is far too many with only 64 years of data.

One approach to the dimensionality problem is to use a multivariate model of a very restricted form. The CARIMA model of deBeer (1985) can be viewed this way. Perhaps the simplest restricted multivariate model results from using univariate models for each series, but fitting them jointly rather than separately (minimizing the determinant rather than the trace of the residual covariance matrix). To illustrate, Table 2 shows the results of fitting univariate (1,1,0) models to $\log(R_{it})$ for i = 20,21,22 both separately and jointly, with regression adjustments for the years 1942-47. The models fit separately imply a moderate positive correlation in the year-to-year changes, which has some intuitive appeal (if relative fertility at age i increased last year, it is more likely to increase than decrease this year). The dramatic changes when the parameters are estimated jointly are disturbing and not very appealing. This occurs because the strong contemporaneous relationships between the series reflected in the residual correlation matrix cause the two fitting criteria to be very different. The joint fit is best in an aggregate sense, and so is not necessarily "bad", though it lacks intuitive appeal. Use of other restricted multivariate models are unlikely to alleviate this problem. Interestingly, deBeer (1985) apparently used a trace rather than a determinental fitting criterion with his CARIMA model.

Fitting a restricted multivariate model to the entire set of 30+ time series might also lead to numerical problems (along with the problem of finding computer software to do this.) The strong contemporaneous relationships could make the joint estimation problem ill-conditioned. Rather than pursue this sort of modeling further, we consider other approaches to addressing the dimensionality and consistency problems.

4.3 Curve Fitting to Reduce Dimensionality

We shall briefly describe an approach to forecasting fertility using fitted gamma curves that is discussed in detail in Bell et al. (1988). A variant of this approach was used in the latest set of national population projections done by the U.S. Census Bureau.

The shifted gamma probability densisty is defined by

$$\gamma_i = \frac{1}{\Gamma(\alpha)\beta^\alpha} (i-A_0)^{\alpha-1} \exp\{-(i-A_0)/\beta\} \qquad i \geq A_0 \qquad (4.1)$$

where A_0 is the starting point of the curve, α and β are the gamma distribution parameters, and $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} \exp(-u)du$ is the gamma function. This was fitted to the relative fertility rates R_{it} (i=14,...,45) in each year t by minimizing the weighted sum of squares, $\sum_i w_i(R_{it} - \gamma_{it})]^2$, where w_i is the weight for age i. Weights of 4 for ages 18 through 32 and 1 for all other ages were used to give more emphasis in the fit to ages with high fertility. Parameters A_{0t}, α_t, and β_t were determined for each year by nonlinear least squares, constraining A_{0t} to the interval $0 \leq A_0 \leq 14$. The resulting fits for two years are shown in Figure 7. The gamma curve provides a good overall approximation to the relative fertility rates, especially in recent years.

The basic idea is to forecast the gamma curve parameters to produce forecasted gamma curves to use in forecasting the relative fertility rates R_{it}. This reduces the dimensionality of the forecasting problem to 4 - the number of gamma curve parameters plus TFR. It also addresses the consistency problem directly since even long-term forecasts will follow the smooth shape of the gamma curve.

Before modeling, the parameters were transformed to the mean and standard deviation of the gamma curve:

$$MACB_t = A_{0t} + \alpha_t\beta_t \qquad SDACB_t = \beta_t \cdot (\alpha_t)^{1/2}$$

These are the gamma curve analogues of the mean and standard deviation of age at childbearing. They are more interpretable and have more stable traces over time than α_t and β_t. A_{0t} took the value 0 or 14 for most years, and so was not modeled but was projected to remain at its most recent value of zero.

A restricted multivariate ARIMA(3,1,0) model was developed for $\log(TFR_t)$, $\log(MACB_t)$, and $\log(SDACB_t)$ adjusted for war year effects. (See Bell et al. 1988 for details.) The resulting forecasts are shown in Figure 3a for TFR, and Figures 8 and 9 for MACB and SDACB. These forecasts can be used to produce forecasted gamma curves.

In its basic form this approach forecasts not the relative fertility rates themselves, but the gamma curves that will eventually be fit to them as the future data become available. For medium- to long-term forecasting this distinction is mostly unimportant since the error in forecasting TFR will swamp the curve fitting error at most ages. However, for short-term ($\leq$ 5 years ahead) forecasting it is also important to forecast the deviations of the fitted curves from the actual rates at each age (the age-specific biases). Examination of the historical time series of biases at each age suggested a random walk model for each bias series was not unreasonable. Thus, the biases are forecast to remain constant at their values in the forecast origin year. Figure 10 shows the results for forecasting 1982 and 1984 fertility rates from 1980. Notice that adding the bias forecasts to the forecasted gamma curves produces a big improvement in the forecasts for 1982, and a lesser improvement for 1984.

The gamma curve approach effectively addresses the dimensionality problem since most of the attention in modeling can be focused on TFR, MACB, and SDACB; the biases can be forecast by simple means. It also addresses the consistency problem, though in the Census Bureau projections adjustments were made in the forecasts for women aged 40 and over (where fertility was forecasted by the model to rise rapidly in percentage terms, though the resulting forecasted rates were still very low.) The behavior of A_{0t}, which is not suited to modeling, is a slight drawback. Another drawback is that because the biases are not modeled jointly with TFR, MACB, and SCACB (which would put us right back at the dimensionality problem), we cannot produce forecast intervals for individual age-specific rates or functions of these with this approach.

The general approach can be used with curves other than the gamma. Rogers (1986) suggests fitting "double exponential" curves; we found these to yield comparable fits to the gamma. The choice between well-fitting curves that depend on only a few parameters is not critical, since no curve is likely to be immune from the bias problem.

4.4 Linear Regression Approximations and the Principal Components Approach

We can obtain useful, dimension reducing approximations to the age-specific fertility rates from the gamma and other curves that are nonlinear in their parameters. It is also useful to consider approximating fertility rates with functions of age that are linear in their parameters, that is, by linear regression on some variables that are functions of age. The goal would be to find a reasonably small set of variables that, when regressed upon, would provide a good approximation to the fertility rates in each year of the data set. We could then model and forecast the time series of regression parameters, and use the forecasted regression approximations in forecasting the fertility rates, analogous to what was done with the gamma curve.

More specifically, let $\underset{\sim}{R}_t = (\log(R_{14,t}),\ldots,\log(R_{46,t}))'$ and let Λ be a $33 \times J$ full column rank matrix of J regression variables that we shall use to approximate each $\underset{\sim}{R}_t$. The approximation is $\Lambda \hat{\underset{\sim}{\beta}}_t$, where $\hat{\underset{\sim}{\beta}}_t$ is chosen for each t by least squares regression, i.e. $\hat{\underset{\sim}{\beta}}_t = (\Lambda'\Lambda)^{-1}\Lambda'\underset{\sim}{R}_t$. We could also use weighted least squares to give more emphasis to certain ages in the fit. Having determined $\hat{\underset{\sim}{\beta}}_t$ for each year t in the data, we model and forecast $\hat{\underset{\sim}{\beta}}_t$ to get $\hat{\underset{\sim}{\beta}}_n(\ell)$ say, and use $\Lambda \hat{\underset{\sim}{\beta}}_n(\ell)$ as a forecast of $\underset{\sim}{R}_{n+\ell}$, assuming the approximation error to be negligible for forecasting purposes.

Two obvious candidates frequently used to approximate smooth functions are Fourier series (where the variables would be sine and cosine functions of age at various frequencies) and polynomials (where the variables would be age, age^2, age^3,...). While we shall not present the results here, we found a Fourier series required a large number of terms for a good approximation. Polynomials worked better; a polynomial of degree 6 providing perhaps a better approximation than the gamma curve, though with 2 additional parameters. We can also ask in general what set of J variables, or what Λ, can provide the best approximation in terms of lowest total sum of squares of the approximation errors over all the years of the data set. The answer turns out to be that the columns of Λ should be the first J principal component vectors (i.e. the J eigenvectors corresponding to the J largest eigenvalues) of the sum of squares and cross products matrix of the data $\underset{\sim}{R}_1,\ldots,\underset{\sim}{R}_n$. We can also define weighted principal components corresponding to a weighted least squares criterion. Details are given in Bozik and Bell (1987). Figure 11 shows approximations to the 1980 relative fertility rates using 4 or 8 principal components. Four components gives a much better approximation than the gamma curve, and still better fits result as more components are added.

There are two ways to view the use of linear regression approximations. The first viewpoint is that taken above, that the goal is to use the regression approximation to reduce the dimensionality of the forecasting problem. This is the view taken in Bozik and Bell (1987) in developing the principal components approach, where consideration is given to how many principal components are needed to provide approximations with errors that are negligible for forecast purposes. (This question was not resolved.) The second viewpoint is as follows. Notice that the approximation, $\Lambda \hat{\beta}_t = \Lambda(\Lambda'\Lambda)^{-1}\Lambda' R_t$, is a linear transformation of the data. If $J < 33$ this transformation is of reduced rank and some information is lost, though principal components minimizes this for any J. If $J = 33$ the transformation is nonsingular and no information is lost ($\Lambda \hat{\beta}_t = R_t$), though dimensionality is not reduced either. Why consider using $J = 33$, then? Because $\hat{\beta}_t$ may have a simpler structure for time series modeling than R_t. (Recall the problems noted in section 4.2 in modeling R_t directly.) This second view is taken in Bozik and Bell (1988), where a model for the full set of 33 principal component series and log(TFR) is developed. While we shall not go into details here, the principal component series turn out to be much simpler to model than the original data: only log(TFR) and the first 4 principal components were modeled multivariately, the remaining principal components followed univariate models with all those after the eighth following independent univariate models that could be fit separately. The resulting model can be used to develop forecasts and intervals for any linear functions of the fertility rates. Figures 12 and 13 show forecasts and 67% forecast intervals for F_{it} from 1984 through the year 2000 for a few select ages, and for all ages for a few select years.

The principal components approach is appealing because for any given reduction in dimension (any J) it provides the linear transformation of the data that can be used to best approximate the original data. Any other linear transformation of the data of rank J must lose some information relative to principal components. The use of birth rates for 5-year age intervals is a weighted average of the single-year-of-age rates if the age-specific female population figures are taken as fixed. Hence, use of birth rates for 5-year age intervals must lose information relative to use of the same number of principal components.

5. Conclusions

In this paper we have tried to provide some guidance on how time series methods can be applied in forecasting age-specific fertility. The way this is done will depend heavily on how much time series data is available. Extensive modeling, such as multivariate modeling of gamma curve parameters or of principal component series, requires substantial time series data. With less data more simple models are dictated. With very limited data the modeling may require a substantial amount of judgmental input.

We shall summarize the main points of the paper. (1) Time series models work better the longer are the series they have to work with, though some consideration should be given to the comparability of the data over time, i.e., does it make sense to assume a single model applies over the entire series? Use of regression terms to account for unusual events and other variables known to affect the series in question can help in this regard. (2) Time series models are most conveniently applied on a period basis. (3) Transformation of fertility rates to TFR and the relative fertility rates, and then to logarithms of these is generally recommended. (4) Forecast intervals from time series models can be used in developing alternative demographic projections, at least short-term. (5) Forecasting fertility rates separately for each age (using univariate time series models or other means) can run into consistency problems in long-term forecasts, while direct multivariate modeling of rates at all ages is difficult due to the high dimension of the system. (6) The dimensionality and consistency problems can be addressed by two techniques presented: approximating the age-specific fertility rates with fitted curves (e.g. gamma curves), or use of principal components approximations to the data.

REFERENCES

Abraham, B. and Ledolter, J. (1983), Statistical Methods for Forecasting, New York: Wiley.

Ansley, C. F. and Kohn, R. (1986), "Prediction Mean Squared Error for State Space Models With Estimated Paramters," Biometrika, 73, 467-473.

Ansley, C. F., Spivey, W. A., and Wrobleski, W. J. (1977), "A Class of Transformations for Box-Jenkins Seasonal Models," Applied Statistics, 26, 173-178.

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

_____(1984), "An Introduction to Forecasting With Time Series Models," Insurance: Mathematics and Economics, 3, 241-255.

Bell, W. R., Bozik, J. E., McKenzie, S. K., and Shulman, H. B. (1986), "Time Series Analysis of Household Headship Proportions: 1959-1985," Research Report 86/01, Statistical Research Division, Bureau of the Census, Washington, D.C.

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

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 No. 88/16, Statistical Research Division, U.S. Bureau of the Census, Washington, D.C.

Box, G.E.P. and Jenkins, G. M. (1970), Time Series Analysis: Forecasting and Control, San Francisco: Holden Day.

Box, G.E.P. and Tiao, G.C. (1975), "Intervention Analysis with Applications to Economic and Environmental Problems," Journal of the American Statistical Association, 70, 70-79.

Bozik, J. E. and Bell, W. R. (1987), "Forecasting Age Specific Fertility Using Principal Components," SRD Research Report 87/19, Statistical Research Division, Bureau of the Census, Washington, D.C.

_____(1988), "Time Series Modeling for the Principal Components Approach to Forecasting Age Specific Fertility," SRD Research Report, Statistical Research Division, Bureau of the Census, Washington, D.C.

Brockwell, P. J. and Davis, R. A. (1987), Time Series: Theory and Methods, New York: Springer-Verlag.

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.

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

Jones, R. H. (1985), "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations," Technometrics, 22, 389-395.

Land, K. C. and Cantor, D. (1983), "ARIMA Models of Seasonal Variation in U.S. Birth and Death Rates, Demography, 20, 541-568.

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.

_____ (1975), "Natural Fertility, Population Cycles, and the Spectral Analysis of Births and Marriages," Journal of the American Statistical Association, 70, 295-304.

Liu, L. M., Hudak, G. B., Box, G.E.P., Muller, M. E., and Tiao, G. C. (1986), The SCA Statistical System: Reference Manual for Forecasting and Time Series Analysis, Scientific Computing Associates, DeKalb, IL.

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

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

Miller, R. B. (1986), "A Bivariate Model for Total Fertility Rate and Mean Age of Childbearing," Insurance: Mathematics and Economics, 5, 133-140.

Miller, R. B. and Hickman, J. C. (1981), "Time Series Modeling of Births and Birth Rates," Working Paper 8-81-21, Graduate School of Business, University of Wisconsin-Madison.

Miller, R. B. and McKenzie, S. K. (1984), "Time Series Modeling of Monthly General Fertility Rates," Research Report 84/16, Statistical Research Division, Bureau of the Census, Washington, D.C.

Passel, J. S., Rives, N. W., and Robinson, J. G. (1977), "A Regression Method for Estimating the Completeness of Registration of White Births for States," paper presented at the annual meeting of the Population Association of America, St. Louis, MO.

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.

Thisted, R. A. and Wecker, W. E. (1981), "Predicting a Multitude of Time Series," Journal of the American Statistical Association, 76, 516-523.

Thompson, P. A. (1987), "A Transformation Useful for Bounding a Forecast," Working Paper 87-84, 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 (1988), "Projections of the Population of the United States, by Age, Sex, and Race: 1988-2080," Current Population Reports, Series P-25, No. 1018, Washington: U.S. Government Printing Office.

Table 1. Behavior of the Forecast Function, $\hat{Y}_n(\ell)$, and Forecast Error Variance Function, $V(\ell)$, for Some Simple ARIMA Models

Model	$\hat{Y}_n(\ell)$	$V(\ell)$	As $\ell \to \infty$: $\hat{Y}_n(\ell) \to$	As $\ell \to \infty$: $V(\ell) \to$
$(1-\phi B)(Y_t-\mu) = a_t$ "AR(1)"	$\phi^\ell(Y_n-\mu)$	$\sigma^2[1+\phi^2+\ldots+\phi^{2(\ell-1)}]$	μ	$Var(Y_t) = \dfrac{\sigma^2}{1-\phi^2}$
$(1-B)Y_t = a_t$ "Random Walk"	Y_n	$\sigma^2\ell$	Y_n	$+\infty$
$(1-B)Y_t = \theta_0+a_t$ "Random Walk with Drift"	$Y_n+\theta_0\cdot\ell$	$\sigma^2\ell$	$+\infty$ $(\theta_0>0)$ $-\infty$ $(\theta_0<0)$	$+\infty$
$(1-B)Y_t = (1-\theta B)a_t$ "Exponential Smoothing"	$(1-\theta)[Y_n+\theta Y_{n-1}+\theta^2 Y_{n-2}+\cdots]$	$V(\ell) = \sigma^2[1+(\ell-1)(1-\theta)^2]$	constant for all ℓ	$+\infty$
$(1-\phi B)(1-B)Y_t = a_t$ "(1,1,0)"	$b_0-b_1\phi^\ell$	$\sigma^2[1+(1+\phi)^2+\cdots+(1+\phi+\cdots+\phi^{\ell-1})^2]$	b_0	$+\infty$

<u>Note</u>: For the (1,1,0) model

$$b_0 = Y_n + \frac{\phi}{1-\phi}(Y_n - Y_{n-1}) \qquad b_1 = \frac{\phi}{1-\phi}(Y_n - Y_{n-1})$$

Table 2. Fitting Univariate ARIMA(1,1,0) Models Separately and Jointly to U.S. Data on $\log(R_{it})$ for Ages i = 20, 21, 22
(The Models Used Regression Terms to Adjust for the Effects of World War II on Fertility.)

	Separate (Univariate) Estimation		Joint (Multivariate) Estimation	
Age(i)	$\hat{\phi}_i$	Residual Variance	$\hat{\phi}_i$	Residual Variance
20	.29	1.88×10^{-4}	-.22	2.42×10^{-4}
21	.54	1.56×10^{-4}	.00	2.14×10^{-4}
22	.45	1.69×10^{-4}	.09	1.96×10^{-4}

<u>Note</u>: $\hat{\phi}_i$ is the estimated AR(1) parameter in the model for age i: $(1-\phi_i B)\nabla\log(R_{it}) = a_{it}$. The residual correlation matrix for the joint fit is

$$\begin{bmatrix} 1.00 & & \\ .91 & 1.00 & \\ .76 & .86 & 1.00 \end{bmatrix}$$

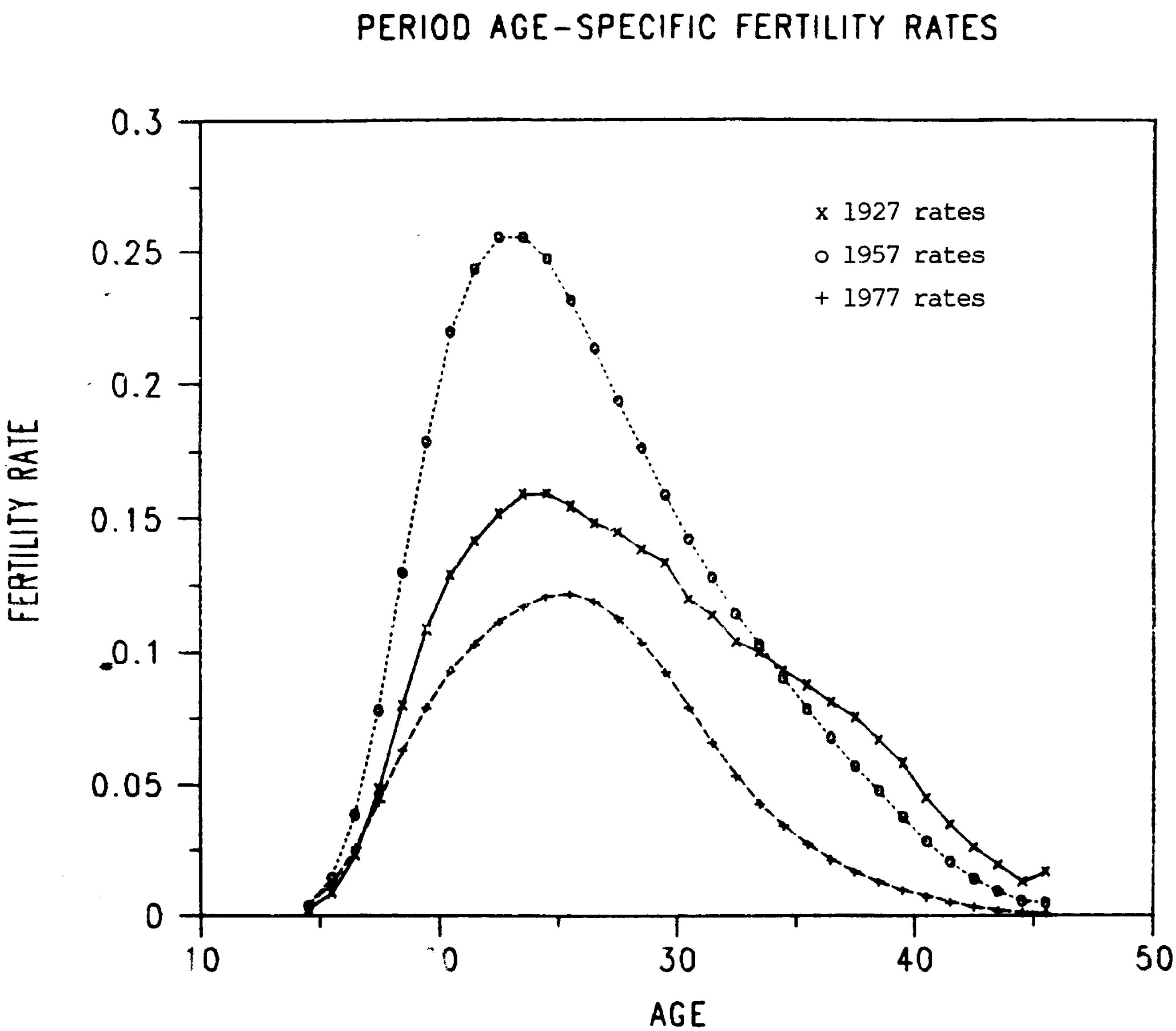


Figure 1a. Period age-specific fertility rates for three years, U.S. white women. Total fertility differs in these years, and the age-specific pattern shifts, but the rates for all three years have a similar smooth shape across age that is well-approximated by a scaled and shifted gamma density. The 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.

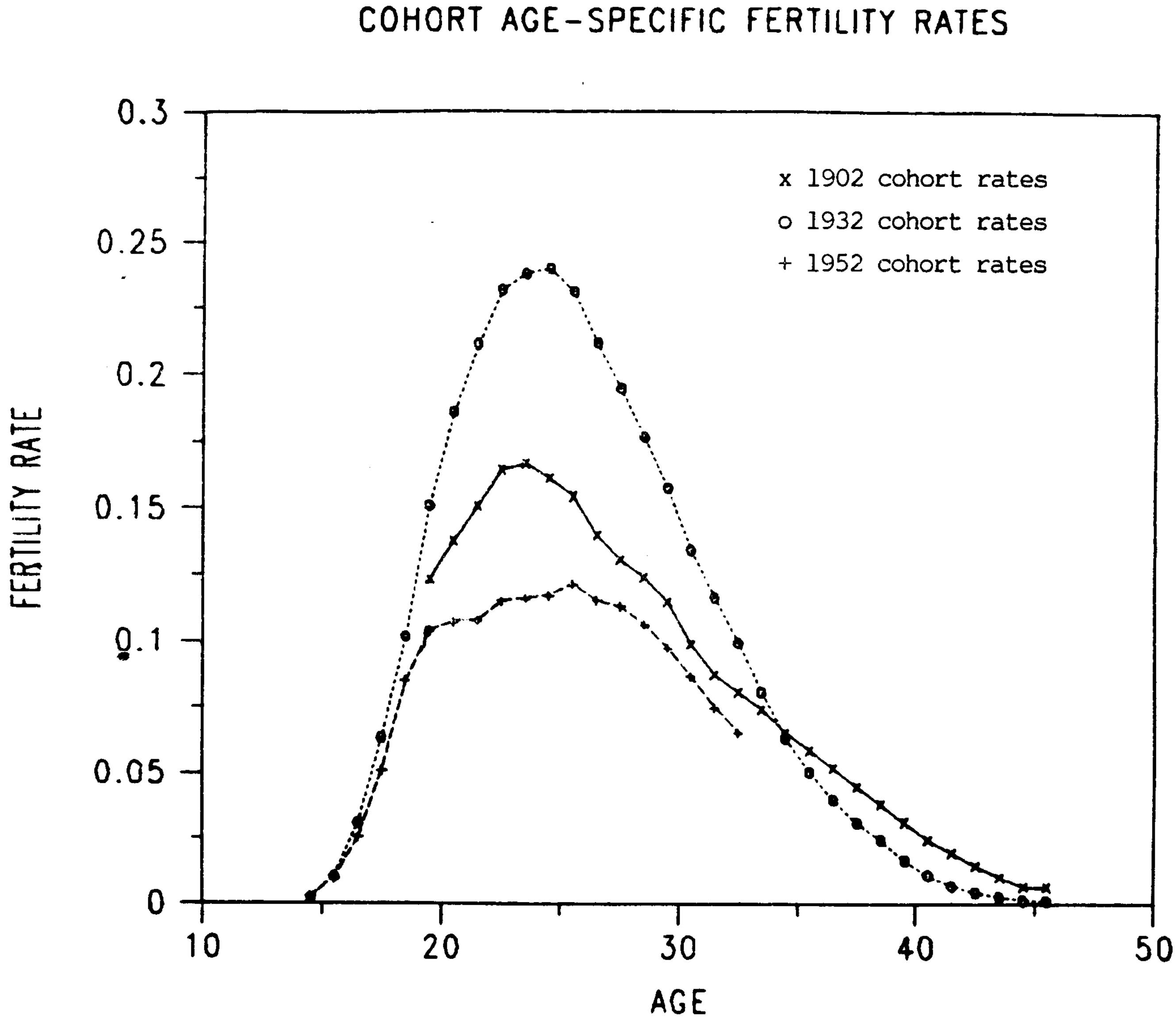


Figure 1b. Cohort age-specific fertility rates for three cohorts, U.S. white women. Since we are only using data for 1921-1984, the 1902 cohort is incomplete at ages 14-19, and the 1952 cohort is incomplete at ages 33-45. In contrast to period rates, cohort rates do not follow such similar smooth shapes 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.

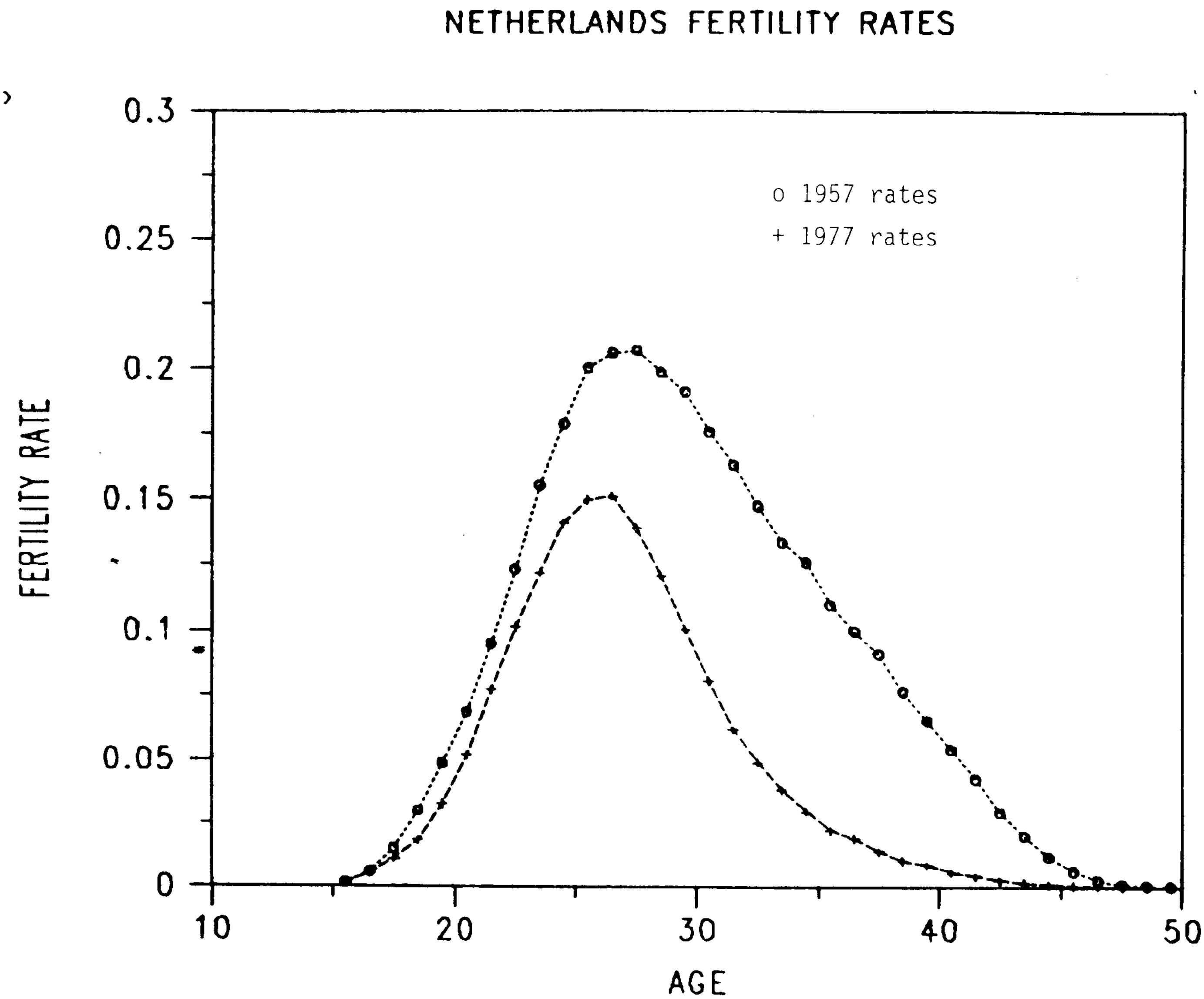


Figure 2a. Period age-specific fertility rates for two years, The Netherlands. The rates have a smooth shape across age similar to that for the U.S. data.

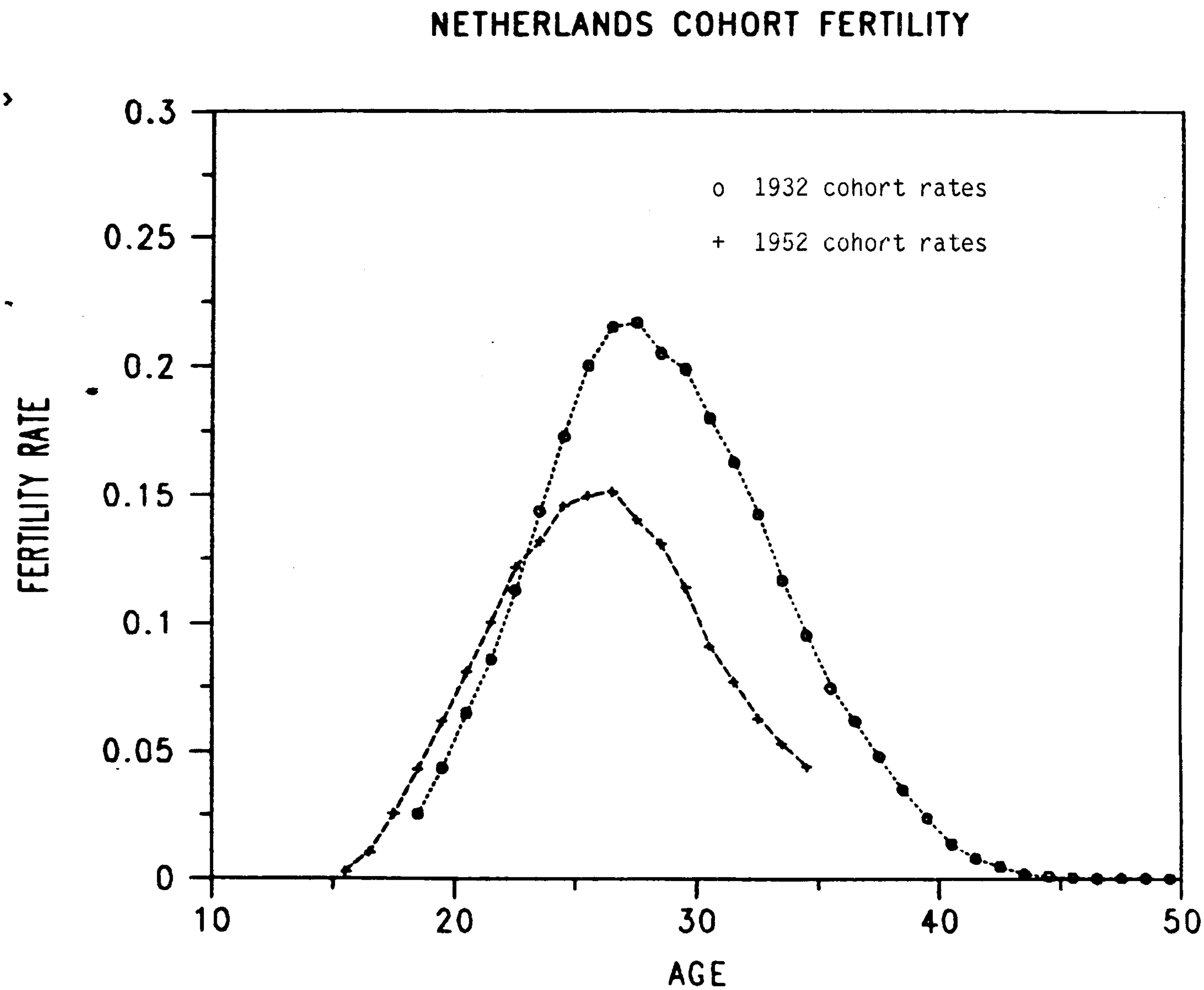


Figure 2b. Cohort age-specific fertility rates for two cohorts, The Netherlands. Since we are only using data for 1950-1986, the 1932 cohort is incomplete at ages 15-17, and the 1952 cohort is incomplete at ages 35-49. The cohort rates are not quite as smooth functions of age as the period rates, though this difference is not as pronounced as in the U.S. data.

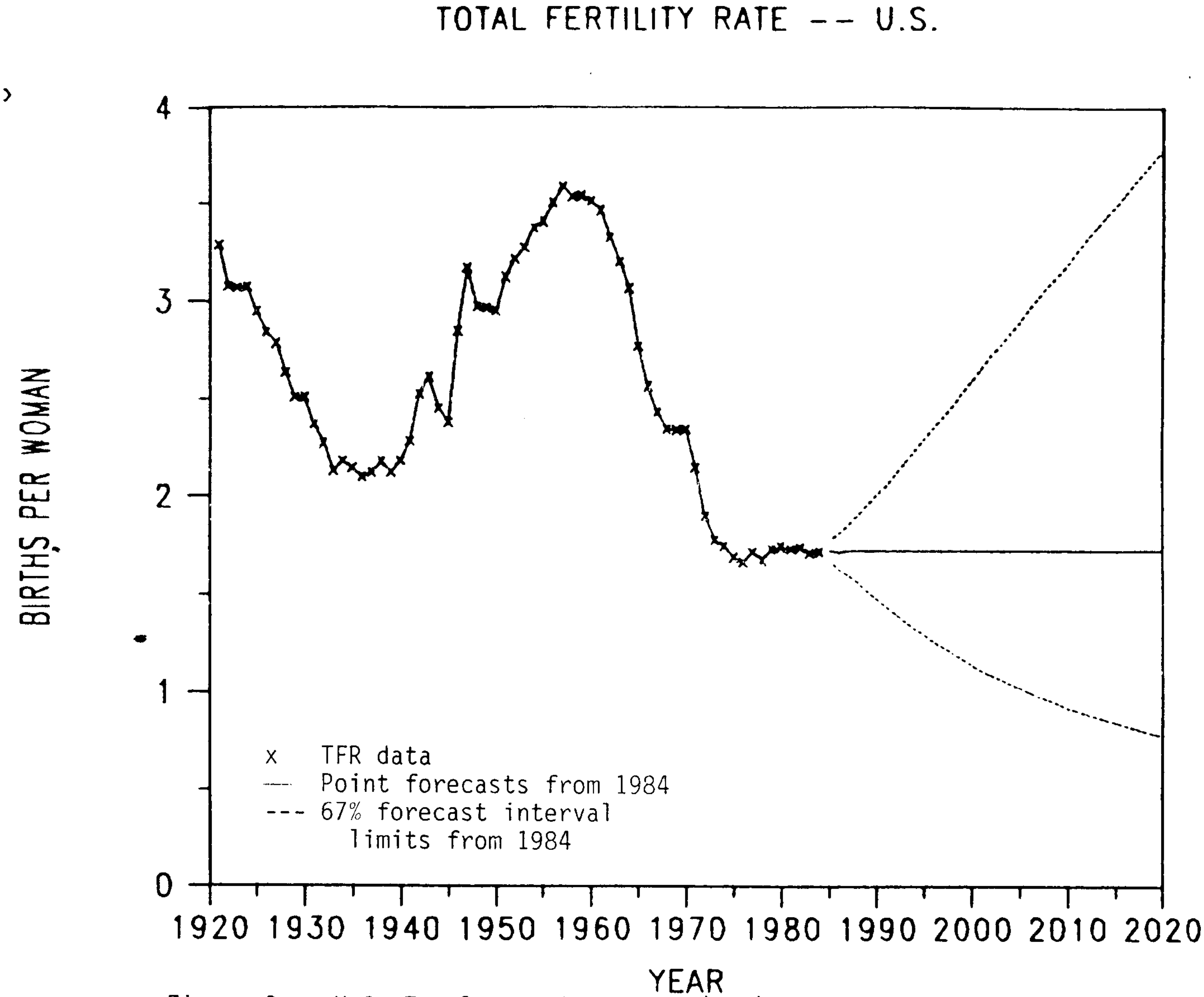


Figure 3a. U.S. Total Fertility Rate (TFR), observed (1921-84) and forecast (1985-2020). Forecasts and interval limits are obtained by exponentiating those in Figure 3b, which prevents point forecasts and interval limits from going below zero. Notice the asymmetry of the resulting forecast intervals. Notice also the unusual behavior in the years 1942-47 reflecting the effect of World War II on fertility.

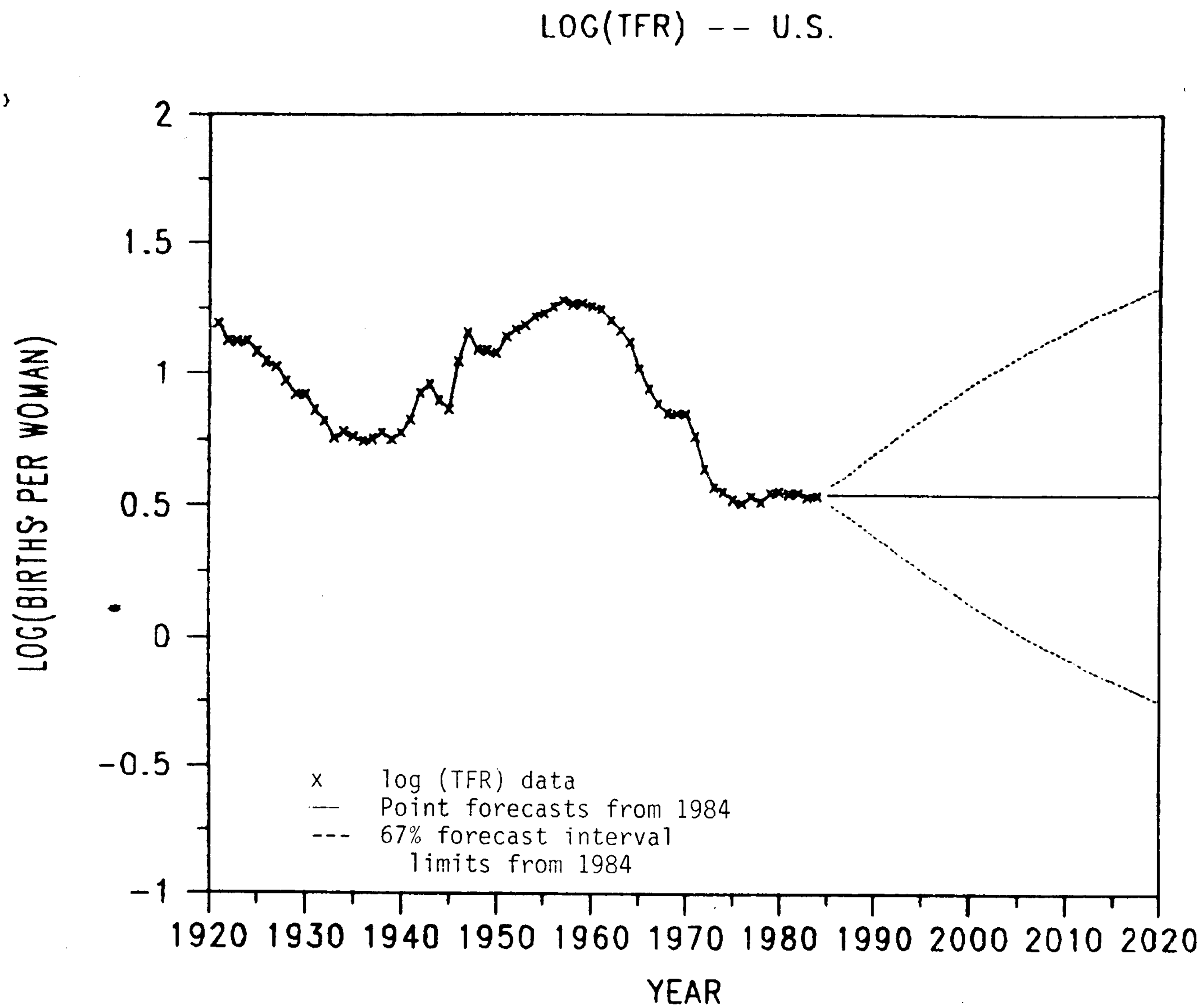


Figure 3b. U.S. log (TFR), observed (1921-84) and forecast (1985-2020). Forecasting was done with a multivariate ARIMA (3,1,0) model that also included log ($MACB_t$) and log($SDACB_t$) developed in Bell, et.al. (1988). The forecast intervals for log (TFR) are symmetric about the point forecasts.

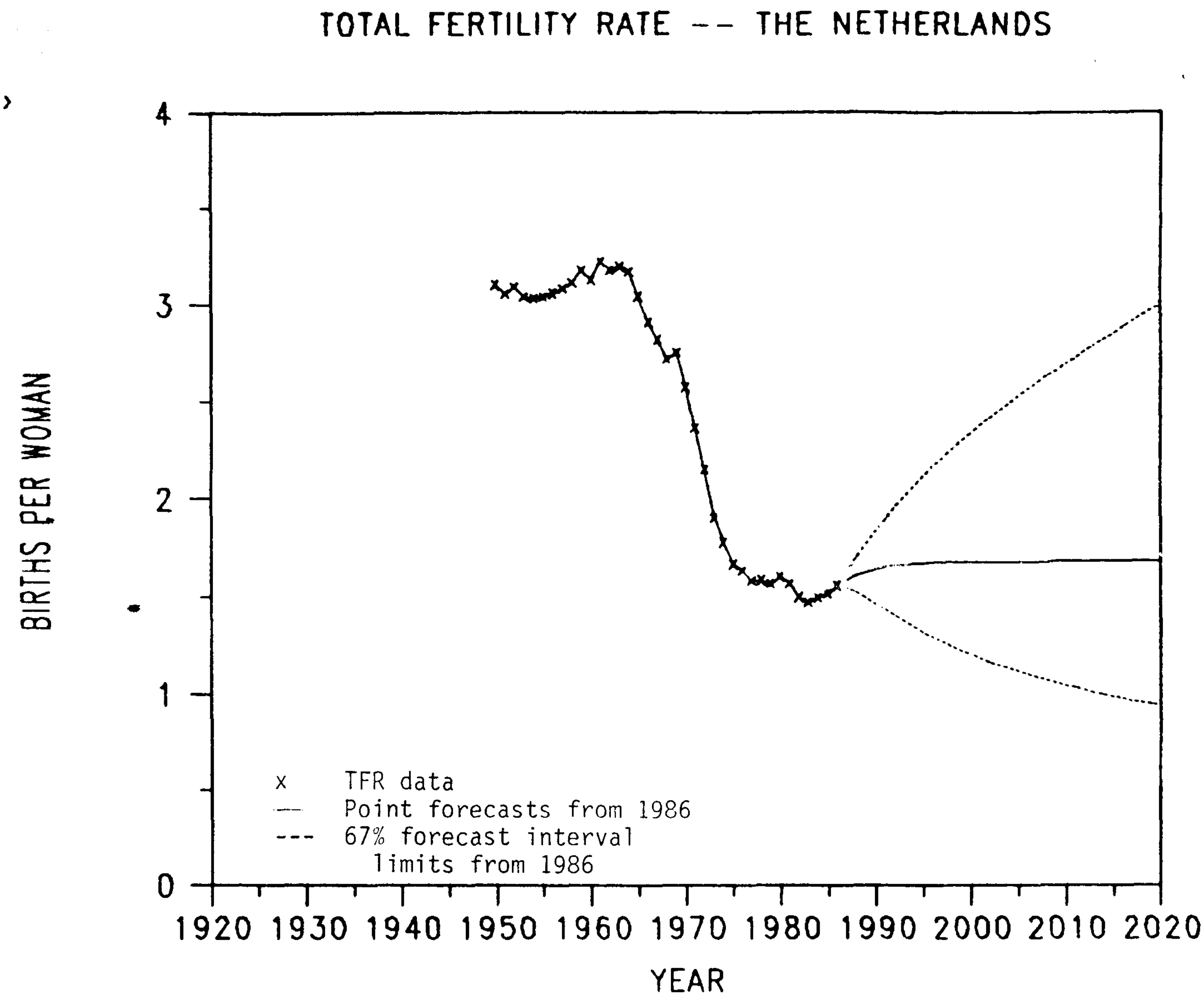


Figure 4a. Netherlands Total Fertility Rate (TFR), observed (1950-86) and forecast (1987-2020). Forecasts and interval limits are obtained by exponentiating those in Figure 4b., resulting in asymmetric forecast intervals.

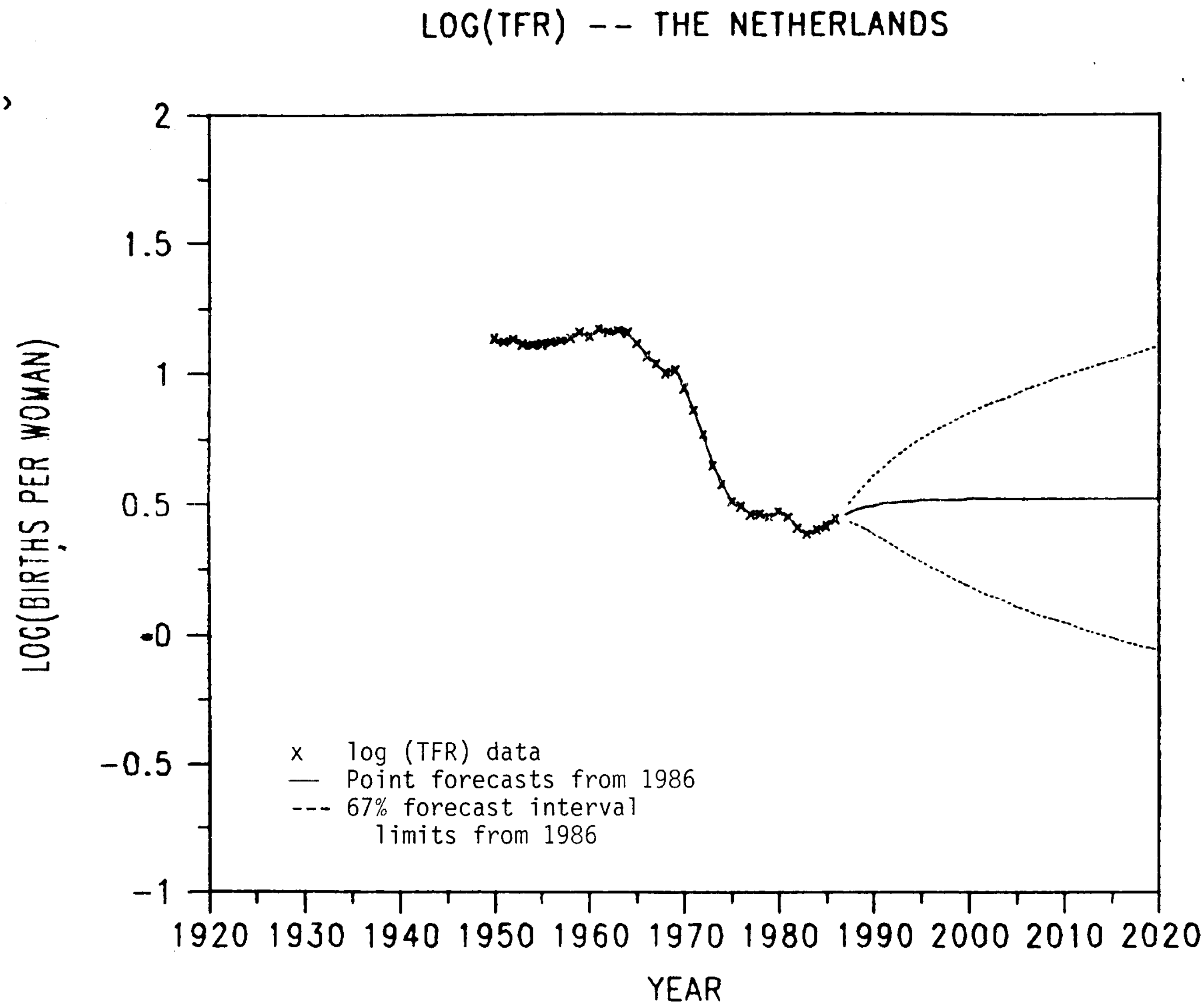


Figure 4b. Netherlands log (TFR), observed (1950-86) and forecast (1987-2020). Forecasting was done with a univariate ARIMA (1,1,0) model. The forecast intervals are symmetric about the point forecasts.

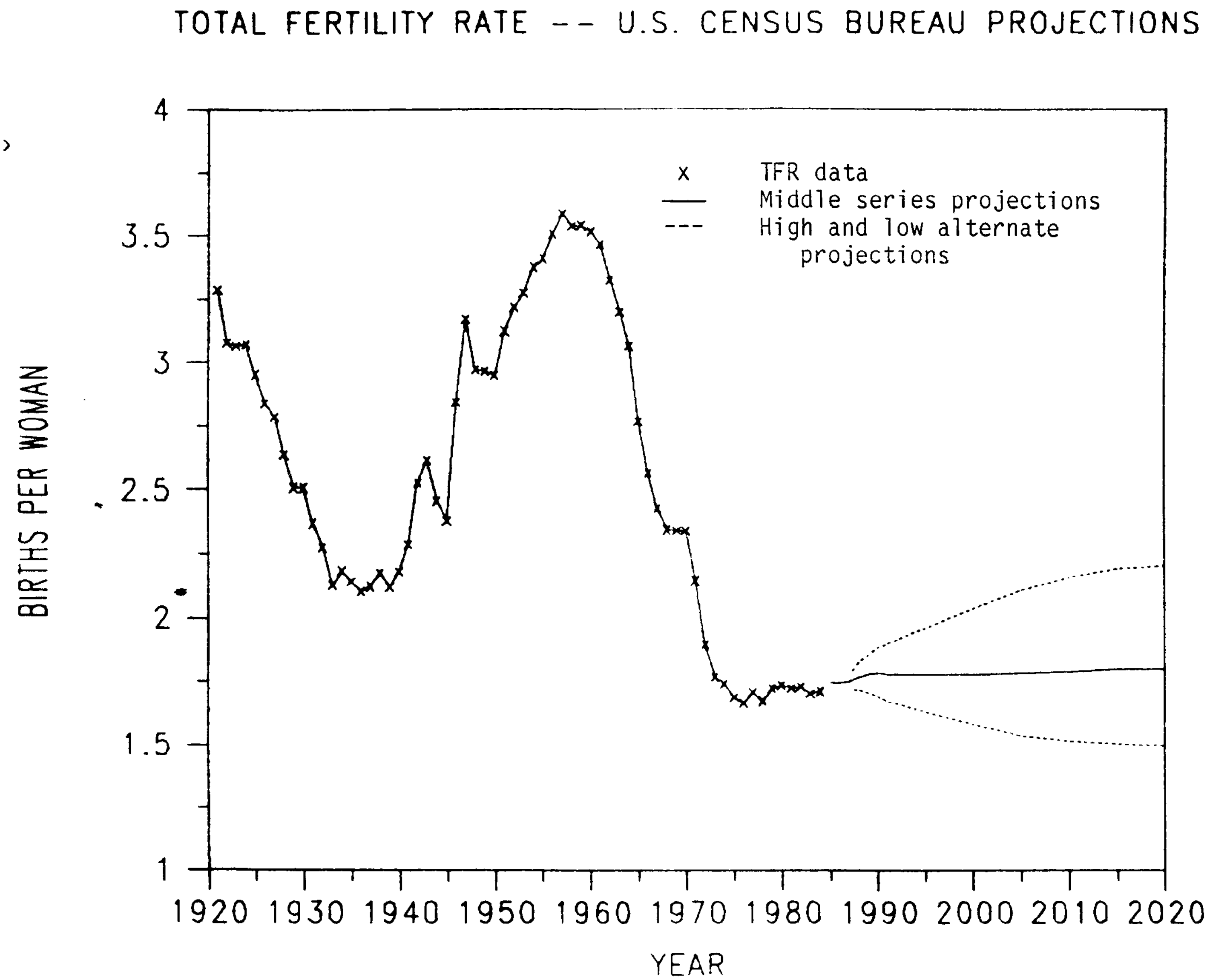


Figure 5. U.S. Census Bureau Total Fertility Rate projections (1985-2020). The projections used results from a time series model through 1990 interpolated to ultimate low, middle, and high values of 1.5, 1.8, and 2.2 which were determined judgmentally. Preliminary information on total births was used to modify the model forecasts for 1985 and 1986. These were then treated like actual data, so there are no high and low alternatives for these years. Notice how narrow are the intervals defined by the high and low alternatives in the long-term, compared to the forecast intervals in Figure 3a.

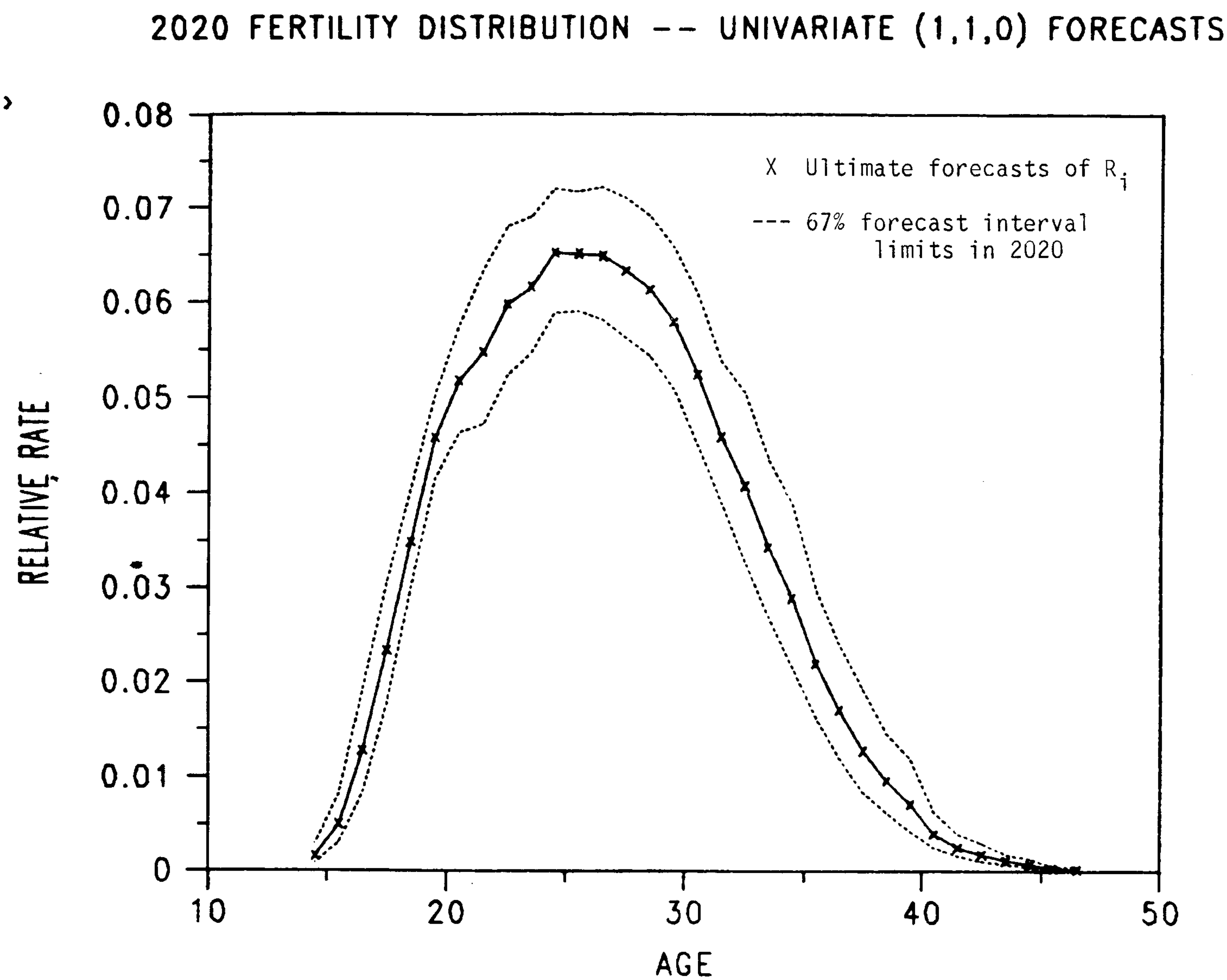


Figure 6. Ultimate U.S. fertility distribution forecast from univariate time series models. Forecasts were developed from univariate ARIMA(1,1,0) models for $\log(R_{it})$ fit separately for each age. These converged quickly to the ultimate values shown. The models were used to produce separate forecast intervals for each age in the year 2020.

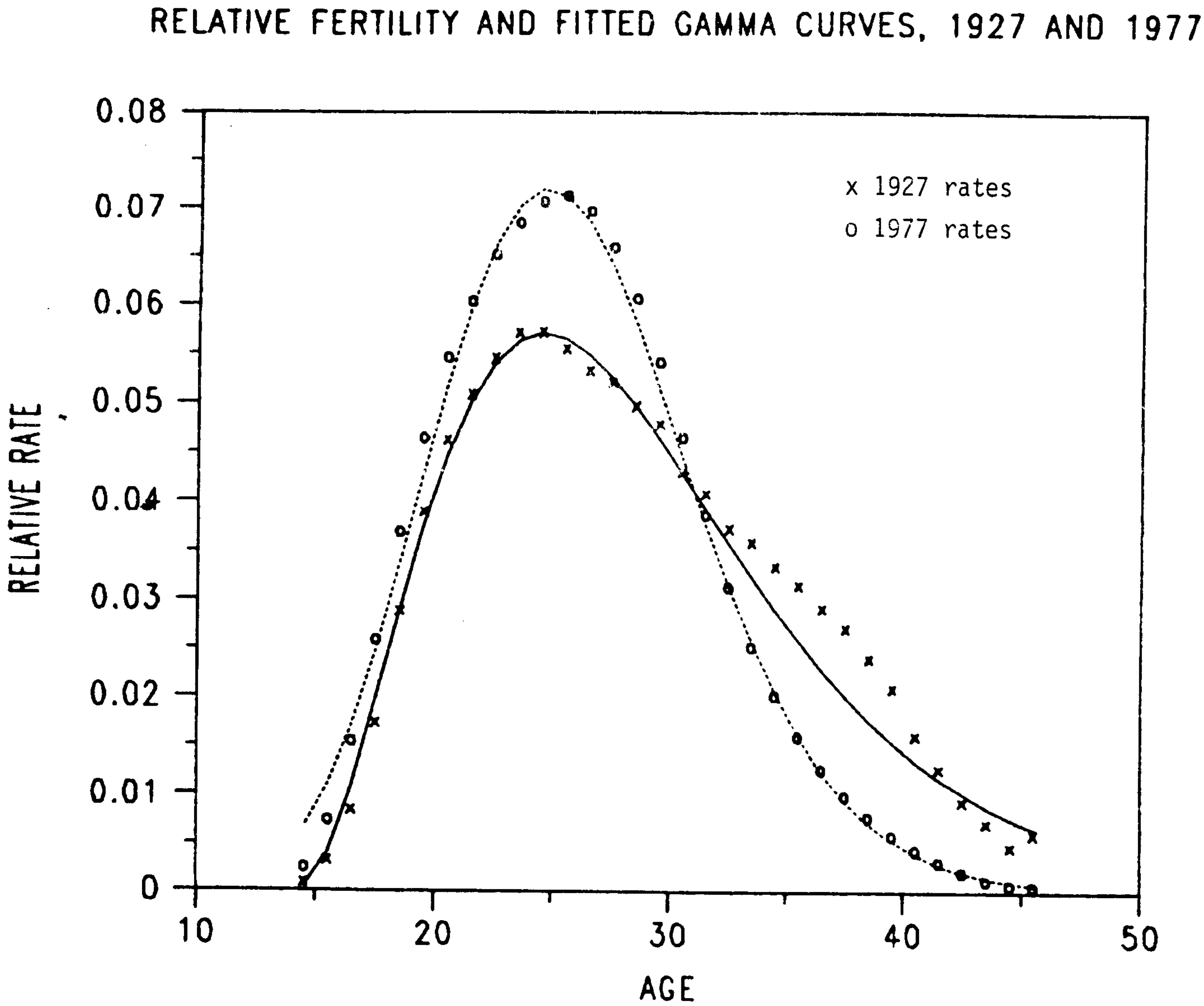


Figure 7. Two fitted relative fertility curves, U.S. data. We show the fitted curves from 1927 and 1977, which have, respectively, the smallest and largest alpha values. The three adjustable gamma curve parameters allow the curves to approximate a variety of age-specific fertility patterns. The observed fertilities in these years are also shown; in general, the fitted curves provide good overall approximations to the data.

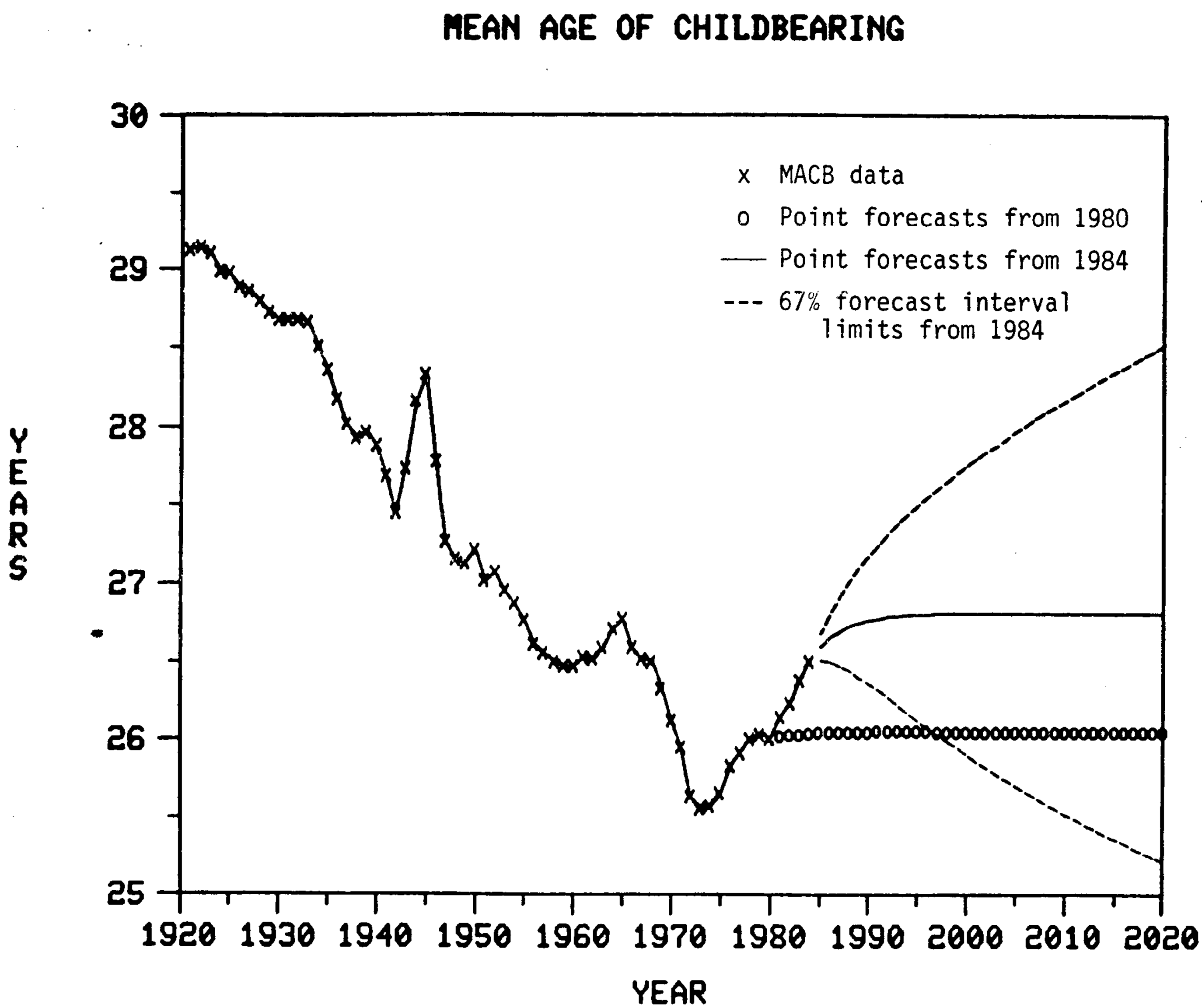


Figure 8. U.S. MACB data with point forecasts from 1980 and point and 67% interval forecasts from 1984. MACB is calculated from a gamma curve fitted to the relative fertility rates, and is analogous to the (empirical) mean age of childbearing. Forecasting was done with a multivariate ARIMA(3,1,0) model for log(TFR), log(MACB), and log(SDACB) developed in Bell et. al. (1988).

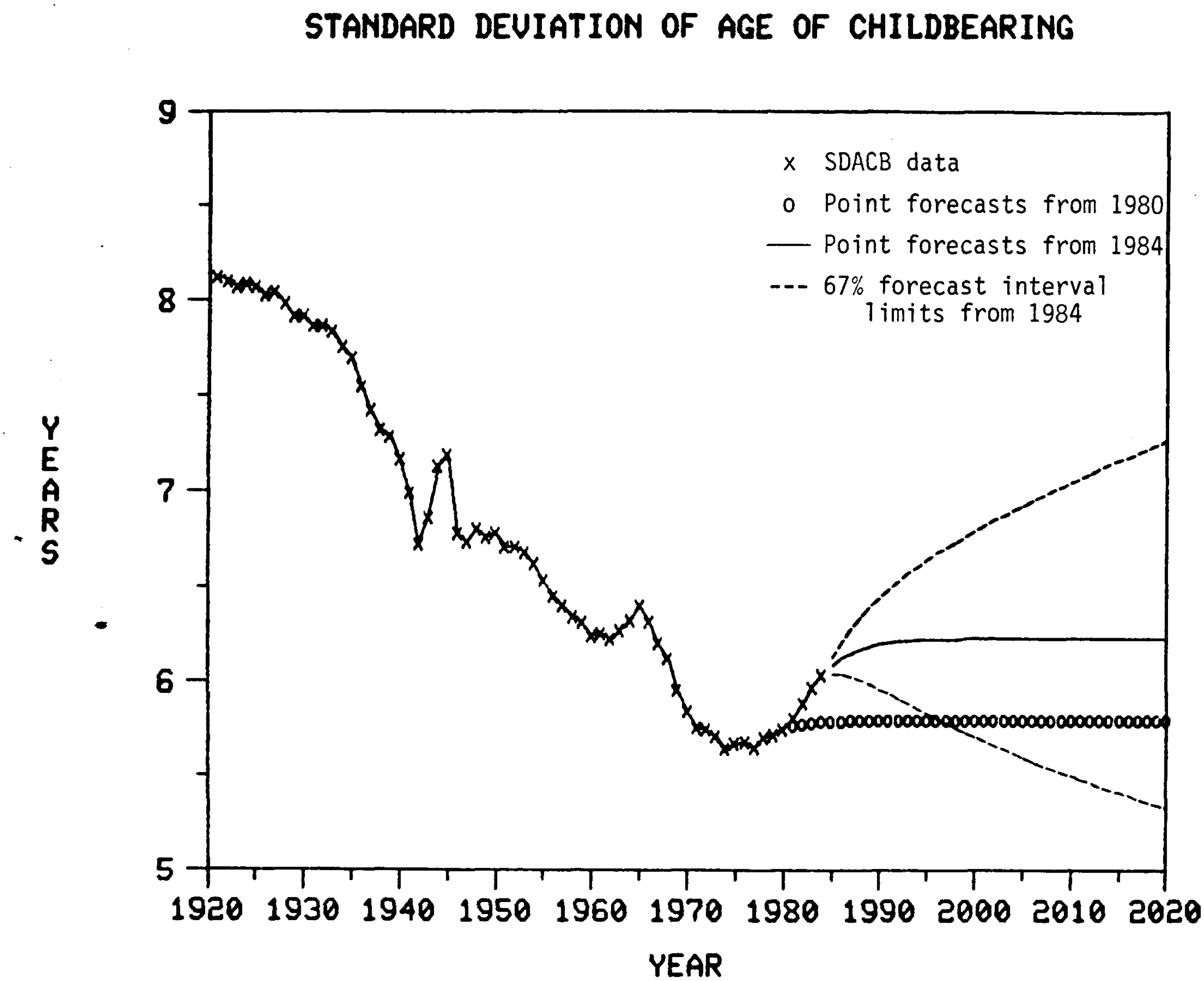


Figure 9. U.S. SDACB data with point forecasts from 1980 and point and 67% interval forecasts from 1984. SDACB is calculated from a gamma curve fitted to the relative fertility rates, and is analogous to the (empirical) standard deviation of age at childbearing. Forecasting was done with a multivariate ARIMA (3, 1, 0) model for log (TFR), log (MACB), and log (SDACB) developed in Bell et al. (1988).

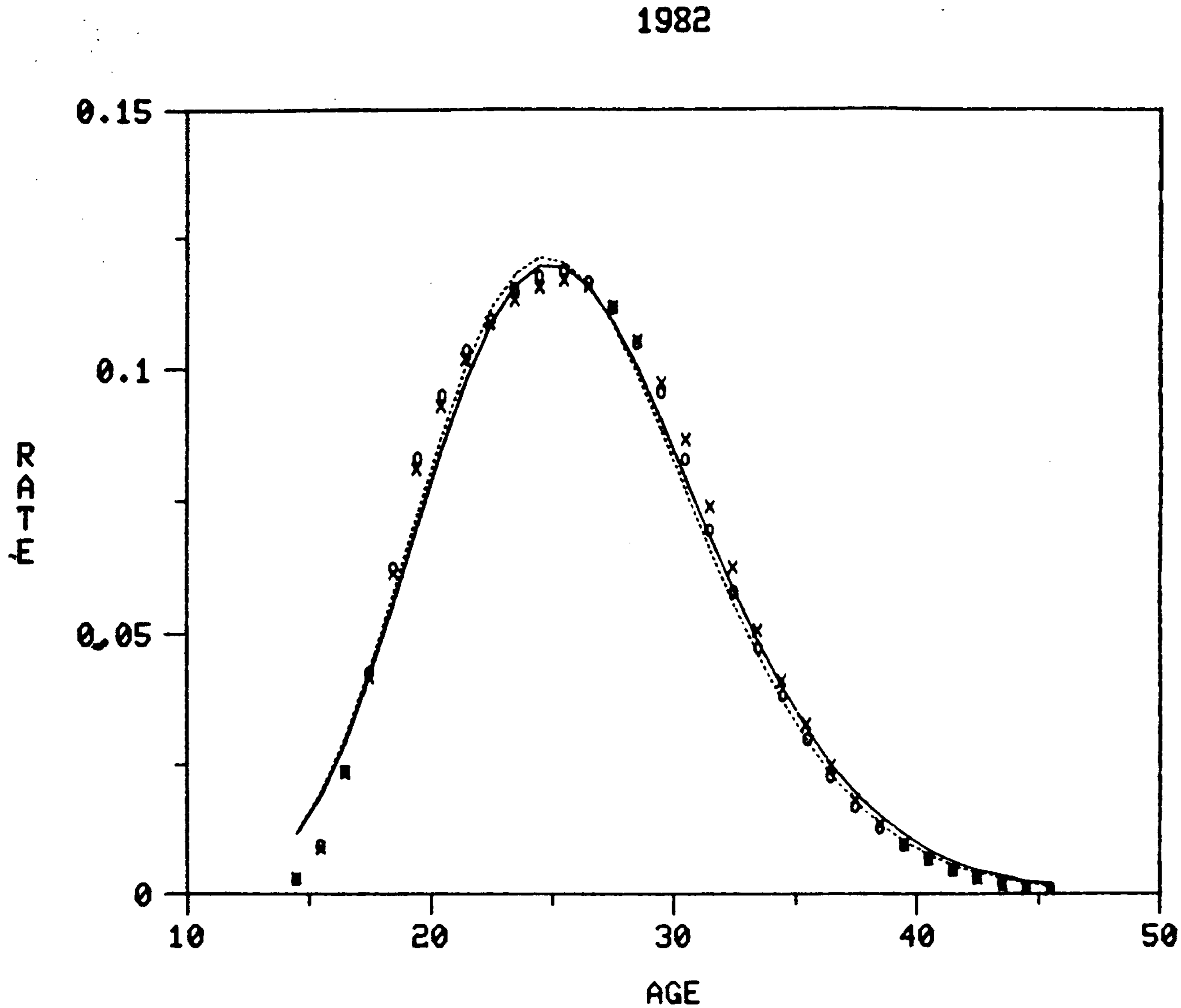


Figure 10a. Actual and forecasted fertility rates; fitted and forecasted gamma curves, U.S. data, (a) 1982 and (b) 1984. The gamma curve parameters are forecasted from 1980 using a multivariate ARIMA model estimated with data through 1980 (see Bell et al. 1988). The forecasted parameters produce forecasted gamma curves (--) which may be compared to the fitted gamma curves (—), obtained when the data for a given year become available. The forecasted curves are then adjusted with "bias forecasts" (see text) to produce forecasts of age-specific fertility rates (O) which may be compared to the actual fertility rates (x). The bias adjustment produces a large improvement in the 1982 forecasts, less of an improvement in 1984, showing the importance of forecasting the biases in the short-term.

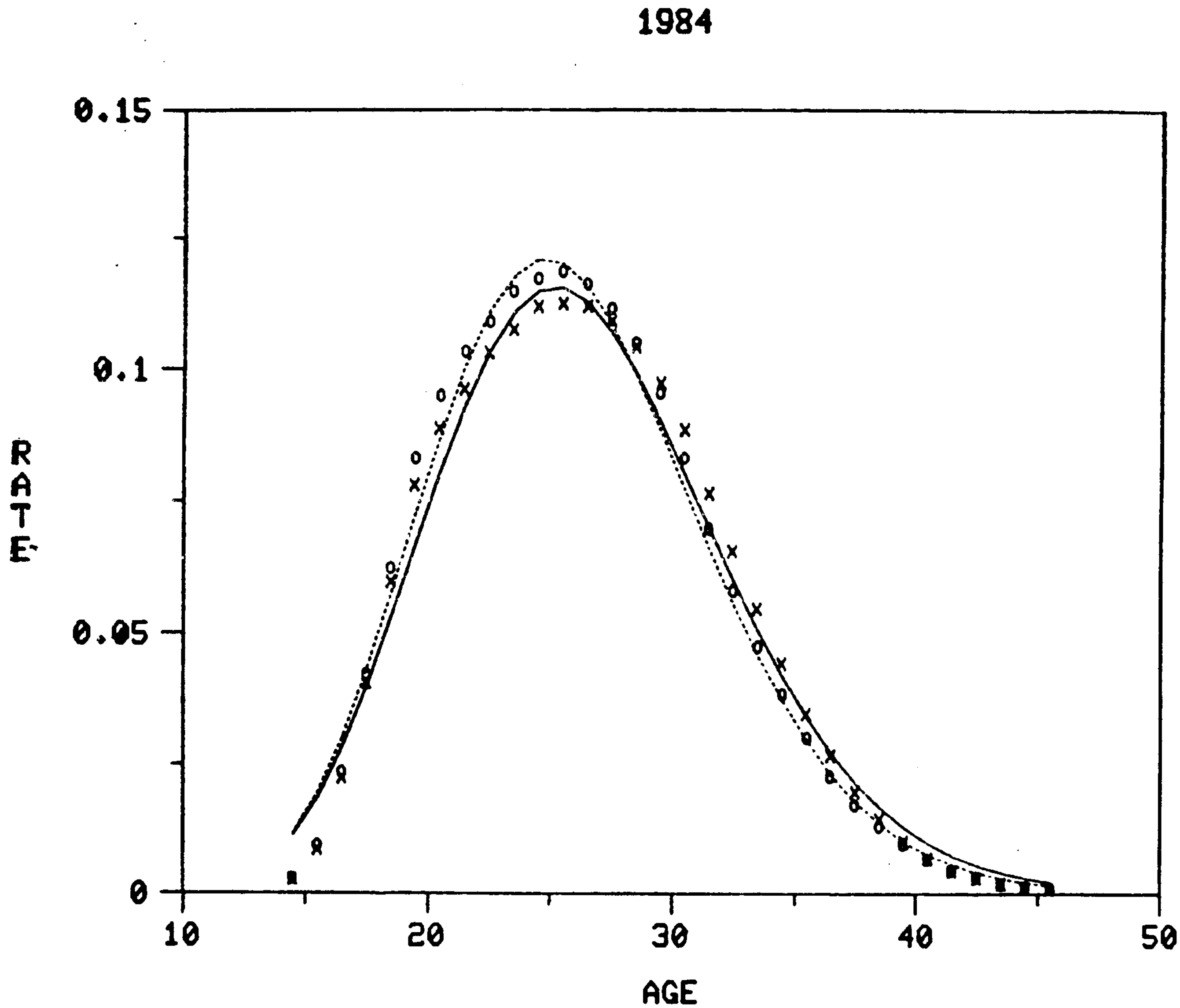
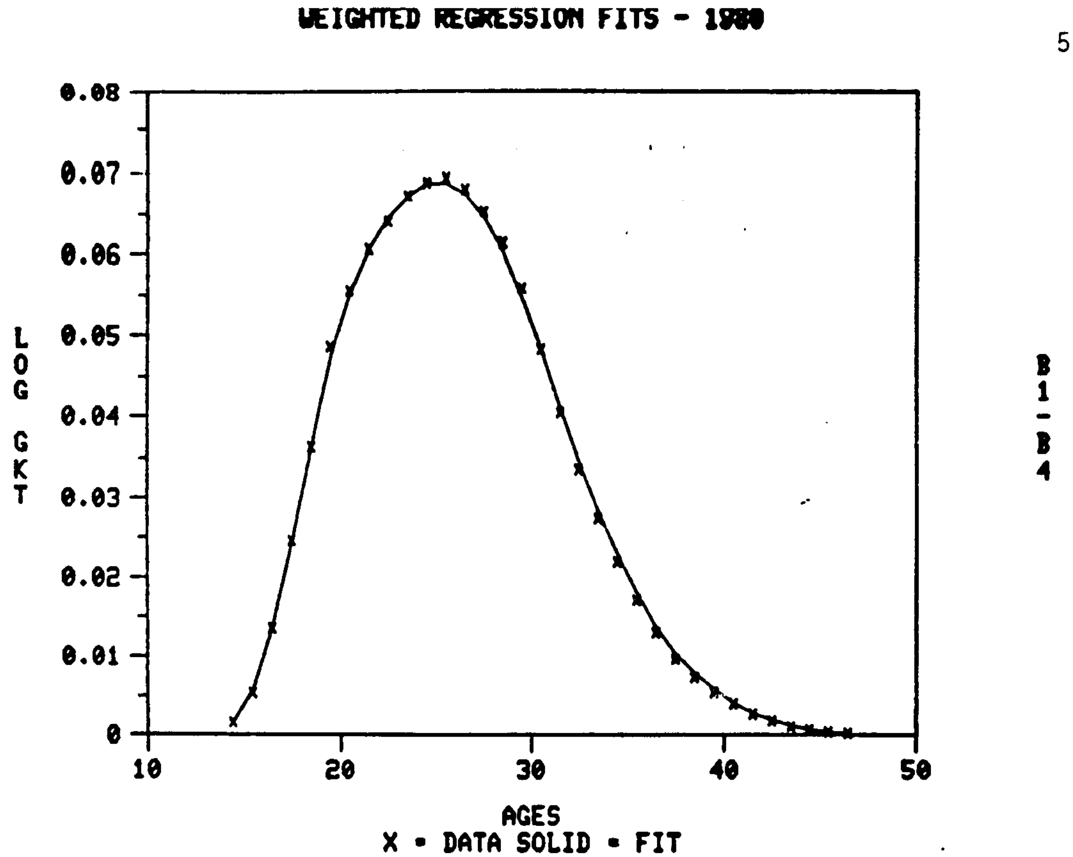
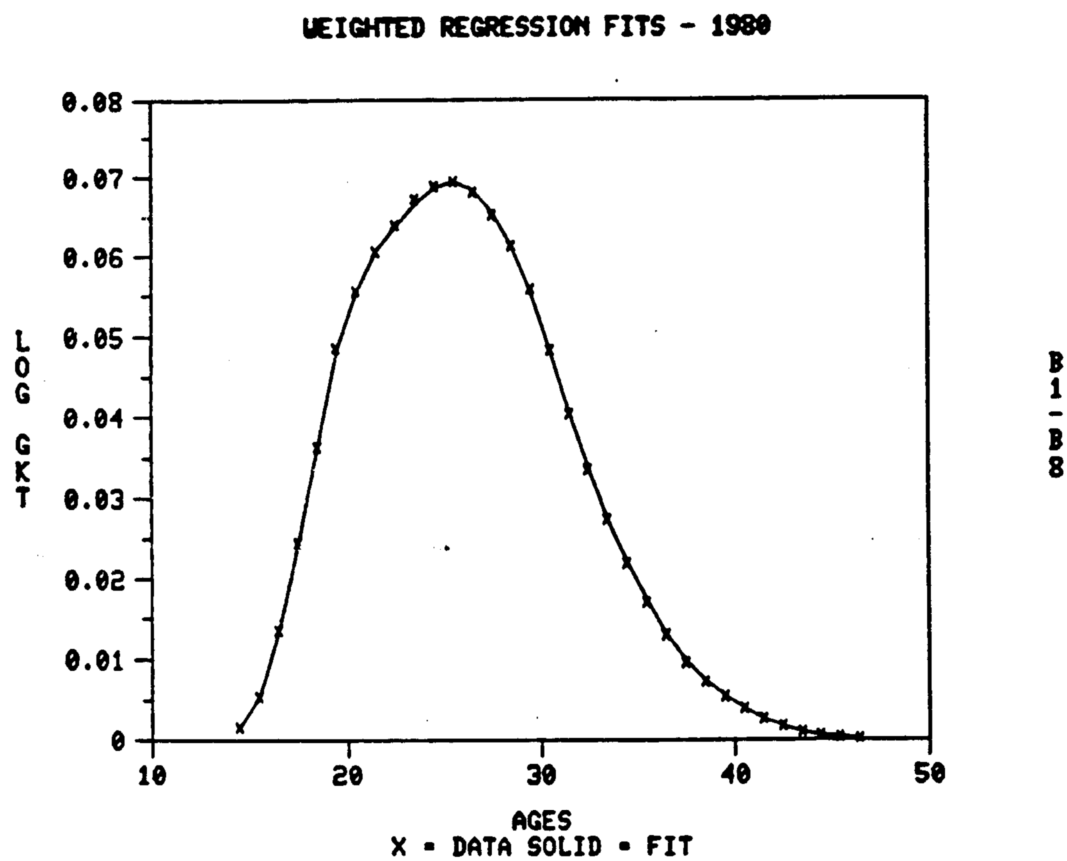


Figure 10b.



(a)



(b)

Figure 11. Principal components approximations (—) to 1980 U.S. logged relative fertility rates (x). (a) Using 4 principal components. (b) Using 8 principal components.

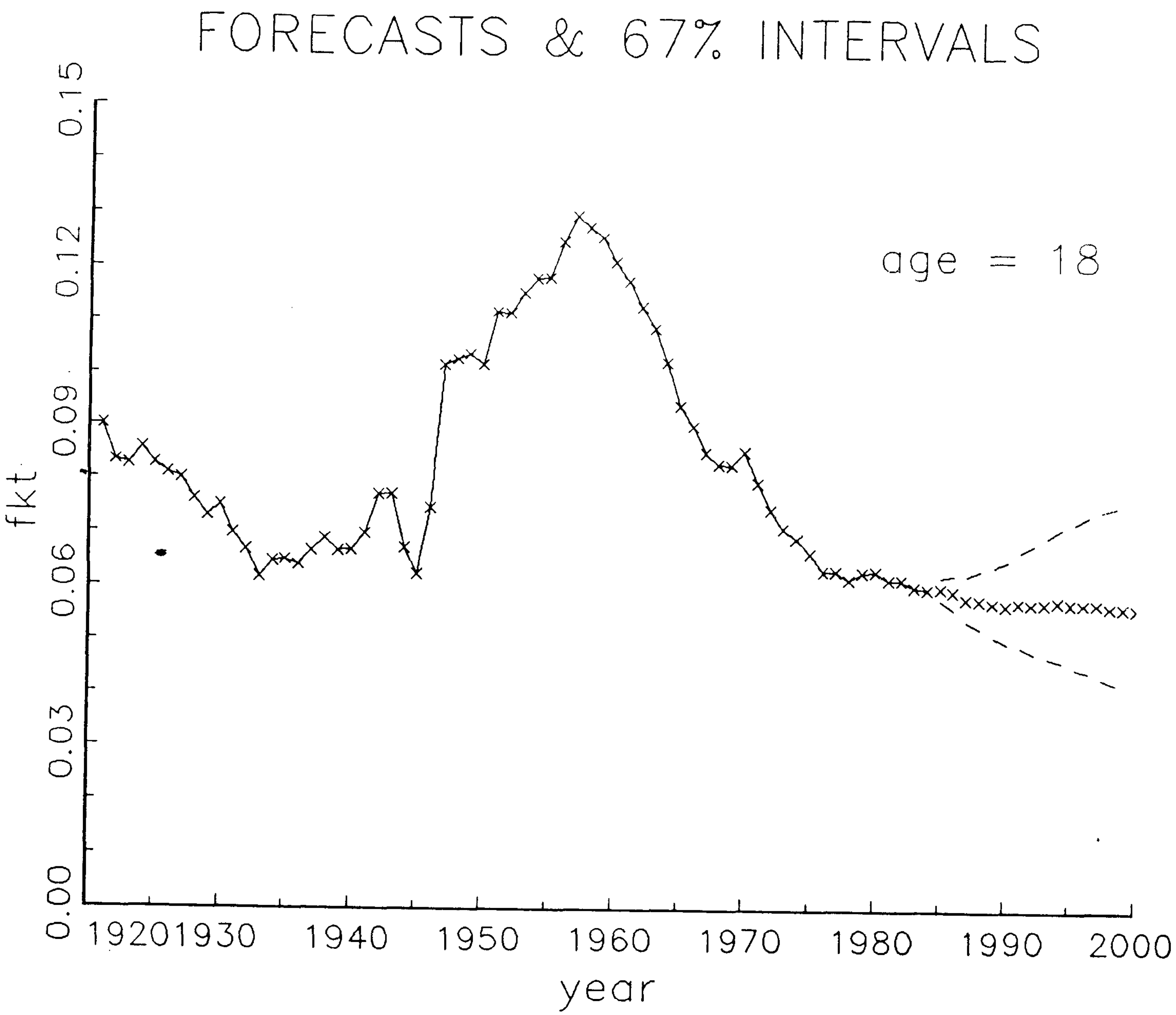


Figure 12a. U.S. age-specific fertility rates (—x—), point (X) and 67% interval forecasts (--) for 1985-2000 from 1984, for selected ages: (a) 18; (b) 25; (c) 30. The point and interval forecasts are from a multivariate ARIMA model for the principal component series developed in Bozik and Bell (1988).

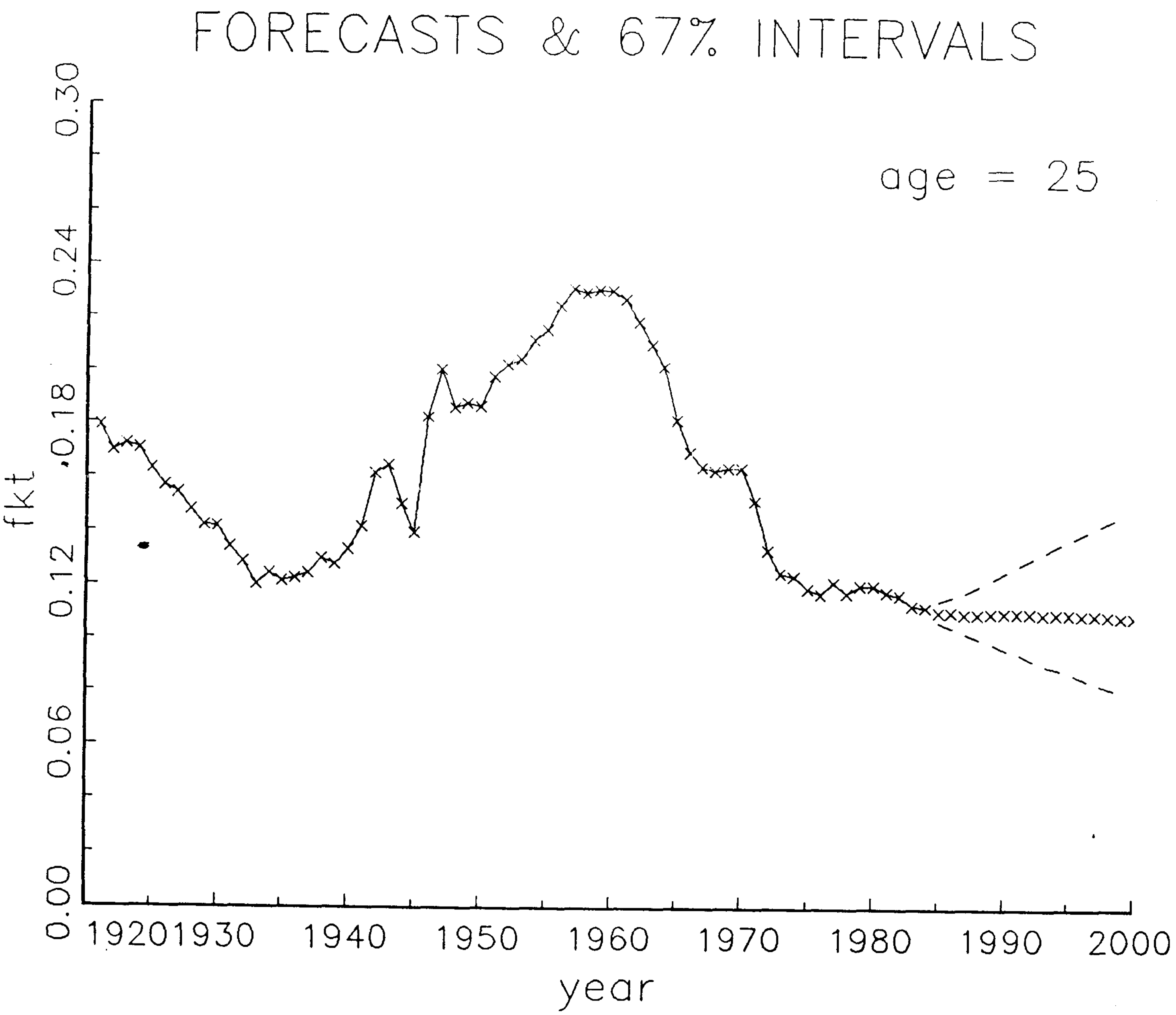

FORECASTS & 67% INTERVALS
age = 25
fkt
0.00
0.06
0.12
0.18
0.24
0.30
1920
1930
1940
1950
1960
1970
1980
1990
2000
year

Figure 12b.

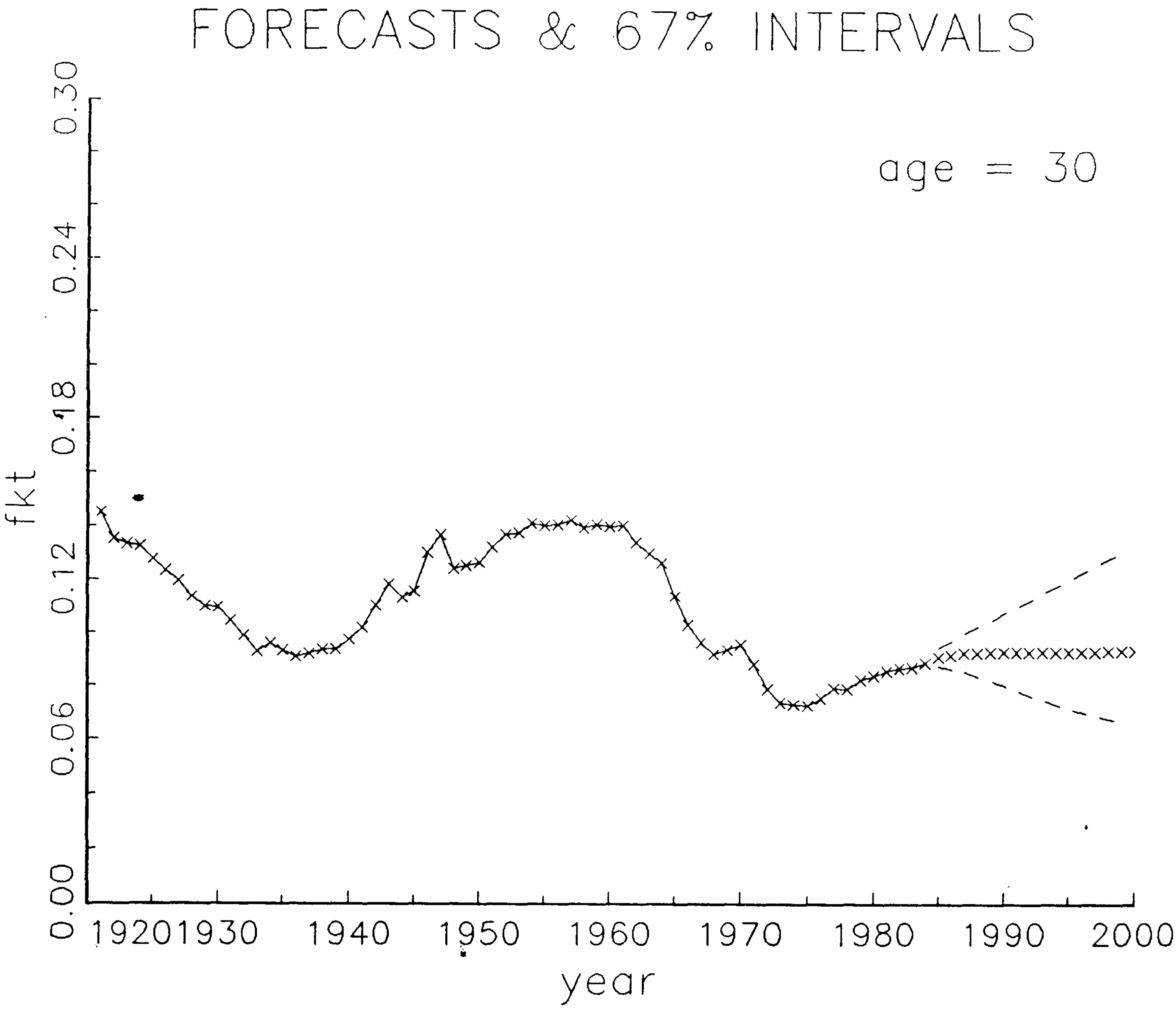

FORECASTS & 67% INTERVALS
age = 30
fkt
0.00
0.06
0.12
0.18
0.24
0.30
1920
1930
1940
1950
1960
1970
1980
1990
2000
year

Figure 12c.

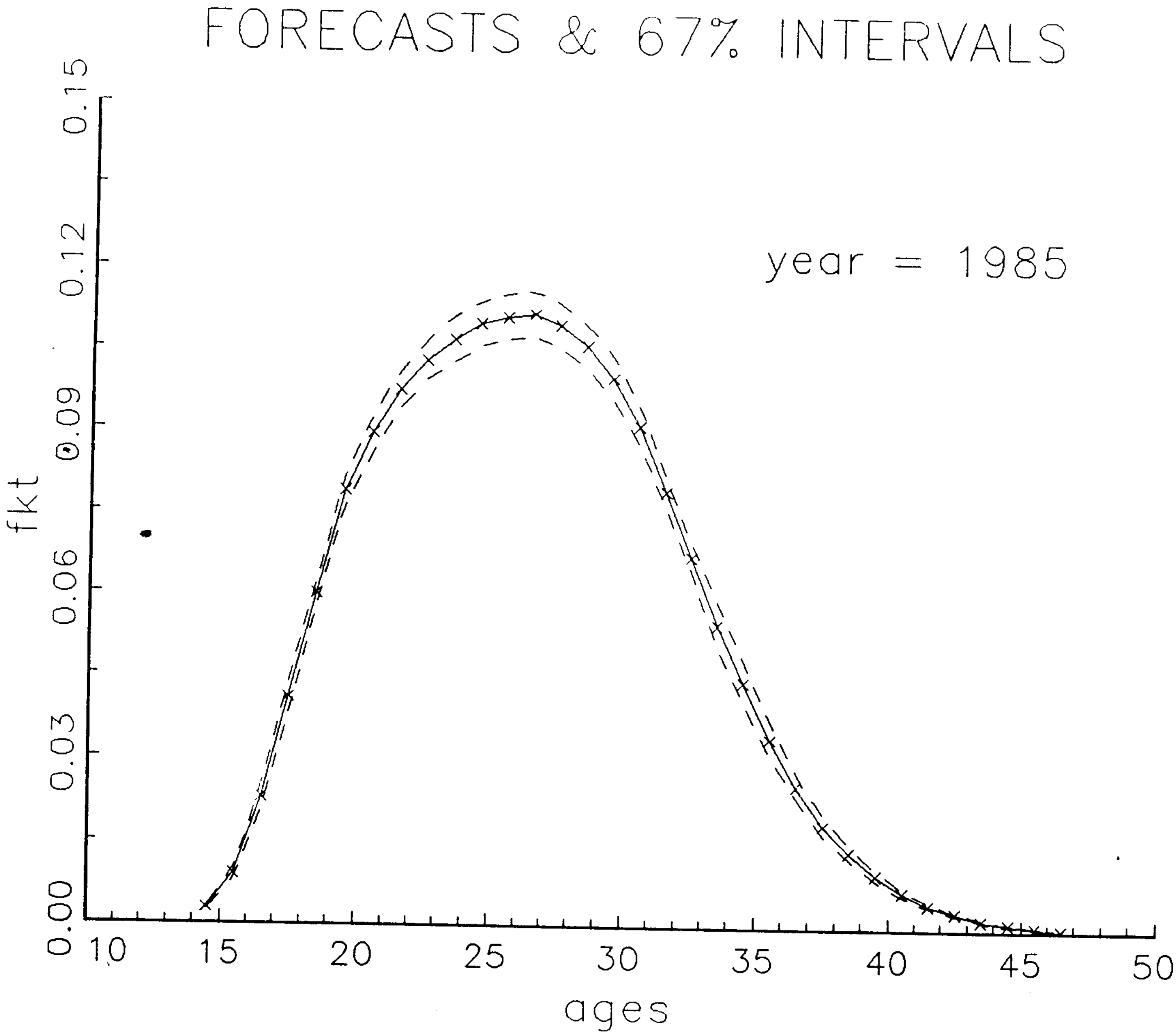


Figure 13a. Point (X) and 67% interval forecasts (--) from 1984 of U.S. age-specific fertility rates for all ages and selected years: (a) 1985; (b) 1988; (c) 1995. The point and interval forecasts are from a multivariate ARIMA model for the principal component series developed in Bozik and Bell (1988).

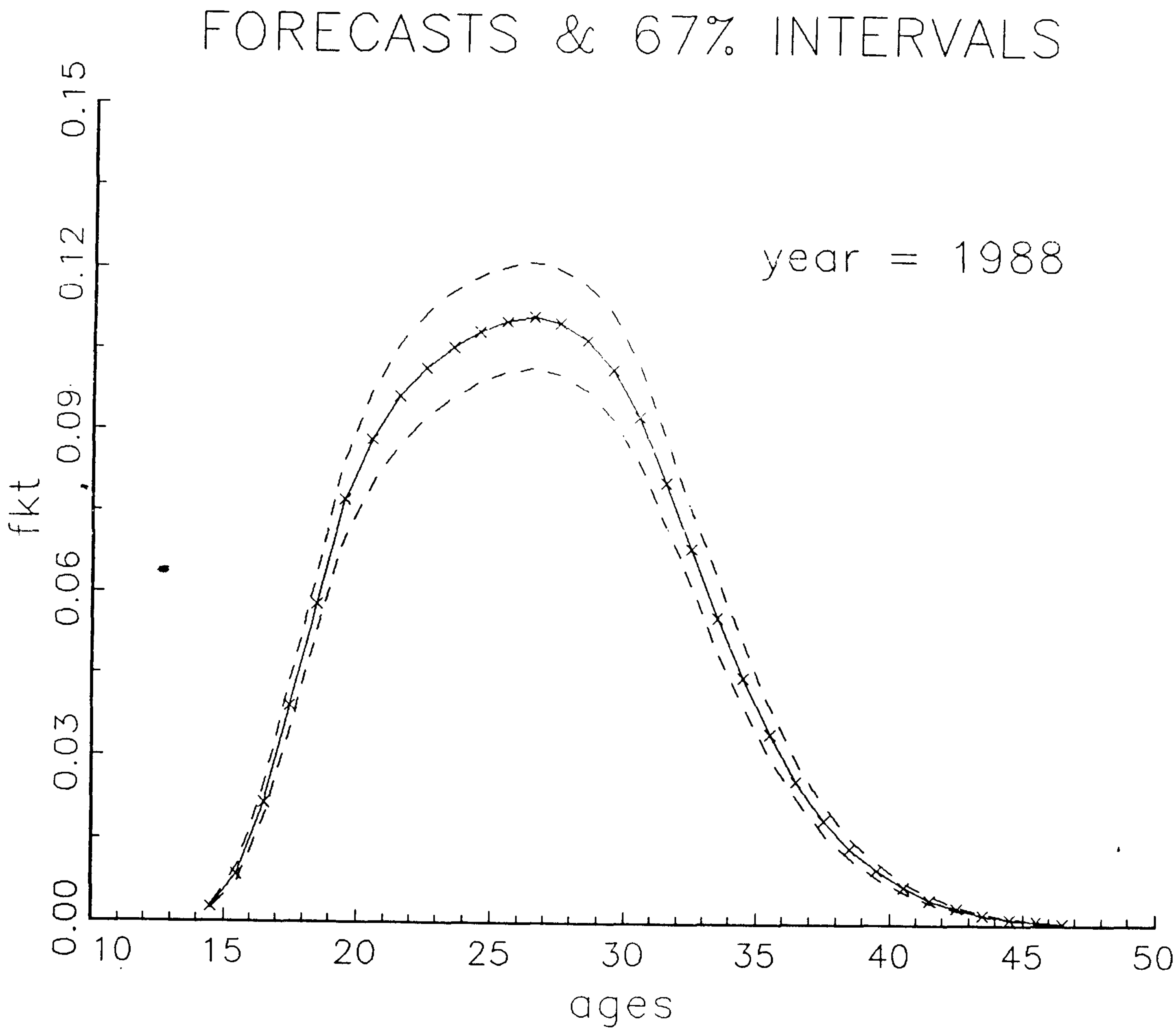

FORECASTS & 67% INTERVALS
year = 1988
fkt
0.00
0.03
0.06
0.09
0.12
0.15
10
15
20
25
30
35
40
45
50
ages

Figure 13b.

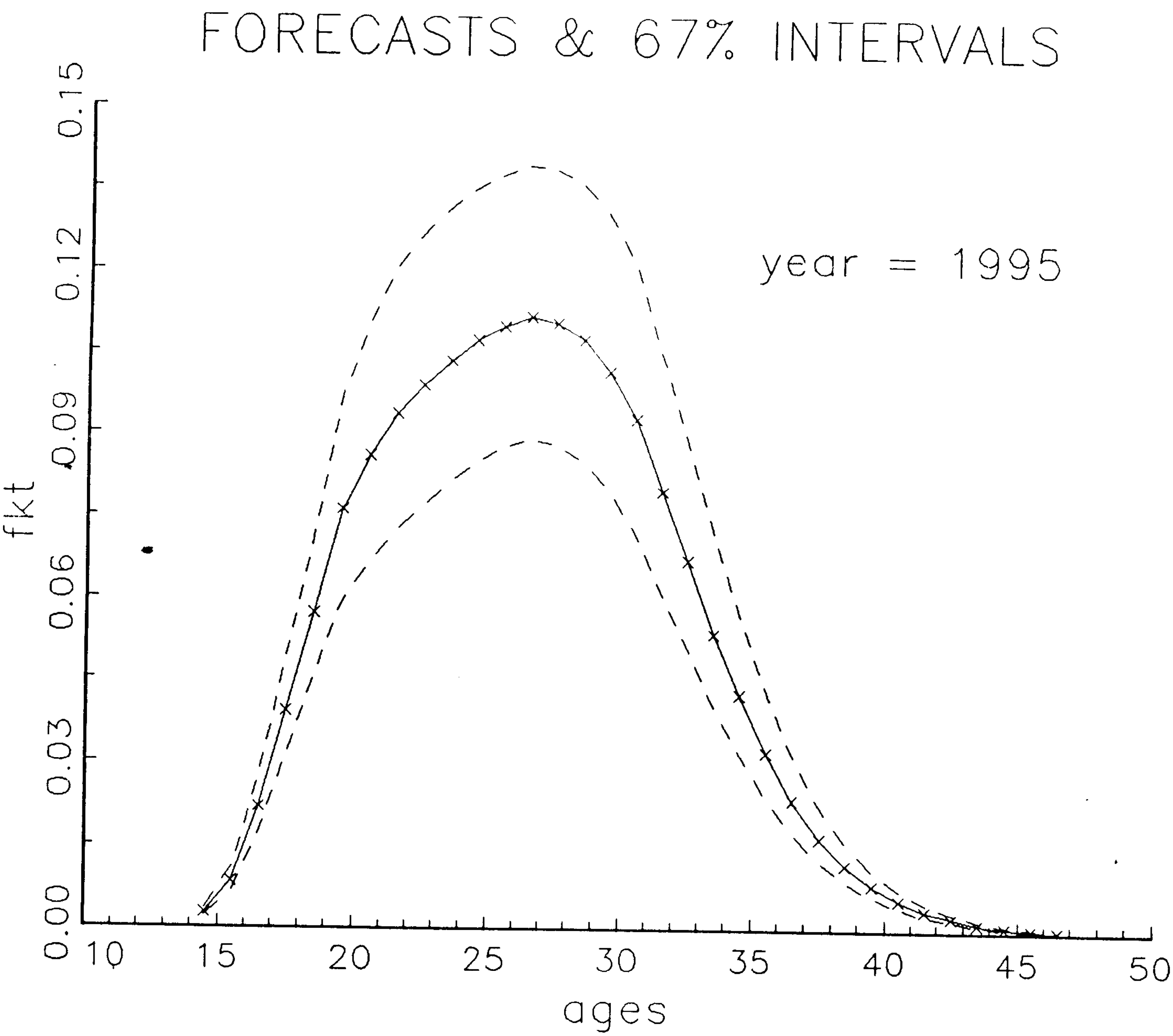

FORECASTS & 67% INTERVALS
year = 1995
fkt
0.00
0.03
0.06
0.09
0.12
0.15
10
15
20
25
30
35
40
45
50
ages

Figure 13c.