A MODEL OF BIRTH PROBABILITIES SPECIFIC FOR AGE AND PARITY OF WOMEN

Harry M. Rosenberg, Carolina Population Center, University of North Carolina
Ralph E. Thomas, Battelle Memorial Institute, Columbus Laboratories

Formal models of aggregate human fertility have focussed, in the main, on the variable, age of woman, and its relation to the probability of having a birth. The literature includes early work such as Lotka's graduation of the maternity function[1] by the normal or Gaussian distribution, Wicksell's fit by a Pearson Type III curve, and Hadwiger's exponential fit.[2] Recent work by Mitra and Romaniuk has used the Pearson Type I curve for graduation of fertility rates by age of woman; Brass has used polynomials; and Murphy and Nagnur have fit Gompertz functions to fertility rates, cumulated by age of woman.[3] Considerably less attention has been given to the variable parity and its relation to the probability of birth, even though some, such as Ryder, have come to emphasize the importance of this variable. Ryder has stated, "The most important fertility variable is birth parity, not merely because it influences fertility, but because it is fertility."[4] Accordingly, in some of Ryder's recent work, birth analyses are carried out entirely in terms of parity indices rather than birth rates specific for age.

Conceptually, we need not adopt the view that parity of women is a more useful perspective, or a more powerful correlate of birth probability than age of women. Instead, parity and age can be treated as complementary perspectives on fertility, combined in a more basic index of aggregate fertility, the birth rate specific for both age and parity, hereinafter called the "birth probability."

Pioneering work with birth probabilities done by Whelpton was reinforced by Karmel's discussion of Whelpton's work.[5] More recently Murphy assessed the stable population implications of birth probability assumptions.[6] The present paper attempts to redirect attention to rates specific for both age and parity of women by constructing a general structural model that captures the major features of a birth probability time series for American women during the years 1917 to 1968.

THE MODEL The attempt has been made in developing the birth probability model to construct one which, while showing good correspondence with the data, also embodies relations and parameters that can be related to certain social and biological aspects of human fertility. Such a model can have value for both estimation and explanation.

*Research upon which this paper is based was supported by Grant No. IRO1 HD05981-01 from the National Institute for Child Health and Human Development (Center for Population Research). Valuable assistance in this study has been provided by Norman Ryder, Office of Population Research, Princton University; John Patterson, Division of Vital Statistics, National Center for Health Statistics; Amy J. Kuntz, Department of Mathematics, University of North Carolina, Chapel Hill; and Roger W. Cote, Battelle Memorial Institute, Columbus Laboratories.

Evaluation of arrays of birth probabilities for individual birth cohorts of women indicates that the likelihood of having a birth increases monotonically with increasing parity n of women, controlling for age x; and decreases monotonically with increasing age of women after age 22 years, controlling for parity.[7] A model that can represent these relations is as follows:

$$b(x,n)=\begin{cases} 1, & x-\lambda n<m_1+e^{-1/A}(m_1-m_2) \\ -A\ln\left(\dfrac{x-m_1-\lambda n}{m_2-m_1}\right), & m_1+e^{-1/A}(m_2-m_1)\le x-\lambda n\le m_2 \\ 0, & x-\lambda n>m_2 \end{cases} \qquad (1)$$

where λ represents the number of years 'consumed' by each additional birth; m_1 represents the age at menarche; m_2, the age at menopause, and the parameter A denotes a scale factor. The end conditions restrict birth probabilities from exceeding 1.0 at the onset of menarche, and the lower restriction prevents birth probabilities from becoming negative after menopause.

Conceptually, parameter m_1 is interpreted as representing the chronological onset of female fertility; m_2 as representing the termination of the fertile period; $x-m_1$, then, represents the number of years since the menarche. The parameter λ is the reproductive 'cost' in years per additional child, and the entire expression in the numerator of equation (1), namely $(x-m_1-\lambda n)$ represents nonreproductive years at age x. The denominator m_2-m_1 can be viewed as a representation of the fertile period.

Thus the ratio $\frac{x-m_1-\lambda n}{m_2-m_1}$, varies between zero and one when $m_1\le x-\lambda n\le m_2$. Because the numerator represents the complete span of her reproductive years, it is seen that the ratio is that proportion of her reproductive life that to that point has <u>not</u> been used for reproduction. A low value for this ratio, which implies a high potential for future childbearing, can result under either of two conditions: (1) when the woman is near age m_1, or (2) when n is large so that the residual $(x-m_1)-\lambda n$ is small. In the first case, the ratio will be small because of the woman's chronological youth; in the second, because her reproductive youth has been maintained through repeated childbearing, reflecting a large value for n. Small values of the ratio, then, are associated with probabilities close to the value one; large values of the ratio are associated with probabilities close to zero.

It is a significant feature of the model that a reproductive history of high fertility is associated with a high probability of having a next child. Another way of stating this is to say "the more children a woman has had, the more she is likely to have, except as she ages."

Another way of picturing the process qualitatively is as a series of monotonically decreasing probability schedules by age of woman, one for each parity class of woman. The schedules increase in height by $-A\ln\left(\frac{x-m_1-\lambda(n+1)}{x-m_1-\lambda n}\right)$

with increasing parity as illustrated in Figure 1, such a woman remaining childless would be associated with the continuous downward trajectory indicated as "I", while a woman having a first birth at age x, shown as path "II", would, at that age, experience a discontinuity as she moved to a higher, but decreasing, trajectory of women in parity class one. Although Figure 1 shows a series of birth probability schedules, they are, in fact, only a single curve, as represented in equation 1.

The parameters m_1, m_2, and λ are assumed to be biological processes that may change slowly over time. In contrast, the parameter A is assumed to reflect mainly aggregate fertility responses of a birth cohort to exogenous social and economic factors. For this reason, the behavior of A is of particular analytic interest and is pursued in another paper. Previous work along these lines has been done by Campbell and others.[8]

PARAMETERIZING THE MODEL Parameters of the model were estimated using data from the Division of Vital Statistics, National Center for Health Statistics, U.S. Department of Health, Education, and Welfare.[9] Birth probabilities are actually derived measures, in which the numerator represents births of a specific birth order; that is, first births, second, third, fourth births, etc., by age of woman.[10] The denominator represents an independent estimate of women at risk, by age and parity, to have a birth of that order. Thus, between ages m_1 and m_2, all parity zero women are assumed to be at risk to have first births; parity one women, at risk to have second births, etc. The numerator and the denominator are each subject to the errors and biases characteristic of their respective sources, namely, the vital registration system and the decennial censuses. These include problems of under-reporting, misclassification of age and parity of women, and misreporting of birth order of child. Moreover, it is likely that historically the extent and type of biases in the data arising in both sources has changed, so that some perturbations over time in data series actually reflect changes in the quality of data rather than changes in fertility.

Furthermore, biases and distortions in the rates are likely to be concentrated among certain age and parity combinations, where the probability of having a birth is very high. Such a group would be young, high parity women among whom the risk of birth is high. For such numerically small groups, rates tend to be less stable and reliable than those of other age-parity combinations.

Parameter estimation, for all cohorts combined (1877 through 1954), involved a two-stage procedure in which, first, λ was estimated, and second, the parameters m_1, m_2-m_1, and A were derived by a linear least squares program,[11] using the estimated value of λ and the following transformation of equation (1): $(x-\lambda n)=m_1+(m_2-m_1)e^{-b/A}$ (2)

Estimates of parameters for all cohorts combined were as follows: λ= 1.83 years; m_1= 8.58 years; m_2-m_1= 30.83 years; and A= 0.209. Our goal is to develop a model whose parameters can be interpreted as social or biological. However, the initial estimates of the parameters do not entirely satisfy the objective. For example, the estimated value for m_1, interpreted as the age of menarche, is about nine years, well below the recorded experience for this country.[12] The length of the fertility period, to which the difference m_2-m_1 refers, is also less than the span indicated by the data set.

In validating the model, subsequent to establishing single values for each parameter based on the combined cohort estimates, we found that a single estimate for the parameter A resulted in consistent underestimates of birth probabilities for parity zero women. As a consequence, separate estimates were made of parameter A for each parity.

In general, the fit between observed and model-generated estimates of cohort birth probabilities is good; it reveals, interestingly, that positive or negative deviations persist over successive ages rather than showing random movement around predicted values between successive ages. Autocorrelation of this kind might represent actual aggregate fertility behavior in which lower parity specific rates are subsequently compensated by higher rates and the reverse; or, it could reflect either data or model bias. For all the parities combined in the 1908 cohort, the relative difference (root mean square expressed as a percent) between the actual and the fitted data was 1.6 percent. This was the best fit achieved for cohorts 1877-1954; the poorest fit was for the 1897 cohort for which the model values deviated from the actual values by about six percent.

MARKOV PROCESS REPRESENTATION As Keyfitz indicated in summarizing the work of E. M. Murphy, birth rates specific for age and parity lend themselves quite naturally to representation as Markov processes.[13] In the Markovian model, the birth probabilities are the likelihood that a randomly-selected woman in a parity class will move, or make a transition, from parity state n to parity state n+1. The birth probability model here presents a particular kind of Markov process that can be characterized as nonhomogeneous and irreversible. Nonhomogeneity refers to the characteristic that the state transition probabilities are conditional on age of the woman, while irreversibility means that women can only move to higher parity classes, and never return to lower ones.

With these observations, the transition matrix M(x), as a function of calendar age (x), may be written as follows:

$$M(x) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7+ \end{array} \begin{array}{c} \begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7+ \end{array} \\ \left[\begin{array}{cccccccc} q_0 & p_0 & \cdots & & & & & 0 \\ & q_1 & p_1 & & & & & \cdots \\ & & q_2 & p_2 & & & & \cdots \\ & & & q_3 & p_3 & & & \\ & & & & q_4 & p_4 & & \\ \cdots & & & & & q_5 & p_5 & \\ \cdots & & & & & & q_6 & p_6 \\ 0 & \cdots & & & & & & 1 \end{array}\right] \end{array},$$

where the p_n, q_n denote transition probabilities, and the parity states are shown by the labels bordering the matrix. Only the diagonal and the super-diagonal have non-zero entries; all other entries are zero. The symbol p_n denotes the probability that a randomly selected woman of age x and parity n will make a transition to parity state n+1 in the coming year; the probability that no such transition is made is denoted by $q_n = 1-p_n$, n=0, ..., 6.

Next, at the beginning of a given year, let $V(x) = (v(0), \ldots, v(7))$ denote a parity distribution of 1000 randomly selected women of age x among the parity states 0 through 7. The expected parity distribution at the beginning of the next year is then given by

$$V(x+1) = V(x)M(x).$$

Similarly, by incrementing age by one year we have

$$V(x+2) = V(x+1)M(x+1),$$

and the replacement of V(x+1) yields

$$V(x+2) = V(x)M(x)M(x+1).$$

Proceeding recursively, it follows that

$$V(x_m) = V(x_n) \prod_{x=x_n}^{x_m} m(x),$$

where x_n and x_m denote the ages of menarche and menopause, respectively. As an example, if 1000 women of age 14 are all taken to be in parity state 0, then V(14) = (1000, 0, 0, 0, 0, 0, 0, 0) and the expected parity distribution at age 49 is given by

$$V(49) = V(14) \prod_{x=14}^{49} M(x).$$

This result shows how the parity distributions can be computed from the matrices of transition probabilities. These transition probabilities may be taken equal to the reported birth probabilities. Alternatively, the published birth rates may first be 'smoothed' by fitting the birth probability model developed in this paper.

THE MODEL FOR YOUNGER WOMEN The model postulates decreasing birth probabilities with increasing age of women. And, in general, the fit between recorded and fitted values is quite good for women in the age range 22 to 47 years. Before age 22 years, however, observed values are considerably below those expected on the basis of the model. For example, the model yields 0.328 for the birth probability of a 15-year-old woman of zero parity, compared with a reported value for this birth birth probability of 0.0087. However, the reported figures do show increasing birth probabilities with increasing parity, controlling for age of women.

The large discrepancies between the observed and the model birth probabilities at lower ages are believed to be attributable, in large measure, to the relatively small number of women who are married at young ages. Proportions of women ever married increase from about one out of 100 at age 14 years to over seven out of ten at age 22, stabilizing for recent cohorts at about 90 percent ever married at age 30 years.[14] Marriage, we know, is neither a necessary nor sufficient condition for childbearing, given the incidence of premarital conception, which may or may not be followed by wedlock, together with the prevalence of marital infertility; however, marriage can serve as a proxy of changes in exposure to the risk of having children at younger ages.

For simplicity, consideration of marriage as an explicit element was deliberately omitted in developing the model in its present, preliminary form. There are, nevertheless, ways in which it could be incorporated into a somewhat more complex model. Keyfitz, for example, suggests that the matrix formulation include parity states specific for nuptiality.[15] But the major restriction on such an approach are data limitations with respect to the legitimacy status of newborns and the marital status of women at risk to have children.

A somewhat less formidable way of approaching the problem, but one also constrained by data limitations, is normalizing reported birth probabilities by estimated proportions of women-ever-married prior to estimating model parameters. The effect of normalizing birth rates by marital status can be illustrated using 1960 data on births and women ever married by age, as shown below:

Age Groups of Women	Birth Rate 1960 [16]	Proportion of Women Ever Married, 1960 [17]	Birth Rate for Ever Married, Normalized Women 1960
15-19 years	0.089	0.118	0.754
20-24	0.258	0.720	0.358
25-29	0.197	0.869	0.227
30-34	0.113	0.891	0.127
35-39	0.056	0.881	0.064
40-44	0.016	0.860	0.019

Normalization by marital status, this illustrates, changes the standard maternity function to one compatible, in its general magnitude, with our birth probability model.

DISCUSSION The availability of fifty years of birth rates specific for the age and parity of American women provides a rich data base for structural and historical analysis. The general structural model

$$b(n,x) = -A(n)\ln\left(\frac{x - m_1 - \lambda n}{m_2 - m_1}\right)$$

approximates the major features of the truncated fertility function cohorts of women born during 1896 to 1945. The basic features of the model are the following: (1) two parameters that relate conceptually, though not quantitatively in our initial work, to the biological states of menarche and menopause; (2) a parameter for the socio-biological process of childspacing; and (3) a scale factor that appears to reflect a cohort's responses to socio-economic forces, both short-term and long-term.

The model serves to organize the age and parity specific birth probabilities in a manner that is intuitively appealing and is useful as an investigative tool. However, certain shortcomings of the model should be noted: (1) some modifications should be made to extend the range of the model downward to age 14 years by procedures such as those discussed here; (2) further adjustments to the parameters identified as predominantly biological may be desirable to bring them more closely into alignment with independently-derived values; and (3) further studies should be made to minimize biases in the model, and to ensure the validity of reported rates, through adjustment if necessary.

Future work will involve detailed historical analysis of the behavior of each of these parameters and examination of the covariation between these parameters and related biological and social variables. In addition, initial work indicates that the model may be helpful in extending Ryder's work on the relation between cohort and period fertility indices.18/ Also, it may be useful in projecting age-specific period fertility rates on a more secure cohort fertility base.

General Model Equation:

$$b(x,n) = -A\ln\left(\frac{x-m_1-\lambda n}{m_2-m_1}\right)$$

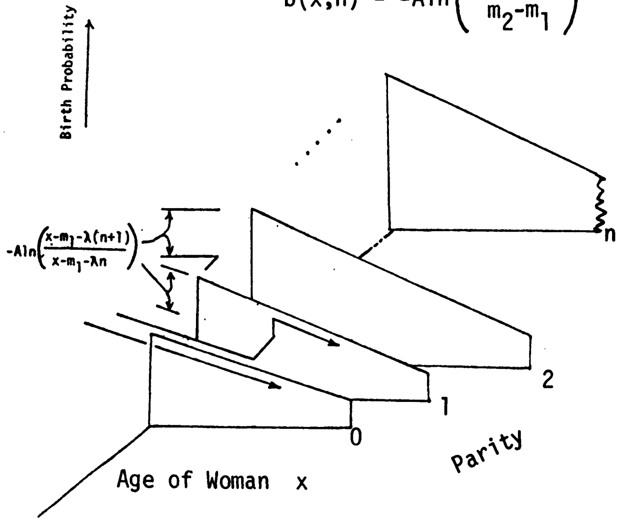

Figure 1. Hypothesized Relation Between Birth Probability and Age of Woman, II.

Footnotes

1/ In developing the birth probability model, we do not explicitly incorporate the force of mortality, ordinarily expressed as p(x) in mathematical treatments of population renewal. We use the expressions maternity function and fertility function interchangeably.

2/ J. Alfred Lotka, *Theories analytique des associations biologiques.* Part II. Analyse demographique avec application particuliere a l'espece humaine, Actualities scientifiques et Industrielles, No. 780. Paris: Hermann and Cie, 1939; S. D. Wicksell, "Nuptiality, Fertility, and Reproductivity," *Skandinaviske Aktuarietidskrift,* 1931, pp. 125-127; and H. Hadwiger, "Eine analytische Reproductionsfunktion fur biologische Gesamtheiten," *Skandinaviske Aktuarietidskrift,* 1940, pp. 101-113. For a concise summary of the earlier work, see Nathan Keyfitz, *Introduction to the Mathematics of Population,* Reading, Massachusetts: Addison-Wesley Publishing Company, 1968. pp. 140-169.

3/ S. Mitra, "The Pattern of Age-Specific Fertility Rates," *Demography*, Volume 4, 1967, pp. 894-906; A. Romaniuk, "A Three Parameter Model for Birth Projections," *Population Studies* Volume 27, Number 3, 1973, pp. 467-478; William Brass, "Graduation of Fertility Distributions by Polynomial Functions," *Population Studies,* Volume 14, Number 2, 1960, pp. 148-162; Edmund Murphy and Dhruva N. Nagnur, "A Gompertz Curve that Fits: Applications to Canadian Fertility Patterns," *Demography*, Volume 9, 1972, pp. 35-50.

4/ Norman B. Ryder, "The Measurement of Fertility Patterns," in Mindel C. Sheps and Jeanne C. Ridley, editors, *Public Health and Population Change,* Pittsburgh, Pa.: University of Pittsburgh Press, 1965, p. 293. Parity refers to the number of births a woman has had; and birth order to whether the child is a first, second, third, etc child. Thus, women who have had no children are referred to as zero parity women; those who have had a first child (birth order one child), as parity one women, etc.

5/ P. K. Whelpton, "Reproductive Rates Adjusted for Age, Parity, Fecundity, and Marriage," *Journal of the American Statistical Association,* Volume 45, 1950, pp. 119-124.

6/ Edmund M. Murphy, "A Generalization of Stable Population Techniques," unpublished Ph.D. dissertation, Department of Sociology, University of Chicago, 1965.

7/ Development and parameterization of the model is restricted to women aged 22-47 years. See the section below on problems of fitting the model to younger women.

8/ See, for example, Campbell's discussion of the short-term covariation between birth probabilities and indices of economic activity in Clyde V. Kiser, Wilson H. Grabill, and Arthur A. Campbell, *Trends and Variations in Fertility in the United States,* Cambridge, Massachusetts: Harvard University Press, 1968, Chapter 12.

9/ For 1959-1968, the data are from unpublished worksheets; for 1917-1958, the data are from P. K. Whelpton and A. A. Campbell, "Fertility Tables for Birth Cohorts of American Women," *Vital Statistics--Special Reports,* Volume 51, Number 1, January 29, 1960.

10/ See footnote 4 above.

11/ P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences,* New York: McGraw-Hill, Inc., 1969, pp. 208-212.

12/ The estimated age at menarche in the United States is between 12 and 13 years. See J. M. Tanner, *Growth at Adolescence,* Springfield, Massachusetts; Charles C. Thomas, 1962.

13/ Keyfitz, *op. cit.*, pp. 328-331, and footnote 6 above. For a general treatment of Markov processes, see J. G. Kemeny, and J. L. Snell, *Finite Markov Chains,* Princeton, New Jersey: D. Van Nostrand Company, 1960.

14/ U. S. Department of Commerce, Bureau of the Census, *1960 Census of Population* PC(2)-4E, "Marital Status," Table 1.

15/ Keyfitz, *ibid.*, p. 328

16/ U. S. Department of Health, Education and Welfare, Division of Vital Statistics, *Vital Statistics of the United States, 1960,* Volume I, "Natality," Table 1E. Birth rates by age of women are equal to the summation of birth rates by age and parity of women weighted by the parity distribution of women at each age.

17/ See footnote 14. Proportions ever married at midpoint of age group, *e.g.*, age 17 years for age group 15-19 years, etc.

18/ Norman B. Ryder, "The Process of Demographic Translation," *Demography*, Volume 1, Number 1, 1964 pp. 74-92; "The Structure and Tempo of Current Fertility," in *Demographic and Economic Change in Developed Countries,* for the National Bureau of Economic Research, Princeton, New Jersey: Princeton University Press, 1960, pp. 117-136.