PEARSONIAN TYPE I CURVE AND ITS FERTILITY PROJECTION POTENTIALS

S. Mitra, Emory University, and A. Romaniuk, Statistics Canada
and University of Montreal

1. THE MODEL FOR AGE-SPECIFIC FERTILITY RATES

The purpose of this paper is twofold: to reduce further the number of parameters used in constructing models of fertility graduation and simulation, utilizing primarily those readily available or easily estimated so as to expand the range of their application; and, secondly, to explore the models potentials for deriving the birth series for forecasting population by the component method. The paper may be regarded as a step forward in the development of a fully-fledged parametric model for fertility projections adaptable, primarily, to requisites of empirical material available in the developed countries. Fertility time series for Canada are used to test the validity of the procedures developed therein.

This model was developed on the basis of certain well known facts about child bearing patterns of women. First, children are mostly born to mothers in the age-group say, 15-50 years and, second, fertility rate continues to increase beyond age 15, reaches a maximum somewhere in the age interval 20-30 years and begins to decline thereafter.

Such a curve fits the description of type I of the Pearsonian family of curves, and this model was attempted earlier (Mitra, 1967) in order to describe the pattern of age-specific fertility rates of a number of countries subject to certain modifications. With origin at mode, the equation of the curve is

$$y = y_o \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2} \text{......(1)}$$

where y_o is the modal ordinate and $-a_1 \leq x \leq a_2$.

Also, $m_1/a_1 = m_2/a_2$ ......(2)

Ordinarily, the parameters of the type I distribution can be obtained by solving equations generated by equating the first four moments of the observed distribution with those (Elderton, 1930) that can be directly derived from (1). For the distribution of age-specific fertility rates it was assumed that the age interval of 15-50 years is generally adequate and could therefore be used as an additional restriction. Thus

$a_1 + a_2 = 35$ ......(3)

together with (2) and the starting point at age 15, reduced the number of independent parameters to only two and the solutions were obtained from the following equations (Mitra, 1967).

$$m_1 + m_2 = \frac{u_1'(a_1 + a_2 - u_1')}{u_2} - 3 \text{......(4)}$$

$$m_1 = \frac{m_1 + m_2 + 2}{a_1 + a_2} u_1' - 1 \text{......(5)}$$

where u_1' is the mean and u_2 is the variance.

Equations (2) and (3) could then be used to determine a_1 and a_2 and for y_o, the following, namely,

$$y_o = \frac{N}{(a_1 + a_2)B(m_1 + 1, m_2 + 1)} \cdot \frac{m_1^{m_1} \quad m_2^{m_2}}{(m_1 + m_2)^{m_1 + m_2}} \text{.(6)}$$

can be used where N is the sum of the age-specific fertility rates and B stands for the Beta function. In fact y_o may be regarded as a multiplier that equalizes the sum totals of the observed and the graduated distributions.

2. FURTHER VERIFICATION OF THE MODEL AND THEORETICAL CONSIDERATIONS

If data are available in greater details, the proposed model can be tested by carrying out the analysis at sufficient depth. With that in mind the authors of the present paper have examined these distributions for Canada which are available by single years of age beginning 1926. As such, these data are too detailed, since there are too many class intervals than are necessary to investigate the nature of the distribution function. According to Pearson, data condensed into twelve or so classes are detailed enough for curve fitting. However, such condensation has not been attempted because that would require the use of somewhat unconventional (two to three year) age-groups.

For all these distributions, the Pearsonian index k has assumed values justifying the use of a type I curve for describing the distribution of age-specific fertility rates. This verifies the theoretical formulation of the pattern of the distribution as postulated above.

As mentioned before, the best fit of a given fertility distribution may perhaps be obtained by estimating all the parameters in the usual manner which in this instance will require the computation of the first four moments. While that may not be too restrictive for fertility distributions given by single year of age, there appear to arise a few operational as well as technical problems that seem to complicate matters during the process of searching for meaningful interpretations of the parameters of the model. For one thing, the parameters of the model should respond in a manner that is consistent with the trend in the fertility rates. Otherwise, variations in the distributions cannot be properly related to the variations in the parameters and therefore will have little practical value. With so many parameters, it is quite difficult to relate the effect of say a reduction in total fertility rate on modal age, on the fertile age span, etc.

For another, it is difficult to explain why the effective fertile interval should be subject to variation from time to time. Yet, this is what must happen when the type I distribution is obtained in a straightforward manner, i.e., without any restriction. Not only can that happen, but the start and end of the curve may be in direct conflict with reality (see Table 1).

The earlier study (Mitra, 1967) took care of these problems by introducing a restriction like the one specified in equation (3). Of course, there is no reason to believe that the definition of that equation is unique in any way. In fact, it seems only appropriate that some such equation should be arrived at in a realistic manner and be used consistently, unless there are valid reasons to do otherwise. The advantages of these restrictions are primarily, a reduction in the number of independent parameters from four to two, which, in turn, renders the study of the pattern of variations and interrelationships of these two parameters relatively simple and perhaps more meaningful.

3. OTHER VARIANTS OF THE ESTIMATING PROCEDURE

I. Of course, there are more ways to estimate the parameters of a type I curve once the initial restriction on the fertile age interval (like ages 15 to 50) is imposed. The method used before and described earlier in this section is based on equating the first two moments of the observed distribution with those of the theoretical distribution. Alternatively, the first moment or the mean, and another measure, say the mode, may be used in a similar manner. The latter is estimated by some conventional method, or simpler still, through the substitution of the midpoint of the single year age interval corresponding to which the age-specific fertility rate is the maximum. In this case, given the value of the mode or a_1, with the origin at the start of the curve, the remaining parameters can be obtained from the following:

$$a_2 = (a_1 + a_2) - a_1 \quad \ldots\ldots(7)$$

$$m_2 = \frac{a_2(a_1 + a_2 - 2u_1')}{(a_1 + a_2)(u_1' - a_1)} \quad \ldots\ldots(8)$$

and $m_1 = \frac{a_1}{a_2} m_2 \quad \ldots\ldots(9)$

II. Another method based on the fertile age-range, modal age and modal fertility also seems reasonable and worthy of investigation. With orgin at start of the curve, and using the relationship

$$\frac{m_1}{a_1} = \frac{m_2}{a_2} = \frac{m_1 + m_2}{a_1 + a_2} = C \quad \ldots\ldots(10)$$

equation (6) for the modal fertility rate can be rewritten as

$$y_o = \frac{N}{(a_1+a_2)B(Ca_1+1, Ca_2+1)} \cdot \left[\left(\frac{a_1}{a_1+a_2}\right)^{a_1}\left(\frac{a_2}{a_1+a_2}\right)^{a_2}\right]^C \quad (11)$$

where N, for the single year distribution, is equivalent to the total fertility rate. The constant C, obtained through iterative procedure, can then be used in (10) to estimate the remaining parameters. Iteration techniques based on descent method of the type suggested by Keyfitz (1968) is still another approach that could be used to improve the fit. Such an approach proved to be useful in fitting the Gompertz curve to Canadian data on cumulative fertility (Murphy, Nagnur, 1972) and could be eventually, with due adaptations, applied to fertility data in the form of age-specific distribution.

4. THE METHOD AND THE DATA

The first three methods outlined above have been tested against the Canadian fertility series from 1926 to 1969. (Methods based on iterative techniques will be presented in another paper). For simplicity in future references, these three methods will henceforth be recalled by the codes shown in parentheses as in the following:

A. Method based on first four moments (4M)

B. Method based on first two moments (2M)

C. Method based on mean and mode (1M)

The numerals 4, 2, and 1 stand for the number of moments on which a particular method is based. It will be interesting to examine the estimates of the parameters obtained from these three models and these have been done next where, for simplicity in presentation, the years, 1926, 1931, 1941, 1951, 1961 and 1969 have been selected that represents the whole range for which such data are available. A priori, one may expect the goodness of the fit to improve as the number of moments used is increased. It follows that (4M) sets some standard against which estimates obtained by other methods can be gauged.

5. MODAL AGE, MODAL FERTILITY AND TOTAL FERTILITY

Of all the parameters of the type I distribution, the one that is most meaningful in this context seems to be the modal age. Since the method (2M) is restricted by a fixed fertile interval, estimate of the modal age will not be unique and will depend on the specific choice of that intercal. The results based on two alternative fertility intervals of 15-50 and 17-50 years by (2M) along with those calculated by (4M) are compared with observed modal ages in Table 1. It should be mentioned that observed values presented in the table have been adjusted for irregularities which were apparently due to random fluctuations and age misreporting. This adjustment has been made by tracing through the points of observed mode ages, a freehand curve. Also it should be remembered that method (1M) has no built-in mechanism for calculation of mode age and that the latter has to be given in order to enable us to make use of this method for derivation of other parameters.

Table 1 ESTIMATES OF MODAL AGE FOR CANADA FOR A FEW SELECTED YEARS

Year	(4M)	(2M) (15-50)	(2M) (17-50)	(1M) (observed)	Total Fertility (per 1000)
(1)	(2)	(3)	(4)	(5)	(6)
1926	28.1 (17.0-48.5)	28.9	27.9	28.0	3356
1931	27.6 (16.9-49.3)	28.5	27.6	27.7	3201
1941	26.3 (17.2-49.6)	27.6	26.4	27.1	2824
1951	25.5 (17.0-49.9)	26.6	25.4	25.3	3480
1961	24.2 (17.2-50.4)	25.7	24.3	24.0	3857
1969	24.0 (16.6-53.0)	25.1	23.7	23.8	2410

The figures in parenthesesin Col (2) are the estimates of fertile age range based on (4M), the lowest (16.6) and the highest (53.0) limits of which are both found in 1969. The lower limit is no larger than 17.2 years and the smallest value of the upper limit is 48.5 years. The latter appears to vary more than the former, and the variation, though not large, is of considerable magnitude, and the actual values are not quite in accord with the real data. Such inconsistencies, as mentioned earlier, are unavoidable when all the parameters are estimated from the moments of the distribution.

Table 2 ESTIMATES OF MODAL FERTILITY FOR CANADA FOR A FEW SELECTED YEARS

Year	(4M)	(2M) (15-50)	(2M) (17-50)	(1M) (15-50)	(1M) (17-50)
(1)	(2)	(3)	(4)	(5)	(6)
1926	178	181	181	152	184
1931	174	175	175	150	181
1941	159	159	160	147	173
1951	205	203	205	172	204
1961	239	231	236	195	231
1969	157	150	155	130	155

6. THE PARAMETERS m_1 AND m_2

Apart from the goodness of fit, the justification of a model also depends on logical and meaningful interpretations of the parameters. The parameter a, which also corresponds to the modal age is simple to understand while the relevance of m_1 and m_2 is not that apparent. However, it can be argued that among other things, the physiological ability to reproduce depends on age of mother which, in turn, can be expressed as the number of years elapsed since the beginning of the fertile interval (15-50 for example), as well as the number of years left to reach the end of that interval. Of these two, the former may be regarded as a positive force that has the effect of raising the fertility rate while the latter acts negatively to hold the rate down. The intensities of the pushing upward and pulling down effects of these two forces, thus defined, and measurable in this instance by m_1 and m_2, can then be regarded as largely responsible for generating a given form of the fertility curve. This is not to suggest, however, that the pattern of fertility is, in fact, controlled in this manner, but merely to indicate the probable consistency of such an explanation of the parameters in this particular context.

The location of the mode (Table 1, cols. 2-5) indicates that the curve has positive skewness (mode is closer to the start of the curve as $a_1 < a_2$) and therefore $m_1 < m_2$, because of equation (2). From Tables 3 and 4, it is quite apparent that m_1 and m_2 must be very much sensitive to the estimation procedure compared to the actual frequency distribution.

As could be expected, the values of the parameters are, and should be, independent of the total frequency, since their derivation are dependent primarily upon the distribution of the relative frequencies. Accordingly, changes in the parametric values are reflections of the changes in the pattern of the distributions themselves. Particularly sensitive to changing age pattern of fertility is m_2. Correlation between this latter and such measures of frequency distribution as mean age, variance, skewness and kurtosis exceeds 90.

Table 3 ESTIMATES OF m_1 FOR CANADA FOR A FEW SELECTED YEARS

Year	(4M)	(2M) (15-50)	(2M) (17-50)	(1M) (15-50)	(1M) (17-50)
(1)	(2)	(3)	(4)	(5)	(6)
1926	.90	1.55	.98	.83	1.05
1931	.95	1.55	.97	.89	1.09
1941	.83	1.49	.89	1.13	1.23
1951	.85	1.45	.82	.75	.81
1961	.70	1.37	.73	.64	.64
1969	.95	1.39	.72	.76	.74

Table 4 ESTIMATES OF m_2 FOR CANADA FOR A FEW SELECTED YEARS

Year	(4M)	(2M) (15-50)	(2M) (17-50)	(1M) (15-50)	(1M) (17-50)
(1)	(2)	(3)	(4)	(5)	(6)
1926	1.67	2.35	2.00	1.40	2.11
1931	1.93	2.46	2.08	1.57	2.28
1941	2.12	2.68	2.25	2.14	2.82
1951	2.43	2.92	2.43	1.81	2.40
1961	2.64	3.12	2.57	1.86	2.39
1969	3.71	3.42	2.80	2.25	2.84

7. GOODNESS OF FIT

The classical approach to measure the goodness of fit is to calculate X^2, the magnitude of which depends partly on the relative difference between the observed and the graduated values. This method was considered unsuitable in the situation characterized by a large number of intervals and, perhaps because of that (4M) estimates were zero in some cases, at the beginning and at the end of the observed fertile interval. Overall index values, comparing an observed with the expected distributions, can be obtained in a number of ways. One such index, namely the index of dissimilarity (Δ) is obtained by reducing the two distributions in percentage form and summing only the positive differences between corresponding percentages. A given value of this index indicates the percentage of observations that need be redistributed among intervals so that the two distributions become identical. It is easy to see that Δ can range from 0 to 100 and its magnitude depends also on the choice of class intervals.

Table 5 Δ VALUES FOR CANADA FOR A FEW SELECTED YEARS

Year	(4M)	(2M) (15-50)	(2M) (17-50)	(1M) (15-50)	(1M) (17-50)
(1)	(2)	(3)	(4)	(5)	(6)
1926	2.74	4.08	2.77	8.94	2.80
1931	2.66	3.78	2.59	7.15	2.27
1941	2.32	4.30	2.39	5.33	4.51
1951	2.28	4.03	2.41	9.42	2.41
1961	2.97	5.38	3.30	11.34	4.03
1969	3.48	5.23	3.99	10.39	4.13

According to this index, the discrepancy was at its highest level at both extremes, in 1961 when the total fertility rate was the maximum for the period, and also in 1969 when the rate was at its lowest. Of these two years, the index values for (4M) were higher for 1969 and those for (2M) were slightly higher for 1961. As could be expected, the index values were uniformly the smallest for the (4M) method, but considering the simplicity and logical consistency of the other methods, the values are not very high and in general, the model seems to be quite satisfactory, at least as a first approximation of a graduation formula for the fertility rates. The sudden increase of the index values in the latter years seem to indicate the possibility of systematic biases, rather than random fluctuations in the series generated by the differences between the observed and the graduated fertility rates.

8. POTENTIALS FOR FERTILITY PROJECTIONS

It seems that the period covered in this study is quite representative of the wide variation in fertility patterns of Canada in recent decades, and in all likelihood the observed ranges of all the principal characteristics will continue to represent their variations in the near future. An examination of the models, in terms of their parametric values and also in terms of their goodness of fit makes it quite clear that for Canada, the accuracy of fertility projections depends more on the accuracy of projecting the total fertility rate than on the relative distribution of these rates by age.

There are a number of approaches that could be followed in projecting total fertility rates. A cohort-orientated approach involving a three-step operation is being considered for use in the population projections in Canada and may be described briefly as follows.

The first step in this approach consists in projecting, for each successive generation of women, completed family size; that is, number of children a woman will achieve upon completion of her childbearing life. For women, who at the beginning of the projection, have already reached a sufficiently advanced stage of family formation, completed family size can be projected without a great risk of error by straightforward graphical extrapolation of the curve of cumulative fertility to date. Acceptable results can also be obtained through the fitting of a Gompertz curve (A. Romaniuk and S. Tanny, 1969). The real challenge for the forecaster is presented by the cohorts of women who have not yet entered childbearing but who only in a few years will form the majority of childbearers. Among various indicators, parity distribution, that is, completed fertility by birth order, offers the most effective basis for forecasting the movement of family size for future mothers. When plotted on the graph, series of completed fertility by birth order offers a clear-cut time perspective that is not apparent in the series of general fertility. The projection basis can further be enhanced by examining specific factors that determine the behaviour of fertility for each birth order. Variables such as age at marriage and proportion of never married as well as childlessness are clearly related to the fertility of the first birth order. The total number of children per woman is then obtained by adding completed fertility rates for each birth order.

As a second step, after future family size has been projected, changes in fertility age pattern should be considered in order to project mean age of fertility. It should be noticed in passing that, although shifts in timing of fertility appear on examination of past trends to be less violent than shifts in family size, their impact on total fertility rate is nonetheless appreciable and, therefore, should receive close scrutiny. It is well documented that variations in the total fertility rate (post-war baby boom and very likely also the recent dramatic fall in fertility) is the result of the combined effect of shifts both in the family size and in the timing of childbearing. Age at first marriage, child-spacing and parity distribution are of prime interest and anticipation of their future trends will help to formulate assumptions with regard to the mean age of fertility.

The final step in this procedure involves derivation of period total fertility rate for future years from the two cohort measures mentioned above -- family size and mean age of fertility. A cohort-to-period translation model of the Ryder type (1969) could be used to perform this operation. The latter model could be improved by a built-in factor that would make allowances for the fact that changes in the timing of fertility are not necessarily linear as implied in the Ryder translation model but might vary in magnitude from one cohort to another.

After the period total fertility rate was obtained all that remains to be done is to distribute it by age of mothers. It is at this stage that the model described in this paper is expected to render most valuable service to the forecaster. Indeed total fertility can be distributed by age by means of formula (1) from a few selected parameters, and the forecaster needs no longer to go through lengthy procedures of projecting fertility rates for individual ages.

In order to generate fertility distribution by formula (1), one has the alternative of projecting dependent parameters a_1, a_2, m_1, m_2, or of projecting independent parameters -- , moments and other measures associated with frequency distribution. Which one of these approaches shall be used is a matter of personal judgement. Maybe measures such as mean age or mode age are more meaningful concepts than the constants $\underline{a}$ and $\underline{m}$ and the rationales for their future movement can be more easily substantiated (Stone, 1970).

Selection of the independent parameters to be projected depends upon which particular procedure, among the three developed in this paper, will be used for calculating the constants implied in formula (1). The best results in terms of goodness of fit are obtained by the procedure involving four distribution moments (4M). But if only an approximate fit is sufficient for the purpose of projecting the births, as this is actually the case (see next paragraph), then the practical considerations, such as the easiness of projecting future movement, become overriding in the selection of specific independent parameters to be projected. In this respect procedure (1M) may be preferred. Instead of four moments, it involves only two relatively simple measures, mean and mode age of fertility. Once the mean age is projected mode age can easily be derived because there is a very high positive correlation between these two measures. Fertility age interval, which in this procedure is assumed to be a fixed one, can be set up in the light of some testing against historical series. It has been shown in this paper that in the case of Canada, the model performs better with 17 than with 15 as the starting age of childbearing.

In the making of population projections by age by means of the component method, the number of annual births for future years is an important input. In order to demonstrate how well our model performs in terms of generating annual number of births, Table 6 below has been prepared. It contains, for few selected years for Canada, ratios of estimated to actual number of births, whereby estimated numbers are obtained by multiplying the number of females in ages 15 to 49 by the age-specific fertility rates derived by means of three variants of calculating parameters in formula (1). One is impressed by the close agreement between actual and estimated figures. Deviation is negligible except in one case where it comes near to two per cent. It is true, this high agreement is achieved at aggregate level only, partly through compensation of errors at individual ages. Yet it should be borne in mind that what is required for population projections, as input, are not births at individual ages, but the total number of births, and in this respect our model performs almost to perfection.

Table 6 RATIO OF ESTIMATED TO ACTUAL NUMBERS OF ANNUAL BIRTHS FOR A FEW SELECTED YEARS FOR CANADA

Year	4M	2M (15-50)	2M (17-50)	1M (15-50)	1M (17-50)
1926	1.001	.986	.999	1.005	1.000
1931	1.001	1.001	1.000	1.010	.998
1941	.998	.997	.998	1.000	.996
1951	.998	.997	.997	.984	.996
1961	1.003	1.005	1.001	1.013	1.002
1969	1.008	1.006	1.011	1.018	1.009

CONCLUSION

Additional research might be needed before a parametric model of fertility projections of the type outlined here can be made fully operational. This paper contains, nevertheless, important ingredients for such a model. It is shown that fertility distribution by age can be derived by mathematical function from only a limited number of parameters that need to be projected. The paper offers alternative ways of calculating these parameters, it tests their results against fertility data for Canada and points to potentials for further developments in the areas of fertility projections. Reduction of age schedule of fertility to only a few meaningful parameters makes possible an in-depth analysis to a degree that cannot be achieved under conditions of conventional procedures used in fertility projection. Furthermore, the model is a powerful labour-saving device, for in defining fertility in terms of a mathematical function, computers can be utilized in performing many of the involved operations.

REFERENCES

Elderton, W.P., Frequency Curves and Correlation, Cambridge University Press, 1930.

Keyfitz, N., Introduction to the Mathematics of Population, Reading Massachusetts, Adison-Wesley, 1968.

Mitra, S., The Pattern of Age-Specific Fertility Rates, Demography, Vol. 4, No. 2, pp. 894-906 1967

Murphy, E.M. and Nagnur D.N. "A Gompertz That Fits: Applications to Canadian Fertility Patterns", Demography, Vol. 9, No. 1, 1972, pp. 35-50.

Romaniuk, A., and Tanny, S.M., 'Projection of Incomplete Cohort Fertility for Canada by means of the Gompertz Function', Analytical and Technical Memorandum, No. 1, Dominion Bureau of Statistics, Census Division, Canada, Ottawa, March, 1969.

Ryder, N.B., "The Emergency of a Modern Fertility Pattern; United States, 1917-1966" in Fertility and Family Planning, S.J. Behrman et al (ed.). Ann Arbor, Michigan: The University of Michigan Press, 1969.

Stone, L.O. "Parametric Approaches of the Age Distribution of Cohort Total Fertility for Projections", Unpublished Manuscript, Dominion Bureau of Statistics, Ottawa, July, 1970.