AN APPRAISAL OF THE MODEL TABLES OF VITAL RATES

S. Mitra, Emory University

1. The Models

The search for parametric models that can adequately describe a particular series of age-specific vital rates began in the Population Branch of the United Nations in the early 50's which resulted in the publication of a set of model life tables suitable particularly for underdeveloped countries (U.N., 1955). From 158 life tables of various countries and periods, the UN analysts noted that a second degree polynomial in ${}_nq_x$ provides excellent approximation of ${}_nq_{x+n}$ where the parameters defining the polynomial can be obtained by the method of least squares. This provided a tool to build up the entire ${}_nq_x$ series from an initial value of ${}_1q_0$ and the successive polynomial approximations. The estimated values of ${}_nq_x$ were then used to generate values of all life table functions.

Coale and Demeny created four categories of life tables (1966), on the basis of greater homogeneity in the patterns of mortality within each category that produced high correlations between successive ${}_nq_x$ values. For these categories labelled North, South, East and West, linear regressions of ${}_nq_x$ and log ${}_nq_x$ on $\overset{0}{e}_{10}$ were obtained by the method of least squares and both of these approximations were used to develop estimates of ${}_nq_x$ values, and thus the engire life table was obtained from an initial value of $\overset{0}{e}_{10}$.

Similar experimentation with age-specific fertility rates (Mitra, 1965), resulted in a set of model fertility tables. Here also, high correlations were observed among successive age-specific rates, between an age-specific fertility rate and the general fertility rate and also between an age-specific fertility rate and the sex age adjusted birth rate. Later, another series of model fertility tables was developed by using Pearson's Type I curve to describe the age-specific rates (Mitra, 1967). The latter model, unlike the former, depends upon more than one parameter, and it was shown that the number of independent parameters can be reduced from four to two or even to one, under certain reasonable assumptions.

2. The Problems With the Use of Model Tables

Although, the model tables were designed to provide estimates of age-specific vital rates, they were far from being flawless. As Gabriel and Ronen (1958) pointed out that estimates of $\overset{0}{e}_0$ from ${}_1q_0$ by graphic interpolation among the life tables used by the UN show that the model life tables tend to overestimate life expectancy (by an average of 2.133 years), or underestimate mortality. Coale and Demeny mentioned the logical problem of using the same ${}_nq_x$ as both the dependent and the independent variable in the regression equations. They also mentioned that the use of any other ${}_nq_x$ as the initial value would have been equally justified but would have produced different estimates of the same life table functions.

Gabriel and Ronen also examined the joint variation of ${}_1q_0$ and $\overset{0}{e}_0$ and experimented with a second degree polynomial in ${}_1q_0$ as an estimator of $\overset{0}{e}_0$ obtained by the method of least squares. A comparison with the model values revealed that the model values were higher as well as lower than their quadratic estimator at higher and lower levels of the continuum of $\overset{0}{e}_0$. However, they did not carry out further their investigation of the discrepancies between model values and their estimates by alternative methods.

If they did, they would have seen that since the model life expectancies increase non-linearly (almost like a quadiatic), with decline in ${}_1q_0$, a property also shared by their quadratic approximations, such two curves may not intersect at more than one point. In that case, the difference between the two estimates will change its sign only once as ${}_1q_0$ moves from one to the other end of the continuum. Their first observation about consistent upward bias in the model values, also seem to have a logical explanation that may follow from a simple illustration. Consider l_5 and l_{10} of the life table expressed as (for $l_0 = 1$)

$$l_5 = (1-{}_1q_0)(1-{}_4q_1)$$

$$l_{10} = (1-{}_1q_0)(1-{}_4q_1)(1-{}_5q_5) \qquad (1)$$

For simplicity, assume that ${}_4q_1$ is estimated from ${}_1q_0$ by a linear regression, and ${}_5q_5$ from that of ${}_4q_1$. Using the superscript ~ to denote estimates,

$$\tilde{{}_5L_5} = \tfrac{5}{2}(\tilde{l}_5+\tilde{l}_{10}) =$$

$$\tfrac{5}{2}(1-{}_1q_0)[(1-{}_4\tilde{q}_1) + (1-{}_4\tilde{q}_1)(1-{}_5\tilde{q}_5)] \qquad (2)$$

where ${}_1q_0$ is the initial value and therefore not an estimate, so that the expected value

$$E({}_5\tilde{L}_5) = \frac{5}{2}(1-{}_1q_0)[E(1-{}_4\tilde{q}_1) + E\{(1-{}_4\tilde{q}_1)(1-{}_5\tilde{q}_5)\}] \quad (3)$$

Note that ${}_5q_5$ is dependent upon ${}_4q_1$ and therefore the expected value

$$E\{(1-{}_4\tilde{q}_1)(1-{}_5\tilde{q}_5)\} = E(1-{}_4\tilde{q}_1)E(1-{}_5\tilde{q}_5) + r\,\sigma_1\sigma_5 \quad (4)$$

Where σ_1, σ_5 and r stand respectively for the standard deviations of ${}_4q_1$ and ${}_5q_5$ and the correlation coefficient between them. Thus (4) can be written as

$$(1-{}_4q_1)(1-{}_5q_5) + r\,\sigma_1\sigma_5,$$

and after substitution in (3),

$$E({}_5\tilde{L}_5) = \frac{5}{2}(l_5 + l_{10} + r\sigma_1\sigma_5) \quad (5)$$

which for r>0 has an upward bias. Thus, $\overset{0}{e}_0$ which is obtained by summing the L_x values will have a tendency to overestimate the actual life expectancy. While such was not the case when the model life expectancies were compared with those derived from the life tables developed by Gabriel and Ronen, the explanation lies in the fact that their life tables were different from the UN model life tables in that the former were derived by assuming a linear relationship between consecutive ${}_nq_x$ values, whereas the estimation formula was assumed to be quadratic in the latter.

There is at least one other fact that has so far remained in the dark and this is concerned with the efficiency, or lack of it, of the methods used in the construction of UN model life tables as well as Mitra's model fertility tables. While it may be true that the correlation between ${}_nq_{x+n}$ and ${}_nq_x$ is higher than that between ${}_nq_{x+n}$ and other q values preceding ${}_nq_x$, the efficiency of the estimate of ${}_nq_{x+n}$ from ${}_nq_x$ is questionable when the latter, in turn, is estimated from the q value preceding it and so on. As an example, consider the following where, for reasons of simplicity, the relationship between the q values, has been assumed as linear. Thus,

$${}_4q_1 = a_1 + b_1({}_1\tilde{q}_0) \quad (6)$$

$${}_5q_5 = a_5 + b_5({}_4\tilde{q}_1) \quad (7)$$

where the variables on the left hand side of the equation are the expected values and those with ~ as superscript indicate the observed values. If ${}_5q_5$ is estimated from ${}_4q_1$ which, in turn, is estimated from the initial value of ${}_1q_0$, then (7) can be written as

$${}_5\hat{q}_5 = a_5 + b_5a_1 + b_5b_1({}_1\tilde{q}_0) \quad (8)$$

It is easy to see that (7) and (8) will produce identical results for all values of ${}_4q_1$ only when the ${}_1q_0$ and ${}_4q_1$ are perfectly correlated. Further, (8) expresses ${}_5q_5$ as a linear function of ${}_1q_0$ which will generate best estimate of the former only when the expression coincides with the linear regression of ${}_5q_5$ on ${}_1q_0$ obtained by the method of least squares. The residual or the error sum of squares in the latter case is proportional to

$$1-r^2_{05} \quad (9)$$

and for (8), which of course is an unbiased estimator, the residual sum of squares is

$$(1-r^2_{05}) + (r_{05}-r_{01}r_{15})^2 \quad (10)$$

where

$$r_{01} = \text{corr}\ ({}_1q_0 \text{ and } {}_4q_1)$$

$$r_{05} = \text{corr}\ ({}_1q_0 \text{ and } {}_5q_5)$$

$$r_{15} = \text{corr}\ ({}_4q_1 \text{ and } {}_5q_5)$$

Thus, unless $r_{05} = r_{01}r_{15}$, (8) will be an inefficient estimator of ${}_5q_5$ than the one provided by the straightforward regression of ${}_5q_5$ on ${}_1q_0$. From these results, it is quite apparent, that the consistent use of the initial parameter to estimate the remaining parameters of the life table is still the best, no matter how low the correlation gets with increase in the age difference and no matter how high the correlation is between two consecutive parameters. This provides the statistical justification of the life tables constructed by Gabriel and Ronen as well as those by Coale and Demeny.

However, Coale and Demeny's choice of $\overset{0}{e}_{10}$ as the initial parameter has more than one oddity as they themselves have acknowledged. But they do not seem to realize that the treatment of ${}_nq_x$ or log ${}_nq_x$ as dependent upon $\overset{0}{e}_{10}$, is a statistical manipulation that cannot be justified on logical grounds. In fact, the use of ${}_1q_0$, but not any ${}_nq_x$, to estimate the rest, seems to be perfectly logical, at least in a generational life table, for which ${}_1q_0$ precedes in time, and therefore, can be regarded as a determiner of the remaining ${}_nq_x$ values. From that point of view, any ${}_nq_x$ can be regarded as dependent upon any, a few, or all of the preceding ${}_nq_x$'s and not vice versa.

It can similarly be argued that since $\overset{0}{e}_{10}$, even in a cross-sectional life table, can be obtained only when all ${}_nq_x$ values, beginning at x equal to 10 are available, any definition of the latter as dependent upon the former suffers from a logical absurdity. Further, Coale and Demeny's use of the regression of ${}_nq_x$ on $\overset{0}{e}_{10}$ at one end, that of log ${}_nq_x$ on $\overset{0}{e}_{10}$ at the other, and the average of the two regressions in the middle of the $\overset{0}{e}_{0}$ continuum does not seem to have been based on sound logical or technical principles. Such a solution to the problem of achieving better fit is neither unique, nor does it follow from a theoretical model. One could as easily divide the range of $\overset{0}{e}_{10}$ into three or four sections and fit a linear regression for each of these, as for these sub-sections, the correlations would definitely turn out to be even higher to justify such a procedure. There is one other point that should also be mentioned in this context. The authors do not seem to be concerned when the ${}_nq_x$ values so obtained could not reproduce the initial $\overset{0}{e}_{10}$ that was the principal, if not the sole pivot of the entire structure. On the other hand, they are satisfied with the observation that the linear regression of calculated $\overset{0}{e}_{10}$ on ${}_nq_x$ fits the observed ${}_nq_x$'s even better.

These authors have omitted from their study, an analysis of the intercorrelations of ${}_nq_x$'s for the four regional categories, although they have done so for the log $({}_nq_x)$'s. The reason for such omission is not clear, but it seems that those intercorrelations, if obtained, would have been higher than the corresponding values of the regions combined into one group, and would have justified the development of regional tables following the method outlined by Gabriel and Ronen.

3. Mathematical Models

The controversies about the dependence independence relationship, or about the selection of the appropriate regression model are somewhat unavoidable as long as attempts to settle the issues revolve only around statistical relationships. The soluntion of this problem seems to rest, therefore, on the development of mathematical models that can describe the vital rates or functions thereof with reasonable accuracy, and when the formulation of such models is based on the observable characteristics of these variables. Such models were proposed in the past by Gompertz (1825), modified later by Makeham (1860) and Perks (1932) in the area of mortality, as well as by Wicksell (1931) and UN (1963) for the graduation of age-specific fertility rates. Needless to say that these models were proven to be somewhat less than adequate and others were proposed for the graduation of life table functions (Mitra, 1965) and of age-specific fertility rates (Mitra, 1967). The latter model, tried specifically on the Canadian data was found to be quite satisfactory (Mitra and Romaniuk, 1973). All of these models are expressions of the functions of vital rates as functions of the most important variable in the area of vital rates, namely, age.

For the fertility rates, it was observed that their pattern of variation with age resembles the mathematical form of that of Pearson's Type I, and under certain plausible assumptions, the number of unknown parameters of this function can be reduced to two or even to one which renders the model very useful in cases where such data are lacking. Mortality rates or other life table functions derived therefrom are somewhat difficult to graduate, however, a polynomial approximation of the logarithm of life expectancy was found to be quite adequate. In fact, a quadratic function of age, determined by an initial value of $\overset{0}{e}_{0}$, and an independent estimate of the maximum value of the life expectancy and the corresponding age (usually less than 5) were found to be sufficient for developing the series of $\overset{0}{e}_{x}$ values (Mitra, 1971), and these can then be used to generate the entire life table. Experiments with a few male model life tables (UN, ibid) demonstrated the usefulness of this method. Presumably, the goodness of fit will improve even further, if the life tables are classified into homogeneous groups, as Coale and Demeny did for developing their series of model tables.

Graduation formulas for other life table functions like force of mortality (μ_x) or the number of survivors (l_x) at age x are yet to be discovered. Because of the difficulty of integrating a function that has a polynomial as an exponent, such graduation formulas are not deducible from the polynomial function found suitable for the logarithm of life expectancies. This does not imply that the mathematical relationships among life table functions should be used as criteria for testing the validity of the models, even in cases where such mathematical operations are difficult to perform. Like any other model, the justifications for these models should also be derived primarily from the logical steps that were used for their theoretical formulations, and finally, from the extent to which they demonstrate their ability to reproduce reality.

4. Tests for Critical Values

More important than finding consistencies among the mathematical models of different life table functions are several tests that can be used to determine the validity of a life table. Preliminary tests can be made by comparing the values of a life table function with the general pattern that is quite well known. Not so well known are some properties possessed by the life table functions, e.g., the expectation of life attains a maximum value at an age (usually before 5) where its reciprocal is also equal to the force of mortality (Mitra, 1971). In general, this age is less than the age at which the force of mortality assumes its minimum value which in turn is less than the age at which the life table survivorship function has a point of inflection, or where its first derivative assumes a maximum or a negatively minimum value (Mitra, 1973). The significance of these three optimum values of the life table functions and their interrelationships need hardly be stressed in judging the adequacy of model life tables.

As is customary in such cases, estimates of life table functions obtained either by correlations or from mathematical models will be subject to sampling errors. For the mathematical models, the parameters will have to be estimated from an available set of life tables, usually by the method of least squares, and as a result, the sampling error of the estimate of any life table function will be determined by the sampling errors of these parametric estimates. Derivations of these estimates will, however, be approximate in most cases, as demonstrated in the following example where the mathematical model is given by

$$\log_e \left(\mathring{e}_x\right) = a+bx+cx^2 \qquad (11)$$

When a, b and c are determined by the method of least squares, the variances of $\log_e\left(\mathring{e}_x\right)$ can be obtained for each x, and from the approximate relation that

$$V\{f(x)\} \approx V(x)\{f'(x)_{x=E(x)}\}^2 \qquad (12)$$

$$V\left(\mathring{e}_x\right) \approx \left(\mathring{e}_x\right)^2 \, V\left(\log_e \mathring{e}_x\right) \qquad (13)$$

substituting observed $\mathring{e}_x$ for the expected value. Similarly,

$$V\left(\frac{1}{\mathring{e}_x}\right) = V\left(\mathring{e}_x\right) \Big/ \left(\mathring{e}_x\right)^4 \qquad (14)$$

Again, $\log_e T_x = - \int \frac{dx}{\mathring{e}_x}$ (15)

Therefore,

$$V(\log_e T_x) = V(\mathring{e}_x) \left[\frac{d}{d\mathring{e}_x} \int \frac{-dx}{\mathring{e}_x}\right]^2 \qquad (16)$$

Now

$$\frac{d}{d\mathring{e}_x} \int \frac{-dx}{\mathring{e}_x} = \frac{d}{dx} \int \frac{-dx}{\mathring{e}_x} \cdot \frac{dx}{d\mathring{e}_x}$$

$$= - \frac{1}{\mathring{e}_x} \Big/ \frac{d\mathring{e}_x}{dx} = - \frac{1}{\mathring{e}_x(\mathring{e}_x\mu_x - 1)} \qquad (17)$$

Where $\mu_x = - \frac{1}{l_x} \frac{dl_x}{dx}$ is the force of mortality.

Substituting in (16)

$$V(\log_e T_x) \approx \frac{V(\mathring{e}_x)}{(\mathring{e}_x)^2(\mathring{e}_x\mu_x-1)^2} \qquad (18)$$

and

$$V(T_x) \approx T_x^{\,2} V(\log_e T_x) \qquad (19)$$

$V(l_x)$ can be obtained from the relationship

$$V(T_x) = l_x^{\,2} V(\mathring{e}_x) + (\mathring{e}_x)^2 V(l_x) \qquad (20)$$

by noting that in the difference equation

$$\Delta T_x = \mathring{e}_x \Delta l_x + l_x \Delta \mathring{e}_x \qquad (21)$$

l_x and $\mathring{e}_x$ can be regarded as independent. These and other results similarly obtained can be used to test the efficiency of the estimates of life table functions.

5. Summary

An analysis of the methods used for the development of model life tables by the United Nations, regional model life tables by Coale and Demeny and model fertility tables by Mitra revealed certain shortcomings and scopes for improvement. The UN model tables were found to be biased and lacking in efficiency by Gabriel and Ronen who had suggested certain modifications. The reasons for some of their deficiencies (also present in the model fertility tables), not elaborated by Gabriel and Ronen have been developed in this paper. Some of the problems associated with the construction of the regional model life tables have also been outlined, and the advisability of searching for justifiable mathematical models have been proposed. Such models, developed earlier by Mitra in the areas of fertility and mortality were found to describe the general pattern of these series reasonably well. Since the parameters of the model have to be estimated by following statistical procedures, the efficiency of any estimate or a function thereof depends upon the standard errors of the estimates of those parameters. Some of these formulas have been obtained for life table functions derived from a simple mathematical model proposed earlier for the construction of model tables.

References

Coale, A.J. and Demeny, Paul, Regional Model Life Tables and Stable Populations, Princeton University Press, Princeton, N.J., 1966.

Gabriel, K.R. and I. Ronen, "Estimates of Mortality from Infant Mortality Rates", Population Studies, Vol. 12, No. 2. 1958.

United Nations, "Age and Sex Patterns of Mortality, Model Life Tables for Under-Developed Countries", ST/SOA/Series A/22,1955.

Makeham, W.M., "On the Law of Mortality and Construction of Annuity Tables", Journal of the Institute of Actuaries, Vol. 8, 1860

Perks, W., "On Some Experiments in the Graduation of Mortality Statistics", Journal of the Institute of Actuaries, Vol. 63, 1932.

Mitra, S., "Model Fertility Tables", Sankhyā, The Indian Journal of Statistics, Vol. 27, Series B, 1965.

Mitra, S., "Graduation of Life Table Functions", Proceedings of the 1971 Conference of the American Statistical Association. Also, see "On Some Properties of the Expectation of Life $\overset{0}{e}_x$" in Contributions by Indian Authors to Second World Population Conference, Government of India Publications, 1965.

Mitra, S., "The Pattern of Age-Specific Fertility Rates", Demography, Vol. 4, No. 2, 1968.

Mitra, S., and Romaniuk, A., "Pearsonian Type I Curve and Its Fertility Projection Potentials", Demography, Vol. 10, No. 3, 1973.