SAMIR GUHA ROY

DEVELOPING MODEL SCHEDULES OF FEMALE PROPORTIONS SINGLE FOR INDIA (*)

Proportion single is an important indicator of the extent of first marriages in a population. These proportions at the younger ages indicate the tempo of recent nuptiality, and those still single at older ages show the prevalence of lifelong non-marriage (1). The Indian censuses provide marital status data cross-classified by sex and age. But the data are often defective and may be quite seriously distorted. In some cases the ratios of proportions single in successive age groups in two consecutive censuses exceed unity because of faulty data (or selective mortality or migration). This happens, for example, between age groups 25-29 in 1941 and 35-39 in 1951, and similarly between 30-34 and 40-44, 35-39 and 45-49, and 40-44 and 50-54. Bias in age reporting may affect the proportions single at different ages. The observed distributions of proportions single are subject to fluctuations that may conceal the true nature of the phenomenon to a large extent. In the present study, we try to smooth out the inconsistent data as much as the proposed logit linear model would permit.

METHODOLOGY

The logit of the proportion p $(0 < p < 1)$ is defined as logit $p = 0.5 \log_e p/(1-p)$. (In demographic applications the factor 0.5 is customary, but it is usually not used by statisticians). Our interest focuses on the proportions single, $S(a)$, at age a so that p is replaced by $S(a)$ in this analysis. The graph of logit $S(a)$ against a, though more linear than the graph of $S(a)$ against a, is still not linear enough. So we use the following "relational" form equation instead. When several schedules of proportions single can be related linearly to the same function $S_s(a)$, this common function may be regarded as a standard schedule. Proportion single schedules have

---

(*) I am very grateful to Robert Retherford and Peter Smith, and the other scientists of the East-West Center, Hawaii, for their critical appraisal of this study. I had useful discussions on the paper with M.V. Raman and C.R. Malaker of the Indian Statistical Institute. I also appreciate the assistance received from Prasanta Kumar Swar in typing work.

(1) P.C. Smith, *Indexes of nuptiality*, Asian and Pacific Census Forum, 5(2), 1978.

been recorded in many countries for a number of censuses, making it possible to search for underlying common pattern. We can thus find $S_s(a)$ such that

$$logit\ S(a) = \alpha + \beta\ logit\ S_s(a) + e$$

or

$$P(a) = \alpha + \beta P_s(a) + e$$

where $P(a) = logit\ S(a)$, $P_s(a) = logit\ S_s(a)$ and e is the error term, the amount by which any individual $P(a)$ may fall off the line implied by the model. Now the linear fit is better. Then α and β give a two-parameter summary of the shape of the $S(a)$ curve. Subject to an appropriate choice of $S_s(a)$, all of the temporal variation may be described by the trajectory of the parameters α and β. By analogy with Brass relational mortality model we see that α varies directly with singulate mean age at marriage (SMAM); β varies inversely with the dispersion around the SMAM.

The principle underlying the logit system originates from bio-assay, and has been used and described extensively for mortality and postpartum variables (2). Only a brief description of the analogous system is therefore in order. At a given age a, each woman may be single or not. The proportion of women remaining in the original state was designated as $S(a)$ for each age a. The shape of $S(a)$ curve may often be described by a function similar to a logistic (3). The derivation of the logit transformation from the underlying logistic distribution has been made elsewhere (4). This transformation has the useful property of preserving (when transformed back) the end points 0 and 1, and is therefore applicable to distributions with these end points (5). It should, however, be noted that the logit fit may not be very good if nuptiality is changing rapidly over time. For example, if age at marriage oscillates, as it did during World War 2 in many countries, the cross-sectional proportions single by age could show some strange patterns.

We considered least squares procedures adequate for our purpose of estimating the parameters in the above model. Since the tails of the distribution are sometimes prone to reporting errors, a more robust procedure of estimation (6), considering only the central portion of the $P(a)$ schedule in the age range 20 to 44, was also used in a few cases on a trial basis. This involves splitting suitably the observations into two portions, A and B. Designating the mean values for each portion

---

(2) W. Brass, *On the scale of mortality*, in W. Brass et al., *Biological Aspects of Demography* (London: Taylor and Francis, 1971); W. Brass, *Methods for Estimating Fertility and Mortality from Limited and Defective Data* (Chapel Hill: Poblabs, University of North Carolina, 1975); K. Hill and T.J. Trussell, *Further developments in indirect mortality estimation*, Population Studies, 31(2), 1977, pp. 313-334; and R.J. Lesthaeghe and H. J. Page, *The post-partum non-susceptible period: Development and application of model schedules*, Population Studies, 34(1), 1980, pp. 143-169.

(3) H. Hyrenius et al., *Demographic Models - DM3* (Sweden: Demographic Institute, University of Goteborg, 1967).

(4) K. Hill and T.J. Trussell, *loc.cit.*, in footnote 2.

(5) W. Brass, *op.cit.*, in footnote 2 and K.Hill and T.J. Trussell, *loc.cit.*, in footnote 2.

(6) W. Brass, *op.cit.*, *in footnote 2 and R.J. Lesthaeghe and H.J. Page*, *loc.cit.*, in footnote 2.

as $\overline{P}_A$ and $\overline{P}_B$ for the observed schedule and $\overline{P}_{S,A}$ and $\overline{P}_{S,B}$ for the standard schedule, the estimates for β and α are as follows:

$$\beta = (\overline{P}_B - \overline{P}_A)/(\overline{P}_{S,B} - \overline{P}_{S,A})$$

$$\alpha = \overline{P}_A - \beta\overline{P}_{S,A}$$

Since the estimates by the two methods did not differ much ultimately, we persisted with the least squares fitting procedure applied to observed data for age range excluding 0-4, 5-9 and 50+ . Yet another alternative procedure is to use median regressions (7). But the number of data points, as given by the age group for which $P(a)$ is reported, being small, this method is difficult to apply.

An approach similar to the one adopted for postpartum variables by Lesthaeghe and Page (8) in the development of a standard schedule may be adopted. The procedure is as follows: an observed schedule of proportions single for a population is first selected. The data are then smoothed by three-point moving averages, and the resulting schedule is converted to logits. Linear relations emerge when logits of other schedules are plotted against this preliminary standard, which is further smoothed graphically to get a standard without irregularities. Since the relational form equation used is not, in practice, exact, the estimates are likely to be affected by the choice of standard. Recently, Thapa and Retherford (9) had shown divergence of indirect estimates (by Brass method) of infant mortality based on different standard life tables. In the case of nuptiality, it may also be useful to try at least two standards instead of just one to see if the results came out about the same.

APPLICATION TO INDIAN NUPTIALITY

The general framework for marriage in India being characterised by early and almost universal marriage with very low celibacy rates, $S(a)$ attain unity near age 10, and 0 around age 50. If we include those Indian "marriages" that are formal but unconsummated child marriages then the assumption of $S(a) = 1$ around $a = 10$ is possibly not realistic. In our formulation we therefore consider the transition from single state to only effective (that is, women attaining reproductive age) marriage state.

In the present analysis, we considered eight observed schedules for the period 1901-71. More or less uniform definitions have been adopted in the classifi-

(7) G.W. Barclay et al., *A reassessment of the demography of traditional China*, Population Index, 42 (4), 1976.

(8) R.J. Lesthaeghe and H.J. Page, *loc.cit.*, in footnote 2.

(9) S. Thapa and R.D. Retherford, *Infant mortality estimates based on the 1976 Nepal Fertility Survey*, Population Studies, 36 (1), 1982, pp. 61-80.

cation of population by marital status in different censuses of India (10). This ensures the comparability of data, and thus provides an analytic basis for study of nuptiality trend. However, only smoothed data for 1931 census were available and Agarwala (11) noted a bias in this census caused by smoothing. We had not explicitly used this schedule for building up a standard, but the result derived from the logit relational procedure for 1931 should be interpreted with caution.

We had chosen two standard schedules reflecting Indian nuptiality patterns rather than relying on any "universal" standard. One (designated as Standard I) is based on the Census 1961, as it is thought to be better organized and nuptiality pattern, as broadly indicated by mean age at marriage, did not change very significantly (as in 1971) at this point. The other (designated as Standard II) is taken as a sort of average of 1901-71 reported schedules to eliminate the peculiarities of the individual years (12). The standard schedules (Table 1), derived from the logit system, describe the essence of observed ones (Table 2).

Preliminary examination reveals the existence of fairly strong linear relations between the observed and the standard. Even for earlier periods in which data irregularities are more likely, these relations hold quite fairly. Figure 1 illustrates the relations between the logits of a pair of observed schedules at an early (1901) and a recent (1971) census dates and that of a standard schedule, and compares them with the lines represented by the model derived later. This figure is complemented with the table 3 that gives differences between observed and fitted proportions single.

An examination of the deviations of observed from fitted provides a means of testing for the presence of errors in the original data or in the fit. In absolute terms, the deviations are not large, and show some systematic patterns. Larger deviations occur almost uniformly in the younger age range (around 20) than above it. Substantial misreporting in the population age distributions have been a common feature in India (13), and are likely to affect the distribution by age and marital status. The common tendency to understate age by single, but marriageable females, probably account for a large proportion of the deviations in the younger age range.

The mean square error of the regression is computed (Table 3) as a measure of goodness of fit of the model to the reported values. Since the statistic does not attain a value of zero, the fit is obviously not perfect. In the absence of an objective criterion, it is difficult to infer on the nature of the fit, but following

---

(10) S.C. Shrivastava, *Census Centenary Monograph No. 1*, 1971 Census (New Delhi: Monograph Series, Indian Census in perspective, 1971).

(11) S.N. Agarwala, *Age at Marriage in India* (Allahabad: Kitab Mahal, 1962).

(12) M.A. Stoto, *Advances in mathematical models for population projections*, Presented at the International Population Conference (IUSSP), Manila, December, 1981.

(13) United Nations, *Methods of Estimating Basic Demographic Measures from Incomplete Data: Manual on Methods of Estimating Population*, Manual 4, Population Studies No. 42 (New York: United Nations, 1967).

TABLE 1

Standard schedules of proportions single

Age (a)	$S_s(a)$ Standard I	Standard II
<10	1.000	1.000
10 – 14	0.695	0.633
15 – 19	0.395	0.315
20 – 24	0.110	0.096
25 – 29	0.013	0.028
30 – 34	0.010	0.016
35 – 39	0.008	0.012
40 – 44	0.006	0.010
45 – 49	0.005	0.009

TABLE 2

Observed schedule of proportions single

Age (a)	1901	1911	1921	1931	1941	1951	1961	1971
10 – 14	0.543	0.555	0.601	0.493	0.755	0.827	0.805	0.881
15 – 19	0.156	0.163	0.188	0.166	0.252	0.280	0.292	0.429
20 – 24	0.042	0.044	0.052	0.042	0.041	0.065	0.060	0.091
25 – 29	0.025	0.022	0.025	0.018	0.014	0.042	0.019	0.020
30 – 34	0.020	0.017	0.019	0.013	0.009	0.018	0.010	0.009
35 – 39	0.018	0.014	0.015	0.011	0.009	0.017	0.007	0.005
40 – 44	0.013	0.013	0.014	0.009	0.009	0.015	0.006	0.005
45 – 49	0.011	0.011	0.013	0.008	0.009	0.011	0.005	0.004

Source: Official census reports.

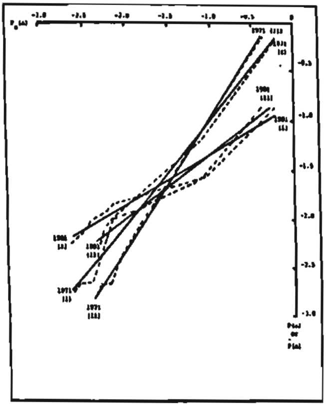
TABLE 3

Age-specific differences between observed and fitted proportions single

Age (a)	Differences between percentages single, S(a) – Ŝ(a) 1901	1911	1921	1931	1941	1951	1961	1971
				Standard I				
15 – 19	2.3	2.5	– 0.5	2.2	– 6.9	3.6	2.4	1.2
20 – 24	– 1.9	– 1.7	– 1.8	– 1.4	– 5.4	– 3.2	– 1.9	– 1.8
25 – 29	0.4	0.3	0.3	0.3	0.0	1.8	0.7	0.8
30 – 34	0.2	0.1	0.1	0.0	– 0.2	– 0.2	0.0	0.0
35 – 39	0.2	– 0.1	– 0.1	0.0	0.0	– 0.1	– 0.1	– 0.2
40 – 44	– 0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0
45 – 49	– 0.2	0.0	0.0	0.0	0.3	– 0.2	0.0	0.0
M.S.E. (*)	0.011	0.006	0.006	0.005	0.044	0.020	0.012	0.016
				Standard II				
15 – 19	2.1	2.3	2.4	2.0	5.8	2.7	1.7	0.1
20 – 24	– 1.5	– 1.2	– 1.3	– 1.0	– 2.7	– 2.5	– 1.1	– 0.5
25 – 29	– 0.1	– 0.2	– 0.2	– 0.1	– 1.0	1.0	0.1	0.2
30 – 34	0.2	0.1	0.1	0.0	– 0.6	– 0.2	0.0	0.1
35 – 39	0.3	0.0	0.0	0.1	– 0.4	0.1	0.0	– 0.1
40 – 44	0.0	0.1	0.1	0.0	– 0.2	0.2	0.0	0.1
45 – 49	– 0.2	0.0	0.0	0.0	– 0.1	– 0.2	0.0	0.0
M.S.E.(*)	0.007	0.005	0.003	0.003	0.020	0.010	0.002	0.002

(*) Mean Square Error = $\sum_{a=1}^{n} (\text{logit } S(a) - \text{logit } \hat{S}(a))^2 / n$, where age agroup index *(a)* 1 stands for 15-19, 2 for 20-24, etc

FIG. 1 - *Proportions single schedules after conversion to logits: relationships between (1) observed and Standards I and II (dashed lines) and (2) model and Standards I and II (solid lines)*



Coale (14) we consider a value of 0.005 as a "mediocre" fit. On this basis, the Standard II schedule, as expected, shows a far closer fit than is the case with Standard I. Even for the former, 1941 and, to a lesser extent, 1951 schedules exhibit a somewhat poor fit.

To understand this anomaly we should consider the disruptive events during these periods. The Second World War, 1939-45, interfered most with the 1941 census operation. The age composition of certain States in India at this census was, in addition, seriously distorted because of the major communities making a determined move to inflate their respective numbers at the enumeration. Moreover, the war had profound effects on the economy and on the way of life of the population (15). Coupled with the possible secondary repercussion of war, the partition in 1947, resulting in communal riots and the movements of millions of refugees, had influenced demographic events subsequently. Because of the war, 1941 may thus have been an atypical year for nuptiality (and for that matter for other demographic phenomena). Also, the logit model cannot possibly remove the

(14) A.J. Coale, *Finding the two parameters that specify a model schedule of marital fertility*, Population Index, 44 (2), 1978, pp. 203-207.

(15) P.B. Gupta, *Demographic Report of West Bengal 1901-61* (Calcutta: Demography Research Unit, Indian Statistical Institute, 1969).

major deliberate errors in the data completely. This is perhaps a part of the reason why the logit fit does not work too well for 1941.

While the fit of the logit model seems to be good to most data on historical series of proportions single in India with almost unchanged marriage custom, it is pertinent to apply the model to a strongly different population. The case of Japan with late marriage pattern is illustrated for the purpose. The standard was chosen, following similar procedure described earlier, from the census series available, and two reported schedules for the years 1920 and 1960 were tested. Since separate points could not be represented on Figure 1, the fit being very close indeed inspite of some changes in nuptiality pattern during 1920-60, the Japanese data are shown only in table 4. It may be worth noting that the age pattern of deviations of observed from fitted values are not similar between the populations of India and Japan, and the deviations are probably more due to errors in the original data than in the fit.

From the above considerations we assumed deviations from the standard as measurements errors, and adjusted the data to generate model schedules as follows. Corresponding to each observed schedule we estimated the best fitting straight line:

$$\hat{P}(a) = \hat{\alpha} + \hat{\beta} P_s(a)$$

where $\hat{P}(a)$ is the estimator of $P(a) = logit\ S(a)$. The model schedules of proportions single $P(a)$, covering a wide range of observed schedules, may be derived from standard ones by varying only two parameters. The estimates of α and β are shown in table 5 for decennial years between 1901 and 1971.

In interpreting the trends in these parameters, it is useful to have an understanding of the nuptiality pattern in the culture. During 1901-71 there has been an increase in proportions single at younger ages and a decrease at older ages, implying rising age at marriage but declining celibacy rates. The upward trends in the parameters α and β appear to correspond to these trends in nuptiality. α's have values less than unity and barring 1971, they are all negative, implying shift of the peaks of the distributions towards younger ages – lowering the average ages at marriage compared to the standard schedules (around 17.6 years). The values of β are all less than unity in the earlier censuses, and tend to decrease the concentrations around the central values. The estimates of these parameters for 1971 notably stand out of the line for the earlier census series. The impact of this departure in raising the age at marriage is likely to be much more telling in the long run.

Table 6 shows the trend in model proportions single. While studying the trends and differentials in nuptiality of the country and their patterns across time, it is important to note that till 1931 no minimum age for marriage was prescribed by law. The Child Marriage Restraint Act or more commonly known as Sarda Act, enacted in April 1930, provided penalties for marriages taken place under 14 years of age of females and under 18 of males. The slight lowering of the mean age at

TABLE 4

Standard, observed and fitted schedules of proportions single: Japan, selected years

Age (a)	$S_s(a)$	1920		1960	
		$S(a)$	$\hat{S}(a)$	$S(a)$	$\hat{S}(a)$
15 – 19	0.966	0.823	0.820	0.986	0.985
20 – 24	0.553	0.314	0.310	0.683	0.708
25 – 29	0.152	0.092	0.096	0.216	0.239
30 – 34	0.057	0.041	0.046	0.094	0.092
35 – 39	0.030	0.027	0.028	0.055	0.047
40 – 44	0.020	0.021	0.021	0.032	0.031
45 – 49	0.015	0.019	0.017	0.021	0.023

Source: Official census reports (cited in Smith, *loc.cit.*, in footnote 1).

TABLE 5

Estimates of α and β across census years for each of the two standard schedules

Year	Standard I		Standard II	
	α	β	α	β
1901	– 0.83053	0.50834	– 0.67750	0.64482
1911	– 0.80213	0.54126	– 0.64229	0.68479
1921	– 0.70445	0.55730	– 0.53854	0.70584
1931	– 0.76047	0.61640	– 0.57840	0.77988
1941	– 0.18488	0.89867	– 0.39630	0.81025
1951	– 0.42715	0.65565	– 0.21420	0.84063
1961	– 0.32102	0.86443	– 0.05710	1.09863
1971	0.05503	1.05290	0.37615	1.33797

TABLE 6

Model schedules of proportions single derived from the standard schedules

Age	1901	1911	1921	1931	1941	1951	1961	1971
				Standard I				
15 – 19	0.133	0.138	0.193	0.144	0.321	0.244	0.268	0.417
20 – 24	0.061	0.061	0.070	0.056	0.095	0.097	0.079	0.109
25 – 29	0.021	0.019	0.022	0.015	0.014	0.024	0.012	0.012
30 – 34	0.018	0.016	0.018	0.013	0.011	0.020	0.010	0.009
35 – 39	0.016	0.015	0.016	0.011	0.009	0.018	0.008	0.007
40 – 44	0.014	0.013	0.014	0.009	0.007	0.015	0.006	0.005
45 – 49	0.013	0.011	0.013	0.008	0.006	0.013	0.005	0.004
				Standard II				
15 – 19	0.135	0.140	0.164	0.146	0.194	0.253	0.275	0.428
20 – 24	0.057	0.056	0.065	0.052	0.068	0.090	0.071	0.096
25 – 29	0.026	0.024	0.027	0.019	0.024	0.032	0.018	0.018
30 – 34	0.018	0.016	0.018	0.013	0.015	0.020	0.010	0.008
35 – 39	0.015	0.014	0.015	0.010	0.013	0.016	0.007	0.006
40 – 44	0.013	0.012	0.013	0.009	0.011	0.013	0.006	0.004
45 – 49	0.013	0.011	0.013	0.008	0.010	0.013	0.005	0.004

marriage to 12.7 in 1931 from 13.7 in 1921 brings out the fact that a large number of child marriages seem to have taken place before the act came into force (16). As a matter of fact, there had not been much time lag between the enactment as well as strict enforcement of the act and the subsequent census, and the people aged below the prescribed minimum apparently went for marriages with more than normal tempo in fear of punishment later. This partly explains that this act, instead of raising proportions single at the younger ages, lowered them initially. The disruptive events that interrupted demographic phenomena around 1941 have already been mentioned earlier. Also, the 1921 estimates are slightly counter to the trend, particularly at ages 15-19. The great influenza epidemic of 1918, believed to have taken a toll of 12 million lives in the country, and many famines prior to 1921 (17) appear to have been reflected in this fluctuation. It is likely that marriages were temporarily postponed for some years during that period, thus causing a bulge in proportions never married.

It is of interest to examine what the proportions single imply in terms of their first differences. The graphs of these differences for the fitted (derived from Standard II) compared with observed (1901 and 1971) are shown in Figure 2. We define the proportions declining between the age groups a-5 and a as

$$D\ (a\text{-}5,\ a) = S\ (a\text{-}5) - S(a)$$

Since everyone does not marry by age 50 (the single condition is assumed definitive by age 50), $D(a\text{-}5,\ a)$ is only approximately equal to the proportion of marriages that occur in each age group. If D is divided by $S(50)$, then this interpretation assumes that the proportions $S(a)$ characterize a synthetic cohort as it moves through age and time. Thus the model works best if nuptiality patterns do not change from one census to the next. If nuptiality changes rapidly, then, as mentioned earlier, there are problems in giving a cohort interpretation to what are essentially cross-sectional data.

In Figure 2 the observed schedules expressed as proportions declining closely follow the same form of distribution as the standard schedules. The distribution has a longer tail on the right implying that the proportions drop off more slowly on the right of the curve than on the left with steep initial rise. In the population with nearly universal marriage, only a few women remain single at older ages. All the schedules give rise to unimodal (and leptokurtic) distributions characterized by rapid transition from single state to married state at the younger ages with a peak at age 15-19. Beyond that point it declines at first rapidly and then slowly.

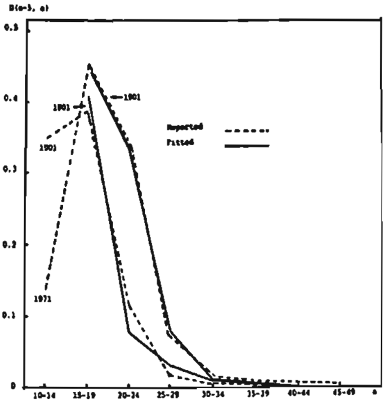
It would be useful to compare the singulate mean ages at marriage (SMAM) computed from the raw data and logit model for each census year. In India, exposure to the risk of conception and child-bearing is confined to period of marriage,

---

(16) Registrar General of India, *The Population of India: 1974 World Population Year*, C.I.C.R.E.D. Series (New Delhi: Ministry of Home Affairs, 1974).

(17) Registrar General of India, *op.cit.*, in footnote 16.

FIG. 2 - *Single proportions declining (D(a-5, a)) by age (a) for selected observed and fitted schedules (derived from standard II)*



and fertility performance below age 15 is assumed insignificant (18). We thus begin at age 15, assuming that no effective marriages occur earlier than age 15. This is, however, not to suggest that formal marriages below age 15 do not occur in India. The SMAMs or more strictly SMAEMs (Singulate Mean Age at Effective Marriage) for one period (19) are found as follows:

	1901	*1911*	*1921*	*1931*	*1941*	*1951*	*1961*	*1971*
Raw SMAM	16.1	16.0	16.2	16.1	16.4	16.9	16.8	17.7
Model SMAM	15.9	16.0	16.3	16.0	16.4	16.7	16.8	17.7

Given the limitations of the reported data, the agreement between observed and model SMAMs can be considered as quite remarkable. It is, however, more appropriate to derive mean age at marriage by calculating proportions remaining single from one decade to the next in order not to assume the census cross section as a cohort. This has been done elsewhere (20), and as stated above, the present purpose is merely to compare census and model estimates.

(18) S. Guha Roy, *Fertility performance and female age at marriage*, The Journal of Family Welfare, 19 (1), 1972, pp. 66-78.

(19) J. Hajnal, *Age at marriage and proportions marrying*, Population Studies, 7 (2), 1953, pp. 111-136.

(20) S.N. Agarwala, *op.cit.*, in footnote 11.

COMPARISON OF THE LOGIT MODEL WITH THE COALE-McNEIL MODEL

Coale and McNeil (21) have developed a closed expression of the first marriage frequency function,

$$g(a) = (0.19465/K)\{exp\ (-0.174/K)(a-a_0-6.06\,K) + $$
$$- exp\ [(-0.2881/K)(a-a_0-6.06\,K)]\}$$

where a_0 is the age at which first marriages begin, and K, the tempo at which first marriages take place. CG(a), the approximate proportion ever married (C being the proportion who will ever marry), can be calculated numerically from $g(a)$. Since a demographic estimate is generally non-unique, a consistent value of $S(a)$ obtained from the relation $S(a) + CG(a) = 1$ may provide an alternative estimate of proportions single. This offers a possible comparison of the logit model with the Coale-McNeil procedure. Considering 1971 data as illustration, the results of comparison are shown in table 7. Following Coale's recommended procedure (22), the parametric values are being estimated as $a_0 = 10.16$ and $K = 0.65$ $(C = 0.995)$, which compare well with the values, $a_0 = 10$ and $K = 0.624$ $(C = 0.996)$, estimated in another study (23).

Upto age 25, the proportions single are corrected in the same direction by both methods. If the data are systematically in error, then may be the Coale-McNeil model actually provides a better fit. That is, the Coale-McNeil model may fit the true proportions single better but the reported proportions single worse than the logit model.

In fitting the Coale-McNeil behavioral model to the empirical data for India one finds difficulty in identifying the stages preceding marriage as conceived in the model. The stages such as entry into the marriage market, meeting a suitable partner, and getting engaged, interpreted in Western cultures, cannot be exactly equated with the stages leading to marriages in the Indian society where marriages are mostly arranged by the parents. However, a very rough correspondence between the two cultures in respect of marriage practices may be assumed. Thus, the sequence of steps antecedent to marriage in India can be stated as the starting of negotiations between the boy's parents and the girl's parents (analogous to entry into the 'state of nubility'), settlement of dowries and other details (similar but not at all identical to the stage between ultimate spouse and engagement in the Western culture), and solemmizing the marriage (corresponding to the stage between engagement and marriage). These stages are not all non-overlapping, especially the first two. It is not surprising that because of the lack of exact behavioral correspondence on the one hand and the irregularities in the basic data on the other, the fit of the

---

(21) A.J. Coale and D. McNeil, *The distribution by age of the frequency of first marriage in a female cohort*, Journal of the American Statistical Association, 67 (340), 1972, pp. 743-749.

(22) A.J. Coale, *Age patterns of marriage*, Population Studies, 25 (2), 1971, pp. 193-214.

(23) P.C. Smith, *loc.cit.*, in footnote 1.

TABLE 7

Reported and estimated (by two procedures) female proportions single in 1971

Age	Census	Logit model	Coale-McNeil model
15 - 19	0.429	0.417	0.4144
20 - 24	0.091	0.109	0.1166
25 - 29	0.020	0.012	0.0308
30 - 34	0.009	0.009	0.0082
35 - 39	0.005	0.007	0.0021
40 - 44	0.005	0.005	0.0005
45- 49	0.004	0.004	0.0001

Coale-McNeil model to the Indian data of 1971 (found to be more so for the earlier censuses) is not good enough. Coale (24) himself warned, even after being successful with French data, that reality is more complex than this model, and the good fit in one instance may not be replicated in the other. Though not wholly incontestable, the logit model allows us to tune the estimates of proportion single to a specific culture subject to a choice of an appropriate standard schedule.

(24) A.J. Coale, *The development of new models of nuptiality and fertility*, Population, 32 (special number), 1977, pp. 131-150.

SUMMARY

The Indian censuses provide marital status data cross classified by sex and age. There are, however, several limitations in the general state of information and in the quality of data on marital status. For instance, the ratios proportions single in corresponding age groups in two consecutive censuses have been found to exceed unity in the absence of any significant migration. In the present study, we try to adjust the schedules of proportions single for 1901-71 by a logit linear model. Inspite of the serious nature of errors in the data, the model appears to work well. In assessing its quality, the technique is applied to more accurate Japanese data, and found to have an extremely good fit. A comparison with Coale-McNeil nuptiality model reveals that for the Indian situation the logit model may be more appropriate.

RIASSUNTO

I censimenti indiani forniscono la distribuzione della popolazione secondo lo stato civile, il sesso e l'età. La qualità dei dati e l'informazione generale, tuttavia, sono spesso limitate. E' stato trovato, per esempio, che il rapporto tra la proporzione di nubili in un certo gruppo di età, calcolato a due censimenti successivi, supera l'unità, nonostante l'assenza di un significativo movimento migratorio.

In questo lavoro si è cercato di correggere le proporzioni di donne nubili (per i censimenti dal 1901 al 1971) attraverso un modello logit lineare. Nonostante gli errori nei dati censuari, il modello sembra funzionare bene. Per verificare l'efficacia di questo metodo, è stata fatta un'applicazione per il Giappone, i cui dati sono più esatti, ed il risultato è molto favorevole. Un confronto tra il modello di Coale-McNeil mostra che, per la situazione indiana, il modello logit è più appropriato.

RESUME

Les recensements indiens fournissent la répartition de la population selon l'état matrimonial, le sexe et l'âge. Toutefois, pour ce qui concerne l'état matrimonial, l'information générale et la qualité des données sont souvent limitées. Par exemple, il arrive que le rapport entre les proportions de célibataires dans un certain groupe d'âge, calculées pour deux recensements consécutifs, excède l'unité, quoique le mouvement migratoire n'ait pas été significatif.

Dans ce travail nous avons essayé d'ajuster les proportions de femmes non-mariées (pour les recensements de 1901 à 1971) par un modèle linéaire "logit". Malgré quelques erreurs dans les données de recensement, le modèle marche bien. Pour vérifier l'efficacité de cette méthode, on a appliqué le modèle aux données japonaises, qui sont plus exactes et le résultat est très favorable. Si on compare le modèle de Coale-McNeil avec le modèle logit, on trouve que ce dernier s'adapte mieux à la situation indienne.