Griffith M. Feeney, a doctoral candidate at the Department of Demography, University of California, Berkeley, will join the East-West Population Institute as a Research Associate effective September 1, 1972. This paper is based on work done during the summer of 1971, when Mr. Feeney held a temporary research appointment at the Institute.

Griffith M. Feeney

A MODEL FOR THE AGE DISTRIBUTION OF FIRST MARRIAGE

Paper No. 23 April, 1972

Corrected Copy - February 22, 1982

The work reported below began with the desire to know the extent to which age distributions of first marriage in human populations conform to a mathematical model derived from some simple sociological notions. Pollard (1969) and Henry (1972) have suggested that the event of marriage is the result of two decisions, the decision that one desires to marry, and, given a possible partner, the decision to enter into marriage. The first, of these decisions may, though it need not be, made independently of the availability of potential marriage partners. The second decision, however, is contingent on this availability. I have been curious for some time as to whether the skewness characteristic of many age distributions of marriage and birth might be generated by a process in which individual decisions to marry or give birth are distributed symmetrically by age with the event of marriage or birth occurring only after a certain lag time, more or less independent of individual control. An opportunity to put this notion to the test came during the Fourth Conference on the Mathematics of Population in Honolulu in 1971, when Ansley J. Coale presented a paper (1971) in which it was shown that age distributions of first marriage in a wide variety of settings exhibit remarkable empirical regularities. I suggested at the conference that these regularities might be consistent with a model in which decisions to "enter the marriage pool" are distributed by age according to a normal distribution and in which the duration of the lag between entering the pool and marrying is distributed according to an exponential distribution. My calculations subsequent to the conference indicated that Coale's "standard schedule" of first marriage (1971:199) could indeed be reasonably approximated by this model, and Coale and McNeil

(1971) have subsequently shown that a natural refinement of the simple normal-exponential model fits the standard schedule extremely well.

An independent interest in the role of changing marriage patterns in United States population growth led me to consider applying the model to United States first marriage experience over the past several decades. Since the model in its behavioral aspect applies to the distribution of first marriage in birth cohorts, a serious data problem arises here. The United States Marriage Registration Area was begun only in 1958 and in 1968 consisted of only 42 states (Carter and Glick, 1970). The recent origin of the Area, with its changing composition, render currently available marriage registration data virtually useless for the purpose of tracing the marital history of birth cohorts. Alternative and potentially much richer sources are the 1960 Census of Population tabulations of a five percent sample of the population showing numbers of never-married persons by age and ever-married persons by age and age at first marriage. These tabulations have been used by several researchers (Saveland and Glick, 1969; Carter and Glick, 1970; Matras and Hirschman, 1971). Since the census is nationally conducted, the problem of limited and changing coverage does not arise. Since the census data are retrospective, available information extends fifty or so years before the date of the census. This historical depth is of course limited by memory and mortality. A person responding in 1960 to a question on his or her marriage some fifty years after the fact may simply not remember the facts correctly, and some of this person's contemporaries will have died in the interim and their age at first marriage will not be ascertained at all. It is possible, however, to analyze the resulting distortions of the data, and they prove small enough for the data to yield useful results.

The 1960 census tabulations of age by age at first marriage contain a great mass of data. A total of 264 birth cohorts are covered, 66 successive cohorts for each of the four subpopulations defined by sex and color. Since the process of fitting the model age pattern to an observed distribution of first marriage involved considerable computation, I selected a relatively small subset of these cohorts for preliminary study. Several considerations entered into the selection. The first derived from the emphasis on cohort rather than period data. For a birth cohort aged 50 or older in 1960, both the proportion of persons ever-marrying and the distribution by age at first marriage of persons ever-marrying are available, negligible numbers of first marriages occurring after age 50. But for a cohort aged, say, 35 in 1960, only the distribution by age at first marriage of persons ever-marrying before exact age 35 is available. Since statistics of marriage patterns for successive cohorts based on differing age spans could lead to spurious comparisons, I decided to consider only the distribution of first marriages occurring within the same age span for several cohorts. I chose the age span 14-34 as long enough to provide a reasonably good representation of the experience of a cohort and short enough so as not to exclude all but a few very old cohorts. (This is, incidentally, the age span for which age at first marriage is given by single years of age in the census tabulations.) This reduced the number of potential cohorts to 180, the number aged 35 or older at the 1960 census date. Other considerations were an interest in the effects of marriage patterns on aggregate fertility, which led me to select females, and a desire to minimize problems of undercounting and misreporting in the census, which led me to select white females. This left 45 cohorts, and from these I chose 15, beginning with the cohort which attained exact age 20 during 1902 and including the corresponding cohorts for every third year up to and including 1944.

For each of the birth cohorts so selected I computed a "partial" or "net" first marriage table (that is, a first marriage table showing what the incidence of first marriage would be in the absence of mortality) by single years of age for the age range 14-34. The observed proportions of women first marrying at each age in these tables were then compared with the model age patterns. According to the model, the entire age distribution of first marriage is determined by four parameters: the mean and standard deviation of the distribution of age at entry to the marriage pool in the cohort, the mean waiting time in the marriage pool for the cohort, and the proportion of the cohort never entering the marriage pool. The idea in fitting the model to the data is to determine, for each of the fifteen birth cohorts, values of these parameters for which the discrepancy between the corresponding model age distribution and the observed (first marriage table) distribution is minimized. The results thus include both an "expected" age distribution of first marriage and estimates of the four model parameters for each birth cohort.

The results may be summarized as follows. A total of 315 deviations of "expected" from "observed" proportions of women first marrying were computed, 21 deviations at ages 14-34 for each of the 15 cohorts. Slightly over two-thirds of these deviations are less than 10 percent of the corresponding observed values, and over nine-tenths are less than 20 percent of the observed values. The more recent cohorts show generally larger deviations. The time pattern of the average deviations across cohorts appears to be interpretable as a response to fluctuations in the business cycle and war-related population movements. Because of difficulties in the estimation of the model parameters, the estimates of these parameters are only rough approximations and should be interpreted only for strong, long-term trends, and then only with caution. The results are nonetheless suggestive. According to the model, mean age at first marriage is the sum of two components, mean age at entry to the marriage pool, and mean waiting time in the marriage pool. It is interesting that the decline in mean age at first marriage between the cohort attaining exact age 20 during 1902 and the corresponding cohort for 1944 is due almost entirely to a decline in the mean waiting time in the marriage pool from about 6 to about 3.5 years. The mean age at entry to the marriage pool is approximately constant at 17 to 18 years of age. This particular component decomposition of mean age at first marriage explains, in model terms, three trends in the observed age distributions of first marriage: a general decline in age at first marriage, an increasing concentration of first marriages within a short age span, and a decreasing skewness of the distribution of first marriage.

THE MODEL

The basic idea of the model is that individuals decide to marry in two stages. The first stage consists of a decision that a single person is ready to marry. This decision might be made freely by a single person, or it might be partially imposed by parents or other persons. It might occur in the form of certain target events before which marriage is felt to be undesirable, as for example graduation from high school or college, or entry to first job. But it is conceived of as being independent of the availability or suitability of particular potential marriage partners. The second stage consists of the delay between readiness for marriage and the event of marriage, if indeed marriage does occur. The length of this delay might be supposed to be influenced by a variety of factors. It is possible that a suitable marriage partner is found before a person is ready to marry, and in this case the delay will be zero. If readiness for marriage precedes finding a suitable partner, the delay will include two components. The lag between readiness for marriage and finding a partner would evidently depend on the diligence with which the search for a partner is pursued, the availability of partners, and the preferences which define a partner as "suitable." The lag between finding a suitable partner and marriage might depend on a number of circumstantial factors, as for example the desire to be married in the month of June. In respect of the age at marriage of females, it might depend on the prospective partner's employment situation.

The idea that two stages are involved in the process of marriage is due to Pollard (1969) and has been elaborated by Henry (1972). It should be pointed out that the model developed here does not make use of Henry's notion of "circles" of marriageable persons.

Under these assumptions single persons in a population at a given time who are ready to marry may be said to comprise a "marriage pool." Suppose that the ages at which women in a particular birth cohort enter the marriage pool are normally distributed with mean μ and standard deviation σ. The parameter μ might be thought of as representing a social norm for age at first marriage, and the parameter σ as representing the heterogeneity of women in the cohort in deviating from this norm. It follows that a woman chosen at random from this birth cohort enters the marriage pool between age x and age $x + dx$ with probability

$$\left[(2\pi\sigma^2)^{-1/2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right] dx. \tag{1}$$

Let the quantity in square brackets be denoted by $N(x;\mu,\sigma)$. Suppose that women leave the marriage pool only by marrying, and that the probability that a woman chosen at random first marries between t and $t + dt$ years after entering the pool is

$$\lambda e^{-\lambda t} dt. \tag{2}$$

The quantity λ^{-1} is then the mean waiting time in the marriage pool and might be thought of as representing the net effect of the delaying factors mentioned in the preceding paragraph. The conditional probability that a woman chosen at random from the cohort first marries between age a and age $a + da$, given that she entered the marriage pool at age x, is

$$\lambda e^{-\lambda(a-x)} da \tag{3}$$

and the absolute probability that this woman marries between age a and age $a + da$ is therefore

$$\left[\int_{-\infty}^{a} N(x;\mu,\sigma)\lambda e^{-\lambda(a-x)} dx\right] da. \tag{4}$$

Stated directly in terms of probability theory, the model says that the age at first marriage of a woman chosen at random from a particular birth cohort is the sum of two independent random variables. The first random variable, age at entry to the marriage pool, has the normal density (1). The second random variable, waiting time in the marriage pool, has the exponential density (2). The density at age at first marriage (4) is the convolution of these two densities. For definitions of the standard probability-theoretic terms random variable, independence, density, and convolution see for example the text of Feller (1966, II).

Let θ denote the mean waiting time λ^{-1} measured in units of σ, $\theta = \lambda^{-1}/\sigma$. Then $\lambda = (\theta\sigma)^{-1}$. By substituting this in (4) and making the change of variable $(a-\mu)/\sigma$ for a, one sees that (4) differs only by location parameters (Feller, 1966:II, 44) from the convolution of a normal density with zero mean and unit variance and an exponential density with mean waiting time θ. The parameter θ is a natural measure of the skewness of this density, for it is symmetric (because normal) when $\theta = 0$ and increasingly right-skewed as θ increases.

If a certain proportion, p, say, of the women in birth cohort never enter the marriage pool, then (4) must be multiplied by $1 - p$. Let the resulting quantity be denoted by $f(a;\mu,\sigma,\lambda,p)$. Under the above assumptions the expected proportion of women in a birth cohort who first marry between exact age x and exact age $x + 1$ is then

$$\int_x^{x+1} f(a;\mu,\sigma,\lambda,p)\,da. \tag{5}$$

This discussion has so far ignored mortality. One might incorporate mortality into the model, but this is undesirable from a technical point of view since the number of parameters would increase. An alternative approach is to construct from the observed data a "net" or "partial" first marriage table (Chiang, 1968:243; Hoem, 1970:8-10) which shows what the incidence of marriage would be in the absence of mortality. The proportions of women first marrying at each single year of age in this table then comprise the "observed" distribution for the cohort to which the model distribution (5) is fitted. In this approach mortality is in effect eliminated from data rather than included in the model.

Age Patterns of First Marriage for Cohorts of United States White Females

Let S_x denote the set of persons in a specified cohort who survive to and are single at exact age x, let M_x denote the set of persons in S_x who first marry between exact age x and exact age x + 1, and let PM_x denote the proportion of persons in S_x who are in M_x. Suppose that no marriage occurs before exact age 14 and no mortality before exact age 35, and let PNM_x denote the proportion of persons in the cohort never married by exact age x. Then $PNM_{14} = 1$ and $PNM_{x+1} = (1 - PM_x)PNM_x$, x = 14, ... , 33. The proportion of persons in the cohort who first marry between exact age x and exact age x + 1 equals $PNM_x - PNM_{x+1}$. Note that the base of these latter proportions is the total number of persons attaining exact age 14. They should not be confused with the proportions PM_x. When these statistics are collected together in a more or less standard tabular format the result is called a "partial" or "net" first marriage table.

Computation of the proportions PM_x requires, by definition, the numbers of persons in the sets S_x and M_x. However the census tabulation can provide at best counts of the numbers of persons in these sets who survive to the time of the census. Let S'_x and M'_x denote, respectively the subsets of persons in S_x and M_x who survive to the census date. If persons in S_x and M_x survive in the same proportions to the time of the census, then

$$\frac{N(M'_x)}{N(S'_x)} = \frac{N(M_x)}{N(S_x)} = PM_x \tag{6}$$

where $N(\cdot)$ denotes the number of persons in the enclosed set. More generally, let p_x^s denote the proportion of persons in S_x who survive to the census date, and let p_x^m denote the proportion of persons in M_x who survive to the census date. Then the estimate $N(M'_x)/N(S'_x)$ may be written $(p_x^m/p_x^s)PM_x$, which shows that the estimate entails a bias determined by the ratio p_x^m/p_x^s. This estimation procedure has evidently been used by Carter and Glick (1970) in the preparation of their Table 3.3. They do not indicate how the entries of this table are computed or mention possible errors due to mortality. The probable magnitude of these errors is discussed in the Appendix below.

The age distributions of first marriage for fifteen cohorts of United States white females are shown in Table 1. The distributions cover the age range 14-34 years and were computed from 1960 census data (United States Bureau of the Census, 1966a and 1966b) using the estimation procedure of the preceding paragraph. The distributions refer to cohorts in the population at the 1960 census date, and therefore exclude persons born to the United States population but not present at the census date and include immigrants to the population present at the census date. The cohorts are referred to as those attaining exact age 20 every third year beginning April 1, 1902 and ending April 1, 1945. Note that "year" here refers to the one year period beginning April 1 of each calendar year.

TABLE 1

Observed Proportions First Marrying in Each Single Year of Age: Selected Birth Cohorts of United States White Females (x10,000)

Age	Cohort attaining exact age 20 during:							
	1902	1905	1908	1911	1914	1917	1920	1923
14	104	144	115	102	102	119	110	97
15	189	223	200	198	238	247	225	233
16	302	328	377	365	383	415	415	439
17	604	489	571	516	589	585	535	665
18	685	670	715	739	773	770	760	818
19	769	824	810	798	805	884	920	904
20	762	835	770	847	814	797	845	867
21	746	780	820	805	785	812	743	810
22	760	664	722	675	759	873	692	704
23	594	564	626	582	590	684	580	595
24	527	583	544	477	535	515	505	515
25	428	437	445	479	565	468	428	418
26	364	397	369	373	399	349	357	341
27	360	313	288	320	300	278	296	238
28	266	273	270	338	261	226	282	188
29	216	232	238	240	216	175	198	169
30	177	197	181	192	164	152	151	149
31	156	158	182	142	132	127	127	141
32	131	141	145	118	95	98	112	115
33	99	125	102	102	97	70	99	102
34	126	110	88	89	76	59	101	87
	1926	1929	1932	1935	1938	1941	1944	
14	126	142	123	120	113	92	95	
15	251	274	261	217	223	216	235	
16	439	445	473	348	414	394	475	
17	682	652	663	562	630	621	775	
18	833	844	742	779	853	924	999	
19	869	877	738	892	855	1160	941	
20	861	840	770	925	908	1220	1020	
21	820	682	832	902	1010	942	1260	
22	714	573	795	773	990	697	1040	
23	619	524	698	665	786	611	709	
24	429	533	584	644	520	704	477	
25	326	454	469	548	351	505	344	
26	303	400	380	429	330	344	250	
27	284	342	340	266	351	240	188	
28	264	263	286	193	249	182	142	
29	223	209	235	183	189	146	120	
30	185	194	153	189	146	106	90	
31	128	160	119	166	104	74	75	
32	120	139	108	107	83	66	59	
33	101	96	119	84	64	52	45	
34	92	78	94	69	57	49	38	

Note: Calculated from 1960 census data. See text for explanation.

Each column of the table gives the proportion of woman in the indicated cohort who first married between exact age x and exact age x + 1, x = 14, ... , 34.

Conformity of the Model to the Data

Table 2 shows the "expected" distributions of first marriage corresponding to the observed distributions given in Table 1. The expected distribution shown for each cohort is calculated from formula (5) above where the parameters μ, σ, λ, and p are chosen so as to minimize the sum over all age groups of the squared deviations of the expected from the observed proportions marrying at each age divided by the expected proportion marrying at that age. A more detailed description of the calculation of the expected distributions is given in the Appendix below.

The magnitudes of the deviations of the expected from the observed proportions marrying at each age are shown in Tables 3 and 4. Certain deviations may be plausibly explained by a form of misreporting of age at first marriage. The census enumeration procedure involved determination of the year and quarter of first marriage and not, directly, age at first marriage. If the date of first marriage was unavailable, census enumerators are reported (United States Bureau of the Census, 1966b:X, "Year of first marriage") to have estimated the number of years since first marriage. The large positive deviations for the 1902 cohort at age 27, the 1905 cohort at age 24, the 1908 cohort at age 21, and the 1911 cohort at age 18 evidently represent a heaping effect on the duration fifty years.

TABLE 2

Expected Proportions First Marrying in Each Single Year of Age: Selected Birth Cohorts of United States White Females (x10,000)

Age	Cohort of: 1902	1905	1908	1911	1914	1917	1920	1923
14	101	103	106	103	113	119	113	90
15	209	212	219	212	232	245	232	209
16	364	369	381	370	403	423	403	394
17	540	548	565	549	595	624	595	613
18	695	705	725	706	761	794	758	801
19	787	799	819	799	856	888	850	900
20	803	815	834	816	865	892	857	896
21	759	770	784	770	808	824	796	818
22	680	690	700	691	714	721	700	708
23	591	600	606	601	612	610	596	598
24	507	515	517	515	516	508	500	499
25	433	439	439	440	433	420	417	415
26	368	374	372	374	362	347	346	345
27	314	318	315	319	303	286	288	286
28	267	271	266	271	253	236	239	238
29	227	231	225	231	212	195	199	198
30	193	196	191	196	177	161	165	164
31	165	167	162	167	148	133	137	137
32	140	142	137	142	124	109	114	113
33	119	121	116	121	104	90	95	94
34	101	103	98	103	87	75	79	78

Age	1926	1929	1932	1935	1938	1941	1944
14	109	153	141	112	102	83	125
15	243	279	264	221	223	207	272
16	441	441	428	379	411	419	496
17	661	610	606	565	640	693	763
18	836	748	759	739	854	951	1000
19	914	823	851	860	990	1100	1130
20	892	829	868	904	1020	1110	1130
21	804	778	822	872	950	1000	1020
22	691	695	737	790	826	841	844
23	581	602	638	685	686	675	664
24	485	513	542	578	556	531	506
25	403	434	456	481	445	415	380
26	335	365	382	398	355	323	284
27	278	308	320	329	283	252	212
28	231	259	267	271	226	196	158
29	192	218	224	224	180	153	118
30	160	183	187	185	143	119	877
31	133	154	157	152	114	93	65
32	110	130	131	126	91	72	49
33	92	109	110	104	72	56	36
34	76	92	92	86	58	44	27

Note: See text for explanation.

TABLE 3

Deviations of Expected from Observed Proportions First Marrying in Each Single Year of Age: Selected Birth Cohorts of United States White Females (x10,000)

Age	Cohort attaining exact age 20 during:							
	1902	1905	1908	1911	1914	1917	1920	1923
14	+3	+41	+8	-1	-11	-1	-3	+7
15	-20	+10	-20	-15	+6	+2	-7	+24
16	-62	-41	-3	-4	-20	-8	+12	+45
17	+64	-59	+6	-33	-6	-39	-60	+52
18	-10	-35	-10	+34	+12	-24	+1	+17
19	-18	+25	-10	-2	-51	-4	+70	+4
20	-42	+20	-64	+31	-52	-95	-12	-29
21	-12	+10	+36	+34	-22	-13	-53	-7
22	+80	-26	+22	-15	+45	+152	-8	-4
23	+2	-36	+20	-19	-21	+74	-17	-3
24	+19	+69	+27	-39	+19	+7	+5	+16
25	-5	-2	+7	+39	+132	+48	+12	+3
26	-5	+24	-3	-2	+37	+2	+10	-3
27	+47	-6	-27	+1	-3	-9	+8	-48
28	-1	+2	+4	+67	+7	-10	+42	-50
29	-11	+2	+12	+9	+4	-20	-1	-29
30	-17	+1	-10	-4	-13	-8	-14	-16
31	-8	-9	+20	-25	-16	-6	-11	+4
32	-9	-1	+8	-25	-29	-12	-3	+1
33	-20	+4	-13	-19	-7	-20	+5	+8
34	+24	+7	-10	-14	-11	-15	+23	+9

Age	1926	1929	1932	1935	1938	1941	1944
14	+17	-11	-18	+8	+12	+9	-31
15	+8	-5	-3	-4	0	+9	-37
16	-2	+4	+45	-31	+3	-25	-21
17	+21	+43	+57	-3	-10	-72	+12
18	-3	+97	-18	+40	-1	-28	-1
19	-45	+53	-113	+32	-136	+56	-193
20	-31	+11	-98	+21	-110	+104	-116
21	+16	-96	+10	+30	+57	-62	+243
22	+23	-122	+57	-17	+164	-144	+192
23	+38	-78	+60	-20	+100	-64	+46
24	-56	+20	+42	+66	-35	+173	-29
25	-77	+20	+14	+66	-94	+90	-36
26	-32	+35	-2	+30	-25	+21	-34
27	+6	+35	+21	-63	+68	-12	-24
28	+33	+4	+19	-78	+24	-14	-16
29	+31	-9	+12	-41	+9	-7	+2
30	+25	+11	-34	+5	+2	-13	+2
31	-5	+5	-38	+13	-10	-18	+10
32	+9	+9	-23	-19	-8	-6	+11
33	+9	-13	+10	-20	-8	-4	+9
34	+16	-14	+2	-16	-1	+6	+11

Source: Tables 1 and 2.

TABLE 4

Percent Deviations of Expected from Observed Proportions First Marrying in Each Single Year of Age: By Size and Cohort Group

Cohort attaining exact age 20 during:	Percent deviation				
	0-4	5-9	10-19	20+	Total
1902-14	54	28	18	5	105
1917-29	54	20	25	6	105
1932-44	30	30	33	12	105
Total	138	78	76	23	315
Percent	44	25	24	7	100

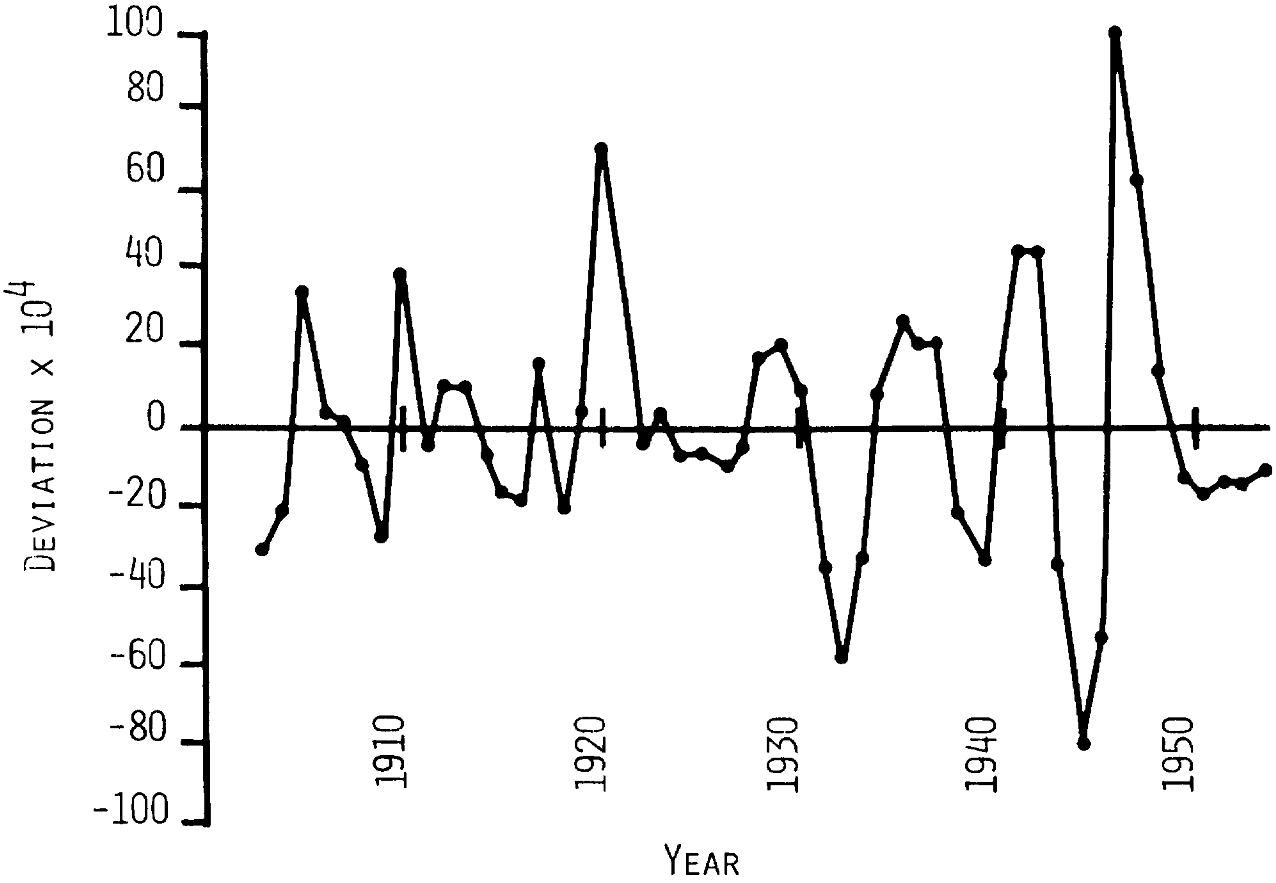
Source: Table 3.

Examination of the deviations suggests that this heaping may occur also on duration forty years.

When the deviations corresponding to each time period are averaged across cohorts, a certain time pattern of average deviations is observed. Likewise, when the deviations corresponding to each age group are averaged across cohorts a certain age pattern of average deviations is observed. The average age pattern of deviations shows positive deviations (observed greater than expected) at ages 22, 25 and 28 and negative deviations of similar magnitude at ages 20, 31, 32 and 33. The deviations at ages 20 and 22 might be explained by a postponement of marriages until after age 21, and the deviations at age 28 and the early thirties by attempts to marry before age thirty. The magnitudes of the deviations are small, however, and these remarks are speculative. The time pattern of average deviations shows magnitudes several times as large as those of the age pattern. The time pattern is graphed in Figure 1. The deviations around 1910 and 1920 probably reflect in part misstatement of date of first marriage in the census data. The remaining deviations are consistent with long recognized effects of the two world wars and the Great Depression (Davis, 1950:243-245). The magnitude and evident interpretability of the average time deviations suggests that elaboration of the model might most fruitfully begin by incorporating economic time series rather than by refining the model age pattern. A natural way to proceed would be to let the rate at which women in the marriage pool marry in a given year be a function of, say, the unemployment rate for that year.

FIGURE 1

Time Series of Average Deviations of Expected from Observed Proportions First Marrying in Each Single Year of Age: United States White Females



Source: Table 3.

Trends in Model Parameters

Table 5 shows the model parameters from which the expected distributions given in Table 2 were calculated. As mentioned in the first section of this paper, these estimates are only rough approximations and should be interpretated only for strong, long term trends, and then only with caution. The details of the numerical minimization procedure which generated the estimates are given in the Appendix below. The most remarkable feature of Table 5 is the relative stability of the parameters μ and σ which define the age distribution of entry to the marriage pool. Mean age at first marriage, estimated by $\mu + \lambda^{-1}$, declined by about 2.5 years between the oldest and the youngest cohort, and this decline was the net result of the nearly constant mean age at entry to the marriage pool and a large decrease in the mean waiting time in the marriage pool. (See Figure 2.) This decline in mean waiting time might be linked to two types of factors. First, developments in transportation and communication during this century evidently increase the effective number of potential marriage partners for each person in the marriage pool and hence decrease the lag between entry to the pool and finding a suitable partner. Second, long term improvements in economic conditions might tend to decrease the lag between finding a suitable partner and marriage.

The process of urbanization is evidently relevant to the time spent finding a suitable marriage partner, and it would be interesting to determine marriage patterns separately for persons in urban and rural areas. A remarkable quantity of relevant data

TABLE 5

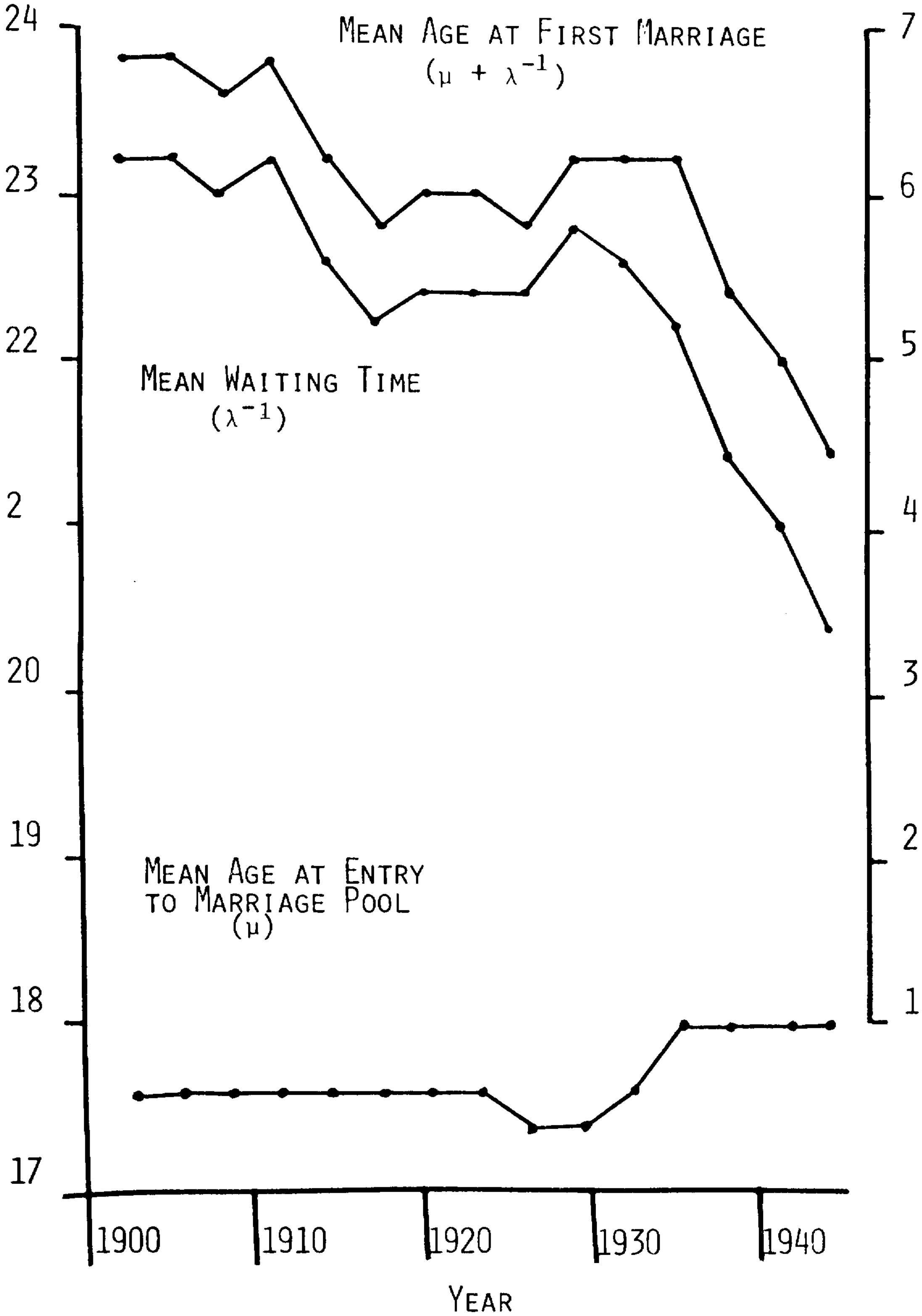
Mean (μ) and Standard Deviation (σ) of Age at Entry to the Marriage Pool, Mean Waiting Time in the Marriage Pool (λ^{-1}), Proportion Never Entering the Marriage Pool (p), and Mean Age at First Marriage: Selected Birth Cohorts of United States White Females

Cohort attaining exact age 20 during:	Mean age at first marriage	Parameters μ (yrs)	σ (yrs)	λ^{-1} (yrs)	p (%)
1902	23.8	17.6	2.2	6.2	10.0
1905	23.8	17.6	2.2	6.2	8.6
1908	23.6	17.6	2.2	6.0	8.2
1911	23.8	17.6	2.2	6.2	8.5
1914	23.2	17.6	2.2	5.6	8.1
1917	22.8	17.6	2.2	5.2	8.7
1920	23.0	17.6	2.2	5.4	10.6
1923	23.0	17.6	2.0	5.4	9.8
1926	22.8	17.4	2.0	5.4	9.1
1929	23.2	17.4	2.4	5.8	6.7
1932	23.2	17.6	2.4	5.6	4.4
1935	23.2	18.0	2.4	5.2	4.6
1938	22.4	18.0	2.2	4.4	5.0
1941	22.0	18.0	2.0	4.0	4.7
1944	21.4	18.0	2.2	3.4	4.9

Note: See text for explanation. Mean age at first marriage is estimated by $\mu + \lambda^{-1}$.

FIGURE 2

Mean Age at Entry to Marriage Pool, Mean Waiting Time in Marriage Pool, and Mean Age at First Marriage: Selected Birth Cohorts of United States White Females



Source: Table 5.

appears in a 1940 census report (United States Bureau of the Census, 1947). The tabulations unfortunately include only women married once with husbands present, and these two selections render the estimation technique used here inapplicable. Whether the technique can be usefully extended to the analysis of the 1940 data is an open question.

The last column of Table 5 shows the percentages of women never entering the marriage pool. The decline in these percentages might be linked to either social or economic factors.

APPENDIX

Analysis of Differential Mortality Bias

The following analysis is based on the multiple decrement theory which originated in the study of mortality by cause of death. This theory has recently been recognized in its mathematical aspect as a special type of continuous time, discrete state space Markov process. (See, for example, Bharucha-Reid, 1960, for a general discussion of Markov processes.) For a textbook treatment and further references see Chiang (1968: Chapter 11).

Hoem (1969) has recently introduced the concept of a "purged" Markov process -- loosely speaking, a process from which certain sample functions have been removed. This concept is precisely what is needed to evaluate the biases involved in estimating first marriage probabilities retrospectively from census data. In this context what has been "purged" is the first marriage experience of women in the various birth cohorts who did not survive to the census date. I generally follow Hoem's (1969, 1970) notation and terminology. Let $\nu(x)$ denote the force of first marriage, $\mu(x)$ the force of mortality for a single female at exact age x, and $\mu(x) - \varepsilon(x)$ the force of mortality for a married female at exact age x. Several formulas will be greatly simplified with no loss of clarity by writing "$\int_a^b f$" in place of "$\int_a^b f(x)dx$" whenever the latter appears in an exponent.

In the absence of mortality, the proportion of women attaining exact age x single who first marry before attaining exact age x+1 may be identified with

$$\int_x^{x+1} e^{-\int_x^a \nu} \nu(a)\,da \qquad (6)$$

which Hoem (1970:10) calls a "partial" probability. This is the quantity we desire to estimate. What we compute from the census data refers only to women surviving to the census date. For these women we compute the proportion attaining exact age x single who first marry before exact age x+1. This proportion may be identified with the conditional probability, for a woman single at exact age x, of first marrying before exact age x+1 given that this woman survives to the time at which the census is taken.

Consider a particular woman single at exact age x, t years before the census. Let $z = x + t$. The conditional probability that this woman first marries before exact age $x + 1$ given that she survives to the time of the census is, by definition of conditional probability,

$$\frac{\int_x^{x+1} e^{-\int_x^a \mu+\nu} \nu(a) e^{-\int_a^z \mu-\varepsilon} da}{e^{-\int_x^z \mu+\nu} + \int_x^z e^{-\int_x^a \mu+\nu} \nu(a) e^{-\int_a^z \mu-\varepsilon} da}. \qquad (7)$$

The numerator is the probability of marriage in the interval $x,x+1$ and survival to exact age z and is obtained by condition on the age a at marriage within this interval. The denominator is the probability of survival to exact age z and is obtained by conditioning on the event of marriage and on the age of marriage if marriage occurs. By rearranging the exponential terms and cancelling a factor of $e^{-\int_x^z \mu}$ (7) reduces to

$$\frac{\int_x^{x+1} e^{-\int_x^a \nu}\, \nu(a)\, e^{\int_a^z \varepsilon}\, da}{e^{-\int_x^z \nu} + \int_x^z e^{-\int_x^a \nu}\, \nu(a)\, e^{\int_a^z \varepsilon}\, da}. \qquad (8)$$

If there is no mortality, there can be no mortality bias, and we see correspondingly that $e^{\int_a^z \varepsilon} = 1$ and hence that (8) reduces to (6).

If there is no differential mortality by marital status, (8) again reduces to (6).

This is a more useful result, saying that no amount of mortality will bias our estimates so long as married and unmarried women are equally exposed -- a plausible result.

We turn now to the effect of differential mortality by marital status. For United States white females in 1940, 1950 and 1960, age specific mortality rates for married women are uniformly lower than those for single women (Grove and Hetzel, 1968:335). In this case $\varepsilon(x)$ will be nonnegative and the factor containing it will be greater than or equal to 1. As $\varepsilon(x)$ increases, for each x, both the numerator and denominator of (8) increase. The increase depends on the magnitude of the factor $e^{\int_a^z \varepsilon}$. The integral in the exponent here may be approximated by the sum of the differences between the married and single age-specific mortality rates over the relevant age groups. For $a = 20$ and $z = 70$, this is .0097 and the factor is $e^{.0067} \approx 1.0067$ for white females in 1940 (Grove and Hetzel, 1968:335), an error of less than 1 percent. When $z = 75$, the corresponding figure is 1.0097, an error still less than 1 percent. The bias appears to increase sharply after age 75. However, for cohorts under this age in 1960 (all but the cohort attaining exact age 20 during 1902 in the present case) mortality bias is of the order of 1 percent.

This analysis does not take account of differential mortality for divorced and widowed females. Since rates for such women are higher than the rates for married women, these differentials tend to cancel out the bias due to the single-married differential. Since rates for widowed and divorced women are also higher than the rates for single women, it is logically conceivable that over cancellation could occur, resulting in a bias in the opposite direction. However, given the level of widowhood and divorce and the size of the differentials this seems highly unlikely.

Fitting Procedure and Estimation of Model Parameters

Consider the observed distribution of proportions of females first marrying in each single year of age. Let $m = \lambda^{-1}$. Let O_i denote the observed proportion for age i ($i = 14, \ldots, 34$), let $E_i(\mu,\sigma,m,p)$ denote the expected proportion, computed from (5), corresponding to the parameter values μ, σ, m, and p, and let $D(\mu,\sigma, m,p)$ denote the sum over $i = 14, \ldots, 34$ of the quantity

$$\frac{(0_i - E_i(\mu,\sigma,m,p))^2}{E_i(\mu,\sigma,m,p)} . \tag{9}$$

The quantity $D(\mu,\sigma,m,p)$ may be regarded as an index of the discrepancy between the observed distribution and the model distribution corresponding to the parameter values μ, σ, m, and p. The "expected" distributions were obtained by numerically minimizing this quantity with respect to these four parameters. The procedure was as follows. First, choose values μ_0, σ_0, and m_0 for the parameters μ, σ, and m. Second, compute $E_i(\mu,\sigma,m,1)$ for each i, set

$$p_0 = 1 - (\Sigma_i 0_i)/(\Sigma_i E_i(\mu_0,\sigma_0,m_0,1)). \tag{10}$$

and then compute $D(\mu_0,m_0,\sigma_0,p_0)$. Repeat this second step beginning with parameter values $\mu_0 + 0.2i$, $\sigma_0 + 0.2j$, and $m_0 + 0.2k$ for various integral values of i, j and k, searching for those values which yield the lowest value of D.

The estimates of μ, σ, and m should obviously not be assumed accurate to better than 0.2 years. This corresponds to relative errors of about 1 percent in the estimate of μ, 6 percent in that of σ, and 10 percent in that of m. The estimation procedure does not suggest a specific error figure for the parameter p, but in view of the levels for the other parameters a figure of 10 percent is perhaps conservative. These errors of parameter estimation are of course in addition to errors in the basic data and biases in the estimation of the observed proportions first marrying.

Description of Input Data

Input data for each cohort consists of twenty-two numbers, each a non-negative integer of six or fewer digits, drawn from two tabulations of the 1960 United States census of population. The first of these numbers gives the total number of persons in the cohort at the 1960 census date as given in the report Marital Status (United States Bureau of the Census, 1966a:13). The remaining twenty-one numbers give the numbers of ever-married persons in the cohort at the 1960 census date who first married while aged 14, 15, ... , 34 as given in the report Age at First Marriage (United States Bureau of the Census, 1966b:32-33). Example: For the cohort of white females aged 56 at the 1960 census date, alternatively identified as the cohort which attained exact age 20 during 1923, the total number of persons in the cohort at the 1960 census date is given as 792,164; the number of these persons who first married while aged 14 is given as 7,708; and the corresponding numbers for ages 15, 16, ... , 34 are given as 18,430, 34,770, ... , 6,883.

ABSTRACT

A model is developed in which a woman's age at first marriage is regarded as the sum of two components, the interval between birth and entry to a hypothetical marriage pool, and the interval between entry to this marriage pool and marriage. The model is fitted to observed age patterns of first marriage in fifteen cohorts of United States white females who reached age 20 between 1902 and 1944. The observed age patterns are estimated from retrospective data in the 1960 census. An analysis of biases due to differential mortality by marital status is given.

REFERENCES

Bharucha-Reid, A. T. Elements of the Theory of Markov Processes and Their Applications. New York: McGraw-Hill Book Company, Inc., 1960.

Carter, Hugh and Paul C. Glick. Marriage and Divorce: A Social and Economic Study. American Public Health Association Vital Statistics Monographs. Cambridge, Massachusetts: Harvard University Press, 1970.

Chiang, Chin Long. Introduction to Stochastic Processes in Biostatistics. New York: John Wiley & Sons, Inc., 1968.

Coale, Ansley J. "Age patterns of marriage." Population Studies 25, 1971:193-214.

_____ and D. R. McNeil. "The distribution by age of the frequency of first marriage in a female cohort." Unpublished, Princeton University, 1971.

Davis, Kingsley. "Statistical perspective on marriage and divorce." The Annals of the American Academy of Political and Social Science 272, 1950:9-21. Reprinted in Joseph J. Spengler and Otis Dudley Duncan (eds.), Demographic Analysis: Selected Readings (The Free Press of Glencoe).

Feller, William. An Introduction to Probability Theory and Its Applications. Volume II, 2nd edition. New York: John Wiley & Sons, Inc., 1966.

Grove, Robert D. and Alice M. Hetzel. Vital Statistics Rates in the United States: 1940-1960. Washington, D. C.: United States Government Printing Office, 1968.

Henry, Louis. "Nuptiality." To appear in Journal of Theoretical Population Biology, 1972.

Hoem, Jan M. "Purged and partial Markov chains." Skandinavisk Aktuarietidskrift 52, 1969:147-155.

_____. "A probabilistic approach to nuptiality." Biométrie-Praximétrie 11, 1970:3-19.

Matras, Judah and Charles Hirschman. "A new look at the marriage market and nuptiality rates, 1915-1958." Demography 8, 1971: 549-569.

Pollard, J. H. "A discrete-time two-sex age-specific stochastic population program incorporating marriage." Demography 6(2), May, 1969:185-221.

Saveland, Walt and Paul C. Glick. "First marriage decrement tables by color and sex for the United States in 1958-60." Demography 6, 1969:243-260.

United States Bureau of the Census. Sixteenth Census of the United States: 1940. Population. Differential Fertility, 1940 and 1910, Part 2, Woman by Number of Children Ever Born and Fertility by Duration of Marriage. Washington, D. C.: United States Government Printing Office, 1947.

______. United States Census of Population: 1960. Subject Reports. Marital Status. Final Report PC(2)-4E. Washington, D. C.: United States Government Printing Office, 1966a.

______. United States Census of Population: 1960. Subject Reports. Age at First Marriage. Final Report PC(2)-4D. Washington, D.C.: United States Government Printing Office, 1966b.