J. Opl Res. Soc., Pergamon Press, 1978. Vol. 29, 4, pp. 349 to 360. Printed in Great Britain

On Microanalytic Simulation and its Applications in Population Projection

ALICE NAKAMURA and MASAO NAKAMURA

Department of Finance and Management Science, The University of Alberta, Edmonton, Canada

Population projections by various variables are often required by decision makers for a variety of planning problems in both the private and public sectors. Mathematical population models have not yet been developed which are suitable for making population projections stratified by large numbers of variables. However, stochastic simulation called microanalytic simulation provides us with a feasible approach for obtaining population projections of this nature. In this paper the structure of a microanalytic population simulation model and its advantages over the transitional matrix method are discussed. Then a general method is presented to condition the course of simulations over historical periods using available aggregate vital statistics data. In this way, deviations of the simulated from the actual population can be controlled to a certain degree allowing us to recreate otherwise unavailable historical time series and cross-sectional population samples with reasonable precision. Numerical results are presented for the population of Alberta, a province of Canada, over the period 1961–1971.

INTRODUCTION

DEMOGRAPHIC variables play an important role in certain planning problems of both the private and public sectors. For example, in the private sector, the family life cycle concept,[1] which has been used to analyse patterns of consumer demand for a variety of products, is based on the premise that each individual, or family unit, can be classified into one of several categories depending on such personal characteristics as age and family status. Population projection by various definitions of the family life cycle would allow a marketing planner to predict the future sales of a product whose demand is known to depend on the family life cycle, and to consider alternative marketing policies over time as related not only to changes in the size of the target population, but also to the proportions of this population in various stages of the family life cycle.

A current public planning problem in Canada is to predict demand for day care facilities. Governments at the federal, provincial and local levels are being called upon to support the establishment of day care facilities and to subsidize their operation costs by allowing differential fee schedules for users depending on their income levels. In order for governments to be able to make predictions over time for the demand for day care, and for day care-related expenditures, population projections are needed for family units by marital status, size, number and ages of children, labour force participation and income, educational levels of family members, etc. Such population projections would also be useful for predicting the future costs of the family allowance programme which has existed in Canada for some time and is subject to frequent changes in the allowances paid to children. Variables such as family earned income and ages of children are particularly relevant since allowances depend on the ages of children and are taxable.

Unfortunately, mathematical population models have not yet been developed which are suitable for making population projections stratified by large numbers of variables.[2] Microanalytic simulation provides us with a feasible approach for such purposes. In a microanalytic population simulation model each micro unit has associated with it a number of characteristics which may include demographic and socioeconomic variables. A population to be simulated over time is assumed to consist of such micro units, and the number of these units in the actual population versus the number in the simulation population in the base year determines the scaling factor. (If we have 500 single men as micro units in our initial simulation population representing an actual population consisting of 2000 single men in some base year, then the scaling factor is 0.25.) Given a file of an initial population in which the distribution of each characteristic over micro units is a replica of the real population of our interest for a specific year, say Year 0, our simulation proceeds as follows. Characteristics of all micro units are updated by Monte Carlo methods according to probabilities called operating characteristics in either a variable time or fixed time increment manner until the simulation clock reaches the end of a specified period, say the end of Year 0 or the beginning of Year 1. During this simulation process some micro units may die (and hence will be excluded from simulation processing afterwards) and some new micro units may be added, say due to birth. At the end of Year 0 distributions of micro characteristics of interest are tabulated. Then the same process is repeated for another simulation year using the file of micro units at the end of Year 0 as the new initial population. This process is continued for as many years as desired.

A household, a family, an individual, a dealer, a firm or any other behavioural unit of interest can be defined to be a micro unit, and a microanalytic simulation model can include more than one type of micro unit. The characteristics associated with each type of micro unit should ideally be chosen so that sufficient data exist to estimate the operating characteristics which are used to update the characteristics. There are several basic decisions to be made with respect to building a microanalytic simulation model. These decisions include (1) determination of types of micro units used, (2) characteristics associated with each micro unit, (3) the unit of time (or that of the simulation clock), and (4) evolution and resolution of interrelationships among micro units (for example, how marriage, divorce, birth and death should be handled). Most of these decisions depend on the kind of applications for which a microanalytic simulation model is proposed.[3–6]

There are two objectives of this paper. First, we will present some numerical results for a microanalytic population simulation model for Alberta, a province of Canada, over the period 1961–1971. This is important since very little has been published with respect to the performance of microanalytic simulation models despite the fact that microanalytic simulation is becoming more popular among model builders. The second objective is to present a method for taking advantage of some of the historically available data which governments collect and publish every year for certain variables. These variables include the total annual numbers of births, deaths and marriages for Alberta by sex, age group and other classifications. A microanalytic population simulation model is particularly suited for incorporating historically available information given at different levels of aggregation. Such information can be used to condition the course of a simulation, so that the expected values of the resulting simulation-generated series will usually be closer to the actual series and have smaller variances. (Without such conditioning the deviation between the simulation-generated and the corresponding actual series may increase as the length of the simulation increases.) In this paper an implementation of such conditioning is discussed. Our numerical results suggest that such conditioning can be advantageously used for a variety of applications. These applications include the diagnosis of misspecified operating characteristics and the reconstruction of historical time series and cross-sectional samples not otherwise available which in turn can be used to study the historical interrelationships between demographic variables and other variables of interest. For example, the demand for children's clothes, toys and funiture is believed to depend not only on economic conditions and the total numbers of children of different ages, but also on the birth orders of these children. Reconstruction of otherwise unavailable historical time series for the numbers of children classified by age as well as birth order would allow more adequate testing of life cycle hypotheses of this sort. A detailed reconstruction of birth histories over some historical period might also facilitate more careful studies of the relationships between childbearing behaviour and economic variables such as age-specific female labour force participation rates, the unemployment rate, and per capita income.

A MICROANALYTIC POPULATION SIMULATION MODEL

We are concerned with simulating a population. The method of simulation is microanalytic with individuals chosen as micro units. Associated with each individual in the simulation population is the following record of characteristics: sex*, date of birth*, marital status (single, married, divorced, widowed), parity**, number of living children**, number of marriages, date of current marital status, date of birth of spouse, date of last live birth**, and number of events, where * denotes no change once assigned a value and ** denotes valid for females only. Parity is the number of children ever born to a female.

Associated with each child (both living and dead) is a child trailer which consists of the following characteristics: date of birth, date of death, if applicable, and sex. The child trailer exists only if parity > 0. While a population is moved forward over time during a simulation run, an individual is subject to risks of birth, divorce, death, marriage, widowhood and death of child. For each event there is an event trailer which consists of type of event, date, marital status of the individual just prior to the event, interval since previous birth (for birth), length of marriage in months (for divorce), age at death in months (for death), interval since last change in marital status (for marriage) and interval since marriage (for widowhood).

There are three types of records: a record of characteristics without any trailers [Type (i)], a record of characteristics with only an event trailer [Type (ii)] and a record of characteristics with both a child and an event trailer [Type (iii)]. These record types are represented in Figure 1, where m is the number of events and n is the parity. [Note that Type (iii) does not apply to males.] When we say a (simulated) population consists of 1000 individuals, the computer representation of the population is a file which has 1000 records of either type (i), type (ii) or type (iii). (All programs used for this study are written in Fortran and their details are found in Nakamura and Nakamura[7] and the references cited there.)

Since there is no machine-readable census file which consists of records of individuals randomly chosen from the Alberta population for 1961 (the census year), we generated an initial population in conformity with published cross-tabulations which is a 1/500

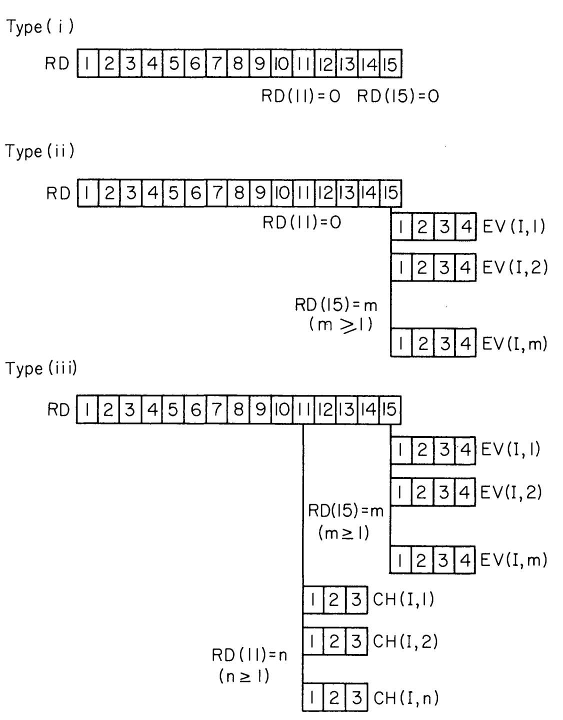

FIG. 1.

TABLE 1. THE SEX–MARITAL STATUS DISTRIBUTIONS OF THE GENERATED AND REAL ALBERTA POPULATIONS FOR 1961 (SCALING FACTOR = 1/500)

	Male					Female					Grand total
	Single	Married	Divorced	Widowed	Total	Single	Married	Divorced	Widowed	Total	
Generated	748	599	8	26	1381	613	594	4	71	1282	2663
Real*	748.8	598.2	31.8		1378.8	612.7	594	78.4		1285.1	2663.8

* From 1961 Canada census.

replica of the real population. Comparisons between the generated the real populations are shown in Tables 1 and 2.

It is not surprising that the sex–marital status distributions shown in Table 1 for the generated and real populations are quite similar, since one of the sets of parameters required to create the generated population is the proportions of the real population in the base year in each sex–marital status classification. The degree of proximity observed in Table 2 for married women by age and number of children under 15 years of age is more difficult to achieve, since the input parameters required in order to generate past birth histories for women are age–parity–marital status–specific birth rates for three time points.

This initial population was then advanced for 10 time periods (years), to enable meaningful comparisons to be made between simulation results and the actual numbers derived from census data (in 1961 and 1971).

To move forward the population let simulated time be represented by TIME. Then the age of an individual is calculated as the difference between TIME and the date of birth. Similarly, the interval from the last live birth to a female is given as the difference between TIME and the date of last live birth. Suppose a population file is to be advanced for one year. Each record is subject to three steps: (1) a record is read in, and a check is made to see if this individual is alive. If not, go to (3). Otherwise (2) the time of the next event is determined by a Monte Carlo method, where probabilities of events depend on age, sex, marital status and possibly other characteristics of the individual. If the time of the next event is within the present simulation year, then this event takes place, and appropriate characteristics of this individual are updated. Then repeat (2). If the time of the next event is not within the present simulation year, then go to (3). (3) The updated record of this individual is written out on to an output file. If a birth took place, a new record is created for the child, and written out also. We note that event and child trailers must always be written out as well after updating. Then go to (1).

The time of the next event is determined in Step (2) on the competing risk basis: suppose the probability of death in the present period for a single male of age 30 is p_1, and the probability of marriage in the present period for the same individual is p_2. Suppose further that these are the only possible events for this individual. Then by drawing two independent random numbers r_1 and r_2 uniformly distributed between 0 and 1, the time of death and time of marriage are calculated respectively to be

TABLE 2. THE DISTRIBUTION OF MARRIED WOMEN BY AGE AND NUMBER OF CHILDREN UNDER 15 YR OF AGE

Age groups for married females	Number of children under 15* 0	1	2+
15–24	21.1 (21)	23.6 (16)	24.4 (25)
25–34	16.2 (16)	24.8 (25)	118.2 (118)
35–44	24.4 (33)	30.2 (31)	93.2 (92)
45–54	56.4 (67)	24.4 (24)	18.8 (19)
54–64	50.5 (59)	3.1 (3)	0.8 (1)

* The first numbers correspond to a generated population, and numbers in parentheses are from 1961 Canada census.

$t_1 = [\log r_1/\log (1-p_1)]$ and $t_2 = [\log r_2/\log (1-p_2)]$. Let $t_2 < t_1$. Then if TIME (of the simulation clock) plus t_2 is less than the end of the present simulation period, a marriage takes place, and the marital status of this individual is changed. Hence the age of spouse is assigned also. (Note that since $t_2 < t_1$, no time of death is set.) A similar procedure (with a different set of possible events) is applied to this individual who is now married with TIME now replaced by TIME + t_2. This is continued for this individual until TIME exceeds the end of the present year.

We apply (1)–(3) to all individuals sequentially in the population file, and then print out summary statistics on the population and the events that took place. By specifying the total number of simulation periods for which the process is to be repeated we get statistics on the simulated population and the events that took place at the end of each simulated year.

The probabilities (operating characteristics) are given as data for death by sex, age and marital status, for divorce by the duration of marriage, for marriage by sex, age and marital status, and for birth (only for females) by age, marital status and parity. When an individual, say, a female of age y gets married, the age (or equivalently the date of birth) of husband is assigned: let x be the the age of husband. Then the model assumes that the variable $Z = (x - 14)/5$ has a log-normal distribution with two parameters μ and σ, where μ and σ can be described as polynomial functions of the age of the woman. A second degree polynomial was found to give an adequate fit. Thus the required input data are the estimated values of α_0, α_1, α_2 β_0, β_1 and β_2, where

$$\mu = \alpha_0 + \alpha_1 y + \alpha_2 y^2,$$
$$\alpha = \beta_0 + \beta_1 y + \beta_2 y^2,$$

and y is the age of the woman.

The parameters of the birth, marriage and death operating characteristics were estimated using data for Alberta. Data for Alberta for divorce, however, are not available in the form required by our programs, and we are presently using parameter estimates based on U.S. data.

PROJECTION BY MICROANALYTIC SIMULATION OR BY TRANSITION PROBABILITIES

Given a breakdown of a population in some base year by variables such as age, sex and marital status, one way of projecting this population forward in time is to calculate transition probabilities from each population group in the base period to each of these groups or the deceased category in the next period. Another approach, however, is to consider sequentially each individual in the initial population and probabilistically determine whether this individual will die, marry, divorce, be widowed, or have a child in a unit time period, given his initial characteristics. This is the microanalytic simulation approach presented in this paper.

As the number of state variables, and the number of values which each of these variables is allowed to assume, are increased the transitional matrix approach becomes unmanageable due to the large number of transition probabilities which must be manipulated. The number of parameters needed to advance a population forward in time can be reduced in two ways by resorting to a microanalytic simulation approach. First, information about the relevant conditional distributions can be stored and manipulated within the model in parametric rather than relative cell frequency form.

For instance, suppose we are trying to represent the conditional distribution of the ages of husbands given the ages of their wives. We could describe this conditional distribution by a set of ratios of the form

$$h_{ij} = f_{ij}/\sum_i f_{ij},$$

where h_{ij} denotes the proportion of women in the jth the age group with husbands in the ith age group. If 14 age groups were used to describe the ages of wives, and

14 more to describe the ages of husbands (i.e. $i = 1,\ldots, 14; j = 1,\ldots, 14$), $14 \times 14 = 196$ relative frequencies would be required to represent this conditional distribution. This number could be reduced by resorting to 10 instead of 5 yr age intervals. However, the extent to which wider classifying intervals can be tolerated clearly depends on the nature of the intended applications. For many purposes smaller intervals would be preferable, which often results in a very large transitional matrix.

By using all of the available data to estimate the parameters of a theoretical distribution function, the problems of handling cells with few observations are avoided. Also, there are no questions of how to discretize the data. On the other hand, the assumption inherent in this approach is that a suitable distribution function can be found.

Second, if an initial population is being advanced over time using a transition probabilities approach, the aging of this population must be expressed probabilistically just as other demographic events such as family formation and dissolution are expressed. Consider an initial population of women aged 30–34 stratified according to their marital status (single, married, widowed, divorced). Suppose we would like to advance this population 1 yr in time. In a year's time some single, widowed and divorced women may marry, some married women may divorce or be widowed, some women may die, and some women may move into the 35–39 year age bracket. Thus a woman in the initial population might be found in any one of the following nine states in the next year:

State	Description
1	30–34, single
2	30–34, married
3	30–34, widowed
4	30–34, divorced
5	35–39, single
6	35–39, married
7	35–39, widowed
8	35–39, divorced
9	Deceased

The following 22 transition probabilities could be estimated and used to describe the conditional distribution of this population in year $t + 1$, given that all members of this population were found in states 1–4 in year t:

$P_{11}, P_{12}, P_{15}, P_{16}, P_{19},$
$P_{22}, P_{23}, P_{24}, P_{26}, P_{27}, P_{28}, P_{29},$
$P_{33}, P_{32}, P_{37}, P_{36}, P_{39},$
$P_{44}, P_{42}, P_{48}, P_{46}, P_{49}.$

No probabilities are listed for a married woman becoming single or a widowed woman becoming divorced, since events of this sort are assumed to occur with a probability of zero.

When a microanalytic approach is used, however, time no longer has to be treated stochastically. Rather each individual adds one year to his age in a year's time.

Thus, in terms of our little example, the following probabilities would be required to advance an initial population of women aged 30–34 one year:

P (marriage | 1), P (death | 1)
P (being widowed | 2), P (being divorced | 2), P (death | 2)
P (marriage | 3), P (death | 3)
P (marriage | 4), P (death | 4).

There are other ways of reducing the number of parameters needed to describe the aging of individuals classified by various other characteristics which are associated with a person's age. Thus the treatment of time in a microanalytic model is not a sufficient reason for adopting this modelling approach; but it is a convenient feature, especially when other time-dependent variables such as the intervals since marriage, since birth of the last child, and since becoming unemployed are built into a model. As the number of intervals used to describe age, and the elapsed time since various events, are increased, or the total number of events whose probabilities depend on these time variables is increased, then this treatment of time becomes essential.

USE OF AGGREGATE HISTORICAL DATA TO CONDITION A MICROANALYTIC SIMULATION—A TRACKING MECHANISM

We have implemented a tracking mechanism which uses aggregate historical data for deaths, births and marriages to condition the course of simulation over a historical period. Tracking by divorce has not been implemented since appropriate data does not exist for Alberta. As an example consider tracking by death. Suppose we have a population file generated for 1962. Assume that the initial population for the base year of the simulation is one five hundredths of the real population for that year (hence the scaling factor = 1/500). Suppose further that we are given the actual numbers of deaths for 1962 stratified by, say, two sex groups, three marital status groups and 12 age groups.

We first calculate the difference D_i between the actual number of deaths divided by 500 (the scaling factor) and the simulated number of deaths for each sex–marital status–age cell i. We also calculate the total number S_i of simulated individuals in cell i who are alive and the total number M_i of simulated individuals in cell i who died during the last simulation year. If D_i is positive, more deaths took place in reality than in simulation and it becomes necessary to "kill" D_i of S_i simulated individuals in cell i. Similarly, if D_i is negative, more deaths took place in simulation than in reality, hence it becomes necessary to "restore" $-D_i$ of M_i simulated individuals in cell i who died in the last simulated year. Since D_i is usually not an integer, we round D_i in the following manner: if D_i is positive, draw a uniform random number r and compare it with $R_i = D_i - \text{Int}\,(D_i)$, where Int (X) is the integer part of a real number X. If $r \leq R_i$, set new $D_i = \text{Int}(D_i) + 1$, otherwise set new $D_i = \text{Int}(D_i)$; if D_i is negative, draw a uniform random number r and compare it with $Q_i = -D_i - \text{Int}\,(-D_i)$. If $r \leq Q_i$, set new $D_i = -[\text{Int}\,(-D_i) + 1]$, otherwise set new $D_i = -\text{Int}\,(-D_i)$. After rounding D_i we first check if D_i is positive, zero or negative. If it is zero, we do not change records of individuals in cell i at all. If D_i is positive, calculate $IJ_i = S_i/D_i$, where / denotes an integer division (as in Fortran) where the remainder is ignored. If D_i is negative, calculate $IJ_i = M_i/(-D_i)$. Then we process a population file sequentially: as we go through a population file, we "kill" every IJ_ith individual in cell i we encounter if $D_i > 0$, and we "restore" every IJ_ith individual in cell i who died in the last simulation year if $D_i < 0$. It is then necessary to update the characteristics of individuals who are either "killed" or "restored" by the tracking mechanism.

Suppose now that a researcher would like to reconstruct the age–parity specific birth rates of married women living in Alberta over some historical period. Data are available for each year for Alberta on the number of births to married women by the age and parity of the mother. Data on the total number of women in each age, marital status, parity group are only available for the census years. These data can be simulated over a historical period, however, by specifying an initial population for some census year, and then conditioning the simulation of the events of death, birth, marriage, and divorce if possible, on the existing Vital Statistics data.

The ability to accurately reconstruct the changes in the structure of a population over some historical time period is also a potentially important aid in the diagnosis of malfunctioning components of our model. This is far from a trivial problem in a large, interrelated model. For one thing, due to the stochastic nature of micro simulation models, even if all aspects of a model are correctly specified, still no one simulation run over some historical time period can be expected to generate an exact replica of the historical experience being simulated. For instance, the number of deaths simulated during the first time period will most likely not be equal to the actual number of deaths during this period. This is in keeping with the uncertainty of the real world, as represented by the nonzero variances of the distributions specified for the event of death. The composition of the simulation population, therefore, will probably be different from the composition of the real population at the beginning of the next time period. For instance, there may be fewer surviving married men 50–55 years of age in the model population than in the real population. If both the model population and the real population are now exposed to the same death probabilities in successive time periods, the discrepancy between the simulation and the actual population may continue to increase.

Using our conditioning programs at the end of each year-long simulation period, however, we can insure that the number and composition of surviving individuals at the beginning of each simulation period is what actually existed in the real population at that point in time. The recorded differences between the number of deaths generated by the simulation in each time period and the actual numbers then should reflect differences resulting from the death probabilities specified in the model being applied to a correct number of surviving males at the beginning of each time period.

A related problem in identifying and diagnosing misspecified components in a microanalytic simulation model is that erroneous output from one component may feed into and cause erroneous output from several other components, and the resulting discrepancies may all interact and compound over time in the manner discussed above. Tracking provides us with a means of breaking this chain of interactions.

AN ANALYSIS OF SIMULATION-GENERATED TIME SERIES AND EFFECTS OF TRACKING

Five independent runs of simulation were performed without tracking and five more with tracking for the historical period 1961–1971 (Table 3).

For the 50 triplets of observations shown in Table 3, the standard deviations for the results with tracking (ST) are found to be smaller than for the results without tracking (SNT) in 41 cases. Also the means with tracking are closer to the actual figures for Alberta in 41 cases; and they are closer in all cases for 1971, the only year in which a full Canadian census was taken. Thus the five simulation populations moved forward with tracking over the period of 1961–1971 appear on the average to provide a better representation of the actual Alberta population over this period than the five simulation populations moved forward without tracking. Both with and without tracking, however, our simulation results tend to underestimate the numbers of married males and married females, and the total population 15 yr of age or older. The explanation of this phenomenon is believed to be our failure to account for migration. During the period of 1961–1971 Alberta experienced rapid economic growth. Net in-migration accompanied this economic growth, and a large portion of those who moved to Alberta during this period were at least 15 year of age. In our subsequent work on this project, we plan to use a program at the beginning of each simulation period to add records to our simulation population for estimated net migrants entering Alberta during the previous year. The categories of married males and married females are also being depleted in our present model because the U.S. age–duration-specific divorce rates which we used in our model appear to be too high for Alberta.

The numbers of single males and single females in our simulation populations are also too large on the average. This is due to the fact that constant age–marital status–parity-specific birth probabilities were used in our model, while the marital birth rates have fallen in general over the period of 1961–1971. The resulting discrepancies are not fully corrected by our tracking program for births.

To explore further the effectiveness of our tracking mechanisms ten more independent simulation runs over the historical period of 1961–1971 were made using the initial population for Alberta and our Alberta model with the marriage probabilities for men

TABLE 3. THE SEX AND MARITAL STATUS DISTRIBUTIONS FOR SELECTED SERIES OF THE GENERATED AND REAL ALBERTA POPULATIONS FOR 1962–1971 (SCALING FACTOR = 1/500)

Year	Single males			Married males			Single females			Married females			Total (≥ 15)*		
	ST†	SNT‡	A§	ST	SNT	A	ST	SNT	A	ST	SNT	A	ST	SNT	A
1962	770.2 (6.77)	770.6 (9.12)	768.8	607.6 (4.80)	603.8 (9.49)	609.2	635.4 (3.32)	635.8 (3.54)	637.8	601.6 (3.54)	596.2 (4.34)	698.6	1761.0 (1.73)	1759.2 (2.83)	1768.2
1963	792.0 (4.53)	791.4 (4.42)	788.6	614.4 (5.25)	608.0 (8.43)	618.8	653.4 (4.01)	656.8 (8.81)	660.2	611.6 (2.36)	603.0 (6.63)	619.8	1791.4 (4.00)	1790.0 (8.60)	1808.0
1964	813.0 (6.44)	813.8 (7.69)	804.8	625.0 (11.25)	613.4 (9.43)	627.6	672.6 (3.33)	674.8 (11.86)	677.6	619.4 (4.96)	608.0 (8.60)	628.4	1824.0 (3.61)	1824.8 (6.63)	1841.8
1965	827.8 (12.67)	833.0 (15.08)	814.8	635.6 (14.65)	619.4 (11.34)	635.0	695.4 (8.06)	693.8 (11.16)	688.4	630.8 (6.68)	616.6 (8.66)	636.0	1859.6 (6.16)	1858.0 (5.57)	1875.2
1966	841.8 (11.73)	850.2 (22.65)	818.9	643.4 (12.63)	626.0 (12.35)	640.8	715.0 (11.31)	712.8 (12.21)	694.4	641.2 (9.24)	622.8 (10.3)	642.5	1900.0 (6.49)	1892.6 (4.80)	1904.9
1967	855.8 (13.01)	870.2 (19.61)	830.0	653.6 (11.56)	633.2 (8.80)	653.6	730.4 (12.02)	718.8 (15.75)	706.4	651.0 (10.12)	638.0 (10.75)	655.4	1935.2 (7.00)	1928.2 (6.08)	1954.8
1968	869.6 (10.57)	888.6 (25.11)	844.8	663.4 (8.98)	642.6 (12.51)	672.2	746.0 (12.61)	742.0 (18.80)	720.2	664.0 (10.17)	649.4 (11.83)	675.0	1980.0 (6.00)	1976.2 (7.28)	2019.6
1969	878.2 (14.47)	906.4 (27.52)	857.4	678.2 (10.67)	649.4 (11.49)	693.6	758.2 (10.88)	767.2 (21.30)	728.8	679.6 (7.49)	658.8 (15.62)	695.6	2024.6 (9.85)	2019.6 (6.78)	2089.2
1970	883.2 (21.26)	930.0 (33.63)	871.2	692.4 (11.69)	657.0 (14.28)	714.2	774.6 (12.56)	787.2 (20.74)	741.4	696.6 (17.32)	671.2 (18.15)	716.0	2069.0 (5.74)	2068.6 (8.89)	2162.0
1971	896.6 (27.85)	964.8 (40.35)	872.2	705.6 (10.42)	662.4 (17.29)	739.0	789.8 (13.74)	808.4 (15.41)	743.4	717.8 (13.57)	683.0 (19.72)	736.0	2117.8 (7.94)	2115.0 (12.21)	2226.7

* The total population 15 yr of age or older.
† Series generated by simulation with tracking. Numbers in parentheses are standard deviations.
‡ Series generated by simulation without tracking. Numbers in parentheses are standard deviations.
§ Numbers for 1966 and 1971 are from Canada censuses. Other numbers are estimated by Statistics Canada.

TABLE 4. THE DISTRIBUTIONS OF THE GENERATED AND REAL MARRIED AND SINGLE ADULT MALE ALBERTA POPULATIONS FOR 1962–1971 (SCALING FACTOR = 1/500) (MARRIAGE PROBABILITIES FOR MEN = 0)

	Single (≥ 15)*			Married		
Year	ST†	SNT‡	A§	ST	SNT	A
1962	270.8	291.8	272.6	606.6	582.6	609.2
	(0.94)	(1.84)		(2.66)	(3.61)	
1963	271.8	313.6	278.2	616.2	568.0	618.8
	(2.81)	(2.35)		(5.37)	(5.43)	
1964	276.6	337.8	283.8	624.4	553.2	627.6
	(6.12)	(3.15)		(7.28)	(6.19)	
1965	280.6	363.2	290.2	634.8	541.4	635.0
	(4.94)	(4.10)		(7.04)	(6.48)	
1966	307.2	389.4	295.4	646.6	529.2	640.8
	(3.72)	(2.44)		(9.96)	(8.03)	
1967	310.4	416.0	305.0	640.0	518.2	653.6
	(3.38)	(1.58)		(12.00)	(7.73)	
1968	314.6	443.8	316.4	662.8	506.0	672.2
	(3.46)	(2.22)		(13.44)	(7.35)	
1969	315.8	473.2	328.0	681.2	495.0	693.6
	(1.37)	(4.34)		(12.88)	(7.31)	
1970	316.6	504.2	339.4	701.8	483.4	714.2
	(3.32)	(4.34)		(14.70)	(7.71)	
1971	318.4	536.4	346.0	729.8	473.2	739.0
	(6.20)	(3.24)		(12.87)	(8.39)	

* The number of males who are 15 yr old or older.
† Means for series generated by simulation with tracking. Numbers in parentheses are standard deviations.
‡ Means for series generated by simulation without tracking. Numbers in parentheses are standard deviations.
§ Numbers for 1966 and 1971 are from Canada censuses. Other numbers are estimated by Statistics Canada.

intentionally set equal to zero. Five of these runs were made without tracking, and five were made with tracking of deaths, births, and marriages. Table 4 shows the extent of correction achieved by tracking for the series "total married men" and "single men 15 yr of age or older".

Suppose now that repeated simulations over some historical period are consistently observed to generate more deaths of single men than actually took place during this period. Perhaps some of the death probabilities for single men specified in the model are too high. Or perhaps the marriage rates for men are too low, resulting in too large a proportion of single men. In Table 5 the mean numbers of deaths of single and married men simulated to take place per year without tracking, and the mean numbers simulated to take place after the application of our tracking programs for the previous year but before application in the current year are compared with the actual numbers of single and married men who died in Alberta each year over the period 1967–1971. Looking at the results without tracking, it would appear that our operating characteristic for death is generating too many deaths for single men and not enough for married men. This result, however, is due in part to the fact that for these simulation runs the probabilities of marriage for men have intentionally been set equal to zero. Thus, in the five runs without tracking, as the simulation period progresses the number of married men becomes progressively smaller than it should, while the number of single men becomes progressively larger than it should, and the numbers of deaths generated in each period are accordingly distorted. In the case of the five runs made with tracking though, in each period the probabilities of deaths specified in the model are applied to approximately the correct numbers of individuals of each type, with the distortions resulting from the incorrectly specified probabilities of marriage for men being limited to the present period. Table 5 shows that the probabilities of death specified in our model are too high for both single and married men.

TABLE 5. DEATHS OF SINGLE AND MARRIED MEN IN ALBERTA (MARRIAGE PROBABILITIES FOR MEN = 0)

Year	Single			Married		
	SAT*	SNT†	A‡	SAT	SNT	A
1967	3.8	5.8	2.0	7.8	6.0	6.6
	(0.83)	(1.30)		(3.19)	(1.22)	
1968	3.8	4.6	2.0	8.6	7.6	6.5
	(2.16)	(2.07)		(1.94)	(1.94)	
1969	5.8	5.6	2.0	6.2	6.0	6.8
	(2.58)	(1.81)		(2.38)	(2.00)	
1970	5.4	4.4	2.1	6.8	5.8	6.8
	(1.67)	(1.14)		(1.48)	(2.28)	
1971	4.0	4.8	2.1	8.8	5.2	6.9
	(2.00)	(1.92)		(2.38)	(2.94)	

* Means for series generated by simulation after application in the previous period of tracking programs for deaths, births and marriages. Numbers in parentheses are standard deviations.

† Means for series generated by simulation without tracking. Numbers in parentheses are standard deviations.

‡ Numbers for 1966 and 1971 are from Canadian censuses. Other numbers are estimated by Statistics Canada.

This is to be expected, since constant 1961 death probabilities are being used in our Alberta model, and the probabilities of death for men of all age groups have declined over the period since 1961.

CONCLUSION AND FUTURE EXTENSIONS

The structure, advantages and use of a microanalytic population simulation model were discussed. Then a tracking mechanism which takes advantage of available historical data and its implementation was discussed. Finally numerical results were presented for the Alberta population for 1961–1971.

Future extensions for which work is under way include (1) addition of socioeconomic variables such as education level, and labour force participation, (2) design of a distributional mechanism to distribute among micro units total national or provincial income and unemployment figures provided by a macro econometric model, (3) estimation of operating characteristics which are time dependent, and (4) addition of immigration and emigration. The final objective of these extensions is to make available a micro-simulation population simulation model coupled with a macro model which will be suitable for use as a tool of analysis for planning problems in both the public and private sectors in which demographic and socioeconomic variables play an important role. We plan to carry out a series of behavioural and operational research studies in marketing, health care delivery, labour force behaviour and housing.

ACKNOWLEDGEMENTS

Portions of this paper were presented at a workshop entitled "Micro-analytic Models: State of the Art and Prospects for the Future" which was sponsored by the National Bureau of Economic Research and held at the Urban Institute in Washington, D.C., May 29–30, 1975, and at the XXII International Meeting of The Institute of Management Sciences held in Kyoto, Japan, July 24–26, 1975. The authors wish to thank the Faculties of Graduate Studies and Research and Business Administration and Commerce of the University of Alberta for providing support for computing for this project. This paper is a part of a larger study entitled "A microanalytic simulation model of the Canadian household sector" being conducted at the University of Alberta.

REFERENCES

[1] P. KOTLER (1972) *Marketing Management: Analysis, Planning, and Control*, 2nd edition. Prentice-Hall, Englewood Cliffs.

[2] N. KEYFITZ (1968) *Introduction to the Mathematics of Population*. Addison-Wesley, Reading, MA.

[3] G. H. ORCUTT *et al.* (1961) *Microanalysis of Socioeconomic Systems: a Simulation Study*. Harper & Row, New York.

[4] A. E. AMSTUTZ (1967) *Computer Simulation of Competitive Market Response*. M.I.T. Press, Cambridge, MA.
[5] D. G. HOROVITZ *et al.* (1970) Simulation of hospital utilization. Proc. Soc. Statist. Section, Am. statist. Ass. Annual Meeting, Washington, D.C.
[6] D. E. YETT *et al.* (1975) A microeconomic model of the health care system in the United States. *Ann. Econ. Soc. Measur.* **4,** 407–433.
[7] A. NAKAMURA and M. NAKAMURA (1975) On microanalytic simulation and its applications to population projection. XXII Int. Meeting, Inst. Mgmt Sci., Kyoto, Japan.