# MICROANALYTIC SIMULATION: A DEMOGRAPHIC MODEL*

Guy H. Orcutt, Gerald E. Peabody, Steven B. Caldwell
The Urban Institute

Work is under way at the Urban Institute on a microanalytic model of the income distribution.[1] In this model, an extension of earlier work by Orcutt, Greenberger, Korbel and Rivlin [6], the behavior of family members that affects the family's income status over time is simulated. To describe dynamically population growth and the formation and dissolution of families, birth, death, marriage, separation, and divorce are simulated, where appropriate, for each person in the sample population. To simulate income activity, the number of weeks worked in the year and the wage rate are imputed for each working member of the family. Transfer payments and income from wealth are added to the total family wage income. Taxes are then calculated and subtracted from the total family income to give disposable income, from which savings are imputed to generate the accumulation of assets. Migration and the educational attainment of children are also simulated. The demographic and education sections of the model are essentially complete, and these operating characteristics, along with the auxiliary macroanalytic model, are described in this paper.

The model is still in the process of development and the description in this paper should not be taken as the final product. In fact, the model is being designed so that it can be easily modified whenever new results can be brought to bear.

This model is designed to provide a dynamic representation of the United States population of individuals and families which will be useful in tracing the effects of various public policies, singly or in combination, on the behavior and well-being of individuals and families over time. However, since this model is focused entirely on individuals and families it cannot present an essentially closed representation of the entire economic system. By itself it cannot be expected to generate unemployment, price level changes, GNP or fractions of GNP going to earned income and to wealth holders. Thus, an auxiliary macroanalytic model has been designed to provide a serviceable interim closure.

Work on the demographic sector of the model has been completed for a simplified family structure model--a persons model. In the persons model marital status is imputed for each adult in the sample but the files of a man and woman who are to be married are not actually merged to form a new family file. This procedure vastly simplifies the computer implementation but at the cost of the loss of most information about the husband in the family. A family model is under development in which the files of the marrying couple are merged to form a complete data file for all members of the family.

The family model will be used to implement the complete model of the income distribution. The persons model, which will be described in more detail below, is being used to test the various demographic components of the model and constitutes a reasonable demographic model in its own right. By simulating marriage, separation and divorce the marital status of each adult in the population is determined. The marital status of women, along with their parity, race and socioeconomic status, provide the necessary input for the birth simulation. With the simulation of deaths we have a complete vehicle for providing population growth predictions. The detailed information about race, social class, etc., of persons in the model gives estimates of the future size distribution by various subgroups of the population.

The persons model contains six operating characteristics: death, marriage, separation, divorce, birth and education. The first five constitute the demographic core of the model while the sixth operating characteristic begins our explicit concern with the intergenerational distribution of status and income. For each pass of the simulation the operating characteristic will generate a probability of the corresponding event occurring to that person during the year. For example, each person will first be subjected to the death routine in which the probability that the person dies during the year will be determined. Via the Monte Carlo technique of drawing a random number and comparing it with the probability of the event, the occurrence, or not, of death will be imputed (if a random number r, $000 \leq r < 999$, is drawn and the probability of death is .05, then death will be imputed if $r < 50$).

For the initial population we are using a self-weighting random sample drawn from the 1967 Survey of Economic Opportunity (SEO) [9]. For

*This research is being conducted under grants from the Ford Foundation and the Social Security Administration. Opinions expressed are those of the authors and do not necessarily represent the views of The Urban Institute or its sponsors.

[1]Our colleagues in this project are John B. Edwards, Harold Guthrie, Sara Kelly, George Sadowsky and James Smith. Edwards achieved the computer implementation of the persons model reported in this paper. Guthrie is co-director of the project. Orcutt, Peabody and Caldwell are listed as authors of this paper because they had responsibility for developing the operating characteristics reported here--Peabody for birth and education, Caldwell for death, marriage, separation and divorce, and Orcutt with Kelly for the auxiliary macroanalytic model.

the persons model most of the family and household structure of the SEO file has been eliminated. A mother and her children constitute the "family" unit, and males are isolated individuals in the file.

In the following sections we will describe the structure of each operating characteristic for the persons model and then discuss the macro model that is being developed to close the model.[2]

Deaths

In the previous work [6] the annual probability of death for an individual was taken to be a function of his race, sex, age and the calendar year being simulated. Corrected central death rates covering the period 1933-54 were fitted to a logarithmic function; i.e., the death probability within each race-sex-age class was assumed to be decreasing by a fixed percent each year toward a limiting value of zero. However, mortality levels since 1954 have proven to be consistently higher than predicted by this two-parameter model.

Rather than re-estimate the same two-parameter model and incorporate subsequent (1955-1968) data, we have respecified the model by relaxing the assumption that mortality is declining toward zero in all classes and included an asymptote as a third parameter. Within each of 76 race-sex-age class we have estimated the following function using uncorrected death rates:

$$P(\text{death}_{it}) = a_o + a_1 e^{a_2(t-t_o)} \qquad (1)$$

where

$P(\text{death}_{it})$ = the central death rate in year t for an individual in age-race-sex category i

and $t_o = 1950$

is an arbitrary base year. The asymptote ($a_o$) could be interpreted as a measure of whatever mortality risk has remained constant over the past 36 years, a "floor" toward which mortality rates are tending.

Data from census years [3] reveal that within age-sex-race categories, substantial mortality differentials exist for the four marital states: single (never married), married, divorced and widowed. Using these data to calculate ratios of marital status-specific mortality within age-race-sex classes, and taking the mean of the ratios for the three points[3] for which we have data, enables us to generate a parameter with which we make the death probability a function of marital status.

The excellent research of Kitagawa and Hauser [5] provides estimates for national socioeconomic differentials in mortality in May-August, 1960.[4] We know of no subsequent re-estimates of these parameters which could serve to give some idea of their temporal stability so we have assumed that these parameters are constant.

Kitagawa-Hauser carried out separate estimates within race-sex classes. They also standardized for age, but presented the estimates only in terms of two age classes (25-64, 65 and over); consequently we do not know if there are significant age differences in the parameters which were not captured by their results. For whites, Kitagawa gives separate estimates for family members and for unrelated individuals, but since the differences are slight [5, p.10], we have chosen to apply the combined estimates across marital status categories. For all persons aged 65 and over, except white females, no clear and significant pattern of education differentials was detected. For all others a strong inverse relationship between years of school completed and mortality was noted and these differentials provide the basis for adding a fifth parameter to the death probability.

Finally, parity differentials were also examined in the above study [5, p.23]. These differentials were calculated for ever married white women 45 years of age and over. No consistent pattern was evident for those 65 and over. However, for those aged 45-64 there was a J-shaped relationship in which women with a parity of three had the lowest death probability. No separate tabulations for non-whites were presented. However, we have decided for the model that it would be less reasonable to leave out parity differentials for non-whites altogether than to assume that the same parameters hold for whites and non-whites, so we have done the latter. The parity parameters were standardized for age and education, to account for the known negative association between fertility and education, but they were not presented in education-specific form.

Thus, in the present version of our mortality operating characteristic the death probability is a function of the person's age, race, sex,

---

[2]More detailed descriptions of the operating characteristics and a discussion of simulation results are in preparation and will be available, upon request, from the authors.

[3]It is difficult to impute time trends to these ratios using only three points; so far we have not done so, though it is a trivial matter to make the parameter a function of time if evidence so suggests.

[4]We have not attempted to correct for possible seasonal biases. For a discussion of this and other possible biases see [4].

education (for all persons aged 25-64, and also for white females aged 65 and over), marital status, parity (for ever-married females aged 45-64), and the calendar year. Full interaction between age, race, sex and marital status is allowed. Education interacts with race, sex and (very roughly) age. Parity is sex-specific, marital status-specific (roughly) and age-specific (very roughly). The time trend is age, race and sex specific.

With respect to all independent variables, but particularly those that are in some degree manipulable by individuals or policymakers (e.g., education, parity, or marital status), the question of causality must be faced. For an individual in the model a change in marital status, years of school completed or parity gives rise to a sudden change in his or her death probability. Only in the case of marital status is this situation likely to generate unrealistic histories, since we do not apply the other differentials until after the age when most people will have completed their education and child-bearing. Undoubtedly the effect due to marital status overstates its causal importance and understates the extent to which the differentials are due to selection. But more important for our purposes is the aggregate question: does a change in socioeconomic composition (by marital status, education, or parity) of a race-sex-age class have an effect on the level of mortality rate for that class, or does it affect only the distribution within the class of the predicted rate. Largely because we have estimated the time trend within age-race-sex classes, we are reluctant to attribute additional temporal variations to changes in any of the other independent variables. Thus a tracking routine uniformly scales the education and marital status parameters up or down each year to arrive at an expected value of deaths consistent with predictions in each age-race-sex class.[5] However, since the parity differentials are relatively minor and we are inclined to attribute some causal importance to changes in parity, we have not scaled the parity parameters. Thus, changing parity composition is allowed to affect aggregate mortality rates, which is not the case for changing marital status and education composition.

Marriage

All unmarried persons over fourteen are assigned a probability of marriage for each year of the simulation. For never-married persons the probability of marriage is a function of race, sex, age, (by single years to age 50) and a trend factor. For divorced persons the probability of remarriage is generated as a function of sex, number of years divorced, and a trend factor. For widowed persons remarriage is a function of age and sex. No matching of marriage partners is carried out in the persons model, but there will be matching, according to various relevant characteristics, in the family model.

The most important missing variable at present in the marriage probability functions discussed above is education. Analyses in progress of SEO and other data should enable us to fill this gap. Such analyses might also yield parameters on significant differences in remarriage rates by number of living children.

Marriage Dissolution

Marriages can be dissolved in our model by divorce or death. Separation, a step prior to divorce that dissolves the family unit, is also simulated. Persons are widowed by assuming that each married person has a spouse of the same race, education, and marital status, of opposite sex and with a two year age difference (male older) and subjecting this shadow spouse to the death routine. In the family model the spouse will be present so we will have the advantage of knowing his or her particular characteristics.

Data concerning separation are scarce (especially for rates) and imperfect; e.g., half again as many women as men generally report themselves as separated [1, p.14] though the two numbers ought to be equal. However, substantial numbers of persons report themselves as separated (over 2 million in 1960) and sharp differentials exist by race. Five or six times as many non-whites, in percentage terms, report themselves as separated. Leaving out separation leads to an over-estimate of the number of intact families, particularly non-white families, and consequently, introduces error when, in the family model, we determine total family income. Thus, it seems preferable to deal with separation, even if ad hoc assumptions must be made, rather than ignore it completely.

Given the lack of data on the rates at which married persons separate and separated persons divorce (i.e., existing data generally skips separation; divorce rates are calculated by using married persons as the base) we are forced to make certain assumptions. We have chosen to use evidence on divorce rate differentials in constructing separation rate differentials. Accordingly, separation rates become a function of sex, duration of marriage, race, number of children, and a trend factor. To account for the fact that a much larger fraction of non-whites report themselves as separated, we assume that non-whites have a longer duration of separation. Thus, the probability of divorce is made a function of race and the length of separation. We do not at present allow reconciliations in the model. Having different functions for

[5] A mixed strategy could easily be implemented by adjusting the scaling factor. For example, moving it one-half the distance toward unity would have the affect of attributing some of the effect to selection and letting the rest reflect a causal influence.

separation and divorce gives us the capability of having some interaction between the two rates. We can use the traditional divorce rates to constrain the separation and divorce probabilities in the model.

Birth

In the previous work [6] the birth probability for a woman was taken to be a function of her age, parity and the interval from her last birth (or marriage). This procedure allocates the proper number of children and distribution of family sizes, but it does not capture the variation of family size with social class. Since our focus is on the distribution of inequality, and since the number of children that a couple has can have a large impact on their financial situation, we have devised a birth simulation that more closely represents the individual couple's approach to, and success with, family planning. We can then more adequately deal with the effects, for example, that unwanted or too closely spaced children can have on the financial status of low income families.

The simulation draws heavily on the various fertility surveys that have been conducted in the past fifteen years, including the Growth of American Family Studies [2], the Princeton Study [8], and the 1965 National Fertility Survey [7]. These studies, along with other demographic research, have provided a wealth of information about the distribution of desired number of children, the circumstances under which (married) women use contraception and how effectively they use it, the intervals between successive births, and fecundity impairments that may limit a couple's fertility. The model incorporates all these features of fertility.

The birth probability in year t for each woman, i, who is not using contraception is simply the couple's fecundability,[6]

$$P(\text{birth}_{it}) = f_{it}\, F(\text{Age}_{it}), \qquad (2)$$

which is a function of the age of the woman and her fecundity, $f_{it}$. For a fecund couple $f_{it} = 1$, and $f_{it}$ assumes values less than one for subfecund couples down to zero for a sterile couple. The monthly probability of conception is about 0.2 for women in their early to mid-twenties and then declines with age to zero at the onset of menopause. Most couples are fecund, but about 30% of American couples are subfecund to some degree, including slightly over 10% who are sterile.

[6] Demographers usually define fecundability as the probability that a fecund couple not in a period of post-partum amenorrhea will conceive in a month; since we will simulate in intervals of a year, we will stretch the definition to be the probability that the couple will have a live birth in a year.

Most couples try to control their fertility through the use of contraceptives. When birth control is being practiced the birth probability of Eq.(2) is reduced by a factor that measures the effectiveness of the contraceptive usage,

$$P(\text{birth}_{it}) = f_{it}F(\text{Age}_{it})[1-\text{Eff}_{it}]. \qquad (3)$$

A couple that is using a perfect contraceptive device will have $\text{Eff}_{it}=1$, while a couple that practices birth control so ineptly that their fertility is not reduced by its use will have zero Effectiveness. Effectiveness includes both the effects of the imperfections of the particular method used and the skill with which the couples use their method. In this country effectiveness is a function of race, education, and purpose (to limit fertility or to control timing of births) and length of use.

For the model we make the assumption that all non-married women use contraception. Illegitimate births will be allowed in the model by assigning an Effectiveness less than one to some unmarried women. The only married couples who will be assumed not to use contraceptives are those whose desired number of children is larger than the number of children they now have and whose interval from their last birth (or marriage) exceeds their desired minimum interval (if they want to control the timing of births). All other married women will be assumed to practice birth control and will have a value for Effectiveness imputed to them.

Thus the couple's desires about family size and spacing of births are used to control the woman's birth probability. To each woman we will impute a desired number of children and the length of the interval after a birth during which the couple uses contraception. Most families in America want two, three or four children. Religion accounts for more of the variance of the distribution of desired family size among couples than any other social variable, but it does not account for much, with Catholics desiring larger families than other religious groups. Socioeconomic status has little impact on desired fertility--except that women who work have much lower fertility than those who don't--and race a slight effect with Negroes wanting fewer children than other racial groups. Social class has more effect on desired spacing with upper class couples spacing their children further apart than lower class couples.

The birth simulation proceeds, then, by assigning to each married woman a desired family size, a desired minimum interval, a fecundity, and an effectiveness of birth control usage. At each pass a decision to use or not use contraception is imputed to each married woman by comparing her desired fertility with her actual fertility and comparing her elapsed interval with her desired minimum interval. Then a birth probability for the year is determined from Eq.(2) or (3) depending upon the use, or non-use, of contraceptives. If a birth occurs it is determined whether or not it is a multiple birth and a sex is assigned to each child born.

Education

The educational attainment of children provides the major mechanism in the model for examining intergenerational distribution of income and status. We deal only with educational attainment, the number of years completed or degree obtained, rather than with educational achievement, the grades obtained or what the student learned, for two reasons. First, it would be very difficult to simulate achievement in a national model, and the necessary data probably do not exist. Second, there is no evidence in the literature that achievement has much effect upon future earnings or status, which is our primary interest. (Achievement is, of course, a very important intermediate variable in determining attainment, but we feel we can capture the important socioeconomic differentials in attainment without including achievement.)

The data that we have accumulated indicates that our present provisions for educating children also turn out to constitute a fairly effective mechanism for maintaining class positions over successive generations. Children of upper class parents are much more likely to graduate from high school, to enter college, and then to graduate from college than are the children of the lower class. The cumulative effect of these inequities leads to a stratification system with far less intergenerational mobility than is commonly supposed.

Our simulation will concentrate on these effects of the parent's status upon the educational attainment of their children. We will also include other family effects such as number of children in the family, race, sex and age of the child. School and community variables have also been shown to be of importance in explaining attainment, but we do not have enough locational detail in the model to be able to incorporate these variables very effectively. We will, however, try to capture the differences between the four census regions and urban-rural differences.

To simplify the simulation we do not pay attention to each grade level that a person passes through. Instead we consider the following levels of schooling:

| n | school level |
|---|---|
| 0 | no schooling |
| 1 | grade school |
| 2 | 8th grade graduation |
| 3 | high school |
| 4 | high school graduation |
| 5 | college |
| 6 | college graduation |

We plan to include two year as well as four year schools in the college sector, and we hope to be able to include vocational schools as another alternative for high school graduates. We may also add graduate school as levels seven and eight.

Children will be entered into the first grade at ages five, six or seven. For children in school, at each level (n) there is a function that gives the probability of being retained at that level for a year, $P^R(n)$, of dropping out of school at that level, $P^D(n)$, or of advancing to the next level, $P^A(n+1|n)$. For children in grade school the retention function will be tested first to determine whether or not the child remains at that level for the next year. If he is retained, we exit from the education simulation. To further simplify the simulation all children will be automatically retained at the grade school level for eight years. At the beginning of the child's ninth year, $P^R(1)$ will be tested to see if he was retained an extra year in grade school. If he is not retained, then $P^D(1)$ and $P^A(2|1)$ are tested (with $P^D(1)+P^A(2|1)=1$). If the person advances to level 2, a test is made in the same pass to see if he enter, high school ($P^D(2)+P^A(3|2)=1$).

The procedure for high school and college is slightly different in that we allow a person to drop out of school each year; hence $P^D(n)$ is tested on each pass. For the first four years of high school or college $P^D+P^R = 1$, so that if leaving school is not imputed, the person is automatically retained for one more year. After four years of high school we will test $P^R(3)$ first to see if the person remained in high school a fifth year. If not, we test $P^D(3)$ to determine whether he dropped out before graduating or he graduated, $P^D(3)+P^A(4|3)=1$. Next we determine whether the person leaves school after graduation from high school or enters college.

If it is determined that the person leaves school, the amount of schooling obtained by the individual to that time becomes a permanent attribute and he never re-enters the schooling simulation. At the present time we do not deal with people who leave school for some period of time and then return to school.

Macroanalytic Model

The value of providing closure between the microanalytic model of individuals and families and the macro economy is two-fold. In the first place the microanalytic models under development need an environment in which to operate. In the second place economists think they know something about the control of some macro variables such as the percent unemployed and the rate of price change. It would be useful to trace out the impact of fiscal, monetary and other policies operated at the macro level on the behavior and well-being of individuals and families. The macroanalytic model being developed is intended to provide both an environment for the microanalytic models and a useful link to variables which can be controlled, or at least influenced, by available monetary and fiscal tools.

The simplest expedient for providing a needed environment for the model of the population of individuals and families would be to treat unemployment, real GNP, price level changes and fractions of GNP going to earned income and wealth holders as direct exogenous inputs. The disadvantage of this approach is that no explicit account is taken of the extensive multicollinearity of these variables. By leaving such variables entirely unconnected, the user of the microanalytic models would be given an entirely unrealistic view of the extent to which outcomes could be influenced by manipulation of policy tools at the macrolevel. The objective behind the auxiliary model was to take a useful step towards capturing the close inter-connectedness of household inputs from the macrolevel and still leave points at which policy assumptions could be entered either by alteration of target unemployment or by alteration of parameter values in appropriate operating characteristics.

In developing an auxiliary macroanalytic model extensive simplification has been achieved by assuming that the federal government can and will cause aggregate demand to vary so as approximately to control the fraction of the labor force which is unemployed. The advantage of this assumption is that if total aggregate demand actually is controlled by the federal government it becomes less critical and probably unnecessary to account for the role of non-household sectors in generating aggregate demand. Their behavior in this area is simply supplemented or offset as necessary to achieve a desired unemployment rate given past price movements. Of course this approach would not do for a model intended to be useful in guiding short-run stabilization efforts. It is hoped and expected to be useful for consideration of long-run consequences of policy measures.

Given the auxiliary macroanalytic model the capability exists for constructing links between micro events and the macro economy. For example, one might argue that age at marriage or the desired number of children are related to the over-all condition of the economy, in excess of the specific monetary effects at the individual level. Given evidence for such connections we could link, for example, the race-sex-age-specific first marriage probabilities to changes in total unemployment, the price index, etc. No such links have been incorporated into the operating characteristics developed to date. However, such questions will be explored as the family model develops.

The auxiliary model should be regarded as a first serious step in establishing useful links from monetary and fiscal policies into micro-analytic models of the population of individuals and families. It also is of interest in that it provides for and makes important uses of outputs of such microanalytic models as inputs into a macroanalytic model. This model has several deficiencies which hopefully can be reduced with additional effort. Perhaps the most serious of these is that the gap between what policy makers might do at the macrolevel and appropriate alteration of parameter values in the auxiliary model is still uncomfortably large.

## Implementation

The persons model has been implemented on a PDP-10 in an interactive mode. We are currently operating with a random, self-weighting initial population drawn from the 1967 SEO that consists of 500 interview units (families or single individuals) and 1553 persons. The sample was drawn in such a way as to control for the distribution of family types (female-headed families, single persons, etc.), and the age, race, and sex distribution of the sample was compared with the total SEO file to check the drawing procedure.

The birth rates, death rates, etc. obtained from the simulations are reasonable, but the initial population size is far too small to give meaningful population projections. We are using the small sample to check the computer logic of the model and the design of the operating characteristics. After completing this first rough stage of the model validation we will simulate with an initial population of approximately 10,000 households. We expect that simulations with this much larger population will give reasonable population projections under different assumptions about the future time trends of the parameters in the model.

## References


1. Hugh Carter and Paul C. Glick, Marriage and Divorce: A Social and Economic Study, Cambridge, 1970.

2. R. Freedman, P.K. Whelpton and A.A. Campbell, Family Planning, Sterility, and Population Growth, New York: McGraw-Hill, 1959.

   P.K. Whelpton, A.A. Campbell and J.E. Patterson, Fertility and Family Planning in the United States, Princeton, 1966.

3. Robert D. Grove and Alice M. Hetzel, Vital Statistics Rates in the United States, 1940-1960. Public Health Service Publication No. 1677, U.S. Government Printing Office, 1968.

4. Evelyn M. Kitagawa and P.M. Hauser, "Educational Differentials in Mortality by Cause of Death: United States, 1960," Demography, Vol. 5, No. 1, 1968, pp.318-353.

5. Evelyn M. Kitagawa, "Social and Economic Differentials in Mortality in the United States, 1960," paper delivered at Conference of International Union for Scientific Study of Population, London, 1969, to be published in the Proceedings.

6. G.H. Orcutt, M. Greenberger, J. Korbel, and A.M. Rivlin, Microanalysis of Socioeconomic Systems: A Simulation Study, Harper, New York, 1961.

7. Norman Ryder and Charles Westoff, to be published.

8. C.F. Westoff, R.G. Potter, Jr., P.C. Sagi, and E.G. Mishler, Family Growth in Metropolitan America, Princeton, 1961.

C. F. Westoff, R.G. Potter, Jr., and P.C. Sagi, The Third Child, Princeton, 1963.

9. U.S. Office of Economic Opportunity, "Surveys of Economic Opportunity," Codebook and other technical documentation.

10. U.S. Public Health Service, Vital Statistics of the United States, U.S. Government Printing Office, Washington, published annually since 1945.