CHAPTER 2
MICROANALYTIC SIMULATION MODELS

July 22, 1997

CHAPTER 2

MICROANALYTIC SIMULATION MODELS

I. INTRODUCTION

Microanalytic simulation models have played a major role in retirement income policy analysis, especially analysis of the levels and distribution of income. (Microanalytic simulation models are often referred to as microsimulation models. This practice is followed hereafter in this report.)

This chapter describes microsimulation analysis and the basic features of microsimulation models. It provides background for the description of each of the microsimulation models reviewed in this report.

The major microsimulation models which have been used in pension policy analysis are the Dynamic Simulation of Income Model (DYNASIM), developed and maintained at The Urban Institute, and the Pension and Retirement Income Simulation Model (PRISM), developed at ICF Inc., and currently maintained at The Lewin Group. The Cornell Simulation Model (CORSIM), a descendent of the original DYNASIM, being developed and maintained at Cornell University, is adding a social security sector and a representation of family assets, and has proposed adding employer pensions. These three models are described in detail in this report.

Several other microsimulation models are relevant to pension policy analysis. The PBGC Pension Insurance Management System (PIMS) model is a microsimulation model of firms, focusing on financial characteristics. Microsimulation models which have been applied to tax and transfer income analysis include TRIM (Transfer Income Model), MATH (Micro Analysis of Transfers to Households), HITSM (Household Income and Tax Simulation Model), and the U.S. Treasury Office of Tax Analysis Individual Income Tax Simulation Model. Microsimulation models and microdata analysis methodologies have also been applied to health care policy analysis. These include AHSIM (Agency for Health Care Policy Research Simulation Model), SPAM (Health Care Financing Administration Special Policy Analysis Model), Congressional Budget Office (CBO) modeling methodologies, Urban Institute TRIM-based health expenditures model, HBSM (Lewin Health Benefits Simulation Model), and models developed at Price-Waterhouse, Peat-Marwick, and the Economic Policy Institute. These models are described and reviewed in this report.

II. BASIC FEATURES OF MICROSIMULATION

The key feature of microsimulation is that it simulates the behavior of individual units, such as persons, families, firms, government agencies, or metropolitan governments. The basic components of a microsimulation model are individual units. The focus is the depiction of behavior of these units.

Most empirically implemented microsimulation models depict persons and families. These include DYNASIM, PRISM, CORSIM, TRIM, MATH, and HITSM. Other microsimulation models depict economic units that are closely associated with families. For example, the OTA model depicts tax filing units, AHSIM and SPAM depict health insurance units (the persons covered by a single health insurance policy, e.g. an individual plus dependents). There are some exceptions. The PBGC Pension Insurance Management System depicts firms. ICF developed a microsimulation model of hazardous waste sites to assess the adequacy of an environmental clean-up trust fund.

While microsimulation models depict individual behavior, they operate on a population database. They depict the state, characteristics, and behavior of each unit in the database. The database may be from a cross-section survey or from a longitudinal survey. A *cross-section survey* collects data about the members of a sample for a single period of time. A *longitudinal survey* collects data on a sample (usually referred to as a panel) of the same individuals for several periods. For estimation of relationships and for calibration and validation, a microsimulation model may use data from a time series of cross-sections, e.g. several similar cross-section surveys taken at periodic intervals (such as the March Current Population Survey for several years).

By simulating the state or behavior of each unit in a population sample, microsimulation models can be used to depict changes in aggregates, distributions, and averages. The behavioral equations and operating procedures of the model characterize the state, changes, and behavioral responses of individuals. For example, to estimate the labor supply response to a tax change, a microsimulation model would simulate the work response of each taxpayer in the sample. The aggregate supply response would be estimated by aggregating over all the (properly weighted) individuals in the sample.

Microsimulation models can estimate aggregate effects or aggregate changes by simulating the effects on and behavioral responses of each individual unit, then summing, tabulating, or aggregating over all the units in the sample. In concept a microsimulation model can provide time series estimates and projections of aggregate variables, if the individual components are represented. For example, a microsimulation model that represents the labor market experience of individuals could be used to project aggregate wage income, for all workers and for specified classes of workers. In addition to providing aggregate estimates, a microsimulation model can provide estimates of distributions of the population by various characteristics and states. Because a microsimulation model simulates and records information for each individual, it provides great flexibility in the specification of the groups, subaggregates, and tabulations which it can analyze.

Microsimulation can often draw on an analytical framework based on microeconomic theory or the sociology of individual behavior. Some argue that a strength of microsimulation is that it can draw on the well developed body of microeconomic theory and can estimate relationships and test hypotheses using the large samples of various independent microdatabases[1], which often provide thousands of observations.

Some microsimulation models are better articulated theoretically than others. Because of their size and complexity, most microsimulation models are developed and estimated one component at a time. The various components, or modules, may be developed by different groups of researchers. The modular approach through which most microsimulation models are estimated and constructed means often that independent theoretical frameworks and estimation techniques are applied to the estimation of each behavioral relationship. Sometimes important potential interactions may not be analyzed and depicted. (For example, an equation used to predict labor force participation may include variables indicating education and numbers of children, but the joint intertemporal decision regarding labor force participation, childbearing, and obtaining additional education may not have been modeled.) Within the same model, some components may be based on a rigorous, well developed theoretical structure, while other components are not. Virtually all microsimulation models contain many relationships which are ad hoc. It is not clear that microsimulation models in general have more rigorous theoretical foundations than more aggregative models. Important relationships in several existing microsimulation models are based on historically observed empirical relationships with little or no articulated bases in economic theory.

Although microdatabases provide large numbers of observations for empirical work, aggregate time series data[2] are also abundant. For many time series there are now at least 35 years of quarterly observations, which provide a large number of observations relative to the number of variables being estimated (degrees of freedom) in most time series models.

III. AGING METHODOLOGY

Microsimulation models operate by simulating the state and behavior of individual units and changes in those states and behaviors. The process of adjusting or assigning changes in states and behaviors is referred to as aging the data. There are two methods of aging: static aging and dynamic aging.

---

[1] A microdatabase is a file of observations on individual units, such as persons or families. Each record in the file may contain many pieces of information about an individual. Most microdatabases are obtained from cross-section surveys or longitudinal surveys. The monthly Current Population Survey and the Decennial Census are examples of microdatabases. A sample of Summary Plan Descriptions is a microdatabase on pension plans.

[2] Time series data refer to series of periodic observations on aggregate variables, such as the monthly unemployment rate over a period of one or more years, or the level of consumer expenditures each quarter for several years.

In **static aging** data records are adjusted by reweighting each record to correspond to projected or estimated changes in aggregate control variables, for example population estimates by age, race, and sex, and perhaps region. Most microsimulation models operate on a sample that is designed to be representative of a larger population. Each record in the sample has a weight, to correspond to the number of individuals in the total population that it represents. The sum of the weights equals the total population. (If the database corresponds to the entire population, such as a complete census, the weight on each record is one.) If the estimated size of the group to which that individual belongs is projected to change between the period for which the data were collected and the period being simulated, the weight of each individual in the group is changed directly to correspond to the change in the group size.

The second method of adjusting data in static aging is to change the data on each record directly, to correspond to projected changes in aggregate data. For example, hours worked and periods of unemployment experienced by each individual in the labor force can be changed to correspond to a projected change in the aggregate unemployment rate. Wages of each worker can be changed to correspond to projected change in average wages. Income from other sources for each individual can be adjusted directly to correspond to projected or estimated changes in total income and inflation. In static aging the data for any time period can be changed to correspond to projections for any other period, even one many periods removed, or to different scenarios for the same time period. Static aging adjusts the data of one cross-section database directly to produce another cross-section database corresponding to a different period or to different economic conditions or policy regimes in the same period.

In **dynamic aging** data are changed by simulating the individual changes in behaviors and conditions that occur for each individual in each period. For example, each year each person may die or age one year; each women of child-bearing age may give birth; each newborn is assigned gender; individual immigrants are added and emigrants are removed from the sample. The combination of all these individual demographic events changes the size and composition of the population each period. Individuals are simulated to enter, leave, or remain in the labor force based on their current and past economic and demographic characteristics; hours worked and lengths of unemployment are estimated for each worker, based on current and past economic, demographic, and labor market characteristics; wage rates are estimated for each worker. Based on the simulated labor market experience of each worker, aggregate labor force participation, unemployment, hours worked, and wages can be estimated for various groups and for the total population in each period. In dynamic aging the data on each record are adjusted for each period, one period at a time. Dynamic aging produces a longitudinal database of histories of the condition and behavior of each individual in each period of the simulation. These data can be reported and tabulated for any given period to produce a cross-section database for that period.

IV. STOCHASTIC FEATURES OF DYNAMIC MODELS

The large dynamic microsimulation models have important stochastic features, i.e. important behaviors and events are modeled probabilistically. Behavioral equations predict probabilities of being in a given state or experiencing a given event (such as birth, marriage, or death) or change (such as changing location, changing job), rather than the outcomes themselves. The equations which predict the probabilities of states or events are sometimes referred to as "operating characteristics." The potential outcomes may be binomial (live or die), multinomial (if married, either stay married, become widowed, become separated or divorced), or may be represented by a transition matrix (e.g., the probability of a worker in each industry in the current period being employed in that or each other industry in the following period).

For cross-section tabulations, having the probability of each event or state is sufficient. (The number of individuals in each state can be estimated by multiplying the predicted probability by the weight for that individual.) However, for updating the population so that it can be used for the following period's simulation, a predicted outcome for each individual for each event is required. Outcomes are assigned or designated to occur stochastically, usually using random numbers. For example, a mortality model may predict the probability of an individual dying in a given year as a function of his or her age, sex, race, marital status, health/disability status, and employment status. For a given individual, the predicted probability of death (the solution to the mortality equation given that individual's characteristics) may be, say, .0040. For that individual, a random number is drawn from a uniform distribution between zero and one. If that number is equal to or less than .0040, the individual would be designated to have died during that period. This procedure is repeated for every individual in the sample. For large groups of individuals with the same characteristics, the proportion of individuals who experience the event would be equal to the predicted probability. This technique is referred to as **Monte Carlo simulation**. For each individual in the sample, for each event or state that is modeled, an operating characteristic equation is solved for the probability of the event, a random number is drawn (generated by a computer) and, by comparing the random number to the predicted probability, the event is designated to occur or not. This rather unsophisticated but powerful technique, which requires numerous operations, is made possible by the use of modern computers.

V. AGGREGATE CONTROLS

By simulating the basic demographic and economic events and behaviors of each individual in a sample in each period and updating the entire sample period by period, a dynamic microsimulation model can simulate and project the aggregate size, composition, and economic condition of the population through time. Often, independent estimates or projections of the population and economic conditions are available. If the simulation is for a historic period, actual data may be available. If it is a projection for future years, independent projections may be available of the population size and composition (e.g. from the U.S. Bureau of the Census or from the Social

Security Office of the Actuary) or of macroeconomic variables (from macroeconomic forecasting models).

Often analysts desire to control or align the aggregate projections of a microsimulation model with such independent estimates.[3] This can generally be done by adjusting each of the initial probability estimates generated by the model for the events or states determining the outcomes to be controlled. For example, if the model initially predicts a population for a given age-sex-race group which is smaller than an independent estimate being targeted, because the predicted average mortality of the individuals in the group is higher than that in the targeted estimate, the probability of death of each individual in the group in the model can be reduced. In general the probability for each individual would be changed in the same proportion as the targeted change in the aggregate outcome.

Individual values may also be adjusted directly to align to aggregate controls. For example, if the average wage level resulting from a microsimulation is lower than the targeted level projected independently, the wage of each worker in the microsimulation can be adjusted proportionally to align the average with the external control.

Almost all dynamic microsimulation models are adjusted (or controlled) in this way to align to external projections of aggregate or group variables when used for policy analysis. This blurs, somewhat, the distinction between static and dynamic models. While dynamic models depict the particular events for each individual which generate demographic and economic outcomes, if the individual outcomes are then adjusted to match a specified external aggregate outcome, the end result may be the same as that of the direct adjustment process used in static aging. This also points out the current limitations of microsimulation as a forecasting technique.

Proponents of microsimulation argue that microsimulation should be relied on to provide distributional information, information about the distribution of outcomes over various groups in the population, even if the estimates for individuals have been adjusted to correspond to aggregate values estimated by more aggregative modeling methodologies. The basic features of the distribution are preserved, even if outcomes for all individuals are modified to achieve a specified aggregate outcome. Proponents argue that only by having independently simulated information for each individual can analysis of interactions among individual characteristics and program rules and features be done. Only by having information for individuals can the effects on various groups of the population be estimated. Such information is vital to analyze programs that affect individuals with different characteristics differently. This information that reflects interactions between individual differences and program features is valid, even if the outcomes for each individual have been adjusted to match a specified aggregate value.

Dynamic microsimulation simulates an individual history of the socio-economic state and behavior of each person. Each of these histories would be adjusted to be consistent with independent

---

[3] This issue arises only for dynamic models. Static models generally adjust the individual estimates directly to match aggregate projections.

aggregate projections of demographic and economic variables. These adjusted individual histories provide information that is useful for analysis of some programs and issues, including many issues concerning pensions and retirement income.

VI. STRENGTHS OF MICROSIMULATION MODELING

The microsimulation modeling approach has major strengths for policy research and analysis.

The basis in individual behavior provides both an analytical framework and abundant data for the estimation of behavioral relationships. It also provides richness and flexibility in the type of information available.

Microsimulation permits the depiction of interactions among program provisions and individual characteristics. This is true for both public sector income transfer and tax programs and private programs such as health insurance and pensions.

For many public policy issues, the distribution of costs and benefits over various groups in the population is important. Microsimulation permits distributional analysis. The individual detail provided permits the analyst to select the specification of distributions to be reported, and to vary that specification after the simulation has been completed.

Because microsimulation retains the covariances that characterized the original data it may capture important interactions which more aggregative or cell-based methodologies may miss.

Dynamic microsimulation simulates longitudinal information, such as individual work histories, which are useful for analysis of pension and social security policy.

VII. LIMITATIONS

Microsimulation modeling has important limitations.

Existing microsimulation models do not provide complete representations of all the important elements of the social or economic systems they depict. Household microsimulation models do not include firms, so they cannot depict the reaction of employers to changes in labor market conditions nor to changes in policy that affect labor costs, such as pension policy, social security or other taxes on labor, health care costs and health insurance. They also do not depict government behavior.

Microsimulation models are not general equilibrium models. Consequently, the solution values may not be internally mutually consistent. Microsimulation models represent only one side of each market. In household microsimulation models, for example, only the supply side of the labor market is represented, so there is no feedback on wages or other forms of employee compensation or job opportunities of changes in labor supply or in factors affecting labor costs. Changes in individual asset accumulation do not affect aggregate saving and investment or the rate of return to those assets. Changes in factors that could affect the demand or utilization of health care do not endogenously affect health care prices. Because of this limitation, it is impossible in current microsimulation models to determine prices or wages structurally. Price and wage equations can be characterized, at best, as *reduced form*, if not ad hoc[4].

Similarly, because there are no endogenous macroeconomic feedbacks in microsimulation models, systematic changes in individual behaviors simulated in the model that produce aggregate changes do not prompt the macro responses that may be expected. For example, microsimulation models cannot be used to simulate the effects of increased pension or retirement saving accumulations on aggregate investment, productivity, and income and the potential feedbacks on wages, profits, and pension costs. Microsimulation models cannot simulate the effects of tax incentives for savings (which may initially reduce tax revenues) on capital accumulation, income, and in turn on tax revenue (which may be increased by these second round macro effects).

The great dependence on micro data can be a liability. Microsimulation models are large, cumbersome, and expensive to develop, although the dramatic reduction in computation costs has greatly reduced simulation costs. The expense and time required to produce comprehensive and representative data bases effectively means that there are long lags between the period when the data were collected and when a model can be developed. While the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP) are frequently released within about a year of the date of collection[5], the complete National Medical Expenditure Survey was not released until nine years after it was collected in 1987. The 1989 cross-section Survey of Consumer Finances was released three years after it was collected, and the 1983-1989 longitudinal file was released six years after the last data were collected. In contrast, many quarterly macro time series are released one or two months after the month or quarter they represent. This, plus their generally smaller size, permits macro time series models to be re-estimated more frequently and perhaps remain more current than microsimulation models. Frequently, by the time a microsimulation model can be developed, it is uncertain if the relationships depicted in the model still are valid.

Nevertheless, for analysis of the distribution of outcomes over groups within the population, and the differential effects of programs on groups with different characteristics, microsimulation models are required. To provide individual histories, such as work histories required for social security and pension policy analysis, dynamic microsimulation models are needed.

---

4 Reduced form and structural models are discussed in Chapter 1.

5 This statement applies to the core SIPP file, containing basic demographic and income data. The SIPP topical modules have frequently been released with longer delays. For example, the topical module on family assets, collected February-May 1995, had not been released as of July 1997. Federal budget reductions have increased the delays in availability of many microdata sets.