

## APPENDIX A

### GLOSSARY OF MODELING TERMS<sup>1</sup>

Terms described in this appendix include: Aging (dynamic and static), Behavioral response (first- and second-round effects), Calibration, Distributional analysis (including gainers and losers tables), Dynamic model, Filing unit, Imputation (including hot deck), Matching (exact and statistical), Microsimulation model, Static model, Weighting. Terms are defined with reference to their application in social welfare policy modeling.

Aging - The process of updating a database to represent current conditions or of projecting a database for one or more years to represent expected future conditions.

Dynamic aging in general terms involves generating a database year by year through applying transition probabilities to the individual records in a cross-sectional database and recording the results of each year's simulation on the records. The result is an enhanced database that contains longitudinal histories; that is, values for each individual for each year of the simulation period. For any one year, the database can provide a cross-sectional representation of the population. The process includes creating new records for people who are simulated to be born and setting variables to null values for people who are simulated to die. For people in the sample each year, the process involves updating their age by one and changing many other variables according to outside, econometrically estimated transition probabilities and dynamic micro equations. For example, DYNASIM2 dynamically ages the following characteristics of the records in the starting database: birth, death, first marriage, remarriage, divorce, work disability, education, migration, wage rate, labor force participation, hours, unemployment, job change, industry movement, and pension coverage. Dynamic models typically calibrate their simulated longitudinal histories using aggregate population and economic growth assumptions from outside sources such as the Social Security Actuary's trust fund model. Dynamic aging is typically used to generate histories over a period of 30 or more years for such uses as modeling retirement income programs; however, the method can be used to generate histories for any length period, short or long.

Static aging in general terms involves adjusting the weights and selected other variables in a cross-sectional database to represent the population in a future year according to outside projections. For example, the aging routines of the MATH and TRIM2 models reweight the records to match projections of the population by age, race, sex, and household composition (projections produced by the Census Bureau or the Social Security Actuary from cell-based models are typically used); adjust the income variables to match projections of inflation and real income growth (projections from macroeconomic models are typically used); and adjust the labor force variables to match expected unemployment rates by age, race, and sex. The MATH and TRIM2 unemployment adjustment algorithm resembles dynamic aging techniques in that employed people are selected to experience unemployment (or vice versa), with other variables adjusted accordingly, according to transition probabilities estimated using panel data. Static aging is typically carried out for a short projection period such as 2 to 5 years; however, the method can

---

<sup>1</sup> This appendix is taken in part from Citro and Hanushek (1991), Appendix, pp. 302-309.

be used to generate a cross-sectional database for any year, no matter how far into the future, provided the needed population and economic projections are available from outside sources.

Behavioral response - The term used for a change in behavior of an individual decision unit, such as a family or hospital, in response to a policy change that, in turn, has feedback effects on program costs and recipients. For example, altering the level of cash or in-kind (e.g., Medicaid) benefits in the AFDC program may affect the work decisions of welfare recipients that, in turn, may increase or reduce AFDC costs and caseloads.

The immediate responses of individual economic units directly affected by the program change are termed "first-round" behavioral effects. There can also be "second-round" behavioral effects of a policy change; that is, effects that alter the nature of factor or product markets or the level and distribution of consumption, production, and employment in the economy or in a sector of it affected by the policy change. For example, a change in a transfer program that alters labor supply may change the wage rate in the labor market and therefore further change labor supply. In addition, in the short run, prior to an equilibrating change in the wage rate, the unemployment rate may be affected and displacement or replacement effects may occur.

Calibration - The process of adjusting simulation outputs to approximate control totals from outside sources. For example, yearly simulations from dynamic models such as DYNASIM2 and PRISM are calibrated to accord with the demographic and economic assumptions incorporated in the projections of the Social Security Actuary. Also, baseline simulations of income support programs from models such as MATH and TRIM2 are calibrated so that the simulated participants approximate selected totals and characteristics of recipients from program administrative data (including the IQCS). Calibration methods vary. For example, MATH compares tabulations of eligible food stamp units from a baseline run with tabulations of recipients from administrative data and selects the needed number of participants in each category up to the maximum number of eligible units (if a category has too few eligible units, excess participants will be selected from another category). TRIM2 uses a probit equation to select AFDC participants from among eligible units on the basis of characteristics such as expected benefit level plus several parameters that are adjusted over the course of several runs so that simulated participants approximate caseloads by state and several characteristics of the national caseload.

Distributional analysis - The term used for tabulations produced by microsimulation models showing the simulation results disaggregated by subgroups of the population, such as households by income class or geographic area. Often, microsimulation models produce tabulations of gainers and losers for alternative policy proposals; that is, which population groups would gain by a policy change and which would lose compared with the current program. For example, the model might produce tabulations of the number of, say, AFDC recipients under current law and one or more proposed alternatives by age of head (or other characteristic), showing, for each age category, the change in the number, plus or minus, for each alternative compared with the current program. Similarly, the model might produce tabulations of the change in average benefit (or tax) amount, up or down, for each category.

Dynamic model - Term used for microsimulation models that generate a database of longitudinal histories for a population sample through means of applying transition probabilities to individual records and then use these histories to simulate alternative policies. Such models are able to trace through the effects of demographic and economic processes and previous and proposed policy changes (such as raising the retirement age for social security). They draw heavily on behavioral research for their many transition probabilities; although, typically, they include relatively few feedback effects of behavioral changes in response to simulated changes in government programs and policies. Also, dynamic models typically represent a hybrid of static and dynamic aging techniques, using dynamic techniques for most but not all variables (e.g., DYNASIM2 includes static routines to supply values for assets, disability, and SSI variables to each record). (See also entry for Aging.)

Filing unit - The term given to the unit of analysis in microsimulation models of income support and tax programs, i.e., the unit entitled to apply for benefits or obligated to file a tax return. Filing units differ within and among programs; for example, tax filing units may include a married couple and their dependent children, or a single person living alone or with others; AFDC filing units may include a single parent and her or his dependent children, or a two-parent family in which the head is unemployed or disabled. Filing units often differ from families and households as defined in surveys; for example, an AFDC mother and her children may reside with other relatives who are not part of the filing unit. (See U.S. House of Representatives, 1990, for detailed information on filing unit definitions and other aspects of the eligibility rules for federal social insurance and public assistance programs.)

Hot deck - See entry for Imputation.

Imputation - The process of assigning values to records in a database that are missing values for one or more variables, because of "item nonresponse," that is, nonresponse to a survey (or nontranscription of an item in an administrative records system), or because the variable was never included in the survey or administrative records system.

Imputation procedures for item nonresponse range from the very simple to the very complex. A simple procedure is to impute the mean value for all people who responded to a particular item to all records that are missing the item. Slightly more complex variants are to impute a mean modified by a stochastic error term, or to impute means, with or without error terms, to categories of nonreporters. The Census Bureau uses very complex item nonresponse imputation methods for household surveys such as the CPS and SIPP, including the "hot deck" and what it refers to as "statistical matching." Hot deck methods assign a "nearest neighbor" value. That is, the data records are sorted by geographic area, processed sequentially, and reported values used to update ("hot deck") matrices of characteristics. A record with a missing item has the most recently updated value assigned from the appropriate matrix (for example, a matrix of earnings for people with specified demographic and occupational characteristics). The Census Bureau's statistical match procedure for item nonresponse (usually for whole groups of items) involves indexing the records by various characteristics that are available for both respondents and nonrespondents and searching for the respondent donor who best matches the nonrespondent host.

Imputation procedures for items not collected also vary from the simple to the complex. A simple procedure is to impute a mean amount for a missing variable based on tabulations from another data source. A more complex procedure is to use another data source to estimate regression equations that include independent variables common to both sources. Then, these equations are run with the estimated coefficients and the values of the independent variables in the records requiring imputation to estimate values of the dependent variable to impute to those records.

Matching - The process of appending entire records (or subsets of variables) from one or more donor files to a host file to obtain values for items not collected for the host file, a procedure that is generally used when large numbers of variables are involved.

Exact matching uses a unique identifier common to the data sets being matched, such as social security number (SSN). Other common information, such as age and sex, is also typically used to validate the quality of the match.

Statistical matching is carried out on two or more data sets when they share variables in common such as age, sex, and income but lack a common unique identifier or come from nonoverlapping samples. In some cases, statistical matches have been performed when it was theoretically possible but not feasible, for confidentiality or other reasons, to carry out exact matches. Statistical matching is a complex procedure that classifies records in two files by variables that they share in common, then uses an algorithm to select the best match from the donor file for each host record and extracts variables from the donor file to attach to the host file records. The validity of a statistical match rests on the assumption of conditional independence, namely that all of the information about the relationship between the variables that are unique to the donor file(s) and the variables that are unique to the host file is contained in the common set of variables.

Microsimulation model - Term used for social welfare policy models that operate at the individual decision unit level, whether that be an individual or family for an income-support program model or a hospital for a medical care payment model or a corporation for a corporate tax model. Microsimulation models essentially conduct program experiments (simulations) on large samples of microdata for individual decision units. In very general terms, the first step, which serves the same function as the control group for an experiment, is to prepare a database representing the situation in the absence of a program change. The second step is to simulate the program change and its impact, while the third step is to summarize the differences between the baseline and alternative program databases. Microsimulation models typically include routines to generate the database, routines ("accounting functions") to mimic the rules of government programs, and routines to produce tabulations of the simulation results (or output files for tabulations by another software package such as SAS). They may also include routines to simulate behavioral responses to proposed program changes. In simulating any type of behavior (whether it be demographic or economic behavior, such as marriage or job change, in generating a database using dynamic techniques; basic program participation behavior; or additional behavioral responses to program changes), microsimulation models are characterized by the use of probabilistic (Monte Carlo or stochastic) rather than deterministic techniques. For example, in implementing a program participation decision, the model would draw a probability for each decision unit at random,

compare that probability to an estimated participation probability for the particular type of unit, and, if the former probability is less than the latter, designate the unit to participate, otherwise not. (See entries for Dynamic model and Static model.)

Second-round effects - See entry for Behavioral response.

Static model - Term used for microsimulation models that operate on a database representing a cross-section of the population at a point in time. Such models typically simulate the direct effects of policy changes, that is, the effects assuming full implementation of the program changes without any feedback effects due to behavioral responses; however, they can also simulate behavioral responses to program changes. Static models may be used to generate estimates for future years, in which case they employ static aging techniques to generate a cross-sectional database representing the baseline program in the future year, subsequently using the aged database to conduct simulations of program alternatives. (See also entry for Aging.)

Weighting - The process of assigning weights (factors) to observations in a sample survey so that the weighted count of all observations will approximate the total population. In order to take account of features of the sample design and to attempt to minimize bias and variance in the weighted estimates, the weights for household surveys such as the CPS or SIPP typically represent the product of several factors, including: a factor for the probability of selection (this factor is the inverse of the sampling fraction); adjustment factors for household nonresponse; adjustment factors to reduce the variance among primary sampling units; and adjustment factors so that the weighted counts approximate estimates of the total civilian, noninstitutionalized population by age, race, Hispanic origin, and sex (developed from the previous decennial census updated by vital records).

**LIST OF ACRONYMS**

AFDC	Aid to Families with Dependent Children [program] (U.S. Department of Health and Human Services).
AHCPR	Agency for Health Care Policy and Research (formerly the National Center for Health Services Research--NCHSR) (U.S. Department of Health and Human Services).
AHS	American Housing Survey (U.S. Department of Housing and Urban Development).
ASPE	[Office of the] Assistant Secretary for Planning and Evaluation (U.S. Department of Health and Human Services).
BEA	Bureau of Economic Analysis (U.S. Department of Commerce).
BLS	Bureau of Labor Statistics (U.S. Department of Labor).
CASE	Computer-Assisted Software Engineering.
CBO	Congressional Budget Office
CGE	Computable general equilibrium [models]; see Chapter 6.
CES	Consumer Expenditure Survey (BLS).
CPI	Consumer Price Index (BLS).
CPS	Current Population Survey (BLS and Census Bureau); see entry above.
DYNASIM2	Dynamic Simulation of Income [model] 2 (Urban Institute); see entry above.
FSA	Family Support Act of 1988; Family Support Administration (U.S. Department of Health and Human Services).
FNS	Food and Nutrition Service (U.S. Department of Agriculture).
GNP	Gross National Product (BEA).
HANES	Health and Nutrition Examination Survey (NCHS).
HBSM	Health Benefits Simulation Model (Lewin/ICF, Inc.); see Chapter 8.
HCFA	Health Care Financing Administration (U.S. Department of Health and Human Services).
HHS	[U.S. Department of] Health and Human Services.
HIS	Health Interview Survey (NCHS).
HITSM	Household Income and Taxation Simulation Model (Lewin/ICF, Inc.); see entry above.
IQCS	Integrated Quality Control System; see entry above.
IRA	Individual retirement account.
IRS	Internal Revenue Service (U.S. Department of the Treasury).
ISDP	Income Survey Development Program (ASPE and SSA); see entry above for SIPP.
JCT	Joint Committee on Taxation (U.S. Congress).
KGB	Kasten, Greenberg, and Betson [model] (ASPE).
MATH	Micro Analysis of Transfers to Households [model] (MPR); see entry above.
MDM	Macroeconomic-Demographic Model (Lewin/ICF, Inc.)
MPR	Mathematica Policy Research, Inc.

MRPIS	Multi-Regional Policy Impact Simulation [model] (Social Welfare Research Institute, Boston College); see entry above.
MS-DOS	Microsoft Disk Operating System [for personal computers].
NCHS	National Center for Health Statistics (U.S. Department of Health and Human Services).
NCHSR	National Center for Health Services Research (now the Agency for Health Care Policy and Research--AHCPR) (U.S. Department of Health and Human Services).
NIA	National Institute on Aging (U.S. Department of Health and Human Services).
NIPA	National Income and Product Accounts (BEA).
NMCES	[1977] National Medical Care Expenditures Survey (NCHSR with NCHS).
NMCUES	[1980] National Medical Care Utilization and Expenditures Survey (NCHS with HCFA).
NMES	[1987] National Medical Expenditure Survey (AHCPR with HCFA).
OASDI	Old Age, Survivors, and Disability Insurance (social security) (SSA).
OBRA	[1981] Omnibus Budget and Reconciliation Act.
OECD	Organization for Economic Cooperation and Development.
OMB	[U.S.] Office of Management and Budget.
OTA	Office of Tax Analysis (U.S. Department of the Treasury).
PBJI	Program for Better Jobs and Income (Carter Administration).
PC	Personal computer.
POPSIM	Population simulation model; see Chapter 11.
PRISM	Pension and Retirement Income Simulation Model (Lewin/ICF, Inc.); see entry above.
PSID	Panel Study of Income Dynamics (Institute for Social Research, University of Michigan).
RHS	[Longitudinal] Retirement History Survey (SSA).
RIM	Reforms in Income Maintenance [model].
SIPP	Survey of Income and Program Participation (Census Bureau); see entry above.
SOCSIM	[Demographic]-social simulation model; see Chapter 11.
SSA	Social Security Administration (U.S. Department of Health and Human Services).
SSI	Supplemental Security Income [program] (U.S. Department of Health and Human Services).
SOI	Statistics of Income [Division] (Internal Revenue Service, U.S. Department of the Treasury); Statistics of Income [Samples of tax returns] (see entry above).
TRA	[1986] Tax Reform Act.
TRIM2	Transfer Income Model 2 (Urban Institute); see entry above.