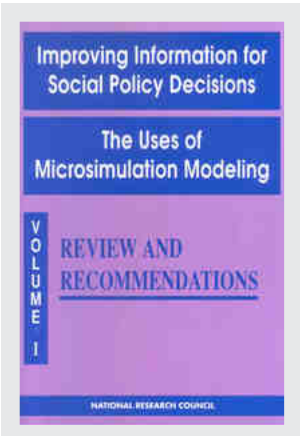
The National Academies of SCIENCES ENGINEERING MEDICINE

THE NATIONAL ACADEMIES PRESS

This PDF is available at http://nap.edu/1835

SHARE



GET THIS BOOK

FIND RELATED TITLES

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review and Recommendations

DETAILS

360 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-04541-4 | DOI 10.17226/1835

CONTRIBUTORS

Panel to Evaluate Microsimulation Models for Social Welfare Programs, Committee on National Statistics, National Research Council

Visit the National Academies Press at NAP.edu and login or register to get:

- Access to free PDF downloads of thousands of scientific reports
- 10% off the price of print titles
- Email or social media notifications of new titles related to your interests
- Special offers and discounts

Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. (Request Permission) Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

Copyright © National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions

The Uses of Microsimulation Modeling
VOLUME I
REVIEW AND RECOMMENDATIONS

Constance F. Citro and Eric A. Hanushek, Editors

Panel to Evaluate Microsimulation Models for Social Welfare Programs
Committee on National Statistics
Commission on Behavioral and Social Sciences and Education
National Research Council

National Academy Press
Washington, D.C.1991

Copyright National Academy of Sciences. All rights reserved.

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Frank Press is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Samuel O. Thier is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Frank Press and Dr. Robert M. White are chairman and vice chairman, respectively, of the National Research Council.

The project that is the subject of this report was supported by funds from the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, and the Food and Nutrition Service of the U.S. Department of Agriculture.

Library of Congress Catalog Card No. 91-62261
International Standard Book Number 0-309-04541-X
Additional copies of this report are available from: National Academy Press 2101 Constitution Avenue, NW Washington, DC 20418

S400
Printed in the United States of America

Copyright National Academy of Sciences. All rights reserved.

PANEL TO EVALUATE MICROSIMULATION MODELS FOR SOCIAL WELFARE PROGRAMS

ERIC A. HANUSHEK *(Chair),* Department of Economics, University of Rochester

DAVID M. BETSON, Department of Economics, University of Notre Dame

LYNNE BILLARD, Department of Statistics, University of Georgia

SHELDON DANZIGER, Institute of Public Policy Studies, University of Michigan

EUGENE P. ERICKSEN, Department of Sociology, Temple University

THOMAS J. ESPENSHADE, Office of Population Research, Princeton University

HARVEY GALPER, KPMG Peat Marwick, Washington, D.C.

LOUIS GORDON, Department of Mathematics, University of Southern California

KEVIN M. HOLLENBECK, W.E. Upjohn Institute for Employment Research, Kalamazoo, Mich.

GORDON H. LEWIS, School of Urban and Public Affairs, Carnegie Mellon University

ROBERT MOFFITT, Department of Economics, Brown University

GAIL R. WILENSKY, Health Care Financing Administration, U.S. Department of Health and Human Services*

MICHAEL C. WOLFSON, Analytical Studies Branch, Statistics Canada

CONSTANCE F. CITRO, *Study Director*

MICHAEL L. COHEN, *Consultant*

CHRISTINE M. ROSS, *Research Associate*

AGNES E. GASKIN, *Administrative Secretary*

* Served until January 1990

Copyright National Academy of Sciences. All rights reserved.

COMMITTEE ON NATIONAL STATISTICS

BURTON H. SINGER *(Chair),*Department of Epidemiology and Public Health, Yale University

NORMAN M. BRADBURN,National Opinion Research Center, University of Chicago

RONALD S. BROOKMEYER,Department of Biostatistics, Johns Hopkins University

MARTIN H. DAVID,Department of Economics, University of Wisconsin

ANGUS S. DEATON,Woodrow Wilson School of Public and International Affairs, Princeton University

CLAUDIA D. GOLDIN,Department of Economics, Harvard University

LOUIS GORDON,Department of Mathematics, University of Southern California

ROBERT M. HAUSER,Department of Sociology, University of Wisconsin

GRAHAM KALTON,Survey Research Center, Institute for Social Research, University of Michigan

WILLIAM A. MORRILL,Mathtech, Inc., Princeton, New Jersey

DOROTHY P. RICE,Department of Social and Behavioral Sciences, School of Nursing, University of California, San Francisco

JOHN E. ROLPH,The RAND Corporation, Santa Monica, California

DONALD B. RUBIN,Department of Statistics, Harvard University

KENNETH W. WACHTER,Department of Statistics, University of California, Berkeley

MIRON L. STRAF, *Director*

FLORENCE E. WOLF, *Administrative Assistant*

Copyright National Academy of Sciences. All rights reserved.

Contents

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Contents Volume II: Technical Papers

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Acknowledgments

At the outset, I wish to thank the members of the Panel to Evaluate Microsimulation Models for Social Welfare Programs for their generous contributions of time and expert knowledge. Several panel members served ably as chairs of working groups on broad issues. Other members participated extensively in the analysis of the panel's Transfer Income Model 2 (TRIM2) validation experiment. Still others contributed papers (in Volume II) or sets of notes on specific topics that greatly helped in preparation of the panel's report. Overall this was an exceptionally hard-working panel that strove unstintingly to grapple with a complex set of problems covering a very broad terrain of policy concerns, models, and data. The members thought and worked hard to develop recommendations that would enhance the quality of policy analysis tools and thereby the quality of national debate about important social welfare issues. It has been a genuine pleasure to work with them.

Of course, anybody who has ever had contact with an effort such as this realizes that the quality of the report—to say nothing about the enjoyment and satisfaction of participating in the project—depends directly on the study director. In this case, speaking both personally and for the panel, I can say absolutely that Constance Citro was the irreplaceable input to our effort. There is little doubt that we would have produced a very different report, one not nearly so strong, if Connie had not been involved. Her substantive knowledge, her prodigious feats of writing at the word processor, and her calm, easygoing nature each contributed to the activity.

The panel is also indebted to Michael Cohen, who served as the panel's

Copyright National Academy of Sciences. All rights reserved.

statistical consultant throughout the project. He prepared papers reviewing statistical matching techniques for developing microsimulation databases, sample reuse techniques for estimating the variance in microsimulation model estimates, the previous literature of microsimulation model validation studies, and (in collaboration with several members of the panel) the results of the panel's TRIM2 validation experiment. Mike participated actively in the panel's deliberations and contributed materially to the quality and depth of the panel's review of model validation issues.

Christine Ross served as research associate for the panel during its first year and made extremely valuable contributions to the study. She prepared materials describing the characteristics of a number of static and dynamic microsimulation models and a background paper comparing two dynamic models. She conducted a number of interviews with agency staff on their use of microsimulation models. Overall, Chris contributed enormously to helping the panel organize its work and gather the information it needed to carry out the study.

The staffs of our sponsors, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) of the U.S. Department of Health and Human Services (HHS) and the Food and Nutrition Service (FNS) of the U.S. Department of Agriculture, were extremely helpful in providing information about their use of microsimulation models for policy analysis and in making available to us the expertise and resources of their modeling contractors. We would like to thank particularly certain individuals for their assistance. Joan Turek-Brezina, Director for Technical and Computer Support for ASPE and the HHS project officer for the panel, contributed generously of her time and support in every aspect of the panel's work. Robert Dalrymple, the FNS project officer for the panel, also devoted time and energy to the project. ASPE staff, including Paul Gayer, M. Eugene Moyer, William Prosser, Reuben Snipper, and Daniel Weinberg (now with the Census Bureau), presented informative case studies of the use of microsimulation models at the panel's first meeting, as did Steven Carlson, Director of Analysis for the Office of Analysis and Evaluation in FNS. Reuben Snipper also provided detailed information about the techniques used to produce estimates for the policy debate that led to the Family Support Act of 1988.

We would also like to thank staff members of the Congressional Budget Office who shared their knowledge and insights about policy analysis techniques. In particular, we owe thanks to Janice Peskin, who provided detailed information about the procedures used to develop estimates for the Family Support Act, and to Nancy Gordon and Sandra Christensen who commented on models for health care policy issues. We would also like to thank James Cilke of the Office of Tax Analysis of the U.S. Department of the Treasury, who provided useful information on microsimulation models for tax policy.

Staff members of firms involved in microsimulation modeling provided invaluable assistance to the panel through briefings and sharing documentation

Copyright National Academy of Sciences. All rights reserved.

and other materials about their models with the panel. We would like to acknowledge in this regard the valuable contributions of Harold Beebout and Pat Doyle of Mathematica Policy Research, Inc., and of David Kennell and John Sheils of Lewin/ICF, Inc. We would like to thank particularly Richard Michel, Director of the Income and Benefits Policy Center at the Urban Institute, and his staff, including Sheila Zedlewski, Linda Giannarelli Paul Johnson, and Margaret Moore. Dick and Sheila not only made a number of presentations to the panel on static and dynamic models, but helped the panel design a validation experiment with TRIM2 that was a centerpiece of the panel's work, providing a wealth of information for the panel's deliberations. The experiment was ably carried out by Linda, Margaret, and Paul.

We are indebted to many other people who made presentations to the panel or prepared background papers at our behest. These include Gary Burtless of the Brookings Institution, Thomas Grannemann of the University of Colorado, and Robert Strauss of Carnegie Mellon University, each of whom prepared a paper on modeling behavioral effects of program changes; Steven Caldwell of Cornell University, who prepared a paper on techniques of aging microsimulation model databases; Deborah Chollet of Georgia State University, who prepared a paper on microsimulation models for health care policy issues; Paul Cotton of Fulcrum Technologies, Inc., and George Sadowsky of New York University, who collaborated on a paper about future computing environments for microsimulation models; Donald Fullerton of the University of Virginia, who briefed the panel on computable general equilibrium models; and Barry Bluestone of the University of Massachusetts-Amherst and John Havens and Alan Clayton-Matthews, both of Boston College, who briefed the panel on the Multi-Regional Policy Impact Simulation (MRPIS) model.

The panel also benefited greatly from a dinner meeting and discussion with congressional staff members who work with and interpret the information developed by policy analysts for the legislative debate, including William Hoagland, minority staff director for the Senate Budget Committee; Wendell Primus, chief economist of the House Ways and Means Committee; and Randall Weiss, former chief economist of the Joint Committee on Taxation and now with Deloitte and Touche.

The panel is very grateful to Eugenia Grohman, Associate Director for Reports of the Commission on Behavioral and Social Sciences and Education, for invaluable assistance in helping the panel organize a large volume of technical material into a coherent and readable report, as well as for her fine editorial work. We would also like to thank members of the Commission on Behavioral and Social Sciences and Education and the Committee on National Statistics who reviewed the report and proffered valuable comments.

Agnes Gaskin served ably as the administrative secretary for the panel. She made admirable logistical arrangements for the large number of meetings held by the panel—nine plenary and six working group sessions over a two-year

Copyright National Academy of Sciences. All rights reserved.

period. She also coped cheerfully and competently with multiple rounds of revisions to sections of this report and to the papers comprising Volume II.

Overall, the panel was ably assisted in its endeavors at every stage of the process by many, many people, whom we are honored to thank.

Eric A. Hanushek, *Chair*

Panel to Evaluate Microsimulation Models for Social Welfare Programs

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Summary

Since the inception of the U.S. federal system in 1789, decision makers in the executive and legislative branches have sought information to help make choices among alternative public policies. However, throughout most of the nation's history, the supply of policy information has been limited and the demand for it sporadic and ad hoc in nature.

Beginning in the 1960s, quantum improvements in data sources, socioeconomic research, and computing technology made it possible to supply information of much greater depth and breadth to the policy process. In turn, the activist posture of the federal government during that period both stimulated the production of policy research and analysis and drew on its results. At one end of the process, policy research helped identify problems and move them onto the federal agenda; at the other end, it contributed to an understanding of the successes and failures of enacted programs. At the middle stage of the process, in which legislative initiatives are debated, the role of information about the costs and benefits of alternative proposals became institutionalized.

Today, the policy community in Washington takes for granted that neither the administration nor Congress will consider legislation to alter any of the nation's expenditure programs or the tax code without looking closely at "the numbers." Often, these numbers are the product of team efforts to apply formal computerized modeling techniques and large-scale databases to the task of estimating the impact of alternative policies. The kinds of formal models that are used for policy analysis, defined as the production of estimates of the budgetary and population impacts of proposed program changes, vary

Copyright National Academy of Sciences. All rights reserved.

widely. They include large-scale macroeconomic models, single-equation time series models, cell-based models of population groups, econometric models of individual behavior, and large-scale microsimulation models (and, of course, these approaches are frequently supplemented, or sometimes supplanted, by a range of less formal means of developing policy information).

Despite the widespread use of formal models to provide information to the legislative debate, neither the policy analysis tools employed nor the estimates they produce have been subject to much explicit evaluation of their utility or accuracy. Two years ago, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) in the U.S. Department of Health and Human Services and the Food and Nutrition Service (FNS) in the U.S. Department of Agriculture asked the Committee on National Statistics at the National Research Council to convene a panel of experts. They asked that the panel evaluate microsimulation based policy models, such as TRIM2 (Transfer Income Model 2) and MATH (Micro Analysis of Transfers to Households). ASPE, FNS, and other agencies have used microsimulation models for many years to estimate the impacts of proposed changes in social welfare programs, including programs for income support for the poor, retirement income support, and provision of health care, as well as in tax laws. Models of this class were first developed for policy analysis in the late 1960s, but have not been the focus of a major evaluation since a study by the General Accounting Office in 1977.

Our panel sought to evaluate microsimulation models within a broad context. Microsimulation models bring important strengths to policy analysis, but they are far from the only useful type of analytical tool. Moreover, while having unique characteristics, they share aspects in common with other classes of models. Most important, some of the major problems confronting microsimulation models today plague other kinds of policy analysis tools as well.

Below we summarize first our findings and recommendations that apply to policy models generally (from Part I of our report) and then those about the current state of microsimulation modeling specifically (from Part II). The text of all of our recommendations follows, keyed to the chapter in which they appear in the body of the report.

IMPROVING THE TOOLS OF POLICY ANALYSIS: INVESTMENT PRIORITIES

We identified two major deficiencies that demand attention if policy models, of whatever type, are to provide cost-effective information to the legislative debates of the future. The first problem (one of long standing) is lack of regular and systematic model validation. Ingrained patterns of behavior on the part of both decision makers and policy analysts have led to systematic underinvestment in the validation task. The second problem (of more recent

Copyright National Academy of Sciences. All rights reserved.

origin) is underinvestment and consequent deterioration in the scope and quality of needed input data for policy models.

Validation

Model-based estimates of the costs and population effects of proposed policy changes, although certainly not the only type of information used in the legislative process, are regularly consulted and often play a pivotal role in the debate. The estimates produced by the TRIM2 model of the high costs of mandating a federal minimum benefit standard for the Aid to Families with Dependent Children (AFDC) program helped kill this provision in the debate over the Family Support Act of 1988. The estimates produced by the tax policy microsimulation model operated by the Joint Committee on Taxation and the Office of Tax Analysis in the U.S. Department of the Treasury were critical in shaping the 1986 Tax Reform Act.

Given the importance of these and other policy estimates, it is essential that the legislative debate have available, in addition to the estimates themselves, an assessment of their quality. Any estimate, whether coming from a rough back-of-the-envelope calculation or produced by one or another type of formal model, will inevitably contain errors and be subject to uncertainty—from sources such as sampling variability and errors in the input data, as well as errors in the specification of model components. An obvious question to ask is, How reliable is the estimate? For instance, can the policy analyst be reasonably confident that the estimated cost of a proposed policy change of, say, $25 billion lies within a relatively narrow bound, such as, say, $22 billion to $28 billion? Or, given the limitations of available knowledge, must the analyst acknowledge that the likely range is much wider—say, from $1 billion to $49 billion—and hence that the estimate is of much less utility as a guide to decision making? Another obvious question to ask is, What is the track record of the model that produced the estimate? Under reasonably similar conditions, has the model done a good job or a poor job in projecting actual outcomes?

Despite the clear need, it is rare, on the one hand, for questions about the quality of policy estimates or the track record of modeling tools to be asked by decision makers and, on the other hand, for information about the uncertainty surrounding policy estimates to be provided to the policy debate. We identified many reasons for this state of affairs.

First of all, the task of evaluating the results of policy analysis is quite hard. There are many sources of uncertainty. Moreover, a basic difficulty is that almost all policy analyses involve conditional rather than unconditional forecasts: that is, they are designed to answer the question of "what if." What will be the effect on program costs and caseloads if a national minimum AFDC benefit standard is set at one or another level? What will be the effect on tax revenues and investment behavior if rich households are assessed a surtax of

Copyright National Academy of Sciences. All rights reserved.

one or another percentage? In many cases, none of the hypothetical policy alternatives that were analyzed during the course of a debate may be enacted, so that no data ever become available against which to check the validity of the estimates. (In contrast, unconditional forecasts, such as those made with macroeconomic models of expected growth in the gross national product and other economic aggregates for the next quarter or year, can be and often are checked against reality.) Even in cases in which an analyzed policy is enacted, it is likely to be difficult to distinguish among sources of error—for example, an error in projected participation rates for a program that occurs because of poor understanding of the behavior of program participants as opposed to bad forecasts of the overall state of the economy.

The policy process itself exacerbates the problem. Typically, the debate on a policy issue takes place in an environment of time constraints and distractions from competing debates. Analysts are usually under great pressure to prepare a large number of estimates for many different program variants. Consequently, they have little time for such tasks as documenting and evaluating the quality of their estimates. Finally, they have little incentive to find time to develop measures of uncertainty, given the desire of decision makers for precise numbers that "add up"—a desire that is particularly strong in the current constrained fiscal environment in which new program expenditures must be offset to the dollar by new revenues or cuts in other programs.

Despite the difficulties in validating estimates of the impacts of proposed policy alternatives and ingrained institutional behaviors, we believe that it is essential for users and producers of policy information to elevate validation to a priority task. Our message to users, including decision makers and their staffs, is that they must systematically demand information on the level and sources of uncertainty in policy analysis work. It is in both their short-run and their long-run interests to do so. In the short run, users need information about uncertainty for several purposes: to evaluate competing estimates, for example, from congressional and executive branch agencies; to determine how much weight to give to the "numbers" in making policy choices; and to determine when it no longer makes sense to fine-tune a policy proposal because the available information cannot reliably distinguish among alternatives. In the long run, they need information about uncertainty to help set investment priorities for policy analysis tools and databases that are most likely to improve the quality of critical estimates.

Recognizing the difficulty of changing the behavior of decision makers, we urge the heads of policy analysis agencies to take the lead in working to ensure that information on uncertainty becomes available as a matter of course for the estimates their agencies produce. Agency heads should set and enforce standards that validation be part of the policy analysis work of their staffs; allocate staff and budget resources to the validation task; support efforts by their staffs to educate the staffs of decision makers about the need for information on the

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

quality of the estimates and how to interpret such information; and back up their staffs when time constraints and demands for certainty threaten to undercut the validation effort.

Policy analysis work, whether conducted in-house or by contractors, should always include some type of validation effort that, at a minimum, develops approximate estimates of uncertainty in the results and the main sources of this uncertainty. For microsimulation and other types of large, complex models, recently developed computer-intensive techniques make it feasible to develop estimates of variance from sources such as sampling variability in the primary database and other inputs. In addition, for large-scale, ongoing research and modeling efforts, the agencies should let *separate contracts* with independent analytical suppliers for evaluation. The reason for independent contracts is to ensure objectivity and minimize the likelihood that the evaluation will be sacrificed to the need for immediate results to feed to the policy debate.

The focus of these independent evaluation studies should be on two major types of validation that provide important information for determining priority areas for future investment in policy analysis tools. The first type is sensitivity analysis, which involves running alternate versions of one or more model components and data inputs to determine the effects on the estimates; the second type is external validation, which involves comparing model outputs with measures of what actually occurred. Given the rarity with which policy changes correspond to actual model projections, external validation must often be accomplished by other means. Thus, one can use the model with an earlier database to project current program law, thereby making possible comparisons with administrative data on actual program outcomes. The panel itself conducted an illustrative validation experiment with the TRIM2 model, including an external validity study combined with a sensitivity analysis.

The effort that we believe is required for systematic, ongoing validation of policy analysis estimates requires attention to ancillary activities as well. Specifically, policy analysis agencies need to allocate sufficient resources for complete and understandable documentation of policy analysis tools and of the methodology and procedures employed in major policy analyses. The agencies also need to require that major analytical efforts be archived so that the models, databases, and outputs are available for future evaluation. Finally, the agencies need to experiment with modes of presenting information about the uncertainty in their estimates to facilitate understanding and acceptance of such information on the part of decision makers.

Better Data

An essential requirement for policy analysis of alternative government programs, whatever the type of estimate and estimating tool used, is that there be data to analyze. Good data are a critical ingredient for models and other analysis

Copyright National Academy of Sciences. All rights reserved.

tools to produce good policy estimates. Data that are of poor quality, scope, and relevance will increase the uncertainty and decrease the validity of model outputs. Poor data also make it harder for models to respond to changing policy analysis needs in a timely and cost-effective manner.

Given the resources that are at stake, a well-considered, ongoing program of investment in data sources for social welfare policy analysis on the part of the federal government is more than justified. Federal expenditures total over $300 billion a year for social insurance programs, such as social security and Medicare, and almost $75 billion a year for public assistance programs, such as AFDC and food stamps. In comparison, the entire statistical budget of the federal government runs under $2 billion in most years.

A disturbing feature of the decade just completed has been declining federal investment in the production of high-quality, relevant data in many areas of ongoing policy concern. Significant cutbacks in budgets and staff resources for the major federal statistical agencies (amounting to a 13 percent overall budget reduction in real dollar terms from 1980 to 1988), although encouraging the demise of some outmoded programs, in most cases had debilitating effects. Important surveys were reduced in sample size and frequency; programs to review and improve data quality were stretched out or canceled; and key concepts and measurements were not revised to keep up with changing social and economic trends.

These problems are not of purely academic interest; they have had real-world policy consequences. Lack of recent trend data resulted in estimates of the cost of covering prescription drugs as part of the 1988 Medicare Catastrophic Coverage Act (since repealed) that turned out to be much too low. Inadequate data greatly hampered the development of good estimates of the likely impact of child support and employment programs for welfare recipients: Congress had to enact these major new provisions of the 1988 Family Support Act in large measure on faith. Decision makers in both the public and the private sectors have made policy choices based on preliminary economic statistics that later turned out to have large errors.

Members and committees of Congress have expressed concern over the deterioration of the nation's information base, and the administration has expressed support for budget increases and reallocations to make it possible to effect improvements in important statistical concepts and data series. We strongly support these developments and recommend increased investment by the federal government in the production of relevant, high-quality statistical data for social welfare policy analysis and other purposes.

In addition to budget and staffing constraints, the federal statistical system has suffered over the past few decades a deterioration in mechanisms for interagency coordination and the ability to draw upon and integrate information from a range of databases, particularly administrative records. This situation has

Copyright National Academy of Sciences. All rights reserved.

also contributed to reduced timeliness, quantity, and quality of policy-relevant data.

With its traditionally decentralized statistical system, whereby one agency collects data on health care financing, another on income, and so on, the United States depends heavily on effective coordinating mechanisms to achieve optimal allocation of data production resources. Yet the principal coordinating mechanism, lodged in the Office of Management and Budget, with no more than half a dozen staff members and limited resources, is today a shadow of its former self. We strongly recommend that the federal government strengthen and increase its investment in the coordination of federal statistical activities.

To better position the decentralized federal statistical system to serve changing data requirements over time, there is a need not only for improved coordination of data production, but for the adoption of more far-seeing strategies of government data collection that emphasize flexibility and breadth of use. In this regard, duplication of selected questions across surveys can be very beneficial and should not be rejected out of hand. The collection of overlapping data—for example, the collection of income data in health-related surveys and health data in income surveys—makes it easier to relate multiple data sources and evaluate their quality. Such overlaps also facilitate the ability of policy models, which are usually based on one primary data source, to respond more readily to changing policy agendas.

Federal statistical agencies should also give more attention to data collection strategies that recognize key interactions among individuals and society's institutions-employers, hospitals, government agencies, and others. Most data collection efforts are focused on a single entity, such as the family or firm. However, the characteristics of service providers as well as beneficiaries greatly influence the operation of social welfare programs. For example, the hours of operation, location of offices, and treatment of individuals by welfare agencies affect participation by eligible people.

Administrative records, such as social security earnings histories, case files from public assistance programs, health care claims, and tax returns, are valuable sources of information with which to augment, evaluate, and improve surveys and censuses. Their use can enhance the scope and quality of available data for policy analysis and other purposes at very low additional cost. Yet developments in the past two decades have greatly undercut the contributions to the nation's information base from administrative sources. A major factor has been the increased emphasis placed by statistical agencies on restricting data access in order to guard against possible breaches of confidentiality. For example, the Census Bureau no longer prepares exact-match files for public release from household surveys such as the March Current Population Survey (CPS) matched with Social Security Administration (SSA) earnings records. Because the available CPS-SSA exact-match files date back to the 1970s, models of future retirement income programs must generate data for 10 or more

Copyright National Academy of Sciences. All rights reserved.

past years before they can begin their projections, which not only increases costs but inevitably impairs the quality of the estimates. We strongly support the need to take appropriate measures to protect the confidentiality of individual data records and to take all due precautions against either deliberate or inadvertent disclosure. However, we believe that mechanisms must be found to make it possible for the rich sets of data that are generated for federal administrative purposes to be used more fully for statistical analysis purposes.

Finally, we support reallocation of effort within federal statistical agencies and between these agencies and users to emphasize the development of the highest quality data for policy analysis and research purposes. More data, although needed in many areas, are not enough. The data must also reflect appropriate and accurate measurement. Yet budget and staff constraints, coupled with the difficulty of convincing decision makers of the value of methodological work, have often forced agencies to emphasize the operational activities necessary to timely data release at the expense of measurement research and assessment of quality. We urge that this imbalance be redressed.

In addition, we urge that federal statistical agencies use their assessments of quality to add more value to the data series they release. Traditionally, statistical agencies have seen their role as preparing survey-specific data files and publications. They have not seen their role as producing integrated databases or the best published estimates for such statistics as household income or poverty that could be developed from multiple data sources. (For example, the Census Bureau adjusts household surveys for nonresponse by people in the sample but does not perform other adjustments, such as correcting income amounts for misreporting, that would involve the use of outside sources such as administrative records.) Currently, policy analysis agencies and other end users must perform many additional adjustments to survey data to make them suitable for modeling and analysis. Users often lack the information as well as the resources to perform an adequate job, and users at one agency frequently duplicate the efforts of other users. We recommend that statistical agencies seek, where feasible, to use evaluative studies and multiple data sources to develop improved databases and published series for policy analysis and other important purposes.

THE ROLE OF MICROSIMULATION AS A POLICY ANALYSIS TOOL

The microsimulation model approach to producing estimates of the effects of proposed changes in government programs involves obtaining inputs from microlevel databases of individual records, mimicking how current and alternative program provisions apply to the individuals described in those records, and maintaining the simulated outputs for each program scenario on each of the individual records. For example, in simulating the effects of changes to the

Copyright National Academy of Sciences. All rights reserved.

AFDC program, microsimulation models process records for families as if they were applying to the local welfare office for benefits, and in simulating the effects of tax law changes, they process records for people as if they were filling out their 1040 tax forms.

Models based on microsimulation techniques are conceptually highly attractive because they operate at the appropriate decision level and take into account the diverse circumstances and characteristics of the relevant population, whether it be low-income families, taxpayers, or health care providers. Such models are able to respond to important needs of the policy process for information about the effects of very fine-grained as well as broader policy changes, the effects of changes that involve complicated interactions among more than one government program, and the effects of changes on specific population groups as well as total program costs and caseloads.

Of course, microsimulation models are by no means the only useful tool for policy analysis. Indeed, the policy analysis community benefits from having available a wide range of modeling tools to provide alternative perspectives and answer a variety of questions, not all of which require greatly detailed information.

Yet, when flexible, fine-grained analysis of proposed policy changes is called for, no other type of model can match microsimulation in its potential to respond. However, the capability for the detailed analysis provided by microsimulation modeling comes at a price. Microsimulation models tend to be highly complex (reflecting the complexities of government programs and individual circumstances) and must usually meld together a variety of data and research results of varying quality (making many, often unsupported, assumptions in the process). As a result, the history of microsimulation model development has witnessed instances in which model development and application incurred extra time and costs; in which the model became inflexible in operation and difficult to understand and access; and in which it was hard for the analyst, let alone the decision maker, to evaluate the quality of the outputs.

A typical response in the past to the problems posed by the complexity of microsimulation modeling was to pare back model capabilities or focus new development on the accounting functions that mimic program rules and to leave aside other, more difficult aspects, such as modeling behavioral response to program changes. However, these kinds of design choices limit the usefulness of the models for the policy debate.

Very little information is available with which to assess the performance of current microsimulation models, including how well their outputs compare with actual policy outcomes or the degree and sources of uncertainty in the estimates —which we suspect may be high. Although we could not make definitive judgments, we identified several causes for concern. We found that the data sources used to construct microsimulation model databases have serious

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

weaknesses and deficiencies. We are concerned that the mainframe, batch-oriented computer processing technology that is used to implement most current microsimulation models is no longer cost-effective and presents barriers to model validation and experimentation and direct use of the models by analysts. We also note that the underlying base of research knowledge that is needed to support modeling behavioral responses to government program changes and in other ways expand the capabilities of current models has important limitations.

Finally, we are troubled by some aspects of the current structure of the microsimulation modeling community—that is, the interrelationships among the agencies that use the models, the statistical agencies that produce needed input data, the contractors that generally operate the models, and the academic research community. The highly decentralized nature of policy analysis and database production in the federal government often adds costs for duplicative work and raises barriers to effective communication among analysts and between them and data producers. In addition, because the policy community that actively works with microsimulation models today is largely limited to a small number of expert staff in a few firms and agencies, there are few avenues for new ideas and perspectives—either from users in the agencies, academic researchers, or others—to lead to improvements in models and the estimates they produce.

We believe that microsimulation models are important to the policy process, and we anticipate that the need for the kinds of detailed estimates that they can best generate will grow, not diminish, in future years. We recommend allocating sufficient resources to the current models to maintain and improve them incrementally where appropriate and cost-effective. Further, we recommend in-depth validation studies of the current models, both to provide information on the quality of their estimates to the policy process and to guide decisions on future model development. However, because there is so little information with which to assess current models and because of the limitations of available databases and research knowledge, we do not advocate expanding the capabilities of existing models in any specific direction at this time.

The ultimate objective, in our view, is to develop a new generation of microsimulation models that incorporate improvements in quality, flexibility, accessibility to a broader user community, and overall cost-effectiveness. To achieve this goal will require significant investments in data, research knowledge, model design and validation techniques, and computing technology. We urge policy analysis agencies, over the next few years, to devote the needed level of resources to investment activities, including validation of current models. If budgets remain tight, the agencies should be prepared to cut back on resources for current applications in the short term, in order to have available improved modeling tools to satisfy the detailed information requirements of policy debates over the medium and long term.

Copyright National Academy of Sciences. All rights reserved.

Databases

In considering the data inputs to microsimulation models, we focused on the March income supplement to the Current Population Survey and the new Survey of Income and Program Participation (SIPP), which currently or potentially provide much of the data required for modeling income support programs, as well as programs in other areas. We found a need for in-depth evaluation of the March CPS and also of administrative records that often are used in models to supplement household survey data. We propose investigation of short-term and long-term alternatives for improving model databases through combined use of March CPS, SIPP, and administrative records data. In order to improve quality and perhaps reduce the total costs of generating suitable data for policy analysis, we propose that the Census Bureau play a much more active role in preparing useful databases for modeling and policy research. We note specifically the need to investigate the impact on microsimulation model estimates of population undercount in censuses and surveys and, should important effects be determined, to develop ways to implement coverage error adjustments in surveys that the models use.

Model Design and Development

In reviewing the structure and capabilities of microsimulation models, we found that a number of basic principles of model design (such as modularity) and implementation (such as prototyping) have often but not always been followed in the past. They are necessary to the development of cost-effective, useful, and usable models that are well positioned to respond to changing policy needs. In particular, future models need to be designed to facilitate the conduct of validation studies that involve altering model components and data inputs.

In considering strategic directions for future microsimulation model development, we identified three components that are problematic or not well developed in many current models: techniques to project (or age) the data forward in time, to simulate behavioral responses to proposed policy changes, and to simulate longer-term effects of policy changes. In the absence of a body of research knowledge on which to base specific recommendations in these areas, we recommend a research agenda designed to identify the most important kinds of enhancements for microsimulation models and the best ways to implement them.

Computing Technology

Several technologies, such as powerful microcomputer workstations, possibly linked to other computers, and new kinds of software, such as graphical user interfaces and computer-assisted software engineering, promise to enhance

Copyright National Academy of Sciences. All rights reserved.

greatly the flexibility and accessibility of a new generation of microsimulation models. We recommend that policy analysis agencies position themselves to take advantage of these new directions in computing, particularly when promising software tools become more standardized.

Validation

We identified three validation techniques that we believe can and must be used to obtain vitally important information about the quality of microsimulation outputs: (1) external validity studies in which model results are compared with data from program administrative sources or other targets; (2) sensitivity analyses that assess the effects of alternative versions of specific model components on the estimates; and (3) computer-intensive sample reuse techniques, such as the bootstrap, that measure the variance in model estimates. We recommend that policy analysis agencies support research on model validation methods and bring together available information about model quality. We also propose that agencies organize cost-effective programs to obtain validation results of two kinds: rough-and-ready information for use in informing current policy debates, and in-depth information, provided by independent organizations, for use in identifying model weaknesses and planning needed investments in future model development.

Documentation and Archiving

In comparing current documentation and archiving practices for microsimulation models with industry standards for software documentation, we found a number of problem areas. We recommend greater attention to documentation, particularly from the perspective of making microsimulation models more accessible and their outputs more understandable to end users. Similarly, to facilitate model validation, we see a need to set higher standards for archiving of microsimulation model databases and all of the inputs to major policy applications of the models.

The Structure of the Community

We found several areas for improvement in the relationships among all **of** the organizations and people involved in developing, using, and applying microsimulation model estimates for the policy debate. Adding to our earlier recommendation that the Census Bureau and other federal statistical agencies play a more active role in preparing usable databases for modeling and policy research uses, we recommend that policy analysis agencies undertake cooperative activities and encourage relevant academic research to further microsimulation model development. Given the particularly fragmented nature of health care

Copyright National Academy of Sciences. All rights reserved.

policy data collection and analysis, we urge the Department of Health and Human Services to establish a high-level group to coordinate microsimulation model development in this area. Finally, we recommend greater use of the models on the part of agency staff.

The Use of Microsimulation for Basic Research

Guy Orcutt, an economist who pioneered the concept of microsimulation modeling for policy analysis, had a dream that microsimulation models would also make important contributions to social science research knowledge. With some exceptions, notably in the field of family demography, the realization of that dream has gone largely unfulfilled. Today, there is the likelihood that improved data, model design, and computing technology will greatly enhance the cost-benefit ratio of microsimulation. Moreover, many contemporary research problems present complexities that microsimulation can potentially address. Hence, we believe the time may be ripe for the realization of Orcutt's dream. In turn, a larger role for microsimulation modeling as a basic research tool, which we encourage agencies to help foster and seek to benefit from, should make possible advances in the usefulness of microsimulation techniques for policy research and analysis and thereby contribute to better information for public policy decision making.

RECOMMENDATIONS FOR IMPROVING POLICY ANALYSIS

Data

3-1 We recommend that the federal government increase its investment in the production of relevant, high-quality, statistical data for social welfare policy analysis and other purposes.

3-2 We recommend that the federal government strengthen and increase its investment in the coordination of federal statistical activities, with the goal of improving the quality and relevance of data for policy analysis and other purposes.

3-3 We recommend that federal data collection strategies emphasize breadth of use and ability to respond to changing policy needs. In this regard, duplication of selected questions across surveys should be encouraged to the extent that such duplication enhances utility and facilitates evaluation of data quality.

3-4 We recommend that federal statistical agencies give more attention to data collection strategies that recognize key interactions among individuals and institutions—employers, hospitals, government agencies, and others.

Copyright National Academy of Sciences. All rights reserved.

3-5 We recommend development and implementation of mechanisms to improve access, under appropriate circumstances, to administrative and survey microdata for statistical research and analysis purposes.

3-6 We recommend that federal statistical agencies increase their investment in evaluation of the quality of survey and administrative data We further recommend that they use the results of evaluation studies to implement corrections, when feasible, to databases and published data series, with the objective of improving the quality and reducing the overall costs of providing analytically useful data for policy analysis and other important purposes.

3-7 We recommend that the Census Bureau conduct a thorough evaluation of population coverage errors in the major household surveys and decennial census and their potential impacts on policy analysis and research uses of the data. Should important coverage errors be identified, we recommend that the Census Bureau develop ways to adjust census and survey data that have wide application for policy analysis and research.

Validation

3-8 We recommend that users of policy projections systematically demand information on the level and sources of uncertainty in policy analysis work.

3-9 We recommend that the heads of policy analysis agencies assume responsibility for ensuring, to the extent feasible, that their staffs regularly prepare information about the level and sources of uncertainty in their work. Agency heads should also support efforts of their staffs to accustom decision makers to request and use such information in the policy process.

3-10 We recommend that policy analysis agencies earmark a portion of the funds for all major analytical efforts for evaluation of the quality of the results. For large-scale, ongoing research and modeling efforts, the agencies should let a separate contract for an independent evaluation.

3-11 We recommend that policy analysis agencies routinely provide periodic error analyses of ongoing work.

3-12 We recommend that policy analysis agencies allocate sufficient resources for complete and understandable documentation of policy analysis tools. We also recommend that, as a matter of standard practice, they require complete documentation of the methodology and procedures used in major policy analyses.

3-13 We recommend that policy analysis agencies require that major analytical efforts be subject to archiving so that the models, databases, and outputs are available for future analytical use.

Copyright National Academy of Sciences. All rights reserved.

3-14 We recommend that policy analysis agencies include information about estimated uncertainty and the sources of this uncertainty as a matter of course in presentations of results to decision makers. The agencies should experiment with modes of presentation to facilitate understanding and acceptance of information about uncertainty on the part of decision makers.

RECOMMENDATIONS FOR MICROSIMULATION MODELS

Databases

5-1 We recommend that the Census Bureau evaluate the Current Population Survey March income supplement in its role as a primary source of data for analysis of the income distribution and economic well-being of the population. The evaluation should be designed with input from the policy analysis agencies that are major users of the data. It should be comprehensive, covering the impact on data quality of every stage of data collection and processing. It should also compare the March CPS estimates with estimates from other sources. The results should be brought together in a quality profile that is published for users and updated periodically as further evaluations are conducted and new findings obtained.

5-2 We recommend that the responsible agencies sponsor in-depth evaluations of the quality of administrative data that are used as primary or supplemental inputs to social welfare policy microsimulation models. Such data sets include the Integrated Quality Control System samples on the characteristics of welfare recipients and the Statistics of Income samples from federal income tax returns. The results of each evaluation should be brought together in a quality profile that is published for users and updated periodically as further evaluations are conducted and new findings obtained.

5-3 We recommend that the Census Bureau, in conjunction with policy analysis agencies, immediately evaluate alternative options for short-term improvements to the data used for microsimulation modeling, and policy analysis generally, of income support and related social welfare programs. Alternatives that should be investigated include:

- proceeding with the current plan to obtain added resources to restore the SIPP sample size and overlapping panels, beginning with the 1991 panel; and
- keeping the SIPP budget at its current level with the 1990 design of fewer, larger panels, while reallocating the added budget to some combination of initiatives, including adding a low-income sample to the March CPS; adding a limited set of questions to the March CPS to ascertain family composition during the income reference year; exploiting the longitudinal information available in the CPS; exploring sophisticated imputations that use SIPP data to improve

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

CPS information on intrayear income, employment status, and other variables; and exploring matches of SIPP and CPS data with administrative records in a form that can be made publicly available.

5-4 For the longer term, we note that the Census Bureau now has studies under way to consider the future design of SIPP. We recommend that these studies focus on improving the databases for modeling and analysis of income support and related social welfare programs. We recommend that the studies review all aspects of the SIPP design (such as the sample size and length of each panel and the extent to which overlapping of panels is desirable) and consider how best to design SIPP to facilitate relating data from the SIPP, the March CPS, and administrative records.

5-5 After current studies of SIPP and the CPS are completed, we recommend that the policy analysis agencies plan accordingly to redesign their income-support program microsimulation models to make best use of the improved data on income and related subjects that should be available after 1995.

5-6 We recommend that the Census Bureau assume a more active role in adding value to databases for modeling and research purposes and for generating published data series. In particular, we recommend that the Census Bureau seek to produce the best estimates of the income distribution and related variables, such as household and family composition. Steps necessary to achieve this goal include evaluating income reporting errors in the SIPP and March CPS, on the basis of administrative records and other information sources, and using data from multiple sources to develop improved estimates.

Model Design and Development

6-1 We recommend that policy analysis agencies set standards for the design of future microsimulation models that include:

- setting clear goals and priorities for the model;
- using self-contained modules that can be readily added to (or deleted from) the model and that are constructed to facilitate documentation and validation, including the assessment of uncertainty through the use of sensitivity analysis and the application of sample reuse techniques to measure variance;
- providing for entry and exit points in the model that facilitate linkages with other models;
- attaining a high degree of computational efficiency of the model and its components consonant with other objectives such as ease of use; and
- attaining a high degree of accessibility of the model to analysts and other users who are not computer systems experts.

Copyright National Academy of Sciences. All rights reserved.

6-2 We recommend that policy analysis agencies set standards of good practice for the development of future microsimulation models that require:

- constructing prototypes and establishing milestones throughout the development process so that design flaws can be identified at an early stage and the agency provided with some analysis capability before the entire model is completed; . preparing fully adequate documentation on a timely basis for the model and its components;
- conducting validation studies of the model and its components, including estimates of variance and sensitivity analyses (the latter should be conducted for each new module, prior to full implementation, by examining its impact on the rest of the model in order to identify any unexpected or dysfunctional interactions or adverse effects on use); and
- subjecting the model to a "sunset" provision, whereby the model is periodically reevaluated, obsolete components are deleted, and other components are respecified to optimize the model's usefulness and efficiency.

6-3 We recommend that policy analysis agencies sponsor an evaluation program to assess the quality of estimates from current static microsimulation models as a function of the aging technique that is used and that they further support such evaluations on a periodic basis for future models.

6-4 We recommend that policy analysis agencies require that future static microsimulation models build in an aging capability in a manner that facilitates evaluation and use of alternative aging assumptions and procedures.

6-5 We recommend that policy analysis agencies devote resources to studies of the relationship between behavioral research and microsimulation modeling, including studies of ways in which research and modeling can complement one another, as well as ways in which the two are alternative modes of deriving answers to policy questions.

6-6 We recommend that policy analysis agencies sponsor studies to determine when behavioral response effects are most likely to be important in different policy simulations and, hence, how investment in developing behavioral response capabilities in microsimulation should be concentrated. On the basis of such studies, policy analysis agencies should commission research to attempt to narrow the range of statistical estimates of behavioral parameters that may be of major importance to critical policy changes. Such research may require additional data analysis, replication studies, and multiple econometric analyses that use different data sets and analytic techniques.

6-7 We recommend that policy analysis agencies commission methodological research to develop methods for systematically assessing the impact on microsimulation model estimates of the degree of uncertainty in the behavioral

Copyright National Academy of Sciences. All rights reserved.

parameters that are used—both the uncertainty arising from the variance of specific parameters and that arising from the range of estimates from different behavioral studies. This work should be tied into the development of similar methods for assessing uncertainty of the estimates produced by microsimulation models without behavioral response.

6-8 We recommend that policy analysis agencies support research on second-round effects of policy changes that may be important to understand. We also recommend that the agencies require that future microsimulation models include entry and exit points that could facilitate linkages with second-round effects models. However, except perhaps for health care issues, we do not recommend investment at this time in building second-round effects capabilities into microsimulation models.

Computing Technology

7-1 We recommend that policy analysis agencies invest resources in developing prototypes of static and dynamic microsimulation models that use new computer technologies to provide enhanced capabilities, such as the ability for a wider group of analysts to apply the models; conduct timely and cost-effective validation studies, including variance estimation and sensitivity analyses; and alter major components, such as the aging routines, without requiring programmer intervention.

7-2 We recommend that policy analysis agencies, after experience with prototypes and reviews of developments in computer hardware and software technologies, make plans to invest in a new generation of microsimulation models that facilitate such design criteria as user accessibility and adequate documentation and evaluation of model components, as well as computational efficiency.

Health Care and Retirement Policy Modeling

8-1 We recommend that the U.S. Department of Health and Human Services establish a high-level, department-wide coordinating and steering body to set priorities for development of microsimulation models and related data collection and research needed for improved analysis of alternative government policies and programs for health care.

8-2 We recommend that the Census Bureau perform a new exact match of social security earnings histories with the March CPS as soon as possible. The Census Bureau should develop a program for periodically conducting matches of social security earnings histories with both the March CPS and SIPP records. Ways should be found to make the matched data files available for research and modeling use.

Copyright National Academy of Sciences. All rights reserved.

Validation

9-1 We recommend that policy analysis agencies commit sufficient resources and accord high priority to studies validating the outputs of microsimulation models. Specifically, we recommend the following:

- Agencies, in letting major contracts for development, maintenance, and application of microsimulation models, should allocate a percentage of resources for model validation and revisions based on validation results. The types of validation studies to be carried out by the modeling contractor should include estimates of variance and focused sensitivity analyses of key sets of model outputs. The goal of these efforts should be to provide timely, rough-and-ready assessments of selected estimates that are important for informing current policy debates.
- In addition, agencies, when practical, should let separate microsimulation model validation contracts to independent organizations or in other ways arrange to carry out comprehensive, in-depth evaluations. The types of studies to be performed by a validation contractor should include external validation studies that compare model outputs with other values and detailed sensitivity analyses. The goal of these longer range efforts should be to identify priority areas for model improvement.

9-2 We recommend that policy analysis agencies provide support, through such mechanisms as grants and fellowships, for research on improved methods for validating microsimulation model output.

9-3 We recommend that policy analysis agencies support the development of quality profiles for the major microsimulation models that they use. The profiles should list and describe sources of uncertainty and identify priorities for validation work.

Documentation and Archiving

10-1 We recommend that policy analysis agencies set high standards for documentation of microsimulation models and their inputs and outputs. Agencies should investigate existing standards, such as those published by the Institute of Electrical and Electronics Engineers, for relevance to microsimulation models and determine what additional standards are needed. The kinds of documentation that agencies should require to be developed for analysts and programmers who use, or expect to use, the models include general informational materials; tutorials; and detailed reference documents for model components that describe their theoretical basis, assumptions, operation, inputs, and outputs.

10-2 In order to facilitate model validation, we recommend that policy analysis agencies require archiving of microsimulation model databases on a regular

Copyright National Academy of Sciences. All rights reserved.

basis. In addition, we recommend that the agencies require full documentation and archiving of major applications of microsimulation models. The archived materials should include the model itself, the documentation of the model, the database and other inputs, the analyst's specifications, and the outputs.

Structure of the Microsimulation Modeling Community

11-1 We recommend that executive and legislative branch policy analysis agencies expand their communications and undertake cooperative efforts to improve the quality of microsimulation models and associated databases through such means as cosponsoring research on model validation methods and other initiatives.

11-2 We recommend that policy analysis agencies have a strict policy that only public-use, nonproprietary microsimulation models—for which documentation, inputs, outputs, and programming code can be freely exchanged—will be considered for agency use.

11-3 We recommend that policy analysis agencies set a goal of increasing the in-house use of microsimulation models by agency analysts, who have the ultimate responsibility of interpreting model results for policy makers.

11-4 We recommend that policy analysis agencies encourage and support the involvement of social science researchers in work that is relevant to microsimulation modeling, and other microlevel policy analysis, through sponsoring regular research conferences. The conferences should highlight pertinent research results that can be used for models, with an emphasis on the synthesis of research findings and the reconciliation of conflicting results. These conferences should also work to develop research agendas to address emerging policy needs. The agencies should prepare and disseminate proceedings from all such conferences.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

1

Introduction

A SCENE IN WASHINGTON, D.C.

It is mid-1988. In the quarters of the Congressional Budget Office (CBO) at the foot of Capitol Hill, exhausted analysts are working overtime to prepare final estimates of the budgetary and programmatic effects of what will become the Family Support Act (FSA) of 1988 (signed by President Reagan a few months later). This piece of legislation includes some of the most far-reaching changes in more than two decades to the Aid to Families with Dependent Children (AFDC) program—a federal-state program that is an important component of the nation's "safety net" for the poor. The act both extends program coverage and provides incentives for families to reduce their reliance on welfare support and be more responsible for their own economic well-being. The act represents a policy compromise among the goals of providing more adequate income support to poor children, strengthening family ties, and strengthening attachments to the labor force. It also represents a budgetary compromise: under the guidelines agreed to by both political parties at the outset, the 5-year projected cost of the legislation cannot exceed $3.0-$3.5 billion.

The process of developing the final form of the FSA has spanned a period of almost 2 years, beginning in late 1986. During that time, CBO analysts have prepared cost estimates for half a dozen major bills, each of which contains as many as 50 separate provisions. They have also prepared estimates of which population groups would be affected by the various bills and whether those

Copyright National Academy of Sciences. All rights reserved.

groups would gain or lose under each proposal, in comparison with alternative proposals and with current law.

To prepare their estimates for the FSA, the CBO analysts have called on a wide variety of data sources and data processing tools. For some provisions, they have used a complex computer model, one of a class of microsimulation models that process large, nationally representative samples of families as if they were applying to the local welfare office for benefits. For each family in the sample, the microsimulation model determines which family members would be included in the assistance unit, their eligibility for benefits under the current AFDC program rules, whether they would choose to participate, and the amount of benefits they would receive. The model then repeats the process using the proposed new rules. Tabulations of program costs and caseloads under the current and alternative proposed programs show the impact of reform.

For provisions that would affect only families already receiving AFDC, the analysts have used a similar but less complicated model that processes administrative data from AFDC case records. For still other provisions, existing models were not designed to handle the task or required too much time and resources to modify. The analysts have developed their own ad hoc computations that use available data and research findings—of varying quality and relevance—from a variety of sources. In these instances, they have often used personal computer spreadsheet programs to develop their calculations so that they can more readily handle interactions among various program provisions, incorporate state differences when appropriate, and so on.

In accordance with the requirements of the 1974 Budget Act, the cost estimates must be projected for 5 years, beginning at the time the reform is anticipated to take effect. To make such projections, the CBO analysts have used a variety of data, including forecasts of employment changes and other economic factors developed within CBO with input from forecasts of large macroeconomic models of the national economy. Finally, the analysts have used spreadsheet and word processing programs to pull all of the estimates together into a package for presentation to Congress and to keep track of the many such packages that have been generated during the course of the legislative debate.

During the same period, analysts two buildings away in the Office of the Assistant Secretary for Planning and Evaluation (ASPE) of the U.S. Department of Health and Human Services (HHS) have been preparing their own estimates of the various FSA welfare reform proposals, using similar data sources and processing tools but with a somewhat different mix. The ASPE analysts have relied more heavily on a large microsimulation model for their estimates than have the analysts at CBO. Analysts at other federal agencies, including the Food and Nutrition Service of the U.S. Department of Agriculture, the Family Support Administration in HHS, and the Health Care Financing Administration (HCFA) in HHS, have also been involved in preparing estimates of the effects

Copyright National Academy of Sciences. All rights reserved.

of the proposed changes on the food stamp, AFDC, child support enforcement, Medicaid, and other programs.

At the same time in mid-1988, other groups of analysts—principally in the U.S. Department of the Treasury and the congressional Joint Committee on Taxation, but also in CBO and ASPE—are hard at work estimating the impact on revenues and tax burdens facing different income groups from proposed changes in the nation's tax laws. Their pace is somewhat less frantic than that of the welfare program analysts because 1988 is a period of relative calm following the frenzy of activity that culminated in the Tax Reform Act of 1986. Although the specific models and databases differ, analysts involved in the estimates for tax legislation use many of the same kinds of tools as the analysts looking at welfare program changes, including microsimulation models operating on administrative and survey data, macroeconomic models, and models designed to estimate "second-round" effects, that is, the effects after the economy has adjusted to the legislative changes.

At the same time, still other analysts working on retirement policy, principally in the Social Security Administration, are assessing the impact of current law and proposed changes on social security revenues and benefits. Because of their need to produce frequently updated long-range projections for 30 years or more, social security analysts have typically relied on "cell-based" models. These models, which estimate program effects for specific population groups (e.g., retired men age 65) instead of individual assistance units, provide less detail than microsimulation models and are typically less complex and less expensive to maintain and run. But social security analysts have also used microsimulation models, both "static" models that use samples of current households to represent the population in future years (through statistical adjustments to the survey weights) and "dynamic" models that represent the future by "growing" individual sample cases forward in time.

Separate groups of analysts in CBO, HCFA, and other agencies are also working hard to assess the cost and social effects of changes in the nation's programs to provide and pay for health care. Health care policy in the United States is of immense and growing complexity: policy issues run the gamut from how to provide health insurance coverage to the working poor and long-term care benefits to the elderly to how to alter the behavior of hospitals and physicians to achieve the most cost-effective medical care. Correspondingly, the world of health care policy analysis is highly fragmented. To estimate the impact of health care policy changes, most individual analysts rely on their own data extracts and special-purpose computer models (frequently, cell based) to address very specific questions affecting particular segments of the health care system.

The results of all of these activities are numbers—numbers of dollars and numbers of participants—that are estimates and projections of what will happen if new programs are enacted or existing programs are maintained, changed,

Copyright National Academy of Sciences. All rights reserved.

or eliminated. And on the basis of these numbers, policy makers throughout Washington are making decisions about the government's role in social policy.

THE TOOLS OF POLICY ANALYSIS

Twenty-five years ago, no one would have anticipated the major role that quantitative information about the effects of alternative proposals now plays in shaping legislation. In the early 1960s, computer technology was in its infancy, data sources were limited, and modeling techniques were primitive. Although analysts in executive branch agencies (there was no CBO until 1975) often made estimates of the impact of policy changes, they were severely limited in the number of different estimates they could prepare, the detail they could provide about the effects on specific population groups, and the time frame in which they could respond to requests. The quality of these "back-of-the-envelope" estimates rested heavily on the experience and judgment of individual analysts, and the estimates themselves played no more than a secondary role in the legislative process.

Today, whatever the policy issue, "the numbers" play a prominent role. Indeed, in the Washington of the early 1990s, neither top administration officials nor members of Congress will move very far to develop legislation in the absence of detailed estimates of the cost and other effects of the proposed changes. Top officials will not necessarily know specifically how the estimates were developed, but they will know that the estimates rely heavily on complex computerized techniques. They treat the estimates not only as informative but often, in the case of costs, as binding.

The cost estimate for mandating a minimum benefit standard for AFDC in all states—produced largely for CBO and ASPE by a microsimulation model called TRIM2 (Transfer Income Model 2)—was one of the factors that led Congress to jettison this provision early in the development of the Family Support Act, despite strong support from many representatives. The reason: the estimated cost would have far exceeded the agreed-upon dollar amount.

At each of three times during the hectic days of drafting the final version of the 1986 Tax Reform Act, the staff of the Joint Committee on Taxation determined that problems in their data or specifications had resulted in the model's overestimation of revenues by $17 billion. Each time, although despairing of success, the members shepherding the legislation felt compelled—and ultimately were able—to find ways to make up the revenue shortfall so that the net impact compared with current law would be zero (Bimbaum and Murray, 1987). Senators and Representatives also paid close attention to the estimates of the impact of proposed tax reforms on different income groups in the population. There was strong support for cutting top tax rates, but not in a manner that would injure the middle class or the poor. The final legislation raised corporate taxes to make it possible to reduce personal income tax rates.

Copyright National Academy of Sciences. All rights reserved.

These examples are by no means unique. Formal computer modeling techniques implemented by teams of analysts and programmers today contribute importantly to the policy analysis function—that is, the production of information evaluating the effects of alternative proposals for legislative change. Yet, despite extensive use over the past two decades, most policy analysis tools have rarely been the focus of any explicit evaluation of their utility or effectiveness. Recently, however, policy analysts at HHS, concerned about the increasing pace and new directions of legislative activity in many areas, determined that it was imperative to take stock of the available modeling tools and databases and to evaluate whether they were up to the job.

In the spring of 1988, the Office of the Assistant Secretary for Planning and Evaluation in HHS and the Food and Nutrition Service (FNS) in the Department of Agriculture asked the Committee on National Statistics to convene a panel. The two agencies asked that the panel be charged to evaluate the large microsimulation models that have been heavily used for estimating the costs and effects on population subgroups of proposed changes in social welfare programs. These models had not been subjected to a major evaluation since an initial study by the General Accounting Office in 1977.

The need for an evaluation of policy analysis tools is evident from the perspectives of both demand and supply. The demand for social welfare policy analysis, after a relatively slack period in the early 1980s, has gained strength in recent years and suggests a boom in the 1990s. One source fueling the demand is the recognition that changes in the U.S. population, economy, and society are generating problems that the federal government cannot ignore (although the solutions need not always involve direct federal action). As just one example, the aging of the population is making it imperative to assess the potentially staggering costs of providing health care and retirement support for the elderly and to determine cost-effective approaches for meeting these needs. Similarly, the huge growth in the labor force participation of women, along with other factors, has generated renewed interest in legislative initiatives for child care, through both tax credits to parents and funding for day care services.

The large federal budget deficits, resulting from the tax cuts of the early 1980s and growth in government spending throughout the decade, are another strong and continuing source of demand for policy analysis. The deficits force Congress to engage in often desperate attempts to find more revenue or produce more cost savings by fine-tuning tax provisions and program regulations. The Joint Committee on Taxation, which produces revenue estimates for every proposed change to the tax code, received 1,290 requests for estimates in 1989; in comparison, it had received only 150 requests yearly in the early 1980s (Haas, 1990).

Yet just as the demand for social welfare policy analysis is on the rise, the supply of effective modeling tools and databases has been impaired by constrained resources over the past decade for modeling, data collection, and

Copyright National Academy of Sciences. All rights reserved.

socioeconomic research. Funding support in the 1980s (in real dollar terms) declined—drastically in some cases—or at best held steady for all three elements of the policy analysis infrastructure. Both Congress and the executive branch have acknowledged the deterioration in the nation's statistical and research knowledge base and the need for retooling. An evaluation of the tools that are widely used for policy analysis is timely to help chart directions for future investment to improve the quality and effectiveness of the information on which policy makers rely.

THE PANEL STUDY

The panel formed by the Committee on National Statistics in late 1988 was asked to evaluate microsimulation models as a tool for policy analysis and to make recommendations about the role for microsimulation modeling in addressing the policy analysis needs of the 1990s. The charge was to focus on the modeling requirements of "social welfare programs," specifically, the programs within the purview of ASPE and FNS. With this mandate, we could and did seek to make our task more manageable by excluding a wide array of policy domains—defense, education, energy, the environment, housing, transportation, and the work force—even though many issues in these areas are also amenable to similar formal modeling and analysis. The remaining policy areas that fell within our scope—support for the low-income population, support for retirees, health care benefits, and taxation (which intersects at many points with social welfare policy)—were sufficiently broad to strain our capacity to carry out a useful study.

In considering the utility and cost-effectiveness for analyzing social welfare policy issues of the particular class of techniques called microsimulation modeling, we found it necessary to look more globally at the policy analysis process. Microsimulation models bring unique characteristics to the policy analysis effort; they also share common aspects with other classes of models. More important, it appeared that some of the major problems confronting microsimulation models today plague other kinds of policy analysis tools as well.

Part I of our report considers the role of information in the policy process broadly and presents recommendations in several areas that we believe are important to address, whether the analysis tool is microsimulation or some other approach. Part II presents recommendations specific to microsimulation modeling as a tool for policy analysis. A companion volume of technical papers provides additional information on topics related to microsimulation modeling.

In our assessment, we examine both the inputs and outputs of policy analysis tools and the structure and design of the models themselves. On the input side, the quality and scope of the available data and research about socioeconomic behavior determine the scope, amount of detail, and quality of the

Copyright National Academy of Sciences. All rights reserved.

estimates that models can provide and affect their cost structure. The capabilities and efficiency of a third input—the technological means for computation—importantly affect the cost and timeliness of model outputs as well as the accessibility of the model to policy analysts and researchers.

On the output side, we are most concerned with the added information that is required for model estimates to be useful in the policy process. The estimates need to be evaluated in terms of their likely variability and sensitivity to key assumptions. Such validation is important both for purposes of informing decision makers about the utility and quality of model estimates and for purposes of improving the models themselves. The models and their inputs and outputs need to be appropriately described, or documented, for future use. Finally, model estimates, together with information about their quality, need to be communicated to the decision makers in the executive branch and Congress in a form that they can understand and use. How well the validation, documentation, and communication functions are performed has profound implications for the quality and cost-effectiveness of the information produced by policy analysis and thereby for the legislative process itself.

We conclude that the policy analysis world needs a "second revolution." The "first revolution" of the past two decades institutionalized the use of detailed estimates of cost and population effects of alternative proposals as part of the legislative process, and contributed to the development and widespread application of large computerized models as estimation tools. The second revolution requires significant investments in data, research knowledge, and computing to improve the quality of these models and the estimates they produce.

Even more important, the second revolution requires a commitment to model validation, with implications for improved communication and documentation as well. The validation function has, for many reasons, been the stepchild of the policy analysis process—chronically short on resources and attention. Indeed, the absence of a validation literature severely constrained the ability of the panel to make definitive recommendations for the microsimulation models that we were charged to examine in detail. Our strongest recommendation to policy analysis agencies, for these and other kinds of models, is to invest the needed resources to make validation an integral part of the policy estimation process. Without systematic and rigorous evaluation of models and their inputs and outputs, no one can know their quality today or make informed choices about how best to allocate scarce investment resources to improve their quality and usefulness for tomorrow.

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Part I

Information for Social Welfare Policy: Toward A Second Revolution

In a typical year, the U.S. Congress considers about 5,000 bills and enacts about 250 pieces of legislation (Bureau of the Census, 1990b:255). These laws mandate new government services, modify or drop existing programs, raise needed revenues, and otherwise meet the constitutional mandate to "provide for the common defense [and] promote the general welfare." Behind each piece of legislation is a complex political process that begins long before a bill is formally introduced into the Senate or House of Representatives and continues long after it is formally signed into law by the President. Major stages in the policy process include:

- identifying a problem, such as rising health care costs or declining productivity;
- putting the problem onto the political agenda, that is, making the case that action is needed and appropriate on the part of the federal government;
- identifying and evaluating alternative policies and programs to deal with the problem;
- building agreement on a program and obtaining sufficient support to enact it formally into law;
- implementing the program; and
- evaluating the program, which often leads to identifying a new problem that may require further governmental action.

Each stage of the process usually involves many actors—decision makers in the executive and legislative branches and their staffs, researchers and policy

Copyright National Academy of Sciences. All rights reserved.

analysts from academia and the private sector, interest groups, and the media—and the entire process may span months, years, or even decades. Sometimes the process works in a straight line from problem identification through legislation to program evaluation, but more often it encounters detours and roadblocks along the way.

Each stage of the process also involves the use of information. One kind of information is political. Political information, which is vital for developing strategies and tactics to facilitate the process, concerns which actors support or oppose certain policy options, how open key actors may be to changing their positions and what factors could induce them to change, trends in public opinion, and similar matters. Another kind of information is policy oriented. It concerns the characteristics of policies and programs and plays a role in helping decision makers choose among options at each stage of the process. This report is concerned with the use of the latter type of information for decision making about social welfare policies.

Policy-oriented information not only enters each stage of the political process, but is often embodied in legislation itself. Formula grant programs allocate many billions of dollars of federal appropriations to states and localities, not in predesignated amounts, but on the basis of their characteristics as determined in the decennial census or other data sources. For example, the Job Training Partnership Act allocates funds on the basis of a formula that uses each jurisdiction's share of the unemployed and economically disadvantaged populations, and the Medicaid program determines the federal matching percentage of state expenditures partially on the basis of measures of each state's per capita income.

By policy-oriented information (which we henceforth refer to as "information" or "policy information"), we mean input to the social welfare policy process that has the following characteristics:

- The information is "objective": that is, it is based on evidence rather than on opinion or preference. Ideally, the evidence is quantitative, such as the results of a survey, demonstration, or experiment; however, in many instances, the available information may be largely qualitative or based on such a small number of observations that it is difficult to draw conclusions from the results.
- The information represents a combination of data and analysis: that is, the information is not simply raw numbers, but the result of an analytical process that assigns meaning to the numbers and assesses their relevance and their limitations. (In this regard, we adopt something parallel to the convention of the Central Intelligence Agency, which defines "intelligence" as "data plus analysis.")
- The information relates to the social welfare aspects of policy choices, such as which population groups will be affected by a policy change and why: that is, for the purposes of our study, we are dealing neither with political information—which can also be objective and combine data with analysis—nor

Copyright National Academy of Sciences. All rights reserved.

with other kinds of policy-oriented information, for example, information about the environmental effects of policy choices.

Part I of our report considers some of the key issues involved in improving the evaluation of alternative legislative proposals—what we term "policy analysis"—whether the means to the evaluation is microsimulation modeling or some other technique such as cell-based or macroeconomic modeling. In Chapter 2 we present a brief history of the first information revolution, which concludes by calling attention to the resource constraints that limited investment in models, data, and research in the 1980s, with adverse effects on the quality and cost-effectiveness of the tools that support the policy analysis function. Policy analysis represents a link between social science research and the political process. We next review some general characteristics of each end of the link, which are important to understand in order to develop recommendations directed to improving the policy analysis function itself. We then illustrate the capabilities and limitations of microsimulation vis-á-vis other policy analysis techniques through a case study of the Family Support Act of 1988, identifying some of the problem areas.

In Chapter 3 we offer a strategy for investment to bolster the capabilities of policy analysis tools and the quality of policy analysis information to address the issues of the 1990s and beyond. We discuss in detail three key ingredients to policy analysis that we believe require attention regardless of the analysis tool that is used: data, validation and documentation of analysis tools, and communication of the results of analysis to policy makers. (The Appendix to Part I provides technical information related to validation and communication of the uncertainty in model estimates.) We recommend improvements in the underlying databases, a great expansion in the effort devoted to evaluating and documenting the estimates produced by models, and a commitment to finding useful ways to communicate the results of validation studies in an understandable form to decision makers in the executive branch and Congress. These actions are all necessary elements of the second revolution that we believe is needed in the world of policy analysis.

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

2

The Search for Useful Information

Since the inception of the U.S. federal system in 1789, decision makers have sought social welfare policy information of the type we describe, and such information has helped shape many important political debates over the decades.[1] James Madison, in 1789, suggested that the content of the decennial census be expanded to provide information for decision making in addition to fulfilling its constitutionally mandated role in congressional reapportionment. He observed that Congress

> . . . had now an opportunity of obtaining the most useful information for those who should hereafter be called upon to legislate for their country, if this bill was extended to embrace some other objects besides the bare enumeration of the inhabitants; it would enable them to adapt the public measures to the particular circumstances of the community.

Congress added questions on age and sex to the 1790 census schedule, and, by the time of the Civil War, the questionnaire included items on citizenship, education, disability, marital status, place of birth, industry, occupation, and value of real estate owned (Bureau of the Census, 1970:4).

Information about industrial practices and their effects on the population that was gathered by the great "muckraking" journalists at the turn of the century, including Ray Stannard Baker, Upton Sinclair, Lincoln Steffens, and Ida Tarbell,

[1] Studies on the role of policy information and policy analysis in public decision making, particularly on the part of the Congress, include Bulmer (1986, 1987); Caplan, Morrison, and Stambaugh (1976); Davidson (1976); Lynn (1978); Primus (1989); Robinson (1989); Shipp (1989); Weiss (1977, 1989); and Whiteman (1985).

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

played a role in many Progressive Era legislative initiatives, such as the Pure Food and Drug Act of 1906, the Meat Inspection Act of 1906, and the 1906 Hepburn Act that strengthened the regulatory powers of the Interstate Commerce Commission (Morris, 1953:267-268). Yet, throughout most of U.S. history, the limitations of available data, research knowledge, and computational methods and machinery greatly constrained the extent of useful policy information, and the demand for such information remained largely ad hoc in nature.

Beginning in the 1960s, quantum improvements in data sources, socioeconomic research, and computing technology made it possible to supply information of much greater depth and breadth for the policy process. In turn, the activist posture of the federal government with regard to social welfare policies—Medicare, the expansion of social security coverage and benefits, and the Supplemental Security Income program for the low-income elderly and disabled are only a few examples of the legislative initiatives of that period—both stimulated the production of policy research and analysis and drew on its results.

Today, policy information plays a much greater role at every stage of the political process than ever before in U.S. history. At one end of the process, the results of research studies documenting social needs typically play a part in identifying problems and moving them onto the federal agenda. At the other end, sophisticated evaluations of government programs often lead to further policy initiatives. In the middle stage of the process, the policy community in Washington takes for granted that neither the administration nor the Congress will consider legislation to alter any of the nation's social welfare programs or the tax code without looking closely at "the numbers," and these numbers, in many instances, are the product of team efforts to apply formal modeling techniques to the task of estimating the effects of alternative policies.

By formal, we mean models that are based on a coherent modeling strategy and set of assumptions, developed for repeated application, and designed to produce consistent estimates for a range of policy proposals within a common framework that is, or can be, well documented and evaluated. By their nature, formal models circumscribe, although they do not eliminate, the role of individual analysts' judgments. Such models—macroeconomic and microsimulation models as well as several other major types—vary in size, scope, and the types of data and modeling strategies they use, but they share the attributes we have listed.

Formal models, as we have defined them, are at one extreme of a continuum of policy analysis tools. At the other extreme are very simplistic and idiosyncratic "back-of-the-envelope" calculations, which an individual analyst might develop at the outset of a policy debate as a very rough guide to the likely effects of a proposed policy. In between are models that are developed by an analyst on an ad hoc basis—often using personal computer spreadsheets—to respond to a specific policy debate. Such models, which vary greatly in complexity and approach, will reflect the analyst's best efforts to use all available

Copyright National Academy of Sciences. All rights reserved.

data to develop the estimates needed for the particular debate, but they are not generally designed with any future application in mind.

We believe that formal models bring strengths to the policy estimation process. Through their facility for repeated application within a consistent, documented framework, they can earn a return on investment in their development, serve as a vehicle for identifying important data and research gaps, and contribute to a cumulative body of knowledge relevant to social welfare policy formulation, as well as the development of improved models. However, in practice, formal models may be inadequate to the task if, for instance, they do not reflect relevant data or research findings or cannot readily be altered in a timely manner to respond to the changing course of a policy debate.

Today, for some social welfare policy topics, notably taxation, the use of formal models has largely displaced more ad hoc estimation methods. For other topics, formal models play an important role, although ad hoc models developed for specific applications by individual analysts very often remain the estimating tool of choice.

THE FIRST INFORMATION REVOLUTION

What we term the first information revolution—the institutionalization of the role of quantitative data in shaping legislation—originated with agencies in the executive branch, which led the way in developing modeling tools for analyzing alternative legislative proposals. Executive branch agencies first pioneered the use of large computerized macroeconomic models, such as those of Data Resources, Inc. (DRI) and Wharton Econometrics. These models are used to generate forecasts of economic aggregates, such as the gross national product and the rate of inflation, and also to evaluate alternative government fiscal policies. Macroeconomic models existed in the 1940s, but their use in the policy process took off in the 1960s when the advent of more powerful computing technology made it possible to develop highly complex models that more accurately reflected the operation of the economy. The forecasts produced by macroeconomic models proved useful for providing the economic assumptions used in developing "current services" estimates of federal spending —that is, projections of spending levels over the next 2-5 years that assume no future legislative changes—against which to evaluate the President's proposed budget and legislative proposals of Congress.

Executive branch agencies also pioneered the use of microsimulation models for more detailed cost and distributional analysis. In the early 1960s, a few far-seeing analysts at the Treasury Department encouraged work by the Brookings Institution to conduct the first detailed analysis of federal tax burdens using a large microdatabase and a model to calculate taxes for each individual filing unit (e.g., a married couple or a single adult) in the sample (Atrostic and Nunns, 1988:43-47). At the same time, analysts at the Department of

Copyright National Academy of Sciences. All rights reserved.

Health, Education, and Welfare and the Office of Economic Opportunity were exploring the use of the microsimulation modeling techniques pioneered **by** Guy Orcutt for analysis of social welfare program policy alternatives. In 1968, the President's Commission on Income Maintenance developed the first operational social welfare policy model—Reforms in Income Maintenance (RIM)—which was used extensively over the next few years to model alternative welfare reform proposals (Orcutt et al., 1980:84-85).[2]

Beginning in the mid-1970s, not only the executive branch but also the Congress became accustomed to requesting and receiving detailed estimates of the budgetary impact and also the anticipated social impact of legislation. In particular, Congress sought information on which groups—the elderly, children, the middle class—would benefit and which would be adversely affected by a program change. The Food Stamp Reform Act of 1977 was a milestone in this regard. Over a 2-year span from 1975 to 1977, the Food and Nutrition Service in the Department of Agriculture used a microsimulation model—the Micro Analysis of Transfers to Households (MATH) model—to produce cost and distributional estimates for at least 200 variations of the proposed legislation (Beebout, 1980). As the congressional debate progressed, tables generated by the model on participants whose benefits were increased by a particular proposal as well as those who lost benefits—"gainers and losers"—became more and more elaborate, with more and more detailed categories. Congress, instead of simply accepting the administration's proposal, used the information generated by the model to shape the legislation to meet its own priorities on geographic effects, income cutoff levels, work incentives, and other factors (Shipp, 1980).

Congress set the capstone to the first information revolution when it passed the Budget Act in 1974. That act specified a formal process for setting budget targets and authorized the establishment of a Congressional Budget Office (CBO), charging it to provide Congress with analyses of the federal budgetary cost impact of every piece of legislation reported by a congressional committee. No longer would Congress rely solely on the expertise of executive agency staff; rather, it would obtain independent estimates of legislative proposals originating from the administration or Congress itself. CBO first opened its doors in 1975; today its staff numbers 225, and it prepares more than 1,200 formal cost estimates annually, as well as many informal estimates of costs (e.g.,

[2] Kraemer et al. (1987), in their fascinating volume *Datawars—The Politics of Modeling in Federal Policy making,* describe and analyze the development and subsequent widespread application in the policy process of microsimulation models—using the Transfer Income Model (TRIM) and the Micro Analysis of Transfers to Households (MATH) model as case studies—and macroeconomic models—using the DRI model as a case study. Technological improvements were a necessary precondition for the models' development. However, given the necessary data, research knowledge, and computational technology, and acknowledging the role of key suppliers in promoting their use, the authors argue that the most important factor fostering these models was the agencies' needs for sophisticated policy analysis of the type they could provide.

Copyright National Academy of Sciences. All rights reserved.

for proposals being considered by committee members and staff) and estimates of the social impact of proposed legislation.

Throughout the 1970s and into the 1980s, both congressional and executive branch agencies made use of a variety of models and analytical techniques for social welfare policy analysis. For example, in the early 1980s both the Social Security Actuary's cell-based model and the Dynamic Simulation of Income 2 (DYNASIM2) microsimulation model played a role in evaluating alternative proposals for altering various aspects of the social security system.

The 1980s ushered in a period of severe budget constraints that impaired the ability of policy analysts to respond to new policy issues and to maintain, let alone improve, the quality of the information provided to the policy debate. All key elements of the policy analysis infrastructure—data collection, policy research, and model development—were affected by budget reductions. The operating budgets of nine major federal statistical agencies that produce key economic statistics—such as the gross national product, consumer price index, unemployment rate, and poverty rate—declined by nearly 13 percent (in constant dollars) from 1980 to 1988 (Wallman, 1988:13). The number of professional statistical staff also declined, as did the number of staff in the Office of Management and Budget who were charged with coordinating federal statistical programs. Some policy analysis agencies experienced even more severe cutbacks. The HHS headquarters policy analysis group—the Office of the Assistant Secretary for Planning and Evaluation (ASPE)—from 1980 to 1988 experienced sizable staff losses and a staggering reduction of 86 percent (in constant dollars) in the budget for research and modeling (Citro, 1989:Table 2).

We discuss in Chapter 3 the adverse effects of budget constraints on the data inputs to policy analysis. High-quality, relevant data from surveys and administrative records are essential for all types of policy analysis, regardless of the type of analytical approach that is used or whether the analysis tool is the largest, most complex model or the simplest, most ad hoc calculation. Also, in our review of microsimulation models in Part II, we note many ways in which budget constraints limited model development and enhancement. These models have generally stayed bound to old computer technology (although their logic was rewritten to optimize the use of this technology); key relationships in the models have been only infrequently updated with results from new data; and, in some cases, more complicated but potentially important model components have been put aside. Institutional mechanisms for coordinating and fertilizing model development, such as a microsimulation users' group sponsored by ASPE in the late 1970s, languished.

Yet even as budgets remained tight, the pressure to produce estimates for legislative initiatives remained strong. Indeed, the pressure increased in the mid-1980s with the introduction on the legislative agenda of such explosive and complex issues as tax and welfare reform and the passage of the Gramm-Rudman-Hollings Act, calling for strict deficit reduction measures and

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

necessitating intense scrutiny of the cost and revenue effects of every piece of legislation. Recently, both Congress and the executive branch have become cognizant of the deterioration in the government's capability for policy research and analysis that supplies critical information for the policy debate and the need for investment to restore and enhance that capability (see, e.g., Boskin, 1990; Council of Economic Advisers, 1990, 1991; Darby, 1990).

In the next chapter, we consider three key areas for improving the tools that support policy analysis for social welfare programs. But first we review some aspects of the relationship of policy analysis to both social science research and politics that are important in framing an effective strategy for improving the policy analysis function. We also provide a case study of the advantages and disadvantages of microsimulation models in comparison with other analysis tools for shaping social welfare legislation that identifies some of the problems most in need of attention.

POLICY ANALYSIS: BETWEEN SOCIAL SCIENCE RESEARCH AND POLITICS

Hanushek (1990:147-148) defines policy analysis as social science research "that is directly linked to the policy process . . . [and that] responds to specific and detailed questions such as those that arise over a bill before Congress or a policy proposal in the Office of Management and Budget." He distinguishes policy analysis from both "disciplinary research" and "policy research." The former is basic research designed to advance the frontiers of knowledge, which only fortuitously may have near-term policy implications; the latter is applied research designed to produce broad guidance for policy makers. Policy research "is similar to disciplinary research in that it gives heavy weight to hypothesis formulation, to rigorous analysis, and to agreed-upon statistical standards of evidence. It differs, however, in that its objective is to produce policy implications that have some hope or expectation of being taken seriously."

Policy analysis, which is our concern, in turn differs from policy research in several important respects (Hanushek, 1990:147-148):

> Its focus is highly governed by the detailed specifications of contemporaneous programs or proposals. It generally has a very short time frame. And, perhaps most important, it is very client-oriented. . . . Producing relevant answers takes precedence over theoretical elegance, statistical rigor, and, perhaps, completely balanced and fully qualified results. Its objective is to bring the best currently available information to bear on very specific questions. It is not a showcase for new and different analytical methods. But that does not imply the work is in any sense easier—just that it is different.

From the perspective of policy analysis, an important concern is how to maximize the flow of relevant and useful knowledge from basic and applied

Copyright National Academy of Sciences. All rights reserved.

social science research to the people who have to evaluate specific alternative legislative proposals. The current information flow is not necessarily optimal. Funding constraints for research, particularly research that is targeted to improving the policy analysis function, obviously play a role. In addition, characteristics of the basic and applied research worlds in the social sciences—such as disciplinary fragmentation and the imperatives of academic publishing, which emphasize path-breaking work over replication and consolidation—often militate against providing the kinds of information that policy analysts most need.

Turning to the other end of the link, we do not naively assume that there is a direct connection between the information generated by policy analysis and the outcome of the legislative process. The information may not be of sufficient quality or relevance to the issue at hand. Or the information may be reasonably good but lead to the conclusion that there is no clearly dominant policy choice—one option may be attractive because of low costs but unattractive for other reasons, and vice versa. More important, policy information is only one input to the decision process. Other inputs include political information and the decision makers' own beliefs and values, which act as information filters. Indeed, policy information may often be consulted simply to justify a decision that has already been made.

Other fundamental characteristics of the modern political process shape and constrain the role of policy information. First, because the process involves multiple issues and multiple actors with different perspectives, it inevitably engenders compromise and, indeed, it is designed to do so. Although such compromises may often represent the fruit of careful consideration of the available information and reconciliation of competing objectives, they may also involve blatant horse-trading or logrolling with little concern for the merits of the various provisions.

Second, because the process confronts decision makers with both multiple issues and heavy time pressures, it inevitably reduces the attention that they can devote to navigating the ocean of information that often engulfs them. This situation strengthens the role of intermediaries. One experienced observer (Hoagland, 1989) has concluded that members of Congress rely on the following sources of information in making legislative decisions, in order of decreasing importance: the member's personal staff and committee staff, constituents, fellow Senators and Representatives, newspapers, newsletters, studies and research, and congressional hearings. Of course, staff members, constituents, and colleagues may rely more heavily on information sources such as studies and research; however, the congressional member must take on trust that a staff member or other intermediary has properly interpreted study and research results.

Heavy time pressures also constrain the analysts who are involved in producing policy information. They often lack the time to do as thorough

Copyright National Academy of Sciences. All rights reserved.

a job as they might prefer or to consult all relevant outside experts. Once a policy debate is under way, they often cannot wait for development of improved modeling tools. If existing formal models are not suited to the task at hand, the analysts must work on a more ad hoc basis to develop the best estimates possible with other methods.

Third, the process often constrains policy estimates to be consistent with other information in a manner that may, in some cases, distort their meaning. For example, administration (and CBO) estimates of proposed legislation are constrained to the administration's economic forecasts, which may be based in part on political considerations. Or estimates may be constrained to be consistent with prior estimates for a similar proposal.

Finally, the process places a premium on certainty. Legislators, operating in an era in which legislative initiatives that add to budgeted expenditures must be offset by reductions elsewhere, want to have numbers that "add up." They do not want to be informed that a cost estimate is shaky or has a likely range of plus or minus several billion dollars. In fact, to our knowledge, information about costs and population effects of proposed legislation is hardly ever transmitted to the Congress in a form other than as a set of specific estimates —"point estimates" in the parlance of statistics. At most, policy analysts may indicate that a particular number is more conjectural than they would like.

One effect of the emphasis on certainty is that legislators frequently engage in fine-tuning policy alternatives in a way that cannot be supported by the available data. They will often request more and more model runs to determine the effects of this or that minor change on this or that specific group in order to develop a legislative package that meets a variety of objectives. They will see no need to limit their demands for information, at least until an external deadline intervenes, such as the end of a fiscal year. Yet the existing information base may be totally inadequate to distinguish reliably among policy alternatives that differ only marginally or that affect small groups.

For all these reasons, actual legislative outcomes may, to the outside observer, bear little relationship to the available information about their merits. In many cases, the legislation may be based on a greater degree of certainty in the numbers than is warranted. In other cases, the legislation may ignore the best —or any and all—information that is available.

Yet we believe that it is completely unfounded to reach the cynical conclusion that policy information is only window dressing for the political process. There is ample evidence that decision makers seek information to help them make good policy choices in the presence of uncertainty; indeed, the role of information in shaping policy has increased enormously in the recent past. Staff members close to the legislative process have testified that members of Congress want both political and policy-oriented information, about more and

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

more aspects of proposed legislation.[3] On the political side, they want information about how a bill will affect them in the next election and how it will affect relationships between the executive and legislative branches. On the policy side, they want to know how much a bill will cost—including who will pay and who will benefit, the effect on vulnerable groups, the effect on recipients, and the impact on basic societal behaviors and values. They also want to know how one piece of legislation will interact with another, how a bill will affect U.S. international competitiveness, and how a bill will affect state and local governments and the private sector.

What we have sought to keep in mind in our work is that policy information generally bears an indirect rather than a direct relationship to legislative outcomes. Recommendations must take into account the political, social, and personal realities that channel and filter the flow of information and must seek as much as possible to build on the strengths, and work around the weaknesses, of the process whereby information is generated and used in shaping legislation.

A CASE STUDY OF POLICY ANALYSIS: THE FAMILY SUPPORT ACT OF 1988

The policy analysts at work today, both congressional or executive agency staff and employees of contract research firms, have a variety of data sources and estimating tools to call on in responding to the needs of decision makers for information about alternative legislative proposals. In this section we illustrate some of the capabilities and limitations of microsimulation models vis-à-vis other approaches for policy analysis through considering how they were and might have been used in the legislative process that culminated in the Family Support Act of 1988.[4] The Family Support Act (FSA) includes several major provisions that alter the nation's support for poor children and their families:

---

[3] Personal communications to the panel, June 27, 1989, from William Hoagland, minority staff director for the Senate Budget Committee; Wendell Primus, chief economist of the House Ways and Means Committee; and Randall Weiss, former chief economist of the Joint Committee on Taxation, now with Deloitte and Touche.

[4] Obviously, one example of policy analysis cannot be viewed as representative of the entire class. The kinds of estimation tools that could be used and the problems encountered will differ, depending on the scope of the proposed legislation, the number of specific provisions, and other factors. Nonetheless, the Family Support Act is a good example for our purposes because of its wide scope and because it illustrates clearly some of the major limitations as well as capabilities of today's microsimulation models and other policy analysis tools.

The discussion in this section features the hypothetical uses of various types of models in producing estimates for the Family Support Act in order to illustrate the advantages and disadvantages of microsimulation versus alternative techniques. References to the modeling strategies that the analysts actually used are based on information obtained from interviews conducted by panel staff in the spring and summer of 1989 with Janice Peskin at CBO, Reuben Snipper at ASPE, and Wendell Primus at the House Ways and Means Committee; see also Congressional Budget Office (1989b, 1989d), Cottingham and Ellwood (1989), Rovner (1988), and Snipper (1989).

Copyright National Academy of Sciences. All rights reserved.

a program to increase the amount of child support that is obtained from noncustodial parents; a program to provide job opportunities and basic skills training to recipients of Aid to Families with Dependent Children (AFDC); a program of supportive services for families who move off welfare, including extended child care and Medicaid eligibility; and amendments to the AFDC program itself, to expand coverage of two-parent families in which the primary earner is unemployed and to increase the incentives for AFDC families to seek jobs by increasing the amount of earnings that is disregarded in computing benefits and raising the allowable deduction for child care expenses. The latter provisions were restricted by the 1981 Omnibus Budget Reconciliation Act (OBRA) that embodied the Reagan administration's program to cut back government spending and to reduce taxes. (Another provision that was originally considered for the FSA but did not survive the legislative bargaining process was mandating a minimum AFDC benefit standard in all states.)

The analysts at CBO and ASPE faced a difficult job in developing estimates for the proposed legislation. Because of the agreement in advance that the costs of a reform package could not exceed \$3.0-\$3.5 billion, the analysts knew that cost estimates would be particularly important to the decision makers. They also knew that the effects of the proposed legislation on particular population groups (often referred to as distributional effects) could be important for some provisions. For example, decision makers were interested in the impact on key states of the proposal to mandate a minimum AFDC benefit standard.

It was immediately apparent that no single formal model existed with the capability to handle all of the proposed provisions. Moreover, for some provisions, no formal model existed that could do the job. The analysts developed the needed estimates using several different formal models together with their own specially tailored calculations. (It is not unusual for policy analysts to use different techniques, for example, to produce estimates of costs as distinct from distributional effects or to handle different provisions of a legislative package; the very disparate provisions of the FSA virtually dictated such an eclectic approach.)

Provisions That Could Be Estimated With Microsimulation Models

We consider first one program feature that the analysts could evaluate with existing microsimulation models: extending program coverage to poor, unemployed two-parent families across the nation. Prior to the FSA, all states provided coverage to single-parent families who satisfied the income and assets criteria and to two-parent families who could not support themselves because of disability, but 22 states did not provide coverage to two-parent families who could not support themselves due to unemployment of the principal earner. The

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

original FSA proposal (and the final bill) mandated coverage of the latter group of families in all jurisdictions.

A key issue is the advantages and disadvantages of using microsimulation modeling rather than more aggregate techniques to estimate the total added costs of this proposal and the number of added beneficiaries. Considering the roughest, most global approach first, an analyst could elect to prepare overall estimates on a calculator or personal computer as follows: from current published administrative data for the AFDC program, compute the ratio of the unemployed-parent caseload (benefits) to the total caseload in the 28 states with the program; use this ratio to calculate the added caseload in the remaining states. The analyst could go a step further to provide more detailed information: from the Integrated Quality Control System (IQCS) database,[5] run tabulations, using a statistical analysis package such as SAS, of characteristics of current unemployed-parent beneficiaries (such as race of head, size of family, and prebenefit income), and use the resulting distributions to estimate the characteristics of the added caseload.

Estimates produced in this manner are relatively easy and inexpensive to obtain, yet an analyst is likely to be dissatisfied. First, the estimates require the strong assumption that the ratio of the unemployed-parent component to the total caseload in states currently covering unemployed-parent families is a reasonable proxy for the ratio that would obtain in the remaining states. Second, the set of characteristics for which the analyst can provide detail is limited to the variables contained in the IQCS, some of which are of doubtful quality (see Chapter 5) and all of which are compromised by the same assumption that the caseload in the states currently providing unemployed-parent coverage provides all the needed information about the likely caseload in the other states. Also, the estimates do not take into account interactions with other proposed changes in AFDC or with other programs, such as food stamps and Medicaid. For example, the new AFDC beneficiaries would virtually all be eligible for food stamps, which would add costs; however, some of them would already be receiving food stamps, so their receipt of AFDC benefits would lower their food stamp allotment. Most important, this approach does not give the analyst a tool with which to respond readily to requests for estimates for alternative versions of the proposal to expand the unemployed-parent program.

A more sophisticated approach to producing the estimates would rely on cell-based modeling techniques. Depending on the complexity of the analysis desired, an analyst could implement a cell-based model using a spreadsheet program or call upon a programmer to develop the model. A simple cell-based model for our example might include a cell for each state, with aggregate state

[5] The IQCS is a large monthly sample of case records, drawn from administrative databases maintained by the states for the AFDC, food stamp, and Medicaid programs.

Copyright National Academy of Sciences. All rights reserved.

administrative data on caseload, benefits, and characteristics for unemployed-parent recipients and others, together with such relevant information as the state unemployment rate. The analyst could pair each state not currently covering unemployed-parent families with a state that does cover such families, has comparable characteristics for the basic caseload, and faces a comparable unemployment situation. Then the model could use the appropriate ratio to calculate the expected unemployed-parent caseload for each currently uncovered state.

Estimates produced from such a model would be made more realistic by treating each state individually. However, such estimates would still require the assumption that the experience of states previously covering unemployed-parent families is directly applicable to other states. Moreover, by basing the model only on administrative data for current recipients, the estimates would ignore the distinction between program eligibility and program participation. Participation rates for many income support programs are well below 100 percent: they are currently estimated at about 77 percent of eligible families for the basic AFDC program (Ruggles and Michel, 1987:34), even lower for the AFDC unemployed-parent program, and at about 60 percent of eligible households for food stamps (Doyle, 1990:12). In addition, they vary significantly by state and family characteristics such as prebenefit income. (States show wide variations, which are not well understood, in estimated participation rates for the unemployed-parent component of AFDC.) Experience in the 1970s demonstrated the pitfalls of relying solely on administrative caseload data to develop projections of future caseloads. Projections of continuing sizable increases in AFDC, based on the explosive growth in the program in the 1960s, were off the mark, because the eligible population was growing at a much slower rate.

Determining program eligibility in a model requires a database, such as the Current Population Survey (CPS) March income supplement, that covers the entire population and includes relevant variables about family composition, income, and other characteristics. An analyst could expand the cell-based model in our example to include cells for groups of families within each state (such as unemployed-parent families in several income classes), develop eligibility estimates for each cell by tabulating the March CPS, apply participation rates based on regression analysis of the CPS data, and finally apply estimates of average benefits by cell.

However, the cell-based approach would quickly lose appeal. The analyst would want the model to produce estimates for current recipients that approximate the administrative data and, hence, would find it necessary to experiment with the specification of the cells and would most likely want to expand their number. The analyst would also find it very difficult to determine eligibility for cells, given the complex nature of the eligibility provisions for the AFDC program, which take into account in a very detailed way family composition, income sources and amounts, expenses, and asset holdings. (Also, because the

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

March CPS does not contain all of the needed information, the analyst would have to find a way to use other data sources.) In addition, with a cell-based model, the analyst would be hard-pressed to model alternative versions of the unemployed-parent program, to model interactions with other provisions or other income support programs, or to produce additional information about the effects on particular groups in response to decision makers' queries.

Microsimulation modeling conceptually solves many of an analyst's problems in developing detailed cost and distributional estimates for expansion of AFDC coverage of unemployed-parent families. Instead of working with aggregated or tabular data, microsimulation models of income support programs operate directly on large databases, such as the March CPS (augmented with other data sources), that contain information for many individual families and people chosen to represent the entire population. The models attempt to mimic the operation of programs such as AFDC, subjecting each sample record in the database to a set of tests to determine which family members comprise a potential assistance unit, whether the unit is in fact eligible, and the amount of benefits to which the unit is entitled. The models then simulate the decision by each unit as to whether or not it will participate in the program. The models go through this process for current program provisions and for each alternative that is specified, retaining the simulated values on each record. Hence, an analyst can readily specify tabulations that compare alternative proposals and assumptions on a wide range of dimensions, limited only by the content and sample size of the March CPS.

In addition, the models can simulate many different kinds of program changes separately or concurrently: for example, they can model the interaction between expanding the unemployed-parent program and increasing the deductions for child care expenses and a fraction of earnings. Moreover, because the major models for AFDC also include modules for other related income support programs, such as supplemental security income and food stamps, they are well equipped to trace through program interactions in a consistent framework.

In fact, both CBO and ASPE staff relied largely on the TRIM2 microsimulation model (Transfer Income Model 2; see Appendix to Part II) to produce estimates for the unemployed-parent provisions of the FSA. They also produced rough estimates to compare with the TRIM2 results: TRIM2 projected a larger increase in the unemployed-parent caseload than did the rough calculation.

However, state-of-the-art microsimulation modeling arguably added to the cost of developing policy impact estimates for this and other provisions of the Family Support Act. Added costs stem from the complexity of the models and their implementation in mainframe, batch-oriented computing technology. Microsimulation models have long had a reputation for high computer costs relative to other types of models. Today the cost for a computer run with a model like TRIM2 is quite reasonable, because the models have been optimized for sequential batch processing and use other strategies, such as operating with

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

a subset of CPS households below a specified income level, to reduce computer costs. Nevertheless, total costs can mount up because the use of these models inevitably requires highly skilled programming staff as well as multiple runs.

Current models, because they include many parameters, can readily simulate many proposed program changes—such as raising or lowering the value of the child care expenses deduction—simply by resetting a switch. However, other simulations would involve changing the computer code, which can quickly add time and costs to the project. Indeed, the FSA version finally adopted for the unemployed-parent program—namely, that the states not previously covering these families could limit coverage to as few as 6 months of the year—could not be modeled in TRIM2 during debate on the FSA. Both CBO and ASPE analysts adjusted the TRIM2 estimates outside the model to develop the final estimates for this provision, using monthly information provided by TRIM2. (After the FSA was enacted, ASPE invested the time and money needed for the TRIM2 programmers to revise the code appropriately to handle this provision inside the model.)

The TRIM2 model could readily handle some other changes to the basic AFDC program—specifically, increasing the child care expenses deduction and earnings allowance—and the ASPE analysts used TRIM2 to estimate the effects of these changes. However, the CBO analysts used TRIM2 to analyze these changes only for families made newly eligible for the AFDC program; for families already receiving benefits they used a "benefit-calculator" model running with an IQCS database.

Benefit calculators are a simplified type of microsimulation model. Like the TRIM2 class of income-support program models, benefit calculators for programs such as AFDC operate at the level of the individual assistance unit, mimic the program rules for determining benefit amounts, and take account of some joint program interactions, such as those between AFDC and food stamps. Because they look only at administrative data for current beneficiaries, they are smaller, less expensive, and much easier to use than the TRIM2-type models that simulate the entire population. Moreover, by definition, benefit calculators faithfully reflect the characteristics of actual beneficiaries (at least as reported to caseworkers and transcribed for the IQCS). In contrast, the CPS-based simulations for the entire population from TRIM2-type models have historically produced higher estimates of the proportion of the AFDC caseload with earnings and in other ways painted a different picture of the caseload from that drawn from the administrative records. (The reasons for these differences are not well understood. They may result from such phenomena as undercoverage of low-income population groups in household surveys like the CPS, as well as differences between the IQCS and the CPS in definitions of variables, question wording, etc.)

On the disadvantage side, however, benefit-calculator models include only current program recipients. Hence, although they may provide good estimates

Copyright National Academy of Sciences. All rights reserved.

for proposed program contractions, they cannot simulate the full impact of proposed program expansions, such as a liberalization of allowable deductions, that may bring new participants into the program as well as raise benefits for current participants. Moreover, in the case of program expansions, the strategy of using a benefit-calculator model for current beneficiaries and a model such as TRIM2 for newly eligible cases may introduce a measure of inconsistency into the estimates.

Finally, TRIM2 could be used to model the provision to mandate a federal minimum benefit standard for AFDC. The modeling task was very difficult because of the need to anticipate the behavior of the states with regard to raising benefits in the absence of a mandate, as well as the need to work out the expected time path of state actions to raise their benefits to meet a mandated standard. The ASPE analysts built very large and complicated spreadsheets to develop appropriate estimates of benefit levels by state and household size for TRIM2 to use in its baseline simulation of current law and its simulation of proposed law, including a mandated minimum benefit. CBO analysts used their benefit-calculator model to develop estimates of benefits for current recipients and TRIM2 to develop estimates of benefits for newly eligible units under a minimum benefit standard. The proposed provision was dropped early in the FSA debate because of its projected large costs.

Provisions That Could Not Be Estimated With Microsimulation Models

Other major provisions in the FSA, which covered transitional support services, child support, and jobs and training, were arguably the heart of the legislation, but no existing microsimulation model could readily analyze them. The available income-support microsimulation models are cross-sectional in nature, designed to produce a snapshot of the caseload in a particular month or on an average monthly basis for a given year. They are not oriented to the type of longitudinal analysis required to assess the effects of providing extended child care and Medicaid benefits to AFDC families who increase their earnings sufficiently to become ineligible for basic benefits.[6] ASPE analysts initially used TRIM2 to produce estimates of the characteristics of those eligible for transitional benefits to feed into their spreadsheet models. However, they discontinued this approach because it proved too difficult to produce estimates that could interact in a consistent way with the spreadsheet estimates for the jobs program.

Some microsimulation models in the 1970s—such as the MATH and KGB (Kasten-Greenberg-Betson) models—included modules for jobs programs.

[6] Several current models—such as DYNASIM2 and PRISM (Pension and Retirement Income Simulation Model)—take a dynamic, longitudinal approach to microsimulation, but these models are focused on modeling retirement income programs over the long term (typically, 20-40 years or more into the future) and are not adapted to modeling detailed short-term changes to income assistance programs.

Copyright National Academy of Sciences. All rights reserved.

The TRIM2 staff are currently developing a module that simulates a labor supply response to changes in income support programs. However, in the mid-1980s, no model had a usable module that could readily be adapted to prepare cost estimates for the jobs and training component of the FSA. Similarly, no microsimulation model could readily simulate the child support provisions. Moreover, even if suitable microsimulation models for the impact of these provisions on individual assistance units had been available, those models would not have been able to simulate the response of the state governments to various administrative changes—such as higher federal matching rates for state expenditures—that represented major elements of the child support and jobs programs in the FSA.[7]

The agencies could have decided to invest in model capabilities and relevant data sources in anticipation of the welfare reform debate. For example, information about both parents of children in single-parent families—not available in any data source today—could permit the development of microsimulation models to estimate the likely offsets to AFDC costs of mandatory withholding of child support payments from the wages of the noncustodial parent. However, resources were exceptionally scarce for such investment in the first half of the 1980s, and the likelihood that welfare reform would be a live topic with serious chances for enactment seemed almost nil as late as 1986. Once the debate was under way in early 1987, there was no time to invest in making substantial modifications to the current income-support microsimulation models to handle the child support, jobs program, or transitional assistance provisions of the FSA. The CBO and ASPE analysts developed their own specially tailored models to accomplish the task of producing policy impact estimates for those important components of the FSA.[8]

Problems in the Estimation Process

In briefly reviewing the FSA cost-estimation process, we noted policy areas for which microsimulation modeling could have been appropriate but for which suitable models and databases were not available. There were also several

[7] Microsimulation techniques are most often used to simulate the behavior of individual households and people, but in concept they can be used to simulate larger entities as well, such as state governments, corporations, or hospitals. However, no such model existed to simulate state government responses to the many administrative changes in the child support and jobs programs of the FSA, nor would it likely have made sense to build such a model, given inadequate data and the varied nature of the proposed changes.

[8] The issue of whether to make investments in anticipation of future legislative debates is a difficult one. Investment may well be wasted if the policy debate takes an unanticipated turn; however, failure to invest may unnecessarily limit the benefits from microsimulation or other formal models. We discuss this issue in Chapter 3, strongly urging investments that increase the capacity of microsimulation (and other) models for flexible, timely response to changing policy needs.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

problems that confronted the analysts in producing the estimates, whether they were using a microsimulation model or another approach.

One problem concerns methods for projecting policy estimates into a future period. Congress requires that cost estimates be made for the current fiscal year and the five succeeding years (projections for changes to such programs as social security must be made over a much longer period). Yet, at best, many of the data available for use in models are 1-2 years out of date and, of course, do not represent future conditions. Techniques exist, and have been used in policy contexts, to "age" microsimulation model databases to future years, prior to conducting policy simulations. One commonly used approach, called "static aging," is to reweight the March CPS (or other) database on several dimensions, such as demographic characteristics and household composition, for which outside projections are available. Another approach, called "dynamic aging," is to simulate processes such as birth, marriage, death, educational attainment, and labor force participation, and generate a new database for each year of the projection. The population projections that lie at the heart of static aging and typically serve to calibrate the output from dynamic aging derive from cell-based models long maintained by agencies such as the Census Bureau and the Actuary's Office of the Social Security Administration (SSA).[9]

None of the existing income-support program models uses dynamic aging techniques. Models such as TRIM2 include a static aging capability, but its use to develop aged databases has historically been a complex, expensive, and time-consuming task, and it may distort relationships among key variables, given the inevitable limitations of the available projections. The usual practice of CBO and ASPE analysts—and the FSA estimates were no exception—has been to use TRIM2 to develop estimates of percentage differences in the costs and caseloads of current versus proposed provisions for income support programs using the latest available database (typically the March CPS for the preceding year). They then apply these percentage differences to 5-year projections of costs and caseloads for the current programs developed by using other methods.[10]

The preparation of current services projections for AFDC (which, for the administration, are developed in HHS by the Family Support Administration—now part of the Administration for Children and Families) is a complicated process that makes use of a variety of information, including forecasts of unemployment and inflation (which, for the administration, are developed by the Office of Management and Budget and take account of forecasts of leading macroeconomic models); information about likely actions by the states to

[9] See Grummer-Strawn and Espenshade (Volume II) for a description and review of the Census Bureau and SSA population projection models.

[10] The Food and Nutrition Service, in contrast, has often invoked the in-model static aging capabilities of the MATH model. The Office of Tax Analysis (OTA) in the U.S. Department of the Treasury also uses in-model static aging routines for tax policy simulation with the OTA model.

Copyright National Academy of Sciences. All rights reserved.

modify benefits; and other relevant information, such as projected growth in female-headed families. Analysts have often made use of time-series regression models in developing current services projections. [11] For example, in a time-series model developed for HSS, the equation for projecting caseloads under the AFDC basic program includes variables for the number of female-headed households, current and lagged unemployment rates, seasonal movements in the caseload, weighted average standard of need across states, and a variable for the effect of the 1981 OBRA changes. The equation for the unemployed-parent caseload includes the same variables except that the size of the labor force in states with the unemployed-parent program substitutes for the number of female-headed households.[12] CBO has also developed and used similar time-series regression models for more than 10 years.

The strategy of applying TRIM2 estimates of percentage differences between alternative and current programs to independently developed 5-year projections of costs and caseloads for current programs can provide the analysts with a number of advantages. For example, the current services projections can reflect the latest information on state actions to change AFDC benefits and, generally, it is probably easier to revise them than to redo the database aging process inside of TRIM2. However, developing the projections for current as well as alternative programs inside TRIM2 could provide much more detailed distributional information within a consistent framework. Whatever method is used, there are many difficult problems in projecting future costs and caseloads that have never been fully addressed, nor have the results of alternative methods ever been subjected to rigorous comparative evaluations.

Another major problem in policy estimation concerns the possible effects of policy changes on behavior that, in turn, might have short-term or long-term feedback effects on program costs and caseloads. Such effects could be profound. For example, the transitional Medicaid and child care provisions of the FSA could well reduce the rate of reenrollment in the AFDC program. The child support enforcement provisions over the long term could well alter such basic behaviors as fertility and divorce.

Although Congress does not generally require that such behavioral effects be estimated, policy analysts will try to do so when they believe such effects may be important. However, current models and data sources provide little help for analysts in this regard. Although microsimulation models are conceptually well designed to model the effects of programs on individual behavior, income-support program models do not today (and rarely did in the past) provide a

---

[11] Single-equation time-series models are very simple cousins of the complex, multi-equation macroeconomic models of the national economy.

[12] The unemployed-parent model proved unstable: that is, the coefficients on the explanatory variables were sensitive to the precise specification and time period used, indicating that the model may not be reliable (see Beebout and Grossman, 1985; Grossman, 1985).

Copyright National Academy of Sciences. All rights reserved.

capability for simulating behavioral responses other than the basic participation decision.[13] The cost and complexity of implementing behavioral response capabilities have been contributing factors; however, the main impediment has been the absence of an adequate base of research knowledge providing estimates of sufficient reliability in a form suitable for modeling. The analysts who worked on the FSA were able to draw on several research studies to estimate the effects of the jobs program on earnings of AFDC recipients and the consequences for AFDC program costs. However, there were no reliable research findings to permit assessing the behavioral effects of the FSA in any comprehensive manner, whether inside or outside a microsimulation modeling framework.

A third problem area concerns assessment of the quality of the point estimates that are typically provided to decision makers. Congress does not require that analysts accompany policy estimates with information about their quality, and, indeed, as discussed earlier, the political process places a premium on certainty in the numbers. The FSA analysts informed decision makers that some of the estimates were more problematic than others. In order to gauge problems, they compared estimates produced by different agencies and used alternative estimating methods when possible (e.g., comparing the TRIM2 estimate for the expansion of the unemployed-parent program with a rough calculation based on the ratio of the unemployed-parent component to the total AFDC program in states already covering unemployed parents). However, the analysts made no systematic assessment of the variability in the estimates or of their sensitivity to the various assumptions, procedures, and data inputs used to develop them, nor did the analysts present "error bounds" to accompany their final "best" estimates. Validation admittedly is a daunting task, particularly for large, complex models such as the current income-support program microsimulation models. However, as we stress throughout the report, validation is one of the most important tasks to undertake if there is to be adequate information on which to base improvements to modeling tools and data sources that can lead to improved information for future policy debates.

[13] Dynamic microsimulation models, such as DYNASIM2 and PRISM, incorporate many behavioral response functions, such as marriage and labor force transition probabilities, that they use to project their databases forward in time. However, as noted earlier, such models are not currently adapted to simulating detailed changes to the AFDC program.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

3

Improving the Tools and Uses of Policy Analysis

A STRATEGY FOR INVESTMENT

The agencies that supply policy analysis for social welfare issues need to improve their databases and modeling tools. Although the climate of support for a well-targeted investment program is more positive on the part of both branches of government than at any time in recent years, agencies still face difficult choices in deciding where best to direct limited investment dollars.

Microsimulation models, in our view, offer important capabilities to the policy analysis process—in particular, the ability to evaluate fine-grained as well as broader policy changes from the perspective of their impact on subgroups of the population that are of interest to the user. However, microsimulation models do not serve all policy analysis needs, and the capabilities they provide typically require highly complex model structures and databases that can be resource-intensive for development and use. Other tools that are available for policy analysis, which may, in particular circumstances, offer appropriate and cost-effective capabilities, include:

- large-scale macroeconomic models based on systems of simultaneous equations estimated with historical times series, which can project the effects of aggregate factors, such as rising inflation or changes in the federal budget deficit, on aggregate outcomes, such as gross national product or unemployment;
- single-equation time-series models, which can use historical experience to project aggregate costs and caseloads for specific programs, such as AFDC

Copyright National Academy of Sciences. All rights reserved.

or food stamps, under varying assumptions about changing family composition, inflation, employment, or other factors;

- cell-based models, which can develop estimates of the effects of proposed policy changes, such as raising the social security retirement age or payroll tax, for the specified population subgroups (such as people categorized by age and sex) that comprise the "cells" in the model;
- econometric models of individual behavior, which can estimate the probabilities that decision units (e.g., families and individuals) will make different program participation choices or otherwise alter their behavior in response to a policy change; and
- second-round effects models, which can develop estimates of the longer-run effects of policy changes, such as the effects of changes in tax laws on long-run changes in the character of economic markets.

Many policy applications require more than one modeling technique, and, indeed, many models themselves incorporate multiple approaches. Some models are explicitly "hybrids"—for example, models that link a microsimulation-based model of the household sector and a macroeconomic-based model of the economy. Other models reflect primarily one approach but make use of the outputs of other kinds of techniques. Hence, agencies will benefit from adopting a broad perspective as they consider how best to improve the tools and associated data they need for policy analysis.

In framing an investment strategy, agencies confront the fact of continual change in the policy landscape even though the basic concerns of social welfare policy have not changed much in the years since the Great Depression and World War II: for example, the current interest in revamping the nation's patchwork system of health care financing carries echoes of similar debates going back at least as far as the Truman administration. However, the relative priorities among issues change, as do the particular features of the debate on each issue.

In looking to the next 5-10 years, it is clear that issues related to health care will more and more occupy center stage as the nation faces escalating needs and costs. Thus, it is obvious that investments should be made to improve the capability for modeling health care policies, but it is by no means clear precisely what form these investments should take. Moreover, it would be unwise of agencies to assume that other policy topics, such as income support for the poor or retirement income, will be quiescent and that they can safely defer investment in modeling capability for those topics.

What agencies *can* assume is threefold: that policy options are likely to involve several topics—for example, the use of tax policy to achieve health care cost containment or income support goals; that changes in the debate within and across topics will occur, sometimes with stunning speed—for example, tax policy debate may well shift from capital gains to energy taxes; and that policy

Copyright National Academy of Sciences. All rights reserved.

makers will certainly not reduce and may well expand their appetite for detailed information about the impacts of proposed policy changes.

These features of the policy landscape rule out extreme strategies. There is little merit in a totally proactive strategy of trying to forecast future analysis needs in great detail and so developing a highly targeted investment program, because the inevitable miscalculations will result in wasted dollars. Nor is there merit in a totally reactive strategy of producing on-the-fly estimates in response to the policy needs of the moment, because this posture throws away any opportunity to develop analysis tools that have a longer-term payoff or that can lead to improvement in the quality of estimates. Instead, in our view, agencies need to accord priority to investments in policy analysis tools that maximize their capacity to respond flexibly to shifts in policy interests and that provide capabilities for evaluating the quality and meaning of the estimates and maintaining high standards of documentation and replicability. Given the current climate of constrained resources, agencies also need to seek strategies that promise to reduce costs of model development and future application. All of this is a tall order, particularly in the case of large, complex models. In our review of the investment needs for microsimulation, we identified several approaches that we believe offer promise of success for this class of models and, possibly, also for other classes.

One imperative is for agencies to act upon the maxim that, in the case of multifaceted policy models that are intended to have a long-term use, it is essential to allocate sufficient resources and attention for good model design and implementation at the beginning. Attempts to achieve "savings" in development and testing are all too likely to have disastrous consequences for the users of the model. (The recent failure of the Hubble space telescope to achieve full functionality offers a highly visible object lesson of this point.) Another strategy, with high potential payoffs for making complex policy models more cost-effective, is for agencies to take advantage of the important technological advances in microcomputing hardware and software that are already having an impact in the business and academic worlds. Still another worthwhile approach is for the agencies to work for changes in the policy analysis community, to foster wider use of complex models by analysts and researchers, to encourage production of research that is relevant to modeling needs, and to improve upon some of the ways in which agencies have traditionally operated, both individually and as a group. We discuss these approaches in detail with respect to microsimulation models in Part II.

In determining investment strategies, whether for microsimulation or other types of models, it is important to focus on the goal of policy analysis, which is not just to produce numbers, but to produce numbers that provide useful guidance for decision making. "Useful" in this context implies many things, including relevance to the issue at hand, timeliness, and multidimensionality, for example, shedding light on distributional as well as aggregate effects. Most

Copyright National Academy of Sciences. All rights reserved.

important, "useful" implies that the numbers are of reasonable quality. Given a known level of quality, analysts and policy makers can debate the merits of investing in the next increment of quality or investing in some other dimension. However, for many extant policy analysis models, the level of quality is simply unknown.

In our review, we found that policy analysis agencies have generally skimped on investment in model validation and related activities, such as archiving and documentation, that support validation. In the absence of systematic validation efforts, agencies are blindly spending precious dollars for model application and development: they can neither assess the return to date on their investment in policy analysis tools in terms of the quality of the estimates nor make rational decisions about future investments to improve quality.

Moreover, in the absence of validation, decision makers are using policy analysis estimates as if they were error-free. In fact, all estimates have some uncertainty associated with them, and the level of that uncertainty, in many cases, may be high. By ignoring the errors in estimates, policy makers may reach decisions, with possibly far-reaching social consequences, that they would not have made if they had realized how uncertain was the available information. Or they may waste time and resources in exploring very fine-grained policy alternatives that cannot be distinguished reliably with the available information. Finally, without information on uncertainty, policy makers cannot determine what investments in databases and models are most needed to improve the quality of the estimates for future decision making.

We vigorously urge investment that will facilitate validation of model estimates on a regular and systematic basis. We also urge investment to improve the underlying databases for modeling and for the applied and basic socioeconomic research on which models rely for many important elements such as behavioral response functions. Although little systematic information is available about the overall quality of estimates produced by many models, there is ample evidence that critical data inputs to models have deteriorated in quality and relevance. Moreover, there is evidence that, in some cases, problems with data have had serious consequences for social and economic policy.

In this chapter we discuss in considerable detail important overall improvements that are urgently needed with regard to the quality and availability of data to support a wide range of policy analysis applications using a variety of modeling tools. We also discuss major changes that are needed in the approach of policy analysis agencies to validating and documenting model results and communicating uncertainties in these results to policy makers. We offer recommendations on each of these topics.

DATA QUALITY AND AVAILABILITY

Policy analysis of alternative legislative proposals is undeniably a "data-hungry"

Copyright National Academy of Sciences. All rights reserved.

enterprise. Although some analyses require relatively few data, many kinds of analyses depend heavily on large amounts of data, and all analyses require data of good quality. Among the widely used policy analysis tools, both macroeconomic and microsimulation models stand out in their voracious data appetites. The major macroeconomic models rely on hundreds of historical time series extending over many decades that capture specific elements of aggregate economic behavior (e.g., public and private spending, industrial production, employment, prices). Microsimulation models require large samples with rich sets of information on each individual record in order to estimate policy effects for detailed population subgroups.

In this regard, policy analysis simply mirrors the larger society: contemporary Americans are avid consumers of information of all types, and the federal government supplies much of the data that the public and private sectors use for everything from entertainment to research and analysis to critical decision making. At the lighter end of the spectrum, media outlets rely heavily on government statistics for a wide range of information about the characteristics of Americans and their predilections and problems. At the weightier end, statistics such as the monthly unemployment rate and consumer price index have consequential impact on the national economy, indirectly through their influence on financial markets and business behavior and directly through their use in indexing some wage contracts, entitlement programs (e.g., social security and food stamps), and federal grant programs to states and localities. In between, federal statistics are used by all levels of government, businesses, nonprofit organizations, and academia for all manner of research, planning, decision support, and evaluation purposes.

Good data are obviously necessary for good analysis and informed decision making; consequently, improvements in data quality and relevance for policy analysis and other purposes represent worthwhile investments on the part of the federal government. Certainly, a well-considered continuing program of investment in data (and modeling tools) needed for social welfare policy analysis seems warranted in light of the resources that are at stake. The federal government spends more than $300 billion annually on social insurance programs (including social security, Medicare, unemployment insurance, and workers' compensation) and almost $75 billion annually on public assistance programs (including supplemental security income (SSI), AFDC, food stamps, and Medicaid); state and local governments spend an additional $43 billion and $46 billion, respectively, on social insurance and public assistance programs (Bureau of the Census, 1991:Table 583).[1] In comparison, the entire statistical budget of the federal government is less than $2 billion in most years.[2]

[1] These figures are for 1988; social insurance expenditures exclude federal and state and local public employee pensions.

[2] The Office of Management and Budget (1990:Table 1) reported fiscal 1988 budget obligations for

Copyright National Academy of Sciences. All rights reserved.

However, we recognize that very difficult resource allocation issues arise in considering, first, the share of the federal budget to devote to data production and, second, the share of the federal data budget to devote to particular data needs. We do not pretend to have the answers to questions such as whether a dollar invested in improved intercensal small-area estimates of population and income for use in federal fund allocation and state and local government planning has a higher payoff than a dollar invested in improved sample surveys on income and health care for federal policy analysis use or than a dollar invested in improved input data for the national economic accounts. We do offer some observations and recommendations that we believe have a broad utility for improving data for policy analysis.

Investment in Data Production

A disturbing feature of the decade just completed has been the declining federal investment in the production of high-quality, relevant data for many policy areas. At the start of the decade, nine major federal statistical agencies experienced sizable cutbacks, amounting to a 20 percent reduction (in constant dollars) in their budgets between fiscal 1980 and fiscal 1983. Subsequently, budgets were adequate for the agencies to maintain and, in some cases, expand the core activities that remained after the initial reductions. However, across-the-board cuts implemented again in 1986 and yet again in 1988 resulted in an overall decline of 13 percent in the expenditures of the major statistical agencies between 1980 and 1988 (Wallman, 1988:13).

With regard to information specifically needed for social welfare policy, we note first that federal and state spending for social insurance and public assistance programs increased by 32 percent from fiscal 1980 to 1988 in real terms (Bureau of the Census, 1991:Table 583).[3] In contrast, spending for the statistical agencies that produce relevant data—including the Bureau of Economic Analysis, Bureau of Labor Statistics, Census Bureau, National Center

all statistical activities of the federal government-including programs of large and small statistical agencies, statistics-related activities of policy research agencies, and the programs of administrative agencies (such as the Immigration and Naturalization Service and the Internal Revenue Service) that generate statistical data as a byproduct of administrative actions—at $1.7 billion, including $0.2 billion for the 1990 decennial census. The 11 principal statistical agencies, including the Bureau of Economic Analysis, Bureau of Labor Statistics, Bureau of Justice Statistics, Census Bureau, Economic Research Service of the Department of Agriculture, Energy Information Administration, National Agricultural Statistics Service, National Center for Education Statistics, National Center for Health Statistics, Policy Development and Research Office of the Department of Housing and Urban Development, and Statistics of Income Division of the IRS, accounted for $0.9 billion. In fiscal 1990, the peak year of spending on the decennial census, total budget obligations of federal statistical programs are estimated at $3.0 billion, including $1.3 billion for the census.

[3] Fiscal 1988 budgets were converted to constant 1980 dollars by using the GNP implicit price deflators for federal nondefense and state and local government purchases of goods and services (from Joint Economic Committee, 1990:2).

Copyright National Academy of Sciences. All rights reserved.

for Health Statistics, and Statistics of Income Division in the Internal Revenue Service (IRS)—increased at most by only 12 percent in real terms from fiscal 1980 to 1988 and actually fell by 6 percent in real terms from fiscal 1985 to 1988 (see Table 3-1). A major new survey was introduced during the 1980s to support improved social welfare policy analysis—the Survey of Income and Program Participation (SIPP). However, repeated cutbacks in the SIPP sample size and length of panels greatly undercut its usefulness.

Although budget constraints in some cases encouraged agencies to scale back or eliminate outmoded programs, they have had many serious consequences. Most attention has focused on quality problems with basic economic statistics, such as monthly measures of retail sales, imports and exports, and the gross national product (see Council of Economic Advisers, 1990; Economic Policy Council, 1987; Juster, 1988; Office of Technology Assessment, 1989b; see also Sy and Robbin, 1990, who consider problems with a broad range of economic and other federal statistics as they affect policy uses of the data). Failure of concepts and measurement to keep up with economic trends (such as the shift from a manufacturing to a service economy), reductions in survey samples and in the availability of administrative records, and inadequate staff resources are among the factors cited for deterioration in basic economic data series.

These deficiencies in the quality and relevance of economic data have had important policy consequences. Schultze (Policy Users Panel, 1988) recounts the experience in the late 1970s and early 1980s, when the consumer price index (CPI) overstated the rise in the cost of living by some 1-2 percent a year, with serious economic consequences for wage escalation and overadjustment of social security and other federal entitlements.[4] Fuerbringer (1990) comments on inadequacies in economic data series that, with disturbing frequency, have resulted in large differences between the preliminary and revised estimates of the gross national product (GNP) and other key economic indicators. The preliminary estimates are heavily used for business decisions, and they influence decisions of policy-setting agencies such as the Federal Reserve Board.[5] Samuelson (1990) cites problems with current business surveys that produced overestimates of wages and salaries in 1989 and hence overestimates of projected federal tax revenues. The revised wage and salary estimates, based on

[4] The overestimate was caused by calculations regarding owner-occupied housing, which was treated in the CPI as an investment good rather than as an element in the cost of living that provided a stream of housing services. Hence, soaring interest rates and house prices in the 1970s led to overestimates of the rise in the cost of housing and thereby in the CPI. In 1983 the Bureau of Labor Statistics changed the measurement of housing costs to a rental-equivalence approach.

[5] David (1990:3) cites an example in which the Federal Reserve Board tightened interest rates in the summer of 1989, at least in part in response to a measure of declining inventories: "The higher rates forestalled home purchases and caused entrepreneurs to delay new enterprises. Subsequent revision of the data showed a large error in the original estimate."

Copyright National Academy of Sciences. All rights reserved.

TABLE 3-1 Trends in Funding for Major Statistical Agencies Providing Data for Social Welfare Policy

	Budget Obligations (millions of dollars)			Percentage Increase (constant dollars)	
Agency	Fiscal 1980	Fiscal 1985	Fiscal 1988	Fiscal 1980-1988	Fiscal 1985-1988
Bureau of Economic Analysis (BEA)	15.8	21.8	23.6	+ 4.9	- 4.5
Bureau of Labor Statistics (BLS)	102.9	170.6[a]	175.3	+19.5	- 9.8
Census Bureau (current programs only; not including censuses)	52.5	84.8[b]	94.3	+25.8	- 2.4
National Center for Health Statistics	43.3	42.8	54.4	-11.8	+11.6
Statistics of Income Division (SOI), Internal Revenue Service (IRS)	14.6	19.0[c]	17.2	-17.2	-20.5
TOTAL	229.1	339.0[d]	364.8	+11.7	- 5.5

NOTE: Analyzing statistical agency budgets over time is difficult. Some budget changes are more apparent than real because of transfers of programs from one agency to another. See notes a, b, and d for estimates of the effects of the largest transfers. Other changes reflect cyclical funding patterns: for example, including the population and economic censuses in the Census Bureau's budget would produce dramatically different trends depending on whether the comparison years represented high or low points in the funding cycle.

[a]The fiscal 1985 budget figure reflects a transfer of programs from the Employment and Training Administration of an estimated $23.7 million. If this amount is excluded from the BLS budget for fiscal 1988 (assuming conservatively that it did not increase in nominal terms), the increase for BLS from fiscal 1980 to 1988 in constant dollar terms is 3.4 percent instead of 19.5 percent.

[b]The 1985 budget figure reflects the introduction of SIPP. (The first SIPP interviews were conducted in fall 1983, and the design of overlapping panels was fully phased in by fiscal 1985.) The 1985 budget figure also reflects program transfers from other agencies of an estimated $2.3 million (Baseline Data Corporation, 1984:Table 1). If the transferred amount is excluded from the Census Bureau budget for fiscal 1988 (assuming conservatively that it did not increase in nominal terms), the increase from fiscal 1980 to 1988 in constant dollar terms is 22.9 percent instead of 25.8 percent.

[c]The fiscal year 1985 budget amount is from Wallman (1986:260). The amount reported in Office of Management and Budget (1986:Table 2) includes IRS field costs not previously assigned to SOI.

[d]The fiscal 1985 total budget figure reflects transfers from other agencies of an estimated $26 million (see notes a and b). If this amount is excluded from the fiscal 1988 total, the increase from fiscal 1980 to 1988 in constant dollar terms is 3.7 percent instead of 11.7 percent.

SOURCES: Unless otherwise noted, net direct budget obligations are from Office of Management and Budget (1986:Table 2) for fiscal 1980 and 1985 and from Office of Management and Budget (1990:Table 1) for fiscal 1988. Constant dollar comparisons were calculated by using the GNP implicit price deflator for federal government nondefense purchases of goods and services (from Joint Economic Committee, 1990:2).

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

more complete information that became available in spring 1990, necessitated downward revisions of projected tax revenues and a sharp upward revision in the forecast of the federal budget deficit revision that presented grave political and economic policy problems for the President and Congress.

Data series that are used for policy analysis of social welfare issues have also suffered deterioration in terms of quality and relevance to policy needs. A prime example is the continued reliance on outmoded concepts for characterizing families and economic decision units in widely used household surveys such as the Current Population Survey (CPS) and Consumer Expenditure Survey (see David, 1990). As Juster (1988:16) comments,

> In today's society, the traditional notion of the stable family as the unit of observation in economic and social statistics is in need of rethinking. For example, unrelated individuals in a modern household may have little or no information on the labor force attachment, or the income and wealth positions, of the other members of the household. Is that why teenage unemployment rates appear to be so high? And members of a household may depend more on transfers of income or wealth from outside the household for food and clothing and shelter, or for access to higher education, than on the income and wealth of household members. Is that why forecasts of college attendance have been too low?

Inability to provide adequate descriptions of today's complex family structures and relationships has made it increasingly difficult to assess many important policy initiatives for social welfare. Thus, analysis of child support enforcement programs, which offer the potential to reduce government income support costs, is hampered in the absence of joint information on the family circumstances of both the custodial and the noncustodial parents. There are other examples of problems in socioeconomic data series (see Baseline Data Corporation, 1984; General Accounting Office, 1984; Wallman, 1988):

- Reduced sample sizes and content of major surveys in a wide range of areas have limited the analyses that the data can support. Surveys with reduced content include the CPS (some supplements were dropped), the SIPP (one or more interview waves were dropped for some panels), and the Health Interview Survey. Surveys with sample size reductions include the CPS (a change that did not necessarily affect national estimates, but did affect state data, which are important for programs such as AFDC and Medicaid); the SIPP, for which the sample reduction—as much as 40 percent for some panels—was particularly drastic; the Health Interview Survey (the sample was later restored for this survey); the National Health and Nutrition Examination Survey (NHANES); and the youth cohort of the National Longitudinal Surveys of Labor Market Experience. As noted earlier, samples for business surveys, such as wholesale and retail trade, that are important for estimates of the GNP, projections of

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

budget deficits, and other policy purposes, were reduced, as were samples of tax returns prepared by the Statistics of Income Division.

- Cutbacks in the periodicity of many surveys, particularly health surveys, have adversely affected the ability of analysts to measure important trends. Thus, the NHANES was cut back in frequency from 5 to 10 years, the National Ambulatory Medical Care Survey from 1 to 3 years, the National Nursing Home Survey from 2 to 6 years, and the Survey of Family Growth from 3 to 6 years. The periodicity of several business surveys was also cut back. The American Housing Survey (formerly the Annual Housing Survey) was made biennial instead of annual.

We note just one example—from health care policy—of the impact of less frequent updates of important surveys. National medical care expenditure surveys were conducted in 1977 and 1980 but not again until 1987. Estimates originally prepared by the Congressional Budget Office (CBO) of the likely costs of covering prescription drug costs under Medicare were determined to be much too low, once the 1987 data, which showed a rising trend in prescription drug use on the part of the elderly, became available (see Chapter 8).

- Important differences in concepts across data series, which agencies have identified but not been able to address, have precluded definitive assessment of the quality of policy-relevant information. Thus, the personal income measures in the National Income and Product Accounts (NIPA), which provide important data for evaluating the quality of income measurement in household surveys, have never been disaggregated to permit appropriate comparisons for the household sector. (For example, the personal income estimates of interest and dividends in the NIPA include receipts of nonprofit organizations as well as households.)
- A range of measurement problems, which agencies have not been able to analyze adequately or remedy, has hampered assessment of economic well-being. Such problems include rising nonresponse rates to questions about income, as well as errors in reporting types and amounts of income in household surveys. Also, there is a lack of adequate data on sources of income as diverse as nonwage benefits, which are estimated to account for more than one-quarter of employer labor costs, and receipts from illegal enterprises, which in one estimate account for one-quarter of the income of inner-city men (Levitan and Gallo, 198$1$2$34,25).

Congress has expressed concern over the deterioration of the nation's information base, and the administration has recently expressed support for budget increases and reallocations to make it possible to effect improvements in important statistical concepts and data series (see Boskin, 1990; Council of Economic Advisers, 1990, 1991; Darby, 1990; Office of Technology Assessment, 1989a). Some of the proposals that are relevant to social welfare policy analysis data needs include conducting research on measurement of poverty and

Copyright National Academy of Sciences. All rights reserved.

income; restoring the SIPP sample size; exploring ways to link SIPP data with information from administrative records while safeguarding confidentiality; and modernizing the labor force component of the CPS. Although we have not examined the merits of specific elements in the administration's package and do not presume to judge among data needs across areas, we want to record our support for increased investment in the federal statistical system.

***Recommendation 3-1*. We recommend that the federal government increase its investment in the production of relevant, high-quality, statistical data for social welfare policy analysis and other purposes.**

Coordination of Data Production

In addition to budget and staffing constraints, the federal statistical system over the past decade has suffered a deterioration in mechanisms for interagency coordination and the ability to draw on and integrate information from a range of databases, particularly administrative records. The consequences have been reduced timeliness, quantity, and quality of policy-relevant data.

With its traditionally decentralized statistical system, whereby one agency collects data on health conditions, another collects data on health care financing, another collects income data, and so on, the United States depends heavily on effective coordinating mechanisms to achieve optimal allocation of data production resources. Yet the principal coordinating mechanism—a statistical policy group (variously named over the years) in the Office of Management and Budget (OMB)—with no more than half a dozen staff members and limited resources is today a shadow of its former self. Indeed, the demise of the statistical coordinating function has occurred over a longer period than the past decade: resources for this office, established in 1933 and located throughout most of its existence in OMB, peaked just after World War II (Wallman, 1988; see also Policy Users Panel, 1988; Sy and Robbin, 1990). The deleterious consequences of this situation include long lags between revisions of important government-wide coding schemes, such as the Standard Occupational and Industrial Classifications, and a reduced ability to evaluate interrelationships among data collection activities, given that the statistical policy group can provide limited or no oversight to the OMB desk officers who are assigned to clear survey questionnaires of specific agencies.

Various interagency and intraagency coordination efforts, of greater or lesser formality, continued or started up during the 1980s. They were typically organized around specific surveys—for example, the federal interagency committee on SIPP, which is chaired by OMB—or around specific topics—for example, the interagency forum on aging-related statistics, which is cochaired by the Census Bureau, the National Center for Health Statistics, and the National

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Institute on Aging. These efforts, although useful and important, do not wholly fill the need for a carefully designed structure of coordinating mechanisms devoted to optimizing the cost-effectiveness of federal data production.

The Paperwork Reduction Act, which governs the OMB role in setting statistical policy, is due for reauthorization by the Congress. The legislative discussion, although focused on OMB's role in reviewing federal agency regulations and reducing the burden of administrative paperwork on the private sector, has also considered ways to strengthen OMB's statistical coordination function. We have not studied the merits of alternative legislative proposals nor have we considered the issue of whether OMB remains the appropriate place to lodge the responsibility for statistical coordination, but we want to record our support for efforts to improve coordination among federal statistical agencies.

Recommendation 3-2. **We recommend that the federal government strengthen and increase its investment in the coordination of federal statistical activities, with the goal of improving the quality and relevance of data for policy analysis and other purposes.**

Broadened Data Collection

The decentralized character of the federal statistical system, coupled with the near certainty of changing data requirements over time, argues, in our view, not only for improved coordination of data production generally, but specifically for adoption of more far-seeing strategies of government data collection. It is important for reasons of both total cost and analytical usefulness that the major surveys on particular topics—health care, income support, retirement income, tax policy—be designed with a broad focus and in ways that facilitate relating the survey data to other survey and administrative data.

Cost concerns are frequently used to argue that surveys be focused in terms of subject matter and that they not duplicate topics covered in other surveys.[6] While not denying that costs are important, we believe that false economy is frequently introduced in large-scale data collection efforts by not also considering the benefits gained from more inclusive survey strategies. First, as alluded to previously, there is considerable uncertainty about what the important policy issues of tomorrow will be. Data collection efforts that are finely tuned to today's issues are likely to be unresponsive to future concerns. Second, investments in analytical tools for policy analysis, such as microsimulation models, are generally tied to specific surveys. The investment payoffs will be limited to the extent that the database constricts the model's ability to respond to new policy questions. Third, the most analytically useful

[6] Under the general heading of "costs" we include such diverse things as the time required by respondents to provide information, the costs of encoding and processing survey responses, and the effects on the complexity of the analysis tasks required to use the data.

Copyright National Academy of Sciences. All rights reserved.

surveys frequently include panel observations stretching over a number of years.[7] In these cases, it is especially important to anticipate the range of potential applications. Although no one survey can or should strive to serve all purposes, these factors suggest that proper decision making should look beyond just the cost of mounting the specific survey or data collection effort. Breadth of subject matter—for example, health care surveys including some socioeconomic variables and income surveys including some health status variables—has clear analytical value.

In addition, inclusion of overlapping variables is very useful for evaluation of data quality (e.g., comparing income reports in the March CPS and SIPP) and is essential when it becomes necessary for analysis purposes to add variables to a primary data set from other sources through some type of imputation or matching technique. Hence, duplication across surveys should not be resisted solely on cost grounds—considerations of data quality and utility must enter the calculation as well.[8] We repeat that not all future needs can be anticipated. It is therefore desirable as well to build mechanisms into surveys, such as supplemental modules with changing content, to enhance timely response to emerging data needs. It is also important that such supplemental modules be designed and incorporated in a timely manner.[9]

Recommendation 3-3. **We recommend that federal data collection strategies emphasize breadth of use and ability to respond to changing policy needs. In this regard, duplication of selected questions across surveys should be encouraged to the extent that such duplication enhances utility and facilitates evaluation of data quality.**

Another way in which federal data collection efforts should be broadened concerns the need for data that relate characteristics of individuals and institutions. There is increasing recognition that the success of social welfare policies depends importantly on complex delivery systems. The precise manner in which services are provided frequently influences—sometimes decisively—the extent to which services are truly available to individuals and the extent to

[7] Panel surveys collect repeated data on the same units (individuals, families, firms, etc.) over time, allowing analysis of how changed circumstances affect changed behavior. Panel data differ dramatically in usefulness from purely cross-sectional data, which include only a single observation for each of the sampled units.

[8] For a similar argument, see the interim assessment of SIPP prepared by the Committee on National Statistics (1989). We note that care must be exercised in interpreting the results for overlapping items obtained from two or more surveys, because it is rarely the case that the items can be duplicated in all respects: wording, placement in the sequence of questions, etc.

[9] Several major surveys, including CPS, SIPP, and the Health Interview Survey, currently have a structure of core questions that are asked at every interview together with supplements that vary in content. For some of these surveys, however, including SIPP, long lead times are needed to include new supplements.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

which eligible individuals choose to avail themselves of the services. Indeed, a wide variety of social policies operate, not directly on recipients but through various institutions. For example, Medicaid reimbursements to hospitals influence directly the health services provided to the poor; nondiscrimination rules of the corporate tax code have an impact on the character of individual retirement benefits; the hours of operation, location of offices, and treatment of individuals by welfare agencies affect participation rates of eligible families and individuals. David (1990:4) contrasts the effort put into measuring participation in public welfare programs, which has produced evidence from the SIPP and other sources of high nonparticipation rates for some programs, with the lack of effort put into measuring the reasons for nonparticipation, which may include misunderstanding, fear of stigma in the community, the costs of applying, and other aspects of program administration.

The interaction of individuals and service suppliers generates a special informational demand: the need to link individuals with a variety of institutions. Good analysis of policy alternatives must in general take the underlying linkages into account. For example, understanding the effects of hospital regulations necessitates data that relate the characteristics of patients and hospitals of different kinds. Similarly, understanding the effects of various regulations with regard to retirement requires joint information about characteristics of both individuals and their employers. In this regard, the Social Science Research Council Advisory Group on a 1986 Quality of Employment Survey recommended that the U.S. Department of Labor sponsor surveys that would provide linked data about employers and employees. The group commented (Kalleberg, 1986:8):

> It is imperative to study the organizational contexts of human resource issues since organizations—small or large—are today the central structures in American society through which changes in the nature of work and industry occur and where policies are enacted. It is organizations that, among other things, sign union contracts, adopt automated equipment, relocate to communities or nations with lower wage rates, make capital investments, create supervisory structures, provide fringe benefits, set salary scales, and create or eliminate jobs.

Linkages of data about individuals and institutions must be incorporated into the design of data collection efforts. Most data collection efforts, however, are focused on a single entity—the family, the firm, or the like—and only tangentially collect information on related institutional data.

Recommendation 3-4. **We recommend that federal statistical agencies give more attention to data collection strategies that recognize key interactions among individuals and institutions—employers, hospitals, government agencies, and others.**

Copyright National Academy of Sciences. All rights reserved.

Linking Survey and Administrative Data

We believe it is imperative to address a serious problem that stems from the decentralized structures of the statistical and administrative data functions in the United States and heightened concern for preserving confidentiality of individual records: namely, barriers to relating survey and administrative data. The United States, although not having developed as fully as some European countries the use of administrative records for statistical purposes,[10] has traditionally relied on administrative data to supplement information obtained from censuses and surveys. Records such as social security earnings histories, case files from public assistance programs, health care claims, and tax returns have provided additional data at reduced costs and also served to evaluate the quality of census and survey data. A principal mechanism for relating survey and administrative records has been to perform an exact match based on common identifiers, such as social security number.[11] Yet developments in the past two decades have greatly undercut the contributions to the nation's information base from administrative sources. In the world of economic statistics, deregulation of key industries has eliminated entire databases because there are no longer regulatory agencies in place to generate records. In the world of social statistics, major restrictions on the availability of administrative data for research and analysis have come from legal and administrative actions to guard against possible breaches of confidentiality.

Federal statistical agencies historically have an exemplary record in protecting the confidentiality of individual data records, whether from surveys or administrative systems. But in the past decade, many agencies, particularly the Census Bureau, tightened their policies and procedures to protect confidentiality and restrict access to both administrative records and survey data. In so doing, the agencies responded directly to legislation, such as the 1974 Privacy Act and the 1976 Tax Reform Act,[12] that reflected heightened public unease about potential abuses of government data and indirectly to a broader set of concerns about growing disinclination to cooperate with government surveys.[13]

[10] For example, Denmark conducted its most recent census by extracting information from administrative data registers rather than by canvassing the population, and Sweden is moving in this direction (Redfern, 1987).

[11] Exact match" is something of a misnomer in that data problems and inconsistencies in one or more of the input files inevitably make it difficult to be 100 percent certain about each match. However, exact matches are far preferable to other kinds of matching and imputation in which it is not possible to link records for the same people; see discussion of exact matching, which has generated useful files for microsimulation modeling, in Chapters 5 and 8.

[12] The 1976 Tax Reform Act extended the limitations on access to tax return data to the social security earnings data that employers report to the IRS and the IRS subsequently passes on to the Social Security Administration for administering benefit programs.

[13] During this period, response rates declined in some government (as well as many private) surveys, which could be attributed, in part, to growing distrust of measures to protect confidentiality. Census

Copyright National Academy of Sciences. All rights reserved.

Such restrictions have limited access not only for policy analysis agencies and the research community, but also for other statistical agencies. The effects have been most pronounced on the quality and breadth of publicly available microdata files containing individual responses, which are a necessary input to both microsimulation modeling and microeconometric research. The consequences are also evident for the quality and breadth of aggregate data series developed by statistical agencies from internal microdata sources. To cite one example, the Census Bureau no longer prepares for public release exact-match files (with all identifiers removed) from household surveys such as the March CPS matched with social security earnings histories. [14] Because the available CPS-SSA exact-match files date back to 1978 and 1973, models and analyses of future retirement income program benefits are in the anomalous position of having to generate data for 10 or more past years before they can begin their projections, not only adding costs but inevitably impairing the quality of the estimates.[15]

To cite another example, access to IRS tax return data for statistical purposes can only be described as byzantine in nature. The IRS Statistics of Income (SOI) Division makes available samples of income tax returns for research use; however, only the Office of Tax Analysis and the Joint Committee on Taxation have access to the full data records for these samples. Other agencies involved in tax policy analysis, such as CBO and ASPE, have access only to the much more limited public-use version of the SOI files. The information from tax returns does not provide all of the needed data for tax modeling, which typically requires linked information from household surveys as well. Thus, analysis of new tax proposals frequently requires information on people who do not file tax returns or who do not itemize deductions. Moreover, meaningful distributional analyses require data on the family composition and **other** characteristics of people who do file returns. Files containing exactly matched CPS (or SIPP) and **SOI** data would be very useful; however, they are not available.[16] The IRS permits Census Bureau employees to have access to a limited set of tax return variables that have been used for evaluating censuses

Bureau officials were particularly concerned about public outcries over privacy issues in some European countries, which led to the cancellation of the Netherlands census planned for 1981 and the West German census planned for 1983 (Butz, 1984).

[14] The Census Bureau recently prepared an exact-match file of SIPP and Social Security Administration (SSA) data, which are available only for 2 years to SSA researchers who are sworn in as special census employees.

[15] The 1973 exact-match file is publicly available. The 1978 exact-match file was never widely distributed, but it was made available to President Reagan's Commission on Pension Policy for use in modeling alternative retirement income policies. See Chapter 8 for a fuller discussion of information needs for retirement income modeling.

[16] Tax policy analysis agencies use statistical matching, which is a much less satisfactory technique than exact matching, to relate CPS and SOI data (see Chapters 5 and 8). The Census Bureau itself must develop its estimates of after-tax income from the CPS by imputing tax return information based on public-use SOI data because, as noted in the text, it has access to only a subset of items for all tax returns.

Copyright National Academy of Sciences. All rights reserved.

and surveys via exact-match techniques, but the Census Bureau does not, in turn, release matched files.[17]

We strongly support appropriate measures to protect the confidentiality of individual data records, taking all due precautions against either deliberate or inadvertent disclosure. However, we believe that mechanisms can and must be found to make it possible for the rich sets of data that are generated for federal administrative purposes to be used more fully for statistical purposes. The current situation of very limited availability, not only for policy analysis agencies and researchers but also for statistical agencies other than the originating agency, adds to the cost and reduces the quality of vitally important databases. We note that the administration recently took a step toward addressing this problem, announcing (Council of Economic Advisers, 1991:6) its intention to seek legislation "to provide a standardized mechanism for limited sharing of confidential statistical information solely for statistical purposes between statistical agencies under stringent safeguards. This will improve the accuracy, consistency, and timeliness of data throughout the Federal statistical system."

A panel of the Committee on National Statistics is currently investigating issues of confidentiality and access to federal data from surveys and administrative sources and exploring mechanisms for improved access by outside users as well as other statistical agencies. Such mechanisms include:

- setting up "enclaves" whereby statistical agencies could share survey and administrative records; such data sharing would reduce costs and enhance quality even if public access remained limited (e.g., the Census Bureau could use IRS data to improve imputations for missing income data in the CPS and SIPP and to develop improved estimates of the income distribution for publication);
- using sophisticated techniques to mask or blur data values so that microdata files containing survey and administrative data could be made publicly available;
- swearing-in analysts as special employees to use confidential data onsite at a statistical agency or in a secured facility; and
- requiring researchers to sign agreements that provide them with access to more complete data sets but also subject them to stiff penalties for any data disclosure that breaches confidentiality.

Other mechanisms are also possible. Each has advantages and disadvantages; some may require legislation as well as changes in policies and procedures. But we believe it is imperative for the federal statistical system to find ways out

[17] The Census Bureau will also not release the bulk of the tax return data collected in the SIPP, although these and other confidential data can be used by outside researchers who come to work at the Census Bureau under its fellowship program and are sworn in as special census employees. See Chapter 8 for a fuller discussion of information needs for tax modeling.

Copyright National Academy of Sciences. All rights reserved.

of the bind that currently puts too many data sources, collected at considerable cost, off limits for research and analysis purposes.

***Recommendation 3-5.* We recommend development and implementation of mechanisms to improve access, under appropriate circumstances, to administrative and survey microdata for statistical research and analysis purposes.**

Adding Value to Existing Data

Increased investment, improved coordination, and broader data collection strategies are all important for improved quality and utility of federal data for policy analysis and other purposes. In addition, we believe there is a need for reallocation of resources within statistical agencies to emphasize analysis and amelioration of data quality problems, together with a realignment of the data production functions of statistical agencies vis-á-vis those of users in policy analysis agencies (and elsewhere). Although important data gaps need to be filled in many areas, simply collecting more data more frequently is not the answer to the problems confronting the nation's information base. As we see in the case of flawed economic indicators that have had serious policy consequences, the available information must reflect appropriate and accurate measurement.

Statistical agencies, of course, are cognizant of the need to devote resources to evaluation of the quality of their data. However, budget and staff constraints, coupled with the difficulty of convincing decision makers of the value of methodological work, have often forced agencies to emphasize the operational activities necessary for timely data release at the expense of research on measurement issues and assessment of data quality. Moreover, evaluation research and the information on data quality that is provided to users have tended to emphasize the easily measured sources of errors, such as sampling variability, and to give less attention to other kinds of errors, such as nonresponse biases and content reporting errors, that may be equally—if not more—consequential in their impact. For example, Census Bureau publications from surveys such as the CPS and SIPP typically devote an entire appendix to sampling errors, but give very little space to nonsampling errors. Although there are encouraging examples of focused attention to data quality issues—such as the extensive research and evaluation program of the Census Bureau for SIPP (see Jabine, King, and Petroni, 1990) and the cognitive research laboratories set up by the National Center for Health Statistics and the Bureau of Labor Statistics to conduct pioneering studies of how respondents perceive survey questions—it is our view that much more attention needs to be devoted to data quality. And the need for evaluation of quality extends to data from administrative records as well as those from surveys. All too often, administrative data are accepted

Copyright National Academy of Sciences. All rights reserved.

uncritically as providing more accurate measures than surveys, when, in fact, they also suffer from a variety of reporting errors.

An important consideration for evaluation and measurement research conducted by statistical agencies is that the analysis be well informed by the policy and research needs for the data. Triplett (1990) argues that agencies have generally been responsive to the policy issues of the moment in making decisions about improved data collection but that they have failed to heed the data needs of policy research and more basic socioeconomic research, which are critical for effecting long-run improvements in the quality of policy analysis and decision making. This inattention goes a long way, in his view, to accounting for the relative slowness with which measurement improvements, such as better data on service industries, family structures, or economic well-being, are identified and implemented in the statistical system.

In addition to assessing data quality, statistical agencies need to add more value to the data series they release than is currently the practice. The statistical agencies have traditionally seen their role as preparing survey-specific data files and published tabulations from those files. They have not seen their role as producing analytical databases or publishing the best estimates—for such statistics as household income or poverty—that could be developed from multiple data sources. For example, in processing surveys like the CPS and SIPP, the Census Bureau has concentrated on such tasks as adjusting the data records for nonresponse by households and individuals and editing item responses for consistency. It has not performed additional data adjustments—such as correcting income amounts for reporting errors or modifying family structure to reflect population coverage errors—that would involve use of administrative records and other external data sources. Yet these and other adjustments are often critical for policy analysis and research uses of the data.

Currently, policy analysis agencies and other users carry out these kinds of data adjustments if they are performed at all. But these users often lack the information as well as the resources to perform an adequate job, and users at one agency frequently duplicate the efforts of users at another agency. Having statistical agencies add value to data would, of course, increase the costs to those agencies, but it could result in overall savings for policy analysis and other important applications because of reduced data processing expenditures by user agencies. Moreover, substantial gains in data quality could be made by having the originating agencies with better access to related data sources produce their best estimates of, for example, family composition, health care costs, or household income. These estimates in turn would provide a measure of consistency for all subsequent analyses.

We recognize that it will not be easy to implement a major change in the relationship of statistical agencies and such users as policy analysis agencies. The users will have concerns as to whether the statistical agencies can provide enhanced databases in a manner that is timely and addresses particular analysis

Copyright National Academy of Sciences. All rights reserved.

needs. The statistical agencies, on their part, will have concerns about whether meeting user requirements for enhanced databases will adversely affect their primary responsibilities for original data collection and processing. Nonetheless, we urge that a dialogue be started.[18]

Recommendation 3-6. **We recommend that federal statistical agencies increase their investment in evaluation of the quality of survey and administrative data. We further recommend that they use the results of evaluation studies to implement corrections, when feasible, to databases and published data series, with the objective of improving the quality and reducing the overall costs of providing analytically useful data for policy analysis and other important purposes.**

Finally, we mention one specific issue that falls under the rubric of evaluating and adding value to data and is of especially widespread importance, namely, the coverage of the population in household surveys and censuses. Decades of research have documented a persistent pattern of coverage errors in the decennial census, amounting to a small net undercount of the population (estimated at 1.4 percent in 1980) and large net undercounts of particular population subgroups (see Fay, Passel, and Robinson, 1988). Research has also documented additional large undercounts in household surveys such as the CPS and SIPP (see Chapter 5 and Citro, in Volume II). The mechanisms and correlates of coverage error are not definitively established, but variables such as age, race, sex, family relationship, and income appear to relate strongly to undercount: for example, net undercount rates in the census are high for black men, black children of both sexes, household members outside the nuclear family, and low-income people generally (see Citro and Cohen, 1985: Fein, 1989).

Coverage errors in the census may have important implications for many analyses. Census data are often used directly for policy analysis and research. In addition, they indirectly affect the quality of many other data sets because of their use in the design of survey samples, in adjusting survey data to match census-derived population controls, as denominators for vital rates and other socioeconomic indicators, and as the basis for postcensal population projections.[19]

[18] See our more detailed discussion in Chapter 5 of realigning data production responsibilities between statistical and policy analysis agencies with respect to the Census Bureau's recently announced intentions to develop a database from the March CPS, SIPP, and administrative records that will support an improved system of income statistics.

[19] In this discussion we consider only the importance of adjustment mechanisms for analytical uses of census data, not the political consideration of geographic adjustments and their potential impact on apportionment, funding formulas, and the like. Clogg, Massagli, and Eliason (1986) review the potential impact of census coverage errors on direct and indirect uses of the data, such as denominators for vital rates and weighting adjustments for sample surveys, and cite several examples of important effects.

Copyright National Academy of Sciences. All rights reserved.

The additional coverage errors in household surveys may also have important analytical implications—for example, for studies of low-income population groups. The use of census-derived population estimates to correct coverage errors in household surveys realigns the survey data only by the age, race, and sex distribution of the population as a whole, not by other characteristics, and does not adjust for coverage errors in the census itself.

Evaluation studies of the 1990 census will provide important information about coverage errors and their correlates. We believe it is critically important to use these and other studies of population coverage to assess the extent and implications of coverage errors in censuses and surveys for important analytical uses of the data and to determine whether and what kinds of adjustments the data may need.

Recommendation 3-7. **We recommend that the Census Bureau conduct a thorough evaluation of population coverage errors in the major household surveys and decennial census and their potential impacts on policy analysis and research uses of the data. Should important coverage errors be identified, we recommend that the Census Bureau develop ways to adjust census and survey data that have wide application for policy analysis and research.**

VALIDATION

We must help those who want all knowledge to be clear, definite, and sure to deal with uncertainty . . . and to judge accuracy.

Janet Norwood (1990:67)

The stock in trade of policy analysis is the production of answers to a series of "what if" questions. If the government wants to require enrollment in job training programs as a prerequisite to receipt of income support payments, what will be the effect on program enrollment and costs? If the Medicare system decides to pay a flat amount to hospitals for treatment of individuals with a given illness, what will happen to costs and quality of care? If food stamps cannot be used for certain types of purchases, what will be the impact on nutrition and on program expenditures?

Answering such questions requires estimates that amount to conditional forecasts, that is, projections of hypothetical future events. If a given policy is put into place, what will be its effects? The users of these kinds of estimates—decision makers in the executive branch or Congress and their staffs—must decide how to evaluate the information that is provided to them. They, of course, would like accurate data with no uncertainty, a standard that cannot

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

be met in the real world. Knowing that the standard of certainty cannot be met, users must consider how to incorporate the estimates into policy decisions. Moreover, because there may be competing estimates, which give different answers, users need to have information for assessing the relative quality, or certainty, of the estimates.

We begin our discussion by posing a simple question: Can users make reasonable judgments about the quality of estimates of the likely effects of a proposed policy change? That is, do they have information about the degree of uncertainty or variability in the estimates? And can they judge the track record of a policy modeling tool in terms of how well its previous estimates have corresponded with what actually occurred? We conclude that the answer to these questions is generally no. Indeed, this sweeping statement is not restricted to the work of specific agencies or government contractors. Nor is it restricted to social welfare policy issues or to the use of particular analytical tools. It is the exceptional analysis that can be assessed for validity, not the typical analysis.

In this section we focus primarily on the estimates produced by policy analysis techniques and models that are developed for more than one-time use. We first describe the magnitude of the validation problem, then provide more explicit definitions of what we mean by "model validation" and the kinds of techniques that are involved, and then discuss ways to facilitate the use of model validation techniques. In the next section we argue for the need for good documentation and archiving systems to support model validation and discuss ways to communicate the results of model validation exercises to decision makers.

The Difficulty of Validation

The problem of assessing the quality of estimates, forecasts, projections, or other analytical results is not restricted to the policy process and to policy analysis. It is faced in most scientific and academic inquiries and in business decision making, and a variety of approaches are available and used for validation. These approaches generally involve isolating the various sources of uncertainty, using observed data when possible to estimate the magnitude of such uncertainty, and generally undertaking a series of sensitivity analyses to assess the effect on the results of alternative specifications and assumptions.

The application of these approaches to policy analysis is particularly difficult. The accuracy of policy estimates depends on many different factors, including the precision of specification of the policy to be considered; the validity of the assumed relationship between program characteristics and outcomes of interest; the accuracy of forecasts of other relevant factors, such as the state of the economy; and the completeness of knowledge of how other factors affect outcomes. Each of these factors—as well as the quality of the database used

Copyright National Academy of Sciences. All rights reserved.

for the estimation task—contributes to the inaccuracy of specific estimates and, as a result, contributes to uncertainty in the estimates. Also contributing to uncertainty and complicating the task of validation is the fact that the development of policy estimates often necessitates the use of more than one model or modeling approach.

Evaluating the results of policy analysis is not more difficult simply because of the number of different sources of uncertainty, however; other areas of inquiry face a similar array of sources of potential error.[20] The difficulty in the policy process is that almost all analyses involve conditional rather than unconditional forecasts: that is, almost all analyses apply to a hypothetical set of policy alternatives—for example, a set of proposals to mandate minimum AFDC benefits so as to bring recipients' income up to specified levels of the poverty line. Because none of the alternatives may ultimately be adopted, reality checks on the quality of policy estimates are necessarily limited. This basic source of uncertainty is then compounded by the uncertainties from other, nonpolicy sources.

It is useful to compare this situation with other cases in which analysts make unconditional forecasts. For example, when economists forecast macroeconomic aggregates, such as the rate of growth of GNP, the inflation rate, or the unemployment rate, it is possible to compare what happens with the forecasts. In these cases, even though the forecasts may have been predicated on certain assumptions about exogenous conditions or about policies, the objective is a forecast of what will in fact occur at a specific time in the future.

Although validating unconditional forecasts requires time—it is necessary to wait until the actual data are available—pursuing such a process is straightforward. Indeed, forecasts of leading macroeconomic models are regularly evaluated against actual events. If the model or analytical method producing the estimates remains stable over time, analysts can compile data on the forecast accuracy of the approach and can use the model's track record as part of developing an estimate of uncertainty for future forecasts. Analyzing the precise sources of uncertainty, which is important for understanding the quality of a model's estimates and planning improvements to the model, is also possible, although the task may be quite complex. For example, in evaluating the forecasts produced by macroeconomic models, it is necessary to determine the contribution to total error not only from badly specified relationships inside the model and erroneous assumptions about external factors such as fiscal and monetary policy, but also from the ad hoc adjustments that forecasters typically make to their models' outputs. Nonetheless, a growing body of literature is attempting,

---

[20] It is generally true, however, that the more complex the model, the more difficult the validation task, and the less likely it is that the task will be carried out. For example, there are only a handful of validation studies of microsimulation models (see Cohen, Chapter 7 in Volume II), but considerable literature evaluating cell-based population projection models (see Grummer-Strawn and Espenshade, in Volume II).

Copyright National Academy of Sciences. All rights reserved.

with some success, to do just that—namely, disaggregate the observed errors from macroeconomic model forecasts by source (e.g., see McNees, 1989, 1990).

Most policy analysis efforts, which involve conditional estimates of the outcomes of specific policy alternatives, are much more difficult to validate and analyze than are unconditional forecasts.[21] First, data may *never* become available on the actual outcome of an analyzed policy. Although any given analysis may consider a range of policy options, the specific policy ultimately enacted is frequently not included among those analyzed: the policy process often begins with a specification of the potential range of policies but ends with a compromise outcome of the legislative debate. Second, even if an analyzed policy is enacted, it is difficult to disentangle the specific sources of errors. For example, in projected participation rates for a program, it is hard to distinguish between an error that occurs because of poor understanding of the behavior of program participants and one that occurs because of bad forecasts of the economic environment. Yet it is often vital to know the source of error in order to evaluate the utility of different policy modeling approaches. Third, the policy process does not generate data on a regular basis. Most programs go through only occasional significant changes, making it hard to analyze how specific programmatic elements affect the observed policy outcomes.

These factors suggest that developing information on the quality of policy analysis efforts is inherently difficult. The policy process, however, makes the task even more complex. Typically, programmatic considerations and deliberations are conducted in an environment of time constraints and distractions from other policy deliberations. The pressures to "do something" generally move any discussion away from consideration of the underlying analysis. These same pressures often leave analysts with little or no time to spend on such tasks as documenting the analyses and developing evidence on the quality of estimates for future use.

Although the validation of policy analysis is inherently difficult, we are still troubled by the few attempts that have been made at it. We conclude that the users of policy analysis have rarely asked for data on the uncertainty of model estimates. This lack of interest has serious ramifications for the long-term quality of the analysis that is used in the decision-making process and, consequently, for the quality of the outcomes of that process. Without information to assess uncertainty, users may make decisions on the assumption that estimates are reasonably accurate when, in fact, they are full of errors.

One egregious example occurred in the decision to amend the tax code in 1981 to offer tax-deferred Individual Retirement Accounts (IRAs). Tax analysts substantially underestimated the revenue losses of this provision: many

[21] In this regard, the use of macroeconomic models for simulation rather than forecasting purposes—that is, to estimate the effects of alternative government fiscal policies as distinct from forecasting the state of the economy—presents the same kinds of validation problems as do other kinds of conditional modeling exercises.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

more taxpayers took advantage of the provision than projected. This is not necessarily a criticism of the analysts, because the data available for estimating IRA participation were very sparse; the point is that the decision process made use of the estimates as though they were certain.

Moreover, without information to assess uncertainty, policy analysis agencies cannot determine the greatest need for their investment dollars in order to improve the quality of future estimates. They must make resource allocation decisions based largely on instinct rather than on a cumulative body of evidence. Inevitably, investment dollars will be misdirected, and progress in improving the quality of policy analysis results will be erratic at best.

Before recommendations are considered, it is important to point out one way in which policy analysis estimates may be *easier* to produce reliably than the unconditional forecasts first discussed. The typical policy estimate, done early in the process, considers a variety of alternative specifications of policies designed to achieve a given goal. The policy deliberations themselves tend to concentrate on differences among the alternative approaches. Therefore, the key information from the policy analysis is the *differential* effects or costs of the alternative approaches. This fact frequently makes the task of validation easier because errors or uncertainties that affect the alternatives in a similar fashion have little importance for the consideration of which alternative is preferable. For example, erroneous forecasts of macroeconomic aggregates, such as the unemployment rate or inflation rate, may have little effect on the projected difference in costs of alternative welfare reform proposals. Even though the level of projected costs may be wrong, it is quite possible that the projected levels for each alternative are "equally wrong," making the analysis somewhat immune to some of the major sources of potential errors.

However, this argument is less likely to apply for relatively long projection periods, and it does not apply at all when assessing the impact of completely new programs, for which the critical estimate is the level of expected costs and caseloads. Moreover, in order to assess errors that do affect estimates of differences among program alternatives, as well as errors that affect estimates of levels for new programs, one must be able to decompose the sources of errors —which, as noted above, is a difficult task.

Kinds of Model Validation

The technical nature of the validation task warrants a brief definition of key terms and concepts (see also the Appendix to Part I).[22] The first concept we need to define is "model." A model is a replicable, objective sequence of computations used for generating estimates of quantities of interest. By

[22] The discussion in this section benefited greatly from a set of notes prepared by Michael L. Cohen, consultant to the panel.

Copyright National Academy of Sciences. All rights reserved.

replicable, we mean that the sequence of computations generating an estimate can be reproduced by anyone who runs the model. By objective, we mean that, given the input data for the model, the output estimates are not a function of any analyst's opinions or assessments.

"Model validation," or, sometimes, "model evaluation," has been used in a variety of ways to describe techniques to assess the strengths and weaknesses of a model and the quality of its estimates. We acknowledge the range of topics that at times have been included under the validation rubric, including assessment of how easy it is to use the model and the completeness of the documentation.[23] However, we use the term model validation in a somewhat narrower sense: the processes for measuring the uncertainty or variability in a model's estimates and identifying the sources of that uncertainty.

External Validation

One principal technique for model validation, as we have defined it, is "external validation," or assessment of the validity of a model's estimates compared with measures of reality. For example, externally validating the cost and caseload estimates produced during the legislative debate on the Family Support Act would involve comparing them with the corresponding costs and caseloads obtained from AFDC program administrative records after the act took effect. By definition, one cannot carry out an external validation of a model's estimates when the change is under consideration. However, after-the-fact assessments of external validity can help identify model weaknesses and contribute to measuring the likely uncertainty in successive estimates produced by the same model.

Sometimes it is possible to carry out an external validation simply by comparing a model's estimates with what actually happened. More often, the policy change will not correspond with one particular estimate, or the estimates, as in the case of the Family Support Act, will represent a combination of outputs from several models. Hence, other approaches will be needed.

A common external validation technique is "ex post forecasting." Using this approach (e.g., see Kormendi and Meguire, 1988), one puts oneself in the place of the analyst who, say, 5 years ago was asked to simulate a program alternative to take effect in some future year. One chooses that "future" year to be in the recent past so that measures of what happened, from

[23] The General Accounting Office (1979) lists features of models to consider in a global assessment, including but not limited to model validation as we have defined it. GAO suggests looking at the completeness and adequacy of the documentation; the validity of the model's theoretical concepts, data, and computer code; the ease of maintenance and updating; the understandability of the model's outputs; public accessibility to the model and data; portability among computer environments; and cost and time to run the model. We consider these features and others in our review of microsimulation models in Part II.

Copyright National Academy of Sciences. All rights reserved.

administrative records or other sources, are available. Correspondingly, one chooses the program alternative to be the actual program rules in effect during the comparison year. The panel, as part of its work, carried out an ex post forecasting evaluation of the TRIM2 microsimulation model, which proved quite informative. (The study, which also involved a sensitivity analysis, compared the TRIM2 estimates developed from a 1983 database of AFDC costs and caseloads under 1987 AFDC law to administrative data for AFDC in 1987; see Chapter 9 and Cohen et al., in Volume II.) Alternatively, in a method called "backcasting," one uses the current model and database to simulate program provisions that were operative at some period in the past and compares the model estimates with administrative data or other measures for that period (e.g., see Hayes, 1982).

In either backcasting or ex post forecasting, differences between the model results and the measures of what occurred may involve economic or social changes that the model could not have been expected to capture, such as an unanticipated recession. Differences may also be due to chance variation. Hence, it is important to conduct external validation studies for a number of time periods, including those that were relatively stable on key social and economic indicators. It is also important, to the extent feasible, to construct measures of variability both for the model output (see discussion below) and for the measures of what actually occurred, which may themselves contain errors (see Andrews et al., 1987, for exploratory work on this topic).

Internal Validation

To understand the extent and sources of uncertainty in a model's estimates of the effects of a proposed policy change, one needs to conduct not only external validation studies but also direct investigations, or internal validation studies, of the underlying model. Internal validation refers to all of the procedures that are part of conducting an intensive step-by-step analysis of how model components work, including the theory behind the various modules, the data used, the computer programming, and the decisions made by the analysts running the model. All aspects of internal validation are important; in the context of our discussion of the measurement of uncertainty in a model's estimates, however, we focus on internal validation techniques—namely, variance estimation and sensitivity analysis—that contribute to such measurement. Both estimation of the underlying variance of the estimates and analyses of the sensitivity of the results to alternative model specifications yield potentially important information that can become a standard part of the model improvement process.

It is useful to think of the uncertainty or variability—"errors"—in the outputs from a model as resulting from four sources: (1) sampling variability in the input database, which is only one of a family of possible data sets that could have been used; (2) sampling variability in other inputs such as imputations,

Copyright National Academy of Sciences. All rights reserved.

regression coefficients, and control totals; (3) errors in the database and other inputs; and (4) errors due to model misspecification. Even though conceptually clear, the complete partitioning and full estimation of a model's uncertainty are generally beyond current capabilities. Indeed, in many complicated analytical situations, even the most rudimentary estimates of uncertainty have been intractable until recently. Nonetheless, a combination of approaches can assess portions of the uncertainty and can pinpoint areas of concern—that is, aspects of a model for which uncertainty is likely to have particularly important effects on the results.

For estimates produced by simple models, standard variance estimation techniques are available to assess the variability due to the first two sources of error noted above. For complex models, these techniques cannot be readily applied, but it has recently become possible to use some relatively new variance estimation methods, called "sample reuse" techniques, that harness the power of modern computers.

Sensitivity analysis is a useful supplement to formal variance estimation techniques. Sensitivity analysis, carried out by developing and running one or more alternate versions of one or more model components, looks at the impact on the estimates of decisions about the structure or specification of a model. For example, the decision to use a particular equation to estimate program participation in a microsimulation model or to use a particular set of vital rates in a cell-based population projection model is properly investigated as part of a sensitivity analysis. A sensitivity analysis typically will not identify the optimal method for modeling a component, but it will provide a rough idea as to the components that matter for a specific result. Sensitivity analysis is, in simplest terms, a diagnostic tool for ascertaining which parts of an overall model could have the largest impact on results and therefore are the most important to scrutinize for potential errors that could be reduced or eliminated.[24]

In the current state of the art, sensitivity analysis is the only way to obtain rough estimates of the variability in model outputs due to misspecification (the fourth source of error noted above). Sensitivity analysis is also often the best approach to estimate the variability from errors in the input data sets (the third source). For complex models, sensitivity analysis may also represent the most feasible approach at the present time to assess the variability from the second source, that is, the sampling variability in data sources other than the primary database. What one gives up when going from a variance estimation methodology to a sensitivity analysis is that the probabilistic mechanism underlying a sensitivity analysis is not rigorously determined. Thus, construction of confidence intervals—a type of formal "error bound"—to express the uncertainty

[24] Another way of learning about deficiencies in a model is to make use of completely different modeling approaches to the entire problem, rather than experimenting with individual components. This form of "global" sensitivity analysis is not effective in operating a feedback loop, but it is effective in providing a rough indication of the level of error in estimates from several models.

Copyright National Academy of Sciences. All rights reserved.

in the estimates is immensely more difficult. Indeed, in complex models the variability in the estimates is often not understood well enough to construct any reliable confidence intervals.[25]

Whatever mix of sensitivity analysis and variance estimation is used, the critical point is the need to obtain measures of the uncertainty in the estimates for proposed program changes that are used in the policy debate. We next discuss ways and means of moving model validation from a rare to a regular part of the policy analysis function.

Investment in Model Validation

We believe that the policy process must include consideration of the quality of the information used in reaching decisions, that is, on the level and sources of uncertainty in estimates about the effects of proposed policy alternatives. In order to achieve such consideration, a series of fundamental changes must be made to the routine production of policy analyses.

First of all, it is clear that information on uncertainty in policy analyses will be produced only if the users insist on receiving it. We have noted that decision makers shy away from estimates of uncertainty, particularly of cost and revenue projections, because it is hard to integrate such information into a decision process in which the numbers must add up. Today's severe budget constraints and the perceived necessity by members of Congress to balance changes in expenditures against changes in revenue down to the last dollar reinforce the predilections of policy makers for certainty—or what appears to be certainty—in the numbers.

Our message to decision makers is that they must demand, as a matter of regular practice, information about the level and sources of uncertainty in policy analysis work. It is in both their short-run and their long-run best interests to do so.

Estimates of uncertainty can be very helpful for legislators who are facing immediate decisions on policy issues. First, if there are competing estimates from two different agencies,[26] information about the uncertainty in each estimate

[25] Confidence intervals are ranges about an estimate, constructed so that one can say with a specified "coverage" probability, such as 95 or 90 percent, that the confidence interval includes the actual value in the population (or, more precisely, the average value that would be obtained from all possible samples of the population). For example, according to the Census Bureau, the estimate of the number of people below the poverty level in 1988, from the March 1989 Current Population Survey, is 31.9 million, with a 90 percent confidence interval of plus or minus 0.9 million. That is, one can be 90 percent confident that the range of 31.0-32.8 million people includes the true value (Bureau of the Census, 1989a:2; see also the Appendix to Part I).

[26] A recent example, with serious policy implications, involves competing estimates of the 1991 federal budget deficit from CBO and OMB (excluding projected costs for the savings and loan bailout). In January 1990, CBO projected the 1991 deficit at $138 billion; OMB projected $101 billion. In September 1990, CBO projected the 1991 deficit at $232 billion; OMB projected S149 billion (Magnuson, 1990).

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

can help determine which of the estimates is better. As an example, an estimate that the added costs of a proposed change are in the range of $10-11 billion is clearly more useful than an estimate for the same proposal that the costs are in the range of $5-20 billion. Second, if all of the available cost estimates have wide error bounds, such as $5-20 billion or, worse, minus $20 billion to plus $20 billion, decision makers would be well advised to give greater weight to criteria other than overall cost, such as distributional effects or agreement with important societal values, in reaching a conclusion about the merits of a particular proposal. Third, if the available estimates of costs and distributional effects for program alternatives are not reliable enough to distinguish among them (because, for example, the alternatives involve small changes or focus on small population groups), decision makers would be well advised to minimize the effort spent to fine-tune the policy proposal.

Perhaps more compelling, it is also in the long-run interests of decision makers to demand estimates of uncertainty. Decisions made on the basis of erroneous information can have large unintended social costs. The goal of completely certain information is illusory; however, with estimates of uncertainty for the information provided by currently available models and databases, decision makers can target funds for policy analysis agencies to develop better information on which to base their policy choices in the future. In other words, developing uncertainty information can serve an important feedback function that leads, over time, to the development of better models and better policy information.

Recommendation 3-8. **We recommend that users of policy projections systematically demand information on the level and sources of uncertainty in policy analysis work.**

We recognize the practical difficulties of changing the behavior of decision makers, who have avoided information about uncertainty in the past and who may, despite arguments about the short-term and long-term benefits, remain hesitant to seek such information in the future. Hence, we urge that the heads of policy analysis agencies assume the challenge of working toward the goal of having information on uncertainty available as a matter of course for the estimates their agencies produce. Agency heads can take several actions. They can set and enforce standards that validation be part of the policy analysis work of their staffs; they can allocate staff and budget resources to validation; they can support efforts by their staffs to educate the staffs of decision makers about the need for information on the quality of the estimates and how to interpret such information; and they can support their staffs when time constraints and demands for certainty threaten to short-circuit validation efforts.

Recommendation 3-9. **We recommend that heads of policy analysis agencies assume responsibility for ensuring, to the extent feasible,**

Copyright National Academy of Sciences. All rights reserved.

that their staffs regularly prepare information about the level and sources of uncertainty in their work. Agency heads should also support efforts of their staffs to accustom decision makers to request and use such information in the policy process.

In some instances, agency staffs perform policy analyses from start to finish; in many other instances, agencies contract out for analytic work. Under either approach, policy analysis agencies must explicitly consider how to develop relevant information about the uncertainty in the results. They must plan to obtain this information at the beginning, when they are under less time pressure.

Policy analysis work, whether conducted in-house or by contractors, should always include some type of validation effort that, at a minimum, develops approximate estimates of uncertainty in the results and the main sources of this uncertainty. In addition, for major analyses that are contracted out, we believe that the agencies should at the same time let separate contracts to independent agencies or firms to conduct thoroughgoing evaluations of the work, including external validation studies and sensitivity analyses. The reason for independent contracts is to ensure objectivity of the evaluation and to minimize the likelihood that the evaluation will be sacrificed to the need for immediate results to feed to the policy debate. We recommend independent in-depth evaluations of this type for major policy analysis work per se—that is, work using models to develop policy impact estimates—and for research and demonstration projects of which the results may have subsequent applicability for policy modeling.

Recommendation 3-10. **We recommend that policy analysis agencies earmark a portion of the funds for all major analytical efforts for evaluation of the quality of the results. For large-scale, ongoing research and modeling efforts, the agencies should let a separate contract for an independent evaluation.**

Information on sources of error obtained from sensitivity analysis, along with the results of external validation, is important for determining the priorities for resources for the improvement of policy analysis tools. The focus of the independent evaluation studies will necessarily be on the feedback process whereby evaluation results give rise to better analysis tools that, in turn, produce better numbers for future policy debates.

At the time of a debate, of course, no comparisons with reality are possible, and the time available for extensive investigation of sources of uncertainty is necessarily limited. Still, information about uncertainty and its sources can and should be provided. Some types of analysis are quite amenable to investigations about the magnitude and source of uncertainty. For example, a simple projection of numbers of program participants that comes from a single regression equation can include information on estimated error variances due to randomness in the estimated parameters of the underlying model. Providing this information is

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

a step in the right direction, although such randomness is only part of the error, and additional information would be necessary to completely describe the potential errors (such as those from misspecification of the underlying model, which may introduce more error than the simple sampling errors that occupy most attention).

Other analyses are less amenable to such quantitative estimates. At one extreme, rough back-of-the-envelope estimates almost defy formal error analysis because they rely so heavily on an analyst's judgment. At the other extreme, estimates from large, complex models are difficult to assess because of the sheer number of inputs. Nevertheless, in the former case it should be possible and routine practice for an analyst to identify major potential uncertainties in his or her estimates, even if they cannot be measured in quantitative form; in the latter case, recently developed computer-intensive techniques are available to develop error bounds for projections from complex models due to randomness in one or more of the inputs (see above; Chapter 9; and Cohen, Chapter 6 in Volume II). Sensitivity analysis techniques, in which data inputs and model components are systematically varied in a series of model runs, can also be used to assess the magnitude and major sources of variation.

We acknowledge the difficulties in developing error estimates, given the complexities of the real world and the policy alternatives that analysts are trying to model, but we believe it is possible to make significant progress with allocation of sufficient resources and a strong commitment to the task. We are also optimistic about the prospects that technological developments in computing will make it possible to conduct validation studies of even very complex models with relative ease.

We note that some policy deliberations occur at regular times and are supported by a consistent set of analyses, which would enable error studies to be carried out on a continuing basis. For example, at the beginning of each budget season, both the Office of Management and Budget and the Congressional Budget Office provide estimates of budgetary aggregates. Although they come in various forms, each contains budget (deficit) projections for the condition of no policy changes. It is then possible, as CBO does each year in its August update of the budget analysis, to consider how changes in economic conditions and changes in actual policies affect the budget projection (see, e.g., Congressional Budget Office, 1989a:38). This analysis provides a model for how to approach the task, but it does not go far enough because it is not revisited after the actual data become available.

Routine estimates are made in a wide variety of program areas. It is important, whenever possible, to match estimates with actual outcomes. Clearly, such an activity is most valuable when a reasonably stable projection method is used, because in such cases the time series of evidence can be used to estimate errors and decompose them into various sources. However, this technique is not restricted to time-series approaches. Some program estimates are made

Copyright National Academy of Sciences. All rights reserved.

on a state-by-state basis, and observed state variations provide another way of inferring the importance of uncertainty in policy information.

***Recommendation 3-11.* We recommend that policy analysis agencies routinely provide periodic error analyses of ongoing work.**

DOCUMENTATION AND COMMUNICATION OF THE RESULTS OF POLICY ANALYSIS

Documentation and Archiving as Aids to Validation

We turn next to the critical role of good documentation practices for the proper use of models that provide estimates to the policy debate and for evaluation of the quality of their outputs. From the perspective of applying policy analysis tools, complete and intelligible documentation is essential for their appropriate and efficient use, whether the model is based on microsimulation, macroeconomic modeling, multiple regression, or some other technique. The larger and more complex the model, the harder the task of preparing adequate documentation, but, at the same time, the more necessary the task becomes.[27] Such models can quickly take on the aspect of "black boxes," which can be fully understood only by a handful of experienced analysts who have invested the time to master the intricacies of their operation. Agencies can become too dependent on the availability of these experienced analysts in order to use complex models. Moreover, inadvertent errors due to misunderstanding the interactions of elements of complex models may occur even on the part of highly experienced users.

The quality of any validation effort is highly dependent on the quality of the documentation, that is, the documentation of the particular analysis that was performed, which is needed in addition to the documentation of the policy analysis tool itself. Documentation of policy analysis exercises (such as assessing the cost implications of a proposed program change) should include information about the specifications for the analysis (e.g., the particulars of the policy alternatives modeled), data inputs, key assumptions, changes that were made to the basic model, analyst inputs, and other information needed to understand what was done and to place the results in context.

The necessity for such documentation underscores the importance of making the evaluation process a regular and expected part of policy analysis work. The best time to document an analysis is during the process of performing the

[27] David and Robbin have written extensively on the necessity of providing "metadata" for complex databases and models, that is, information that helps users work with them appropriately. David and Robbin have outlined design concepts for information management systems to facilitate the production of complete documentation and the generation of audit trails that keep track of users' applications on an automated basis (see David, 1991; David and Robbin, 1989, 1990).

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

work; if the documentation effort is deferred for very long, the result is likely to be the loss of key information.

We have noted that it is often difficult to provide information with which to judge the usefulness of a given analysis. In particular, complete information on uncertainties in the estimates may be unavailable. However, when documentation of the methods used is available to other analysts, an internal validity check is possible. For instance, other analysts can ascertain whether reasonable scientific methods were followed or whether the best current information was used. Although the scope and scale of the analysis must be considered in deciding how much documentation is needed (it is not likely to be feasible or sensible to document each and every policy analysis result), the provision of adequate documentation should be a broad objective and requirement of policy analysis work.

***Recommendation 3-12*. We recommend that policy analysis agencies allocate sufficient resources for complete and understandable documentation of policy analysis tools. We also recommend that, as a matter of standard practice, they require complete documentation of the methodology and procedures used in major policy analyses.**

The task of validating policy analysis estimates and building a cumulative body of knowledge about the merits of particular analytic approaches and tools is also dependent on the continued availability of previous versions of models and databases that were used for analyses, as well as the results of those analyses. When a legislative change has enacted one of a set of proposed policies, validation studies that look at what actually happened need access to the estimates that were made at the time of the debate. Such studies also need access to the model and data that were used to determine the sources of errors in the estimates. In particular, such studies need to separate errors due to the model per se and errors due to conditional assumptions about exogenous factors, such as the state of the economy, that did not turn out as assumed. The availability of complete documentation will make it possible, at least occasionally, to rerun the model with the erroneous assumptions removed.

When none of a proposed set of policies has been implemented but some other legislative change has occurred, validation studies can simulate the estimates that would have been made if the analysts at the time had been asked to simulate what was enacted. Such studies can use a current model, but they need access to the original database. Hence, it is important that at least large-scale analytical efforts based on underlying quantitative models be archived in a form that allows them to be used in the future for validation purposes.

***Recommendation 3-13.* We recommend that policy analysis agencies require that major analytical efforts be subject to archiving**

Copyright National Academy of Sciences. All rights reserved.

so that the models, databases, and outputs are available for future analytical use.

Communicating Validation Results to Decision Makers

Developing information on sources of uncertainty represents a formidable undertaking and is only one part of the task of providing information on uncertainty to the policy debate. The other equally important part, with substantial problems of its own, is presenting the information to decision makers in a manner that they can understand and use. There is very little experience to build on in this area: to our knowledge, presentations of policy estimates are rarely if ever accompanied by formal error bounds; at most, there may be oral or written statements identifying those estimates that the analysts believe to be most problematic.[28]

A self-fulfilling prophecy may be seen at work here. Policy analysts are convinced that decision makers will not accept, let alone welcome, information about uncertainty. Hence, such information is not provided, and decision makers are not educated to the need for it.

The prospects for changing this situation are not entirely bleak. In the media, it is now standard practice to provide error bounds due to sampling error for estimates from public opinion polls. Articles based on government statistical reports sometimes cite error bounds as well. To make the job easier for the media and other users, the Census Bureau recently instituted a practice in its reports of including error bounds (90% confidence intervals) for each estimate mentioned in the text, in addition to appending technical information about errors and how to calculate errors at the end of the report (see, e.g., Bureau of the Census, 1989a). Such error reports generally pertain only to sampling error and not other, often more important, sources of uncertainty, but they represent a step forward.

The experience with tax reform in Wisconsin in the late 1970s also provides an encouraging precedent for accompanying estimates of the effects of alternative proposals with error bounds. The confidence intervals provided to the legislators (shown graphically, in most instances) were not resisted or disdained. Rather, they served the useful purpose of eliminating discussion of numbers that could not be made precise because they pertained to rare populations or events (Wisconsin Department of Revenue, 1979).

---

[28] This statement applies to estimates that are delivered during the course of the policy debate. Subsequently, agencies often prepare more detailed descriptions that include attempts to identify important assumptions and sources of error: see, for example, the write-up of the estimates for key provisions of the Family Support Act in a study released by CBO in January 1989 (Congressional Budget Office, 1989d). However, even these analyses rarely include estimates of error bounds or the results of formal sensitivity analyses.

Copyright National Academy of Sciences. All rights reserved.

Recommendation 3-14. **We recommend that policy analysis agencies include information about estimated uncertainty and the sources of this uncertainty as a matter of course in presentations of results to decision makers. The agencies should experiment with modes of presentation to facilitate understanding and acceptance of information about uncertainty on the part of decision makers.**

The question of precisely how to communicate uncertainty to users of various degrees of technical sophistication, particularly how best to express uncertainty in terms of a single measure such as a confidence interval, is a difficult one for which we offer no specific recommendations. Instead, we present below several approaches that might be adopted in different situations, along with their advantages and disadvantages (see also the Appendix to Part I). As measures of uncertainty—including confidence intervals—are provided to various audiences, we believe that the most effective methods will become apparent over time. Of course, whatever the measures of uncertainty that are used, they should always accompany—and not replace—the point estimates themselves.

- Policy analysts can provide formal error bounds (i.e., confidence intervals) for their estimates that represent the variability due to the sampling variance of the input data. In technical reports they can also include information about uncertainty due to model misspecification and other factors. This approach gives an overall impression of the uncertainty of estimates from a model. The major disadvantage of this approach is that the variability with the greatest visibility, namely, sampling variance, is likely to be the least important source of error.
- Policy analysts can provide error bounds for their estimates that represent *total* uncertainty by presenting the widest range obtained through the variety of techniques used in the evaluation, including sensitivity analysis and variance estimation. This approach strongly—perhaps too strongly—communicates the total uncertainty to model users. Its disadvantage is that constructing the broadest range is dependent on the ability of the analyst to perform sensitivity analyses of all important components of the model, which can be difficult to do. More important, it will generally not be possible to state the probability with which such a range includes the actual value in the population, in the way that one can do with a 95 or 90 percent confidence interval.
- Policy analysts can provide error bounds for their estimates based on combining the results of previous external validation studies of similar uses of the model. In a static modeling environment, this approach is a highly appropriate method for conveying the variability in a model's estimates. Its major disadvantage is that the modeling environment is likely to be dynamic in one or more respects so that the current application of the model may not resemble the applications included in the external validation studies. For

Copyright National Academy of Sciences. All rights reserved.

example, important elements of the database may have changed, important elements of the model itself may have been rewritten, and important aspects of the policy question may have altered.

There are other issues that need to be addressed in considering how best to express the uncertainty in policy analysis estimates. First, there are few incentives for analysts to pursue the difficult problems involved in developing appropriate measures of uncertainty for their estimates. Models that have accurately estimated confidence intervals are likely to suffer in comparison with models that have confidence intervals that are wrongly estimated to be too narrow. Second, it will be difficult for analysts to communicate to an unsophisticated audience the extent to which the commonly available confidence intervals are conditional—that is, predicated on the assumption that sampling variability is the only source of error in the estimates—and hence the implications for how the estimates should be interpreted in the policy debate.

Yet even with all these difficulties, we remain convinced that the objective of conveying information about the uncertainty in policy analysis results through some type of error bound is critically important. From the perspective of improving models through feedback, the availability of confidence intervals can be very helpful in distinguishing statistically significant differences that may have to be addressed from nonsignificant differences.

From the perspective of the policy process itself, there are several important uses of confidence intervals or, more generally, statements of uncertainty. First, as noted above, decision makers can readily use measures of uncertainty to assess the quality of two competing estimates. Also, information about uncertainty can help decision makers decide how much weight to give to the estimates of costs and of winners and losers that are produced by policy analysis tools vis-á-vis other important considerations for the policy debate.

Finally, decision makers can make use of measures of uncertainty to help judge the utility of allocating additional resources to the improvement of policy analysis models and databases. Today, many policy makers recognize the growing deficiencies in data series and modeling tools that support the policy analysis function, but they are not able to relate those problems to the quality of the resulting estimates of costs and distributional effects that are of concern in the policy debate. Having measures of uncertainty available for policy analysis estimates would enable decision makers to target issues with a high degree of policy importance and a high degree of uncertainty for concentrated investment of resources.

In summary, regular, systematic evaluation of the tools used for policy analysis is critical for improving the quality of their estimates. And decision makers need information about the quality of the estimates to be able to weigh them appropriately in making the critical choices that shape the nation's public policies.

Copyright National Academy of Sciences. All rights reserved.

Appendix Models, Uncertainty, and Confidence Intervals

This appendix examines technical issues pertaining to models, model validation or measuring the uncertainty of model projections, and representing the uncertainty in some sort of summary form.[1] As in the body of the report, we make use of definitions for terms such as model, model validation, and variance of a model's output that may differ somewhat from the definitions used elsewhere. This is unavoidable because there is currently no standard terminology for these concepts and because the situation in which we use them is somewhat special.

MODELS

Models often can be expressed as the sequential application of components or modules. The precise algorithm or data set used for certain components is often only one among several possibilities, the choice being arbitrary because of the lack of precise information about the effectiveness of the different choices. If a component is a regression model, the precise form of the regression model, the covariates to use, and the data set on which to estimate the regression coefficients may be arbitrary, at least to some extent. Examples of this appear in population projections and macroeconomic models, in which the selection of projected fertility and mortality rates in the former, or the selection of projected inflation rates and productivity indices in the latter, is somewhat arbitrary.

The correspondence of two models can range from being near replicates to essential independence. Two models can differ from one another due to use of

[1] This section was developed by Michael L. Cohen, consultant to the panel.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

different regression coefficients, due to use of different covariates, due to use of a different regression specification (e.g., log-linear or linear), through use of an approach other than regression for a particular component, or by having such different structures that identifying a parallel component is impossible. It is important for the discussion that follows to point out that some of the possible choices have classes of alternatives that can more easily be considered part of a sample space with subjectively assigned probabilities than others. For example, the usual theory underlying regression analysis provides the sample space of alternate vectors of regression coefficients, due to sampling variability, along with their associated densities. In that case, the sample space and the associated probabilities are easily supported. However, the class of approaches other than regression, including time-series analysis, models from linear programming, etc., is difficult to consider as part of a well-defined sample space and is also difficult to attribute probabilities of being "right." There is currently some disagreement about the extent to which subjective probabilities can be assigned to alternative model specifications. We point out that Rubin, in a series of papers on multiple imputation (see Little and Rubin, 1987), has provided a rigorous formulation for the variability due to alternative imputation models, which is closely related to the issue discussed here.

As we stress in Chapter 3, all sources of uncertainty in a model's output need to be communicated to policy analysts and the modeling community. The question is whether uncertainty arising from choices about model structure should be incorporated in a confidence interval, which specifies a precise coverage probability, or whether this uncertainty is more justifiably represented and communicated as part of a sensitivity analysis (in addition to the confidence interval representing the variation due only to sampling error). To many analysts, this is the same question as whether the difference between two models is too extreme to consider the two models to be versions of one another.

UNCERTAINTY OF AN ESTIMATE

The uncertainty of an estimate is an umbrella term for the quantification of the differences between a model's estimates and the truth. As noted in Chapter 3, estimates generated by models will differ from the truth for a variety of reasons, which can be summarized in four broad categories: (1) sampling variability in the input database, which is only one of a family of possible data sets that could have been obtained; (2) sampling variability in other inputs, such as imputations, regression coefficients, and control totals; (3) errors in the database and other inputs; and (4) errors due to model misspecification.

The first two categories of uncertainty listed above—sampling variability and errors from imprecise estimation of other model inputs—are most easily estimated and summarized. These estimates are frequently labeled mean square error. (As discussed further below, the concept of mean square error properly

Copyright National Academy of Sciences. All rights reserved.

includes the third and fourth sources of uncertainty—input data errors and model misspecification—as well. However, in practice, measures of mean square error almost always ignore the fourth source and often ignore the third source as well.)

For simple models, mean square error is measured with standard techniques for variance estimation.[2] For relatively complicated models, it has recently become possible to use nonparametric sample reuse techniques for this purpose. Available sample reuse techniques include the jackknife, bootstrap, balanced half-sample replication, and cross-validation; in particular, the bootstrap has shown good flexibility and utility (see Efron, 1979). Simply put, the bootstrap measures variability by using the observed sample distribution in place of the unobserved population distribution. The strength of this approach is that variance estimation becomes a function of computing power rather than an exercise in solving multidimensional integrations for complex estimators and distributions.

For complex models, sample reuse techniques can generally be used to measure the uncertainty in a model's estimates due to category 1. They can also be used to measure the uncertainty due to category 2, as can related parametric resampling techniques. The typical measurement of uncertainty due to categories 1 and 2 is the standard deviation, with the associated confidence interval for estimates that follow an approximately normal distribution.

Uncertainty due to category 3 is often measured by using a sensitivity analysis, because the error in the inputs may not be understood well enough to be approximated with a probabilistic model. This is accomplished by identifying a small number of methods for reweighting or "correcting" the data. One can then rerun the model with these alternate data sets and measure the impact on model projections. Sometimes, discrepancies due to category 3 can be given a probabilistic structure: in the case of forecasts of the inflation rate, for example, the sample space is well defined.

For discrepancies due to category 4, a well-defined sample space can usually not be defined, or, if defined, elements of that sample space cannot be given probabilities (even subjective) of being "correct," because it is difficult to collect all of the various methods that might be used to model a quantity of interest and to determine their associated uncertainties. Therefore, it is very difficult to impose a probabilistic structure on alternative model specifications. Certainly, sensitivity analyses can be performed for components of interest when alternative specifications for these components are suggested, and such analyses

[2] The simplest example of variance estimation comes from the standard output of regression analyses. When the modeling relies on a single equation estimated by standard techniques, direct estimates of the variance of individual parameters and of predicted outcomes are available. These rely on individual estimates of the error variance in question and assume that the models are correctly specified. In such a case, in which the models produce unbiased estimates, the estimated variances are equivalent to mean square error estimates.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

should be performed on a regular basis for components that are suspected to be suboptimal. However, many analysts believe that it is not possible to provide a confidence interval, with even approximate coverage probability, for uncertainty due to model misspecification.

Some members of the panel believe that the last statement can be relaxed in situations in which a number of versions of a model can be created by making use of equally likely a priori alternatives for a small number of modules. In this situation, the distinction between using alternative econometric forecasts and using alternative modules is not always clear, and it is possible that higher and lower limits defined by the values an estimate takes for these different model versions would provide a reasonable assessment of uncertainty due to model misspecification. Certainly, these ranges need to be communicated as part of an error profile of a model, but it is not completely clear how they could be incorporated into a confidence interval with known coverage probability, because the multivariate density corresponding to certain joint selections of alternatives of model components may not be known.

To frame this discussion in another way, the variability of an estimator about the "truth" can often be decomposed into two terms, variance and bias: variance is a measure of the variability of an estimator about its mean; bias is a measure of the difference between an estimator's mean and the truth. The variance plus the square of the bias is the formal definition of mean square error, which is also the average squared difference between an estimator and the truth. The square root of the mean square error, or the root mean square error, is often used the same way as is the standard deviation: to create confidence intervals, by adding and subtracting, say, two times the root mean square error from the estimate to form a 95 percent confidence interval.

By using the above taxonomy, model outputs can be biased as a result of errors in an input data set or as a result of model misspecification,[3] and therefore the root mean square error is often a better summary of the performance of a model's estimates than the standard deviation. [4] External validation directly measures root mean square error (if there is a well-defined experiment) and is therefore directly useful in assessing an estimator's uncertainty. In contrast, a sensitivity analysis provides information about the size of the bias, but only indirectly, because there is no knowledge of the truth, and therefore a sensitivity analysis does not directly measure root mean square error.

[3] Also, a model can purposively make use of a biased estimator. This is done for a variety of reasons, often because it results in a decrease in variance.

[4] It is important to point out that the distinction between bias and variance is not well defined, because it is often possible to think of bias as resulting from some underlying process and therefore as not fixed. This vagueness in definition is closely related to the vagueness in the definition of a model when considering bias due to model misspecification.

Copyright National Academy of Sciences. All rights reserved.

CONDITIONAL VERSUS UNCONDITIONAL CONFIDENCE INTERVALS

The issue of how best to communicate the uncertainty in a model's estimates to decision makers raises a number of difficult questions. Through application of sample reuse techniques and sensitivity analysis, policy analysts will have some quantitative and some qualitative information about the uncertainty in a model's estimates. Often, they will have quantitative information about the variability in the estimates due to sampling error (in the primary database and other sources) and qualitative information about variability due to errors in the data and model component misspecification. But if the quantitative measure of uncertainty is presented alone in the form of a confidence interval that assumes that other aspects of the model are correct, unsophisticated users may interpret the confidence interval in an unconditional rather than a conditional sense and hence ignore many potential sources of error.

An example of the tendency to misinterpret conditional confidence intervals is the communication of sampling error in public opinion and political preference polls. Although providing this kind of information is an important step forward, readers are all too apt to assume that the confidence interval expresses the total error in the poll, instead of just the sampling variability, and hence to overlook the effects of nonresponse, order of the questions, form of the questions, interviewer methods, and other nonsampling errors on the quality of the results.

Another problem in the use of conditional confidence intervals is that the contribution to uncertainty from components other than sampling error, although generally difficult to quantify, very often (once quantified) turns out to be much larger than the contribution from sampling error. Therefore, there is a very real danger that decision makers will form an overly optimistic picture of the quality of the estimates. In addition, even when information about the nonsampling error components can be qualified roughly through the use of sensitivity analysis, it is often impossible to incorporate this information into an unconditional confidence interval with a known coverage probability. Therefore, the problem for unsophisticated users is how to present them with information combining different levels of probabilistic rigor.

To make clearer what we mean by conditional confidence intervals, consider the case of a multiple regression model that is being used to estimate some quantity of interest. There is a well-defined theory that provides a confidence interval of specified coverage probability due to sampling variability for the fitted values. The theory is based on various assumptions, including that the expectation of the quantity of interest is a specific function of the covariates used. This assumption is often difficult to verify. For example, assume that another individual models the same quantity with a different set of covariates. There is an associated confidence interval for the second model, and if some of the underlying assumptions for either of the models do not obtain, there

Copyright National Academy of Sciences. All rights reserved.

is no reason that these confidence intervals need to overlap. That is, two reasonable models can produce confidence intervals for the same quantity that have no values in common. Two such confidence intervals do not necessarily overlap because they are both conditional on the assumptions underlying the two models. At least one model's assumptions are wrong, possibly both. There is no problem in using confidence intervals that are conditional as long as analysts are aware of the assumptions being made.

The use of sensitivity analysis is an attempt to begin the process of developing unconditional confidence intervals, or at least less conditional ones, to develop a better understanding of the possible modeling approaches and their effectiveness. The ultimate objective is to narrow the alternatives to a single best method, with no contribution to uncertainty from model misspecification. A model is developed by an individual or individuals with a perspective on how certain steps of the modeling process can be accomplished optimally. If experts are truly divided about which way a step should be modeled, a research program should be carried out so that the uncertainty due to model misspecification is reduced.

Returning to the regression example, one way in which to develop an unconditional confidence interval is to include the contribution of variability from the model development process in the confidence interval; this has been done when modelers work from the same data set through use of the bootstrap (see Brieman, 1988).

Many of the problems in communicating uncertainty to decision makers are overcome when an analyst has information from several external validation studies of the same type of policy analysis. It is then straightforward to construct a measure of overall error—specifically, the root mean square error—by taking the square root of the average of the squared differences between each estimate and the applicable measure of what actually occurred. This statistic then provides the basis to form, at least approximately, unconditional confidence intervals for the current set of estimates. However, this approach is easily misused. First, the external validation studies will necessarily pertain to different time periods and policy initiatives, so it is not clear to what extent the information applies to the current analysis. Similarly, it is not clear to what extent the various studies are themselves replications of any defined experiment, and thus it is difficult to interpret the root mean square error. Also, the confidence interval formed from the root mean square error does not communicate the uncertainty in the estimated root mean square error itself. That is, was it based on a single external validation study or 10 or 20? [5] Thus, even in the uncommon situation in which policy analysts can make use of external

[5] In addition, the extent to which the estimates in the validation study follow an approximately normal distribution remains unknown, making the usual confidence interval formed by adding and subtracting twice the root mean square error somewhat suspect.

Copyright National Academy of Sciences. All rights reserved.

validation studies to generate confidence intervals for current estimates, there are serious problems that need to be researched.

AN ILLUSTRATIVE DIAGRAM

To assist in understanding some of these ideas, Figure A-1 describes a simple situation, with five models. The estimated ranges for models 1, 2, and 3 —none of which includes the true value (*x*)—are 95 percent confidence intervals for three different model versions that incorporate only sampling variability. (We have represented the densities of the estimates above the confidence intervals as approximate normal distributions, in which case 95 percent confidence intervals are formed by multiplying an estimated standard deviation by 1.96.)

The differences among the ranges for models 1, 2, and 3 are due to, say, different macroeconomic forecasts. Therefore, range 6 is a range of uncertainty due to sampling variability in the input data set and of uncertainty due to the macroeconomic forecast. It is generally impossible to provide range 6 with a rigorous estimate of coverage probability, but it might admit to a probabilistic interpretation if a probabilistic model for the various macroeconomic forecasts can be developed. (Range 6 also might be considered by some to represent root mean square error, with contributions of variance from sampling variability and bias from use of incorrect macroeconomic forecasts.)

Similarly, the ranges for models 4 and 5 are 95 percent confidence intervals for two versions of a completely different model, with the variability represented due to sampling variability. (Neither of these ranges includes the true value.) The difference between the model 4 and model 5 ranges is due to, say, the use of different imputation routines. Range 7, therefore, is another uncertainty range incorporating sampling variability and uncertainty due to imputation method. Some might consider range 7 more difficult to treat as a confidence interval with a known coverage probability than range 6, because the probabilistic structure of potential imputation routines might be more difficult to specify than that of macroeconomic forecasts.

Finally, range 8 is what one might hope to communicate as an estimate of total uncertainty. It is, in this diagram, the only interval that actually contains

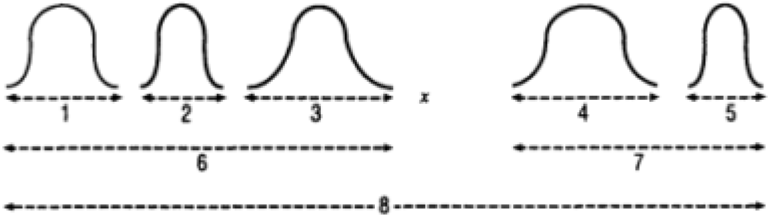


FIGURE A-1 Ranges of estimates of x (truth) from five models.

Copyright National Academy of Sciences. All rights reserved.

the truth. However, this range is difficult to interpret because there may be yet another modeling approach that would further broaden the range of uncertainty. In addition, range 8 cannot be given an associated coverage probability, which greatly weakens its utility. However, it does provide an indication of the amount of uncertainty due to sampling variability, macroeconomic forecasts, imputation routines, and modeling approaches.

Copyright National Academy of Sciences. All rights reserved.

Part II

The Role of Microsimulation as a Policy Analysis Tool

In this part of our report we turn from a broad investigation of the role of information in the policy process to examine with both a narrower and a deeper focus the role of microsimulation modeling as a tool for analysis of social welfare policy issues. Our sponsors—the Office of the Assistant Secretary for Planning and Evaluation (ASPE) in the U.S. Department of Health and Human Services and the Food and Nutrition Service (FNS) in the U.S. Department of Agriculture—asked us to address the basic question of how the agencies should use microsimulation modeling in the 1990s and future decades to support their policy-oriented missions. They further asked us to evaluate the role played by microsimulation models in the legislative debates of the past and to consider advances in databases, research knowledge, computing technology, and statistical evaluation methods that could improve the ability of microsimulation models to support the legislative debates in the future.

We address some of the questions about the historical role of microsimulation modeling in our review of the policy process in Part I. We describe the "first revolution" in the use of numbers to guide policy choices, which fostered the growth in importance over the past 20 years of formal modeling tools as a source of detailed information for legislative decision making. Unfortunately, underinvestment in needed data, research, and model development in the 1980s led to declining, or at best, stagnant capabilities of microsimulation and other types of models to support the policy debate. We have identified two particular areas in which new investment is sorely needed: investment to improve the quality and relevance of the input data used by models and policy research generally, and investment to add value to the outputs of models through systematic

Copyright National Academy of Sciences. All rights reserved.

validation, documentation, and effective presentation of the results to decision makers. We note particularly the need for thorough evaluation of the quality of the information produced by models, both because policy makers need to be aware of the extent and sources of uncertainty in the estimates and because policy analysis agencies need validation results to develop cost-effective strategies for investment in future model development. Thus, we call for a "second revolution," in which both the numbers and associated measures of their quality inform the legislative process and contribute to the development of improved policy analysis tools.

Although our recommendations in Part I pertain generally to the entire range of models, in almost every instance they apply with particular force to microsimulation. The complexity and large scale of most models of this type make it both more than usually difficult and more than usually important to effect improvements in model inputs and to develop systematic programs for validation of model outputs.

Our discussion in this part draws most heavily on experience with the major models for income support programs, which have long been central to our sponsors' interests; however, we also consider models and modeling issues specific to health care, retirement income, and tax policy. (The health care policy area, in particular, is rapidly becoming of critical importance to decision makers.) Among the models that we reviewed are

- TRIM2 (Transfer Income Model 2), MATH (Micro Analysis of Transfers to Households), and HITSM (Household Income and Tax Simulation Model), which are static models of income support and tax programs;
- DYNASIM2 (Dynamic Simulation of Income Model 2) and PRISM (Pension and Retirement Income Simulation Model), which are dynamic models of retirement income programs;
- the submodel added to PRISM to simulate alternatives for financing long-term care of the elderly;
- the tax policy model maintained by the Office of Tax Analysis; and
- MRPIS (Multi-Regional Policy Impact Simulation), which is a hybrid income support and tax policy model that uses microsimulation, input-output, and cell-based techniques.

We first describe the components and operational steps involved in microsimulation, briefly review the important stages in the development of microsimulation models for policy analysis both in the United States and in other countries, and present our general findings about the utility of the microsimulation approach (Chapter 4).

In the remainder of Part II we address in detail the critical issues in microsimulation models: databases (Chapter 5); design principles and practices, and possible expansions in model capabilities (Chapter 6); computing environments for models (Chapter 7); special problems in health care, retirement

Copyright National Academy of Sciences. All rights reserved.

income, and tax policy modeling (Chapter 8); validation of model results (Chapter 9); documentation and archiving of models (Chapter 10); and the structure of the microsimulation modeling community, including the potential of microsimulation for use in social science research (Chapter 11). Each chapter includes recommendations. The Appendix to Part II provides background information on the main features of the microsimulation models listed above, the primary databases used by the models, and technical terms used in microsimulation modeling.

The reader will note that our recommendations for future development of microsimulation models are quite often general in nature and include calls for further study. In particular, we are not able to provide much guidance about desirable changes to microsimulation model design and capabilities—for example, whether model development should emphasize the incorporation of added behavioral elements and whether current static models should adopt more dynamic approaches—because we have found that there is simply no literature evaluating microsimulation models. Very few rigorous studies have been carried out to assess the sensitivity of models to alternative assumptions or specifications or to compare the accuracy of model projections against known values. (As part of its work, the panel conducted a limited validation study of the TRIM2 model, combining an external validation with sensitivity analysis.) Likewise, there have been almost no attempts to assess the variance in model outputs due to such sources as sampling error. Hence, despite obtaining much helpful information and ideas from knowledgeable people in the microsimulation and research communities, we cannot responsibly support specific recommendations about the priority of adding this or that capability or making this or that design change to microsimulation models. We also do not provide recommendations directed to specific models, due both to the paucity of validation studies and to our belief that a broad-based assessment of microsimulation modeling is more important than a detailed evaluation of any particular model. Our frustration on both matters underscores our view that an overriding priority for policy analysis agencies must be to provide adequate resources and support for systematic evaluation efforts that can guide cost-effective development of microsimulation models in the future.

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

4

Microsimulation Models: Then and Now

BASIC ELEMENTS OF MICROSIMULATION MODELS

Defined very simply, the microsimulation approach to evaluating alternative legislative proposals involves modeling the impact of government programs at the level at which they are intended to operate. That is, instead of modeling the impact of program changes on aggregates, such as the national economy or demographic subgroups of the population, microsimulation looks at the impact on individual decision units, which may be families in the case of income support programs, hospitals and doctors in the case of health care cost reimbursement programs, or corporations in the case of changes to corporate-based taxes. Given the diversity of the population, the complexity of most government programs, and all of the factors that need to be taken into account in developing an appropriate microlevel comparison of current policy with one or more hypothetical alternatives, the microsimulation approach inevitably entails a large number of steps.

In Figure 4-1, we sketch the major operations involved in microsimulation modeling, using income-support program models, such as Micro Analysis of Transfers to Households (MATH) and Transfer Income Model 2 (TRIM2), as the prototype. Not every microsimulation model includes all of the steps that are diagrammed, nor do all models implement the steps in the order shown; nonetheless, the chart helps to convey the complexity that characterizes most microsimulation models.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

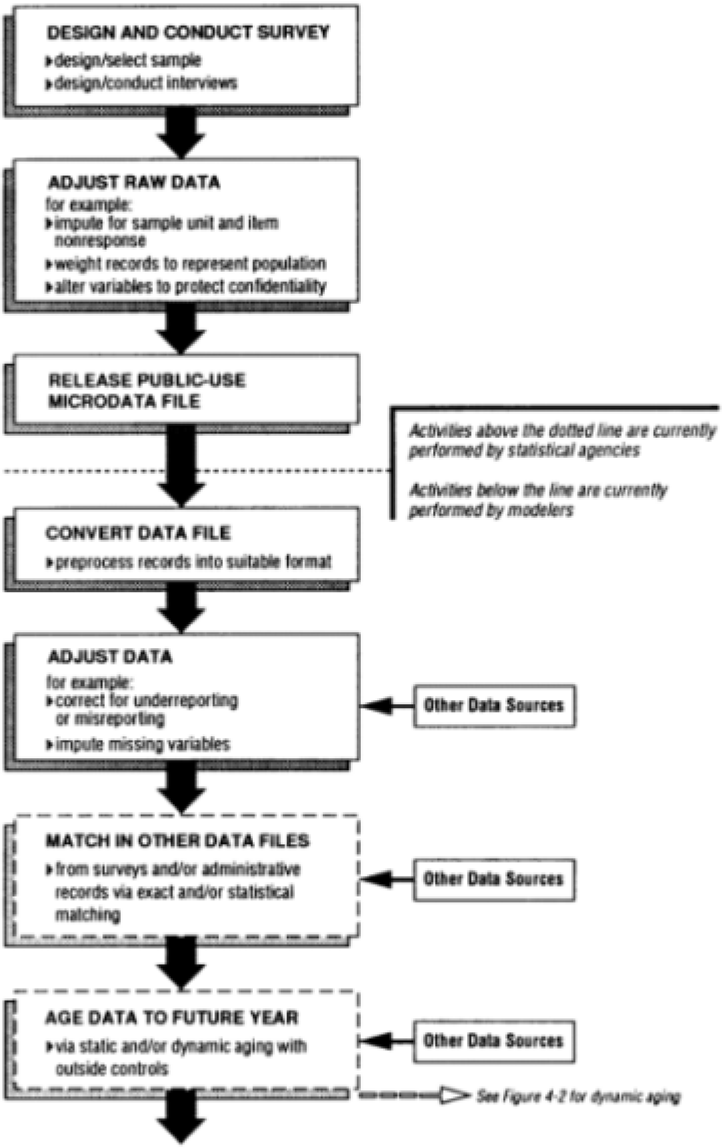
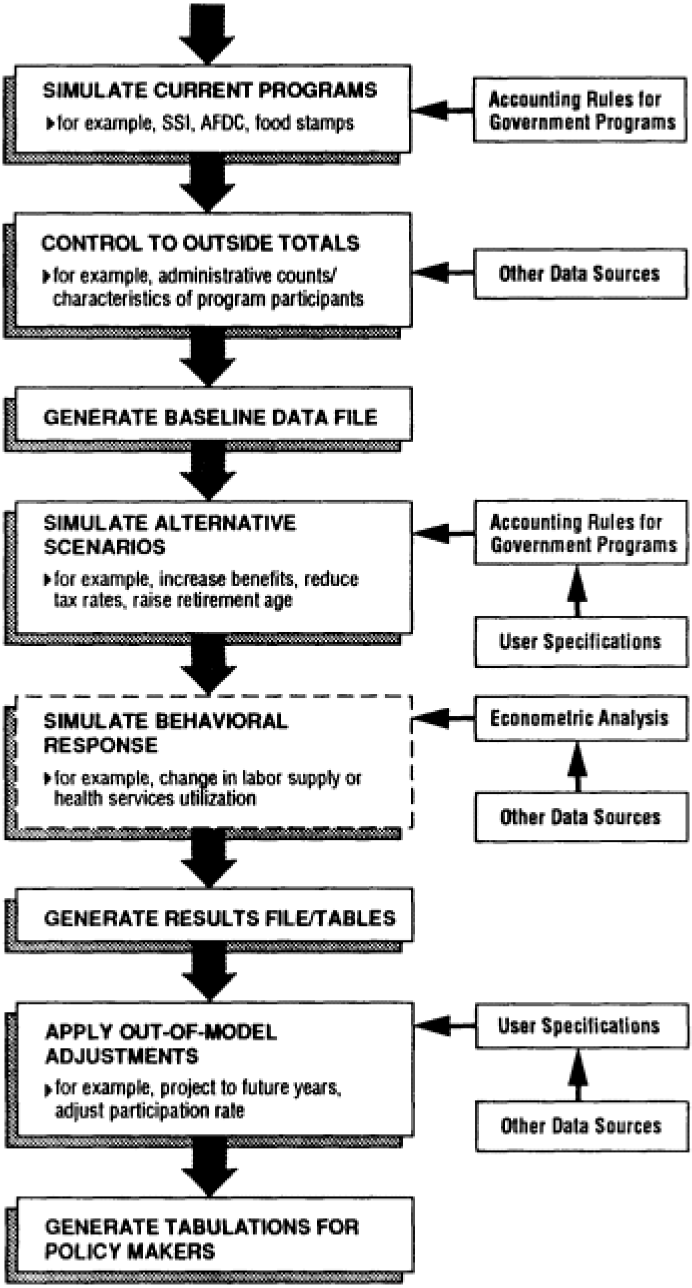


FIGURE 4-1 A microsimulation model schematic.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...



NOTE: *Dotted-line boxes indicate optional components; other components, such as data adjustments, calibration to outside control totals, and out-of-model adjustments, are theoretically optional but almost always included.*

Copyright National Academy of Sciences. All rights reserved.

The chart begins with the operations involved in generating the main database for input to a microsimulation model. For many models, this database is a household survey, such as the March income supplement to the Current Population Survey (CPS). For other models, the main database is a set of administrative records, such as the Statistics of Income (SOI) sample of tax returns. In either case, the originating agency, such as the Census Bureau, goes through a series of operations to generate a microdata file. For a household survey such as the March CPS, these operations include survey and questionnaire design, data collection, data processing, data adjustments (such as imputing values for missing responses and weighting the records to represent the population), and preparing the data for public release. All of these steps have an impact on the quality and utility of the data for microsimulation purposes.

Despite the many operations completed by the originating agency, most microsimulation models go through a number of additional steps to create a database that is suitable to the model's purposes. These steps typically include converting the file to a format that suits the model's hardware and software (e.g., converting a character file to a binary format or converting a household-family-person record structure to a family-person structure) and generating recodes of variables. An important set of recedes for many models is to determine the members of a "filing unit," that is, those people within a household or family who constitute the unit eligible for benefits from a program such as Aid to Families with Dependent Children (AFDC) or the unit that will file a tax return. Models such as TRIM2 and MATH develop a large number of filing unit recodes in a single operation in order to provide flexibility for analysts in specifying policy alternatives. Modelers also typically adjust the database in a variety of ways to make the data more useful for their purposes. Such adjustments include correcting income amounts for underreporting—using control totals from outside data sources such as the National Income and Product Accounts—and imputing missing variables needed in simulations (e.g., imputing various kinds of expenses to the March CPS, such as child care payments that are allowable deductions for programs like AFDC). More elaborate adjustments sometimes involve matching entire data files to the main database, either through exact matching or statistical matching techniques. The major tax models routinely match SOI and CPS records to create a richer database for the entire population; the major retirement income models start with exact-match files containing CPS data and Social Security Administration records of earnings histories. (Exact matches make use of unique identifiers, such as social security numbers, that are present in two or more files to match records for the same people. When exact matching is not possible, statistical matches can come into play, combining records for people who resemble each other on a set of variables contained in two or more files.)

All of these adjustments make use of other data sources that also reflect many operations on the part of the originating statistical or administrative

Copyright National Academy of Sciences. All rights reserved.

agency to collect the data and put them into a form suitable for policy analysis and public release.

Another set of steps that is often, although not always, part of generating a suitable database for modeling, involves projecting, or "aging," the data forward in time. At best, the input data available to modelers are 1 year out of date; usually, they are several years old. At the same time, policy changes are usually proposed for implementation at least 1 year into the future, and Congress typically requires cost estimates for 5 years from the anticipated implementation date. For some policy proposals, such as changes to the social security benefit structure, the desired projection period may extend 30 years or more into the future. Hence, analysts need to make the database more closely resemble anticipated future conditions.

The major income-support program and tax models use "static" aging techniques to project the database. This method involves reweighting the data records to match outside control totals on selected characteristics. Typically, the models adjust the weights to agree with population projections by such characteristics as age, race, sex, and household composition: for example, changing the weights of people aged 24-35 in 1990 to represent the number of people who are expected to be in that age category in, say, 1995. (The population projections themselves are usually the output of cell-based models.) The models may also scale income amounts to agree with inflation forecasts of macroeconomic models and adjust employment status.

The major retirement income models use "dynamic" aging techniques: the model itself generates the expected future composition of the population by applying probabilities for birth, death, marriage, employment status change, and other processes to the individual records; Figure 4-2 sketches the basic steps typically involved in dynamic aging. These models "grow" the people in the database, for example, simulating an unmarried adult age 24 in 1990 to turn age 25 and marry in 1991, to turn age 26 and have a child in 1992, and so on. The aging portion of dynamic models (unlike that of static models) is central to their structure and operation.

After file conversion, data adjustment, and aging, the modeling approach follows a series of steps that involve what one might think of as the model itself. The first step in this sequence typically is to simulate current programs (create a baseline), by using the portion of the model that endeavors to replicate the program accounting rules, and to add to the data records values for eligibility status, the benefit to which the unit is entitled, and so on. A critically important part of the "baseline" simulation involves simulating the participation decision for those units that are determined to be eligible for program benefits. Often, the baseline file is created as soon as a new database is obtained and brought on line so that it is ready for use when the agencies request runs on alternative programs.

An integral part of creating a baseline data file involves adjusting or

Copyright National Academy of Sciences. All rights reserved.

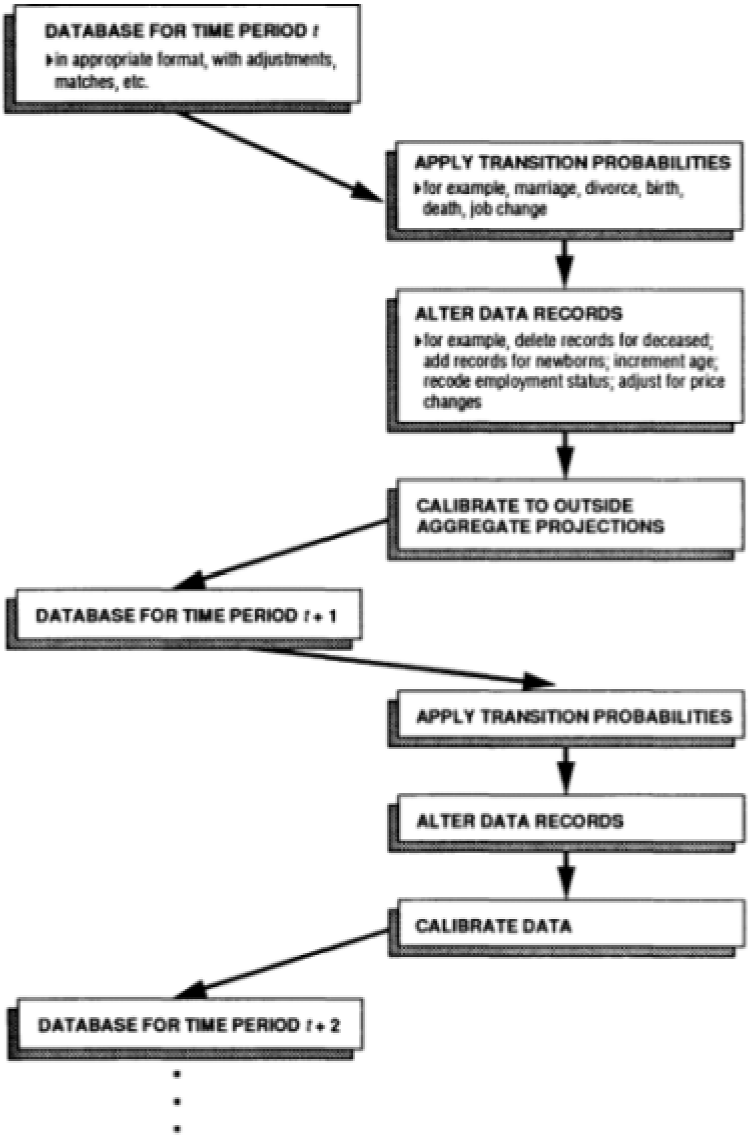


FIGURE 4-2 Dynamic aging in microsimulation models. NOTE: Only the database aging is shown.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

"calibrating" one or more aspects of the simulation so that the simulated values agree as closely as possible with available control totals. For example, the TRIM2 modelers calibrate the baseline simulation of participation in AFDC so that the number and characteristics of units selected to participate in the program match administrative reports of average monthly recipients by state and a few characteristics of the national caseload. The projected database produced by dynamic models is also typically calibrated to outside economic and population projections. Considerable judgment by an analyst enters into the determination of how much effort to devote to the calibration process, because it is rarely possible to achieve complete agreement with all of the desired controls.

The next step is to simulate one or more program alternatives according to the specifications provided by agency analysts. It is often possible to simulate alternatives, such as a change in the amount of an allowed deduction, simply by resetting a parameter switch in the model; in other cases, the model code must be rewritten to accommodate one or more features of the proposed change. If the model simulates behavioral responses to program changes, such as the impact on labor supply of changing the AFDC benefit amount or the subsequent feedback effects of a labor supply response, these components of the model would be the next steps. In practice, however, because the complexities of simulating behavioral responses and second-round effects are an order of magnitude greater than the previous steps, these capabilities are infrequently or only very crudely implemented in today's microsimulation models.

The final step involves tabulating the output for the baseline program and the various proposed alternatives that were simulated. Typical output includes tables of "gainers and losers" under each alternative in comparison with the baseline file. At that time, an analyst may also make "out-of-model adjustments" to the results. For example, analysts in the Office of the Assistant Secretary for Planning and Evaluation (ASPE) project TRIM2 output to the appropriate future years at this stage, instead of using the TRIM2 set of static aging routines. Another type of out-of-model adjustment could include changing the overall level of participation for one or more simulations on the basis of the analyst's assumptions about behavioral response.

DEVELOPMENT OF MICROSIMULATION MODELING FOR POLICY ANALYSIS

Although microsimulation techniques were originally developed by social science researchers, they have been used much less for basic research purposes than for policy analysis. There are some exceptions, notably the field of family demography, which has a 25-year history of development and application of microsimulation models. (We review this history in Chapter 11, as part of a discussion of the potential of microsimulation models to make useful contributions to a broad range of social science research problems that, in turn, could

Copyright National Academy of Sciences. All rights reserved.

well improve their utility for policy analysis.) In this section we give a capsule overview of the key stages in the development of microsimulation modeling as a tool for social welfare policy analysis on the part of the federal government. Our summary is not exhaustive: we cover some of the major models and their applications, which are referred to throughout our report.[1]

Before beginning this overview, we note that microsimulation techniques have also had widespread application for policy analysis on the part of state governments and on the part of government agencies abroad. A recent survey indicated that half the states use microsimulation models, in addition to other techniques, to estimate revenues and assess the impact of tax policy changes (Peat Marwick, 1989). State agencies have also used microsimulation models to analyze a range of social welfare policies.

With regard to microsimulation modeling efforts in other countries, [2] the (preunification) Federal Republic of Germany became active in model development in the early 1970s. German microsimulation models include the Wohngeldmodell, first developed in 1974-1975, which has been used extensively to analyze housing allowance schemes; the Frankfurt model, which was developed over the 1970s primarily for academic research use but also provided an analysis of alternative public pension policies; the BAFPLAN model, first developed in 1976-1977, which has been in continuous use since that time to evaluate public training assistance programs; and the APF model, which was developed in 1984-1985 to analyze family allowance plans.

Canada has also been an active developer and user of microsimulation models since the late 1960s. Recently, Statistics Canada developed the Social Policy Simulation Database/Model (SPSD/M), which is a static model of Canadian tax and transfer programs. The model, which is publicly distributed, is implemented on a personal computer and designed for ease of use by nontechnical analysts. Research and public interest groups, as well as government agencies, have worked with the model.

In Sweden, the Industrial Institute for Economic and Social Research developed a microsimulation model of the corporate sector, the MOSES model, in the mid- to late 1970s. Recently, Statistics Sweden has begun exploring the development of microsimulation models for analysis of public policies affecting

---

[1] There are a number of excellent references on the development of microsimulation modeling for use by federal policy analysis agencies. Turek-Brezina (1988) provides a general overview of the past 20 years of microsimulation modeling history; Haveman (1987) also provides a good historical review. Kraemer et al. (1987) provide a case study of the "datawars" involving the use of the MATH and KGB models in welfare policy analysis in the late 1970s. Webb, Michel, and Bergsman (1990) review the history of TRIM2; Zedlewski (1990) does the same for DYNASIM2. See also Orcutt et al. (1980) for information about the early days of social welfare policy microsimulation modeling and Atrostic and Nunns (1988) for information about the early days of tax policy modeling.

[2] Hellwig (1989a,b) summarizes major microsimulation models developed in the United States and abroad and provides extensive references.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

the household sector. The Hungarian Statistical Office recently completed development of a multipurpose microsimulation model, which to date has been used to analyze income tax and family allowance programs. The model is also intended to generate databases, by means of correcting and merging survey data, that will enable the office to cut back on the frequency of specific surveys. In addition, the Hungarian Statistical Office is planning to develop microsimulation models of the health care system and the corporate sector.

Finally, most members of the Organization for Economic Cooperation and Development (OECD), even if they are not heavily involved in the use of microsimulation techniques otherwise, maintain tax policy microsimulation models. Recently, a working group of the OECD began a program to coordinate development of improved microsimulation models for public policy analysis on the part of interested member countries.

Origins: 1950s to mid-1960s

In the 1950s, Guy Orcutt, an economist, developed the concept of analyzing national economic and social policies by simulating behavior for individual decision units, using microlevel databases and probabilistic (Monte Carlo) techniques (Orcutt, 1957; Orcutt et al., 1961).[3] Such "microsimulation" models, in his view, would provide a richer and much more realistic tool for policy analysis than could the macroeconomic models of the national economy on which he had previously worked. Orcutt concentrated his development efforts on fully dynamic models. In the early 1960s, researchers at the Brookings Institution, encouraged by analysts in the Treasury Department, developed a static, cross-sectional microsimulation model to analyze the impact of changes in federal tax policies on revenues and various categories of taxpayers (Pechman, 1965).

Development of Static Models: Late 1960s to Early 1980s

In the late 1960s, Gail Wilensky and her colleagues on the staff of President Johnson's Commission on Income Maintenance Programs (the Heineman Commission) developed the first federally sponsored operational microsimulation model for analyzing alternative welfare policies. The Reforms in Income Maintenance (RIM) model was used to simulate many variations of a negative income tax and other welfare reform plans (McClung, 1970; Wilensky, 1970). Subsequently, RIM was used by the U.S. Department of Health, Education, and

[3] Monte Carlo or stochastic techniques involve simulating individual decisions on a probabilistic rather than a deterministic basis, for example, choosing eligible units to enroll in a program or married couples to have a child on the basis of a random draw.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Welfare (HEW) and the U.S. Office of Economic Opportunity to simulate welfare policy options considered during the debate on President Nixon's Family Assistance Plan.

In the early 1970s, analysts and programmers at the Urban Institute, under contract to HEW, attempted to rewrite RIM to make the model more user friendly and cost-effective. Failing to do so, they abandoned the effort and developed the Transfer Income Model, the first working version of which was released in 1973. TRIM, like RIM, was a cross-sectional model that used static aging techniques. The Office of the Assistant Secretary for Planning and Evaluation (ASPE) in HEW provided support to the Urban Institute to maintain and further develop TRIM and commissioned simulations of a number of welfare reform schemes. At about the same time, Social Security Administration staff developed their own in-house static microsimulation model, the Simulated Tax and Transfer System (STATS) model, to analyze the distributional impacts of alternative social security policies. STATS was used to produce estimates for the debate that led to the enactment of the Supplemental Security Income program for the elderly and disabled in 1972.

In 1974 analysts involved in the development of TRIM left the Urban Institute to join Mathematica Policy Research, Inc., where they developed a new static tax and transfer policy model based on TRIM: the Micro Analysis of Transfers to Households model. During 1974-1976, the MATH model was used, under contract to the Food and Nutrition Service of the U.S. Department of Agriculture, to analyze more than 200 variations of proposed reforms to the food stamp program; the effort culminated in the 1977 Food Stamp Reform Act. Subsequently, in the mid- to late 1970s, Mathematica administered a subscription service for the MATH model: sponsor agencies funded a program of maintenance, development, documentation, training, and dissemination for the model. At its height, agencies in the Departments of Agriculture, Energy, Health and Human Services, Labor, and Treasury, along with the Congressional Budget Office (CBO) and the Congressional Research Service, were subscribers, receiving regular updates of the model, database, and documentation.

Also in the early to mid-1970s, analysts at the Brookings Institution continued their efforts to apply microsimulation techniques to analyzing the aggregate incidence, by income class, of the entire U.S. tax system—federal, state, and local—as well as the tax burden implications of proposed changes in the tax system. Their databases—originally referred to as MERGE files (see Minarik, 1980)—represented statistical matches of samples of income tax returns from the Internal Revenue Service (IRS) with household surveys such as the March CPS. Meanwhile, the Treasury Department's Office of Tax Analysis brought a version of the tax policy model in-house and undertook continuing development, maintenance, and application of the model for policy analysis purposes. A little later, in the late 1970s and early 1980s, ASPE organized a TRIM users' forum and then a broader based microsimulation users' forum to

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

provide an opportunity for agency and contractor analysts to exchange ideas and information related to the use of microsimulation modeling for policy analysis purposes.

Early in 1977 the Carter administration undertook a major effort to design and enact a welfare reform program that included job training and employment components, known as the Program for Better Jobs and Income (PBJI). Both the Departments of Labor (DOL) and Health and Human Services (HHS) were involved: DOL sponsored the development of modules in the MATH model to simulate acceptance rates for a public service jobs program and labor supply responses to the proposed welfare changes; ASPE staff in HHS constructed their own model to simulate PBJI. The model, known as KGB after the authors' last names (Kasten-Greenberg-Betson), included a jobs component and simulation of labor supply responses (see Betson, Greenberg, and Kasten, 1980). The development of KGB was completed on a crash basis in 5 weeks. Both MATH and KGB estimates were used in the policy process as the two departments hammered out a proposal satisfactory to President Carter. However, the initiative failed in Congress, and by 1980 the KGB developers had left ASPE, and the model had fallen into disuse.

Also in the late 1970s, the Urban Institute received support from ASPE to redesign TRIM in order to effect major improvements in its operating efficiency and usability. (A primary reason for the development of KGB had been the belief that TRIM, as then implemented, was too difficult and costly to modify.) Somewhat later, the MATH model was also redesigned to improve computational efficiency.

Meanwhile, in the late 1970s and early 1980s, proposals were floated to develop comprehensive microsimulation models for the health care sector. However, except for limited capabilities to simulate the Medicaid program that were added to models such as MATH, health care policy analysts were using cell-based models. For example, ASPE used the Health Financing Model, which is primarily cell based, to cost out alternative national health insurance schemes considered by the Carter administration.

Development of Dynamic Models: 1970s to Early 1980s

Dynamic microsimulation models proved more difficult to develop and apply than their static cousins. Orcutt and his colleagues at the Urban Institute received funding in 1969 to develop the Dynamic Simulation of Income Model (DYNASIM), the first version of which was completed in 1975. DYNASIM was designed to simulate a wide range of demographic and socioeconomic life events and their interactions with government policies over time. Interest waned in using the model to analyze welfare policies, but interest grew in

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

applying DYNASIM to generate earnings histories and examine the long-range implications of social security policies.

In the late 1970s, the Urban Institute received funding from DOL and CBO to effect a major redesign of DYNASIM. The goals of the redesign were twofold: first, to make the model much cheaper to run with large samples and able to be moved to government computers; second, to focus the model on retirement issues, including improving labor force histories and the social security module and incorporating a private pension simulation module. Over a 3-month period in 1980, Lewin/ICF, Inc, developed a primitive version of the Pension and Retirement Income Simulation Model (PRISM) for the President's Commission on Pension Policy, to analyze the effects of alternative ways of integrating public and private pension systems.

Retreats and Advances: The 1980s

During the early 1980s, use of the static tax and transfer program models, such as TRIM2 and MATH, languished, because of constrained resources for model development and use, and because the Reagan administration was primarily concerned with cutting back welfare programs and entitlements. This period saw the development of benefit-calculator models, which run on samples of welfare program administrative records and can readily simulate changes that reduce eligibility and benefits. However, ASPE took advantage of the improved cost-efficiency of TRIM2 to have the Urban Institute improve a number of modules, including a module to allocate yearly income and employment values to monthly values (to match program eligibility rules) and modules to simulate state and federal income taxes. CBO also brought TRIM2 in-house and used the model for studies that helped pave the way for later welfare reform initiatives (see, e.g., U.S. House of Representatives, 1985). The MATH model was also used for a study sponsored by the Congressional Research Service to sort out the effects of the 1981 Omnibus Budget Reconciliation Act cutbacks in welfare programs from the effects of the recession on the low-income population (Citro and Beebout, 1984).

In the early to mid-1980s, PRISM was further refined and enhanced (e.g., a macroeconomic model link was added in 1982-1983). It was used for a number of analyses of retirement income, including simulation of the effects of alternative private pension vesting plans and of the major changes to the social security program that were enacted under the Reagan administration. At the same time, DYNASIM2 was used to analyze the implications of retirement policy changes, including legislation to limit mandatory retirement and the 1983 Social Security Act Amendments. DYNASIM2 was also used for other studies, such as the effects on welfare program costs of alternative rates of teenage childbearing.

In the mid-1980s, Lewin/ICF, Inc., developed the Household Income and

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Tax Simulation Model (HITSM), as a proprietary model—an indicator, perhaps, of reviving interest in microsimulation techniques for policy analysis. Several years later, the same firm developed the Health Benefits Simulation Model (HBSM), the first major microsimulation model for analyzing alternative health care policies as they affect the household sector. The primary database for HBSM is the 1980 National Medical Care Utilization and Expenditures Survey. In the mid-1980s, the Brookings Institution, with foundation and HHS support, worked with Lewin/ICF, Inc., to develop a submodel in PRISM to analyze alternative schemes for financing long-term health care of the elderly. Subsequently, with funding from ASPE, the long-term care financing submodel was revised and developed as a public-use model.

Over the 1980-1988 period, ASPE made a major investment in a model to simulate the second-round effects of policy changes, such as the effects on regional employment of welfare or tax reform: the Multi-Regional Policy Impact Simulation (MRPIS) model. It was developed by the Social Welfare Research Institute at Boston College and included microsimulation, input-output, and cell-based components. Meanwhile, the Office of Tax Analysis and the congressional Joint Committee on Taxation continued to develop and maintain tax policy models that received extraordinarily heavy use in the policy debate culminating in the Tax Reform Act of 1986. At the same time, ASPE analysts used TRIM2 to analyze the distributional effects of tax reform proposals. In the late 1980s, TRIM2 was used heavily in the policy debate that led to the Family Support Act of 1988. Subsequently, TRIM2 was used to evaluate alternative child care tax credit schemes. The Medicaid module in TRIM2 was revised and applied to analyze the impact of expanding Medicaid coverage of the low-income population.

Throughout the 1980s, the MATH model was used to simulate a range of proposals to modify the food stamp program. For the fiscal 1984 budget, both MATH and the food stamp benefit-calculator model were used to simulate alternative proposals to simplify the program rules while minimizing the effects on current recipients. Those two criteria proved impossible to satisfy at the same time, so that the effort failed in Congress (see Carlson, 1989). Subsequently, MATH was used to simulate the impact of changes in the minimum wage on the food stamp program and the scale of the program if there had been no changes since 1980.

In the mid- to late 1980s, the Food and Nutrition Service (FNS) supported efforts by Mathematica Policy Research, Inc., to investigate the potential for using data from the new Survey of Income and Program Participation (SIPP) as a microsimulation model database. Mathematica built the FOSTERS (Food Stamp Eligibility Routines) model for the food stamp program with data from the 1984 and 1985 SIPP panels. This model was used to simulate congressional proposals to raise the asset value of vehicles that program recipients can own

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

and still remain eligible; the high price tag estimated for this change deterred congressional action.

Finally, in the late 1980s, both ASPE and FNS sponsored studies to evaluate aspects of the TRIM2 and MATH models, including the study by this panel.

In this once-over of the history of microsimulation modeling in the United States, we have by no means captured the full extent of the contribution of microsimulation models to the policy analysis function; we have glossed over many important policy uses and also many models developed for more limited purposes.[4] However, we hope to have conveyed the flavor of the wide-ranging and important uses that microsimulation models have had in the policy process over the past 20 years.

ROLE AND CURRENT STATUS OF MICROSIMULATION: FINDINGS

On the basis of our review of the experience to date in using microsimulation models for policy analysis, and considering the comparative merits of other kinds of models, we present our general findings about the current role and status of the microsimulation approach. The rest of Part II provides findings and recommendations about the various aspects of microsimulation models that we have addressed—beginning with their data inputs and concluding with their potential uses for basic social science research.

Overall Value

The microsimulation modeling approach to estimating the impact of proposed changes in government programs offers important conceptual and operational benefits to the policy process. Microsimulation models operate at the level of the individual decision unit, by taking into account the diverse characteristics and circumstances of the relevant population, whether it be low-income families, taxpayers, or health care providers. They obtain input from microlevel databases of individual records, mimic the way in which government programs apply to the individuals described in those records, and maintain the outputs of simulated variables for current and alternative programs on each of the individual records. As a result, the models have the capability to respond to important information needs of the policy process:

[4] See, for example, Maxfield and Rucci (1985). which describes a microsimulation model developed for ASPE to simulate the effects of targeted employment and training programs for long-term welfare recipients.

Copyright National Academy of Sciences. All rights reserved.

- First, microsimulation models can simulate the effects of very fine-grained as well as broader policy changes. For example, a microsimulation tax model can estimate the effects of a proposed change to the tax code that applies only to taxpayers with certain kinds or levels of income or expenses, as well a proposed increase or decrease in tax rates across-the-board.
- Second, microsimulation models can simulate the impact of proposed changes that involve complicated interactions among more than one government program. For example, a microsimulation model of income support programs can simulate the net effects of a proposed change to AFDC that also alters the calculation of food stamp benefits.
- Third, microsimulation models can simulate the effects of proposed changes on subgroups of the population, in addition to aggregate estimates of program costs and caseloads. For example, a microsimulation model of physicians' services can simulate the effects of changes in Medicare fee schedules on different types of medical specialties and geographic areas; or a microsimulation model of health insurance programs can provide detailed distributional information about the effects of changes in insurance coverage and benefits on specific types of families.

Besides offering flexibility in examining alternative programs, microsimulation models—in common with many other modeling techniques—provide a framework that ensures consistency of estimates across a wide range of proposals.

The orientation of microsimulation models to the individual decision unit is conceptually attractive. It is, after all, individual parents who decide whether or not to apply for AFDC or to take or quit a job; it is individual taxpayers who decide to itemize or not to itemize deductions or to move assets from taxable to nontaxable instruments in response to tax law changes; it is individual doctors who decide to increase or decrease the number of diagnostic tests of patients in light of government cost reimbursement policies. Although individuals may behave like other parents or taxpayers or doctors with similar characteristics, accurately portraying their individual circumstances and the factors entering into their decisions appears crucial to analyzing the complex issues in social policy.

We conclude that no other type of model can match microsimulation in its potential for flexible, fine-grained analysis of proposed policy changes. Large-scale macroeconomic models, which are designed to estimate the aggregate effects of policy and program changes, such as the implications for the deficit and for national economic growth of a President's proposed budget, rival microsimulation models in size and complexity. However, these models use entirely different data and modeling techniques: evaluation of systems of simultaneous equations estimated with aggregate time series (such as the relationship of public to private spending and investment). Their outputs are for aggregates, such as all consumers in the nation (or in a region or state), and they are in no way able to estimate the impact of changes in particular programs

Copyright National Academy of Sciences. All rights reserved.

on particular groups, such as the effects on the working poor of mandating the AFDC unemployed-parent program in all states.

Simpler macrolevel models, which estimate a single equation on the basis of a few aggregate time series, are often applicable to analyses of particular programs. For example, such a model might estimate growth in AFDC costs and caseloads on the basis of changes in unemployment, inflation, and average benefit level. However, single-equation time-series models are very limited in scope and do not provide any real capability for analyzing complex program alternatives or for sorting out the detailed effects of program changes.

Cell-based models, which develop estimates for subgroups or "cells" that make up the population of interest (for example, an AFDC model might comprise cells for case type by state), can provide more detailed information on policy effects than macroeconomic models, but they, too, are limited in comparison with microsimulation models. Cell-based models, whether they contain several thousand or only a handful of cells, make the critical assumption that all elements within a subgroup will behave in the same way. Should a policy change affect members within cells in different ways, or should policy makers want information for different groupings, a cell-based model must be rebuilt.

Microeconometric multiple-regression models, which produce estimates of the impact of a set of variables on some aspect of individual economic behavior, resemble microsimulation models in their use of microlevel data and their ability to provide disaggregated as well as aggregated results. For example, regression models of welfare program participation—which might include explanatory variables for family size and type; family income and expected benefit level; age, race, and sex of family head; and other characteristics—can be run on a microlevel database to produce participation probabilities for individual families. In turn, these probabilities can be aggregated for subgroups or for the total population. However, the key variable for analyzing the impact of a proposed program change with such a model, namely, expected benefit level, must be supplied by some other means. Indeed, some microsimulation models use a regression-based approach to determine program participation, after they have calculated program eligibility and expected benefits, by applying the detailed program operating rules to each family's record.

A Range of Policy Analysis Tools

Microsimulation models are by no means the only useful tool for policy analysis. Indeed, the policy analysis community benefits from having available a wide range of modeling tools to answer a variety of questions and provide alternative perspectives. As we just noted, microsimulation models are distinguished by their capability for fine-grained analysis of proposed policy changes; however, this capability is not always needed. There are many policy issues and questions for which it is neither necessary nor advisable to crank up

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

a full-scale microsimulation model, particularly in view of the costs of large databases and complex modeling routines.

When policy makers are primarily concerned with the impact of proposed policy changes at an aggregate level—such as the interrelationships among federal spending and revenues, on one hand, and economic aggregates such as the inflation rate, on the other—it seems perfectly appropriate and cost-effective to use aggregate modeling techniques. And in cases for which limited distributional information is desired, it may be most cost-effective to carry out the estimation with a cell-based spreadsheet model.

However, in those many instances in which the policy process is likely to require detailed information for each of a number of proposed alternatives, microsimulation is potentially the only approach that can satisfy the analytic needs. Hence, we believe that microsimulation models merit continued support on the part of federal agencies as an important, if not the only, tool for estimating the impact of proposed program changes.

Analytical Costs

The capability for detailed analysis provided by microsimulation models comes at a price. Although we support the use of microsimulation models for policy analysis, it is important to recognize that the complex nature of such models entails costs. Microsimulation models are highly complex for a number of reasons: they typically require large amounts of data; they must model many complex features of government programs; and they are pressed to provide more and more elaborately detailed information.

Because of their complexity, microsimulation models can be resource intensive to develop and apply, and difficult to understand and evaluate. Moreover, because microsimulation models must usually meld together a variety of data and research results of varying degrees of quality and, in the process, make many unsupported assumptions, there are potentially serious implications for the quality of the resulting estimates. And there are likely to be compounding effects of the errors introduced at each of the many steps in the simulation process.

Indeed, we are gravely concerned that the history of microsimulation model development to date has witnessed too many instances in which costs have proved disproportionately large in comparison with benefits. In our view, the tendency to pile complexity upon complexity has all too often led to a situation in which the modeling task—whether it be for development or application—incurs added time and cost; in which it is difficult for the analyst, let alone the decision maker, to evaluate the quality of the output; and in which the model, instead of providing a capability for timely, flexible response to changing policy needs, becomes sluggish and inflexible in operation.

A typical response in the past to the problems posed by the complexity of

Copyright National Academy of Sciences. All rights reserved.

microsimulation models has been to pare back the capabilities of the model or to focus new development on the model's "accounting" functions that mimic program rules and leave aside other, more difficult aspects, such as modeling behavioral response. However understandable in many instances, these kinds of choices limit the usefulness of the models for the policy debate.

In our review, we accorded high priority to identifying strategies with the potential to improve the quality, flexibility, accessibility, and overall cost-effectiveness of the next generation of microsimulation models without compromising their ability to provide the fine-grained policy information that is their prime reason for being. We believe that such strategies exist: for example, new computer technologies are very promising in this regard. An important implication of our recommendations is that policy analysis agencies must be willing, over the next few years, to allocate a higher percentage of available resources to investment in microsimulation models rather than to current applications (unless, of course, overall budgets can be increased). As we discuss further below, investments are urgently needed to improve the data, research, and computational inputs to models. Investment is even more urgently needed to evaluate the quality of model outputs and to build capabilities into models that will facilitate systematic validation in the future.

Model Uncertainty

The overall uncertainty of the estimates produced by existing microsimulation models is virtually unknown at this time. Although in theory the microsimulation models in use today provide better estimates of distributional impacts and at least as good estimates of overall costs and caseloads as other kinds of models, it is not known if this theory is true in fact. There is very little evidence with which to assess the validity of microsimulation model results, that is, how well they compare with actual policy outcomes. In addition, there are almost no measures available of the degree of uncertainty or variability in the estimates or the major sources of variation. However, we suspect that the level of uncertainty, given the large number and varying quality of microsimulation model inputs, is high.

We believe that analysts and policy makers can have considerable confidence in the quality of the computer models per se, that is, in the accuracy with which the computer code replicates the model specifications. Microsimulation modelers have long made a practice of devoting time and resources to computer model verification, through such activities as examining individual test cases to determine that the code to simulate a program change is working properly. Another check against egregious errors in the computer code is the long-standing practice of analysts from various agencies, in both Congress and the executive

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

branch, to get together periodically over the course of developing major legislation to compare models' outputs and to search vigorously for explanations of discrepancies.

However, very little systematic study has been conducted of the quality of the estimates produced by microsimulation models during their 20-year history of use in the policy process. The dearth of analysis includes external validation studies that compare model output with measures of truth; internal validation studies that assess the sensitivity of model results to the input data, the specifications for individual modules and their interactions, and other components of the simulation process; and studies that assess the variance of model estimates due to sampling error in the primary database and other sources.

Microsimulation models are not alone in lacking systematic validation of their outputs. As we note in Chapter 3, information about the uncertainty in estimates of the effects of proposed policy changes is largely absent from the policy debate, regardless of what type of modeling tool has been used. The conditional nature of almost all policy analyses makes the task of validation difficult. The many different factors involved in most policy analyses are also a hindrance to validation, as is the resistance of decision makers to information about uncertainty. Given the highly complex nature of microsimulation models, it is perhaps not surprising that the validation literature for their outputs is so scant. Yet we believe strongly that the impediments to model evaluation can and must be overcome. Otherwise, policy makers will continue to make decisions based on numbers that may be of highly variable quality, and the agencies that provide support to decision makers will lack information on the most cost-effective ways to invest in improved microsimulation models for the future. Given the high costs of microsimulation model development, it is particularly important to have good information on which to base investment decisions. We are encouraged that, recently, both ASPE and FNS have supported major microsimulation model evaluation studies, in addition to that undertaken by this panel (see, e.g., Doyle and Trippe, 1989; Kormendi and Meguire, 1988); as discussed further in Chapter 9, much more work needs to be done.

Data Adequacy

There are serious questions about the adequacy of the data sources used to construct microsimulation model databases. Much of the computer code and sizable fractions of staff resources for current microsimulation models are devoted to reprocessing and manipulating available input data, not only to produce databases that are more efficient to process, but also to try to compensate for deficiencies in data content and quality. Examples of important deficiencies for modeling income support programs include underreporting of income receipt and undercoverage of population subgroups, particularly low-income minorities,

Copyright National Academy of Sciences. All rights reserved.

in household surveys such as the March CPS. SIPP was designed to address some of these problems, but it does not currently have sufficient sample size and is not timely enough to be a satisfactory substitute. For data on health care, there are serious gaps, difficulties in linking available data sources together, and problems with timeliness. For data on retirement income and tax policy, impediments to linking survey and administrative data cause serious problems for models. In our view, improvements in data quality, together with a shift in the data production function to place more responsibility for producing useful databases on the originating agencies, represent high priorities that promise substantial dividends in terms of reduced cost and improved relevance and quality of model estimates. Again, although we here emphasize the linkage of data quality and microsimulation modeling, we point out that all analytical approaches to the development of policy estimates rise and fall with the quality of the data.

Research Knowledge

There are serious questions about the underlying base of research knowledge that supports modeling individual behavior and other model capabilities. Although predicated on the desirability of simulating individual decisions as they are affected by and affect government programs, current microsimulation models are very limited in this regard. This statement applies not only to models that are avowedly "benefit calculators," such as the administrative records-based models of AFDC and food stamp recipients, but also to models that simulate program effects for the broad population. Except for the basic decision of whether to participate in a new or modified program, the models rarely simulate other behavioral responses, such as the response of income support beneficiaries to work incentives. They also rarely simulate second-round effects of a policy change, such as the impact of raising or lowering health care benefits on consumption of medical services and, consequently, on employment in the health care sector in relation to the rest of the economy.

An important factor in this paucity of behavioral components in microsimulation models in addition to high cost and complexity is the weakness of the underlying research knowledge base. There are no generally agreed-upon estimates of key behavioral relationships, and the form of the available parameter estimates is often not readily suited to implementation in a microsimulation context. We do not anticipate rapid progress in ameliorating this situation, given constrained budgets for research and aspects of the academic research culture that militate against the kinds of research that can most benefit the policy analysis process. However, we offer a number of recommendations for the agencies to spur the production of policy-relevant research. We also recommend model design and development practices that we believe are most cost-effective for incorporating new research knowledge.

Copyright National Academy of Sciences. All rights reserved.

Computer Technologies

The adequacy of the computer hardware and software technologies used to implement current microsimulation models is questionable. The major social welfare policy microsimulation models that are widely used today are designed for mainframe, batch-oriented computing environments that represent yesterday's technology and limit the models in important ways. Computing costs for a single simulation run are much lower for today's models than for the models of the 1960s and 1970s. However, other costs, such as the combined staff and computer costs of rewriting portions of the model code—often needed to simulate innovative policy proposals—remain high. The current computing environment for microsimulation modeling discourages experimentation, either substantively or for validation purposes, and puts barriers in the way of direct access by analysts to the models.

Some model developers have explored the potential of microcomputer technology to support more flexible and accessible models with promising results. Other hardware configurations, such as some combination of linked micro and mainframe computers, may also provide improved capabilities. New developments in software, such as graphical user interfaces and computer-assisted tools for design of software, are also very promising. We strongly recommend that agencies position themselves to build the next generation of microsimulation models around new computer hardware and software technologies that can enhance the cost-effectiveness of this important class of policy analysis tools.

Microsimulation Modeling Community

The current structure of the microsimulation modeling community is costly. Several aspects of the interrelationships among the policy analysis agencies that use microsimulation models, their modeling contractors, and academic researchers are troubling. One set of problems stems from the highly decentralized and fragmented nature of policy analysis in the federal government. While having positive features, the involvement of many different agencies frequently imposes costs of duplication of effort and often isolates groups of analysts who could benefit from a higher level of communication and an exchange of ideas and viewpoints. Our suggestions of useful ways to enhance interagency cooperation are oriented to microsimulation, although the problems in this area also affect policy analysis based on other types of modeling tools.

Another set of problems stems from the very circumscribed nature of the community that is actively involved in developing and applying microsimulation models. As in the past, there are today a handful of private firms that operate the major microsimulation models for social welfare programs on behalf of their federal agency clients. The agencies, which typically have only a few or no staff who are able to use the models themselves, are very dependent on

Copyright National Academy of Sciences. All rights reserved.

their contractors for support. In our observation, these firms have performed responsibly and capably in responding to agencies' needs. Nonetheless, we believe that it would be beneficial to expand access and use of the models on the part of agency analysts. It would also be useful to expand access and use of the models on the part of academic researchers, who in most disciplines have played a relatively minor role heretofore in applying, refining, and evaluating this class of models. Having more people who are knowledgeable about microsimulation models and adept in using them can only help in the development of improved models and in the vital process of validating model results.

Future Directions

In sum, we believe that microsimulation models are important to the policy process, and we anticipate that the need for the kinds of detailed estimates that they can best generate will grow in future years. However, because of the lack of evidence with which to assess the performance of the current models and the limitations of available databases and research knowledge, we cannot responsibly advocate substantial investments that would expand the capabilities of existing models in any specific direction. We strongly support allocating sufficient resources to the current models to evaluate their capabilities, maintain them, and improve them as appropriate and cost-effective. The validation and maintenance functions, together with incremental improvement, are critical to the ultimate objective of developing a new generation of microsimulation models after investments in data, research, and computing technology have borne fruit. Maintaining a cadre of knowledgeable and experienced users and producers of the current models will enable new models to be built much more expeditiously and efficiently. We urge the relevant agencies to make the investments that are required to ensure that a new generation of models is developed in a timely manner to meet the policy needs of the future.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

5

Databases for Microsimulation

Within the data-hungry world of policy analysis, microsimulation modeling stands out as an unusually voracious consumer. These kinds of models require microlevel databases with large numbers of records and large numbers of variables on each record in order to provide the detailed outputs that are their hallmark.

The federal statistical system currently provides a wide range of microdata on which models can draw. Static models of income support programs, such as TRIM2, MATH, and HITSM, have traditionally used the March income supplement to the CPS as their primary database, [1] with information from other surveys and administrative records systems to fill gaps and improve data quality. The Survey of Income and Program Participation (SIPP) was designed to correct many deficiencies in the March CPS and to provide an enhanced database for modeling government transfer programs such as AFDC and food stamps (SIPP was also designed to facilitate modeling tax policies). However, to date, SIPP has been plagued with problems that have hindered its use in microsimulation.

Dynamic models of retirement income programs, such as DYNASIM2 and PRISM, have also relied on the March CPS. Because they require earnings histories over time to calculate entitlement and benefits from social security and private pensions, they have used exact-match files of the March CPS with Social

[1] Some of these models have used other databases in the past, such as the decennial census public-use samples, the 1967-1968 Survey of Economic Opportunity, and the 1976 Survey of Income and Education. However, the March CPS has remained their database of choice, principally because it is updated every year, has a reasonably large sample size, and contains many needed variables.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Security Administration records for the sample individuals. Only one such file has been made widely available—a 1973 exact-match CPS-SSA file, which is the database for DYNASIM[2]. A 1978 exact-match CPS-SSA file was obtained by President Reagan's Commission on Pension Policy; the commission's contractor, Lewin/ICF, Inc., developed a database for PRISM by matching the 1978 file to the March and May 1979 CPS. (The May survey provides detailed information about pension coverage to supplement the employment and income information in the March survey.)

Tax policy models use a combination of data from the March CPS and the Statistics of Income (SOI) samples of tax return records. The SOI provides information (for tax filers) about income reported to the IRS, deductions claimed, and taxes paid; the March CPS provides needed information about family and socioeconomic characteristics of tax filers and the nonfiling population. Because exact-match files of CPS and IRS data are not publicly available, tax models must implement various kinds of imputation and statistical matching techniques to relate the CPS and SOI files.

Existing health care policy models are generally targeted to specific issues, such as extending insurance coverage or modifying policies for reimbursement of hospital costs. Thus, they rely on different specific databases. Some health models have used the health insurance data in the March CPS; other models have used data on health care services and spending from medical care expenditure surveys conducted in 1977 and 1980; still other models have used administrative data sources such as Medicare claims records.

The statistical agencies currently carry out many operations on their data prior to release—including recoding, editing, and weighting—that enhance the quality and utility of the information for modeling and other kinds of research and analysis (see the boxes in Figure 4-1 above the dotted line).2 However, the modelers in their turn typically must implement many additional steps to generate a suitable database for simulation purposes (see the boxes in Figure 4-1 just below the dotted line). A number of these operations would be required in any case—for example, converting a public-use file into the internal format that the particular modeling software is designed to read. Other steps—such as adjusting income amounts for underreporting and misreporting—are implemented to correct problems with the data that the originating agency did not address. Still other steps—such as imputing values for allowable deductions from income in determining program eligibility—are implemented to provide needed information not contained in the primary input file. The result is considerable duplication of effort across models that use the input data and the need for large sections of code in each model for data processing prior

[2] See Citro (in Volume II) for a chart of the steps taken by one model, TRIM2, to create a new baseline file each year from the March CPS. The effort occupies several months of calendar time, and it accounts for a significant share—about one-sixth—of the Urban Institute's total contract funds from ASPE for maintaining and using the model.

Copyright National Academy of Sciences. All rights reserved.

to invoking any of the simulation modules per se. (If it were just duplication across the relatively few existing microsimulation models, we would not be as concerned as we are. Unfortunately, many of these data issues confront a wide range of users of the data: policy analysts and researchers employing different methods and addressing many different questions.) Moreover, even after all of the preprocessing of the data, by both the originating agency and the microsimulation model, data quality problems remain.

In this chapter, we consider the data quality problems that confront current microsimulation models and the kinds of strategies that have been employed to deal with problems of missing, erroneous, and inappropriately specified data, and we present our recommendations for improving data quality in the future. Because of the prominence of the March CPS for microsimulation modeling, our discussion focuses on data quality problems with this survey, particularly in its application for modeling income support programs. We also consider the potential of SIPP to enhance or replace the March CPS as a modeling database.[3]

We conclude that, for the foreseeable future, a mixed strategy is preferable, in which the March CPS continues to be the primary database for models such as TRIM2 and MATH, while other data sources, including SIPP and administrative records, are used to supplement and adjust the CPS data. We further conclude that the overall cost-effectiveness of policy analysis could be improved if statistical agencies, particularly the Census Bureau, evaluated key data sets more thoroughly from the perspective of the policy uses of the data and made use of evaluation results and information from a range of sources to develop enhanced data sets. That is, in terms of Figure 4-1, we propose moving down several steps the dotted line that demarcates the data processing functions of the originating agency from those currently embedded in microsimulation models. Finally, we note that our recommendations for needed improvements in microsimulation model databases—because of the breadth and depth of information that microsimulation requires—are likely to benefit many other kinds of research and analysis as well.

DATA QUALITY: THE MARCH CPS

The databases used by current microsimulation models are the product of substantial expenditures of resources, first by the originating agencies such as the Census Bureau, and then by the modelers themselves. Yet important data quality problems remain. Moreover, the procedures that the statistical agencies

[3] Our discussion of the March CPS and SIPP as microsimulation model databases for income support programs benefited greatly from a paper prepared for the panel by Citro (in Volume II), which also provides extensive references. Key references include: Allin and Doyle (1990); Bureau of the Census (1989a, 1990a); Committee on National Statistics (1989); Doyle and Trippe (1989); Jabine, King, and Petroni (1990); and Vaughan (1988). We discuss data problems for modeling health care, retirement income, and tax policies in Chapter 8.

Copyright National Academy of Sciences. All rights reserved.

and the modelers use to correct various problems are themselves sources of both variability and bias in the resulting databases. This section reviews some of the data quality problems in the March CPS income supplement.

For more than two decades, the March CPS has served as the premier database for modeling income support programs such as AFDC and food stamps. Briefly, the CPS is a continuing monthly survey of the U.S. civilian noninstitutionalized population designed to provide estimates of employment and unemployment for the nation and large states. The sample size is about 60,000 households containing about 120,000 people aged 15 and older. Each March, the survey includes an income supplement that asks about labor force experience and income in the preceding calendar year. The Census Bureau releases a public-use file from the supplement about 6 months after the data are collected. This file is used for many kinds of social welfare policy modeling and analysis, as well as microsimulation, and is also used heavily by academic researchers.

Periodically, in response to the needs of microsimulation and policy analysis generally, the March supplement has been modified to provide more useful data. For example, the number of income sources identified in the supplement was greatly expanded, and questions on health insurance coverage were added. Yet the March supplement exhibits many data gaps and problems from the viewpoint of modeling social welfare programs: (1) problems resulting from the survey design and data collection, (2) problems resulting from inadequately detailed variables compared with modeling needs, and (3) problems of needed variables that are missing entirely from the survey.

Survey-Based Problems

Coverage

The March CPS, in common with other household surveys, fails to cover the entire population. This conclusion is based on comparing the weighted survey counts (after adjusting for known nonrespondents) with population estimates based on the last decennial census, updated by administrative records on births, deaths, and net immigration. Net undercoverage rates in the CPS, which amount to about 7 percent of the total population, vary widely: from only 1 percent of elderly white women to 27 percent of young black and Hispanic men.

The Census Bureau adjusts for undercoverage in the CPS and other surveys by increasing the household weights to match population control totals by age, race, and sex. However, this adjustment does not take into account the estimated net undercoverage in the decennial census itself, which, for the 1980 census, was a little over 1 percent of the total population and perhaps about 15 percent of middle-aged black men.[4] Moreover, the undercoverage adjustment that is

[4] The 1980 census undercount rates for black men aged 35-54 were originally estimated to be as high as 16-18 percent. However, recent work evaluating birth registration data has determined that the undercount rates for this cohort may be several percentage points lower (see Robinson, 1990).

Copyright National Academy of Sciences. All rights reserved.

used assumes that uncounted individuals represent a random sample of each age-race-sex subgroup; it does not take account of the estimated variation in coverage by other variables that are important for social welfare program modeling, such as household relationship and income.

Response Rates

Relative to many other surveys, the CPS obtains high response rates. However, some households contacted for interview—about 5 percent on average—fail to respond to the CPS, and another 9 percent of people in otherwise interviewed households fail to respond. In addition, a considerable number of people, although responding to the basic CPS labor force questionnaire, do not respond to the March income supplement. Nonresponse to the supplement is treated together with other cases of failing to answer one or more specific questions (see discussion below). To adjust for whole household nonresponse to the basic CPS, the Census Bureau increases the weights of responding households; to adjust for person nonresponse, it imputes a complete data record for another person with similar demographic characteristics. These procedures assume that respondents represent the characteristics of nonrespondents. This assumption has not been tested adequately.

In addition to household and person nonresponse, there is substantial item nonresponse in the March CPS. The Census Bureau imputes as much as 20 percent of the total income in the CPS. For some income sources, imputation rates are even higher—as much as one-third of nonfarm self-employment income, interest, and dividend payments are imputed (see Table 5-1).[5] The Census Bureau supplies values for missing income and other items through use of sophisticated techniques that find the closest match for each nonreporter in the file or use values from a similar neighboring record.[6] Even after imputation, however, estimates of recipients and amounts for many income sources in the March CPS fall short of control totals from administrative records. For example, the CPS estimate of AFDC income is only three-quarters of the estimate from program data. The Census Bureau provides estimates of net income underreporting in the March CPS but does not adjust the data in any way. The latest detailed analysis was conducted for March 1983; see Table 5-1.

[5] About half of the value of imputed income in the March CPS is attributable to people who do not respond to the income supplement at all. The proportion of all respondents with missing information for at least one income item in the March CPS has increased substantially over the past decade: from 5 percent in 1948 to 18 percent in 1978 to 28 percent in 1987 (Levitan and Gallo, 1989:14).

[6] The Census Bureau refers to its closest-match technique as statistical matching (although, in more common usage, the term is restricted to a match involving two separate data files) and to its nearest-neighbor technique as hot-deck imputation; see the Appendix to Part II for definitions of these and other technical terms.

Copyright National Academy of Sciences. All rights reserved.

TABLE 5-1 March CPS Income by Type, by Percentage Reported and Allocated [Imputed], and as a Percentage of Independent Estimates, 1983

	CPS Income			Independent Estimate	
		Percentage			
Source of Income	Amount ($ millions)	Reported	Allocated	Amount ($ millions)	CPS as Percentage of Estimate
Total income	$2,201.2	79.9	20.1	N.A.	N.A.
Total income, independent estimates	2,164.9	80.0	20.0	2,402.5	90.1
Sources with independent estimates					
Wages or salaries	1,616.3	82.1	17.9	1,632.2	99.0
Nonfarm self-employment	119.8	67.1	32.9	104.1	115.1
Farm self-employment	10.3	78.6	21.4	8.5	121.3
Social security/ railroad retirement	142.3	79.5	20.5	155.2	91.7
Supplemental Security Income	7.6	82.4	17.6	9.0	84.9
Aid to Families with Dependent Children	10.5	87.2	12.8	13.8	76.0
Interest	99.4	66.0	34.0	220.9	45.0
Dividends	27.3	66.4	33.6	60.2	45.4
Net rent and royalties	16.5	77.9	22.1	34.3	48.1
Veterans' payments	8.8	82.6	17.3	14.0	63.3
Unemployment compensation	19.7	80.9	19.1	26.1	75.5
Workers' compensation	6.6	75.0	25.0	14.1	47.0
Private pensions and annuities	34.6	76.1	23.9	54.7	63.3
Federal government and military retirement	31.8	75.7	24.3	34.9	91.2
State and local government retirement	13.3	80.3	19.7	20.5	64.7
Sources without independent estimates					
Estates and trusts	6.7	71.8	28.2	N.A.	N.A.
Alimony and child support	8.3	84.7	15.3	N.A.	N.A.
Contributions from persons not living in household	5.4	78.4	21.6	N.A.	N.A.
Other public assistance	2.4	80.5	19.5	N.A.	N.A.
All other money income	13.6	77.7	22.3	N.A.	N.A.

NOTE: N.A., not available.
SOURCE: Bureau of the Census (1989b:Table C-1).

Copyright National Academy of Sciences. All rights reserved.

The income reporting problem is complex. The survey responses are a combination of underreporting, overreporting, and misreporting errors, such as reporting a general assistance payment as AFDC or vice versa; they can originate from respondents, proxy respondents, or interviewers. Moreover, in comparing survey reports with aggregate values from administrative records, conceptual differences and quality problems with the latter information can bedevil the analysis. For example, estimates of total wage and salary income for the National Income and Product Accounts, produced by the Bureau of Economic Analysis in the U.S. Department of Commerce, include imputed amounts for food and lodging provided as part of compensation to civilian employees.

Currently, some models adjust for underreporting of some or all nontransfer income sources, but others do not. For transfer income sources, all of the models make a complete adjustment in that they simulate benefits from AFDC, SSI, and other such programs and virtually ignore the reported amounts (except, in some cases, as a factor in choosing participants from the eligible population). In creating a baseline file, the models also calibrate the simulated number of participants to accord with administrative control totals.

Sampling Error

Even though the CPS is one of the largest federal surveys of the household sector, sampling error is significant for the population of interest to models of income support programs. The sample is designed to overrepresent smaller states in order to increase the reliability of their unemployment estimates and, in the March supplement, includes a small additional sample of households headed by Hispanics. However, the sample is not designed specifically to improve estimates for low-income people or any other segment of the income distribution.

Hence, estimates for such populations as AFDC recipients, which account for less than 5 percent of the total, are based on only about 2,000 cases—not a large number to support detailed analysis. Estimates for AFDC units with earnings—a group of considerable interest to policy makers but one that accounts for less than 10 percent of the total caseload—are based on only a couple of hundred cases. CPS weighting procedures help reduce both bias and variance in the estimates, but only to a limited degree. The Census Bureau regularly publishes estimates of sampling error and methods for users to determine sampling error for particular estimates. The modelers currently do not produce estimates of variability in their databases.

Missing Detail

A very troubling set of problems involves missing detail about income and

Copyright National Academy of Sciences. All rights reserved.

family units.[7] Income support programs such as AFDC, which are designed to help people experiencing temporary as well as long spells of hardship, operate on a monthly accounting basis. However, the March CPS collects income and employment data on an annual basis, pertaining to the previous calendar year. Hence, the models must allocate income and employment variables by months across the year. Each of the major income-support program models performs this allocation, by using methods that are similar in broad outline but differ in many details (see Citro and Ross, in Volume II).

The procedure whereby the March CPS ascertains prior-year income and employment for the household members present at the time of the interview causes several problems for modeling income support programs. The survey excludes the income received during the preceding year by persons who left the survey universe (for example, through death or emigration). Moreover, by virtue of ignoring changes in household composition during the year, the survey portrays inaccurately the economic situation of many people. For example, a female-headed family in March that is classified as poor for the previous year on the basis of the woman's income alone may not have been poor if she was married for all or part of that year. The models do not attempt to address these kinds of situations.

Income support programs often limit eligibility for benefits to subgroups of household and family members (for example, an elderly person or couple living with other people). Similarly, people listed on tax returns may exclude some household and family members who file their own returns. However, the CPS provides data for traditionally defined households, primary families, and subfamilies. A major task performed by the models in processing the input data is to create recodes that identify all conceivable types of eligible subgroups (called program filing units), as best as can be done with the available information.

Several studies underscore the importance of accurately characterizing household and family relationships in modeling income support programs. For example, Ruggles and Michel (1987) found that a marked drop in the simulated participation rate for the basic AFDC program—from 90 percent in 1980 to about 80 percent in subsequent years—was largely due to a seemingly small change instituted by the Census Bureau in coding subfamily relationships on the CPS. This change added a million potentially eligible subfamilies to the AFDC population, which had much lower participation rates than other eligible units.

[7] A missing detail problem that was corrected recently relates to the available income information in the March CPS. In 1980 the number of income sources identified in the questionnaire was expanded considerably; however, the Census Bureau did not implement a revised processing system to record the income detail on the public-use files until 1988. For files prior to that year, the models had to allocate combined amounts to specific sources in order to obtain the information needed for simulating AFDC, food stamps, and other programs.

Copyright National Academy of Sciences. All rights reserved.

Data Omissions

Income support programs uniformly apply some sort of asset test in determining eligibility for benefits, but the March CPS does not obtain data on asset holdings of households. The models address this problem in a number of ways, for example, by applying an estimated rate of return to reported interest and dividend income to simulate the value of a household's financial assets (often but not always after adjusting the income amounts for underreporting). Similarly, income support programs uniformly allow some kinds and amounts of expenses, such as day care or work-related expenses, to be deducted from a household's nontransfer income to determine eligibility and benefit amount, but the March CPS does not obtain data on such expenditures. The models address this problem in several ways, for example, by estimating imputation equations for child care expenses from the Consumer Expenditure Survey or SIPP. The March CPS does not contain other kinds of information needed for modeling specific features of income support programs, such as whether a woman is pregnant with her first child and hence possibly eligible for AFDC. Currently, the models rarely address these kinds of special problems.

The March CPS does not lend itself readily to simulating the interactions of traditional income support programs (for which the filing unit does not extend beyond the household) and programs that require information on the extended family—such as child support enforcement, which requires information on both the custodial and the noncustodial families. Because the CPS is a survey of households defined as the residents at a particular address, there is no attempt to interview nonresident family members, such as absent parents or other relatives who share (or could be expected to share) economic resources with the resident household members.

Finally, the March CPS does not contain many of the variables or provide the longitudinal perspective needed to simulate linkages of income support programs with other kinds of assistance—such as job training and employment programs, transitional health and day care benefits, or child support enforcement—in which there is increasing policy interest and for which data are needed that trace people's actions over time.

Microsimulation modelers have long been aware of these various data quality problems and the possible implications for estimates of the low-income and welfare-eligible populations from the March CPS; they have generally only been able to speculate about the level of error in the estimates and the contribution to error from each source. But it is clear that the simulated eligible populations for the AFDC and food stamp programs developed from the March CPS differ from the caseload as portrayed in administrative data from the Integrated Quality Control System (IQCS) on a number of characteristics: for example, many more simulated eligible units report earnings than do program recipients. These differences may be due to several factors, including errors

Copyright National Academy of Sciences. All rights reserved.

in the IQCS data, differences in procedures and concepts between the IQCS and the March CPS, and behavioral differences among eligible units (e.g., eligible units with earnings may be less likely to participate than other units). However, their magnitude suggests that errors in the March CPS also play a role. In turn, these differences have made it difficult for the models to calibrate the simulated participant population to match administrative control totals. Comparisons of reported participants in the March CPS with the IQCS show similar discrepancies: for example, higher percentages of reported AFDC participants have positive income and earnings and comprise larger filing units in the March CPS than in the IQCS (see Citro, in Volume II).

STRATEGIES FOR TREATING MISSING AND ERRONEOUS DATA

From the above discussion, it can be seen that data quality problems in the March CPS require that the data be modified for use in social welfare policy microsimulation. Those modifications take four forms: disaggregation, imputation for item nonresponse, imputation for items not collected, and calibration or adjustment to outside control totals. For each of these modifications, there is a range of methods from which to choose, varying in cost, complexity, and the realism of their assumptions. In this section we briefly describe these methods and the experience with their use in producing suitable databases for models. We note that, of the four, the Census Bureau to date has undertaken only item nonresponse imputations. In other words, the agency has viewed its role as producing a completely filled-in record for each sample case but not as supplying additional variables or adjusting the data in light of other information.[8]

Disaggregation

Disaggregation is used when aggregated variables need to be distributed into detailed categories through an allocation procedure. Allocations can make use of very simple formulas, such as dividing an annual income amount on each record by 12, or they can invoke complicated procedures, such as those carried out in TRIM2, MATH, and HITSM, to determine monthly employment status for each adult.

Obviously, simple allocation formulas are easy to execute, but they rest on dubious assumptions about lack of variability. Data from the Income Survey Development Program 1979 Research Panel and the 1984 SIPP suggest that

[8] We have not described one other important data modification procedure that the Census Bureau carries out, namely, weighting the records to agree with population totals. The weighting process is designed to compensate for nonresponse on the part of households and individuals. It also includes several steps that attempt, inevitably with only partial success, to reduce the variance and bias in the survey estimates.

Copyright National Academy of Sciences. All rights reserved.

patterns of employment and income receipt exhibit considerable intrayear variation. In response to these findings, more complex monthly allocation procedures have been implemented in all of the major static social welfare program models. Results from the panel's validation experiment (see Cohen et al., in Volume II) indicate that the more elaborate monthly allocation procedure in TRIM2 results in aggregate estimates of units eligible for AFDC very similar to those of the simpler procedure used in TRIM2 prior to 1984. However, the effect on some characteristics of the eligible population is substantial, particularly the proportion of units with earnings: the more elaborate procedure produced 13 percent more eligible units with earnings than the simpler procedure. Lubitz and Doyle (1986) found that the more complex monthly allocation procedure implemented in the MATH model produced 7 percent more eligible food stamp units than the earlier simpler procedure.

Imputation for Item Nonresponse

When values are missing on some but not all records for one or more variables, an imputation procedure is required to supply the missing values.[9] Imputation procedures range from the very simple to the very complex. A simple procedure is to impute the mean value for all reporters to all records that are missing a particular item. A slightly more complex variant is to impute a mean modified by a stochastic error term, so as to preserve some degree of variance in the data as well as the central tendency of the distribution. A yet more elaborate variant is to impute means, with or without error terms, to categories of nonreporters. However, all such procedures rest on the strong assumptions that the nonreporters (overall or within categories) are like the typical reporter and, moreover, that the variable being imputed does not exhibit correlations with other variables that may be important in subsequent analytical use.

The Census Bureau currently applies very complex procedures, which it refers to as statistical matches, to impute values in the March CPS for whole groups of variables such as income and employment-related items.[10] The records are classified by a number of characteristics, and the record that is the best match is selected as the "donor" to supply the missing values to the record requiring imputation (the "host"). David et al. (1986) compared the Census

[9] An alternative to imputation would be to delete households with missing data items and reweight the remaining households. However, this strategy would probably greatly reduce the number of sample cases; moreover, it assumes, as in the case of imputing mean values, that the nonreporters would all have furnished the same response as the mean value for the reporters.

[10] The Census Bureau's statistical matching procedures have, over the years, replaced somewhat less complex hot-deck imputation procedures for more and more items. In the hot-deck method, the data records are arrayed by geographic area and processed sequentially, and the reported values are used to update matrices of characteristics. A record with a missing item has the most recently updated value assigned from the appropriate matrix. See Citro (in Volume II) for more details.

Copyright National Academy of Sciences. All rights reserved.

Bureau's imputations with a regression-based imputation for earnings (using IRS data from a 1981 exact-match CPS-IRS file as the measure of truth) and found that the CPS methods performed quite well in reproducing the overall shape of the earnings distribution. However, they and other analysts have determined that the CPS imputations are less successful for small subgroups, such as minorities and specific occupations (Coder, no date; Lillard, Smith, and Welch, 1986). To our knowledge, the impact of the CPS imputation procedures from the viewpoint of modeling the low-income welfare-eligible population has never been evaluated.[11]

Imputation for Items Not Collected

In some cases values are missing entirely for one or more variables for all records because the survey questionnaire did not include the needed item(s). Again, a simple procedure could be used, such as imputing a mean amount for a missing variable based on tabulations from another data source, but such a procedure assumes that the survey respondents are like the typical respondent in the other survey. It also assumes that there are no important relationships between the missing variable and other respondent characteristics.

Currently, social welfare program models such as MATH and TRIM2 most often use regression-based procedures to impute specific variables that are not contained in the March CPS records, such as child care expenses. These more complex imputation procedures are designed to reproduce the variability that one expects to find in the real world and to preserve relationships among key variables so that, for example, child care expenses are a function of number of children and work hours of the parents. However, with the mainframe-based computing environments used to operate the models and to evaluate the large data sets that provide the basis for the equations, it has proved costly and time-consuming to develop and implement those complex imputations. Hence, the typical practice has been to use the same functional form and coefficients for a number of years before making the investment to reestimate the equations and reprogram the model to use them. No assessment has been conducted of the effect of a particular type of regression imputation on the model estimates.

Exact Matches

In some cases, matching procedures have been used to obtain values for missing items, generally when large numbers of variables are involved. Exact matches

[11] Studies of the hot-deck imputations in the 1984 SIPP panel have revealed anomalous results for participants in the food stamp program because the imputation matrices for program-related variables such as benefit amount and assets did not include measures of low income or participant status. Allin and Doyle (1990) found that a sizable proportion of households that reported food stamps but were simulated to be ineligible for the program had imputations of inappropriately high assets or income values in their records.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

of two or more data sets have been performed to obtain variables from the donor file(s) that are not available on a host file. Such matches use **a** unique identifier common to both files, such as social security number. Exact matching procedures are obviously preferable to other kinds of matching and imputation because of the high quality of the resulting combined data set. Exact matches are not error free, however, and can be complex to execute due to mistakes in recording the common identifier on one or both files and the necessity **to** process large numbers of records (see Subcommittee on Matching Techniques, 1980).

One extensively used exact-match file is the 1973 CPS-SSA match, which forms the core database for the DYNASIM2 model.[12] The PRISM model builds on a 1978 CPS-SSA exact-match file, but that file was never made widely available. No exact-match files were publicly released in the 1980s, due partly to resource constraints, but more importantly to agency concerns about protecting confidentiality.[13] The requirements of dynamic models for longitudinal earnings histories mean that the lack of a more recent exact-match file places an undue burden on the models themselves to generate many years of earnings histories before even beginning their future-year projections (see Chapter 8 for further discussion of this point).

Statistical Matches

Statistical matches have also been 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 order **to** obtain variables from the donor file(s) that are not available on a host file (see Cohen, Chapter 2 in Volume II). 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. A considerable number of statistical matches have been conducted in the past two decades for use in policy estimation in areas as diverse as health care, taxation, income support, and household energy consumption in the United States.

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. Typically, the validity of a statistical match rests on the assumption of conditional independence, namely,

[12] The 1973 exact-match file also supported several useful analyses of the quality of the income reporting in the March CPS at that time; see Citro (in Volume II).

[13] The Census Bureau has performed exact matches for internal use, but not for release to outside researchers except under special circumstances. The analysis by David et al. (1986), using a 1981 exact-match CPS-IRS file, was performed when they worked at the Census Bureau as special sworn employees under a fellowship program.

Copyright National Academy of Sciences. All rights reserved.

that all of the information about the relationship between the variables (*Z*) that are unique to the donor file(s) and the variables (*Y*) that are unique to the host file is contained in the common set of variables (*X*).

Largely because of high costs, social welfare policy modelers performed statistical matches less frequently in the 1980s than previously. However, the Office of Tax Analysis (OTA) regularly performs a statistical match of the March CPS with a sample of tax returns to form the database for the tax simulation model used by OTA and the Joint Committee on Taxation. The HITSM model also uses a database that contains statistically matched records from the March CPS, the Consumer Expenditure Survey, and the SOI tax return sample. Statistics Canada has been actively engaged in various types of statistical matching to form microsimulation model databases.

Adjustment to Outside Control Totals

Even after other data modification steps such as imputation have been carried out, aggregate values estimated from a database may not match control totals from an outside source, and so the data need to be adjusted. Several methods are available for "calibrating" the data. A scaling factor can be applied to a particular variable, for example, to inflate all reported amounts of a particular income source. A more complex variant that can be used when detailed control totals are available is to apply different factors for different types of persons or households. Weights can also be modified by applying a simple or complex set of scaling factors, for example, to match population control totals and thereby adjust for undercoverage relative to the decennial census. Another kind of calibration that is used routinely in microsimulation models for income support programs is to adjust the function that selects participants from the pool of simulated eligible units on a baseline file so as to generate simulated numbers of AFDC, SSI, or food stamp cases that closely match the administrative counts.

Calibration techniques vary in complexity and cost. They also rest on a critical assumption, namely, that the control totals used are conceptually appropriate and reliably measured. But in some cases, the control totals are known to be problematic. In others, information is not readily available to assess their accuracy. Also, in making a calibration, a modeler must assume that other important relationships in the data are not being materially altered.

Results from the panel's experiment in simulating the AFDC program with TRIM2 (see Cohen et al., in Volume II) and its review of the program participation functions in TRIM2, MATH, and HITSM (see Citro and Ross, in Volume II) suggest that the calibration process should be carefully evaluated and documented for users. As noted above, it is often difficult to achieve a close approximation of the size and characteristics of program caseloads because of differences between the characteristics of simulated eligible units and those of participants as reported in the administrative records. Furthermore, such

Copyright National Academy of Sciences. All rights reserved.

differences vary from year to year, so the calibration process may introduce errors into the analysis of program participation rates over time. In general, there may be a high level of uncertainty of the simulated baseline participation rates (which are often used for simulations of program alternatives) because of errors in both the control totals and a model's simulations of eligible units.

THE PROMISE OF SIPP

The Survey of Income and Program Participation traces back to an interagency committee, sponsored by the Office of Management and Budget in the early 1970s, that reviewed the deficiencies in the personal and family income statistics derived from the March CPS income supplement and recommended that a research program be developed on the best way to improve the collection of income data from households. Policy analysts and researchers in the Social Security Administration and ASPE provided impetus for a new income survey that would furnish improved data for decision making. Experience with the new microsimulation models that were being developed at that time to evaluate alternative designs for government tax and transfer programs underscored the problems with the March CPS data on income and program participation and gave a further push to the need for a new survey to serve as a database for modeling and policy analysis generally.

In 1975, ASPE initiated a major testing and research program, the Income Survey Development Program (ISDP), to help with the design of the new survey. In fall 1983, after surmounting several setbacks along the way, the first interviews were fielded by the Census Bureau for the initial 1984 SIPP panel. Subsequently, new panels were introduced in February of each year, with the people in each panel followed over time and interviewed at 4-month intervals for a total of about 2-1/2 years. SIPP was designed explicitly to remedy many of the deficiencies of the CPS March income supplement: for example, the SIPP questionnaire obtains monthly measures of income and household composition and periodic measures of assets. To this end, the design for the survey is highly complex and has many innovative features (see Committee on National Statistics, 1989).

Evaluation studies show that SIPP provides higher quality data on many dimensions than does the March CPS. For example, item nonresponse rates for income amounts are considerably lower in SIPP for most income sources; SIPP data are also closer to administrative control totals for many income sources, although underreporting is still evident in most cases. By design, SIPP also provides many more of the variables needed for modeling and evaluating government tax and social welfare programs than does the March CPS, and SIPP provides needed data not elsewhere available to analyze and model the short-term dynamics of program participation and movement into and out of poverty.

Copyright National Academy of Sciences. All rights reserved.

However, on some dimensions, SIPP is proving no better, or even worse, than the March CPS. First, the population undercoverage rates in SIPP are as high as in the CPS. Second, income underreporting and misreporting are still present in SIPP, and the discrepancies in characterizing the welfare population that are evident in comparing the March CPS with administrative data also exist in SIPP. Third, some needed variables that are not available in the March CPS, such as the composition of the extended family, are not provided by SIPP, either. Fourth, the Census Bureau has had serious problems with processing and releasing SIPP data in a timely fashion: release of public-use data files has lagged as much as 3 years behind data collection. Moreover, users have experienced considerable difficulty in processing the complex SIPP data files. Finally, the sample size in SIPP was smaller than the CPS to begin with—20,000 households in each panel—and because of budget cuts, the Census Bureau both progressively reduced the SIPP sample size to about 12,000 households in each panel and cut back the number of interviews.[14]

The Census Bureau has under way several activities to overcome some of the deficiencies in SIPP. For the 1990 SIPP panel, the Census Bureau increased the sample size to 21,500 households, although at the cost of discontinuing the 1988 and 1989 panels in midstream. Concurrently, the administration was seeking funding to restore the original design of overlapping panels introduced each year, beginning with the 1991 panel, and, over the next few years, to restore the SIPP sample size for all panels to 20,000 households. (Through an appropriation together with an internal budget reallocation, the Census Bureau was able to field a 1991 panel of 14,000 households.) The Census Bureau is also striving to reduce the time lag in releasing data, with the ultimate goal of making files available within 12 months of data collection. It also intends to restructure the cross-sectional files to facilitate their use.

Looking to the longer term, the Census Bureau is planning to redesign the SIPP sample to overrepresent poverty households on the basis of results of the 1990 census. The new design will be implemented in 1995, at the same time as revised sample designs for the CPS and other household surveys based on the census are introduced. At that time, the Census Bureau will make other changes to the SIPP questionnaire, design, processing system, and other characteristics that are recommended from its own evaluation studies and those by outside groups.

The panel concludes that SIPP provides a wealth of information for policy analysis and research. However, as discussed below, we do not favor immediate action for it to replace the March CPS as the primary database for modeling income support and other government programs. We are even doubtful that such

[14] By combining SIPP panels, users can obtain larger sample sizes. However, even with the size of each panel originally at 20,000 households, attrition over time has meant that users would be unlikely to obtain any more than 35,000 households by combining data from the second year for one panel and the first year for another.

Copyright National Academy of Sciences. All rights reserved.

a strategy would be wise in the longer term. We believe that the policy analysis agencies and the Census Bureau should carefully review the best allocation of the available funds for producing the most useful databases for the government's research and analysis needs. We argue the merits of a strategy that builds on the strengths of both the March CPS and the SIPP and also brings to bear relevant information from administrative records and other data sources.

RECOMMENDATIONS FOR IMPROVING DATA QUALITY

We have analyzed in detail the problems of data quality in the March CPS income supplement that confront the major microsimulation models of income support programs. A similar analysis of the survey and administrative input data used by retirement income models, the administrative data used by many tax models, and the various data sources used by health care policy models would undoubtedly turn up a long list of data quality problems as well (see Chapter 8).

Very little work has been done to evaluate the impact on model estimates of the problems with the input data, either before or after steps are taken to adjust the data. However, the studies that have been done, including the panel's own validation experiment (see Cohen et al., in Volume II), suggest that the impact is significant. Moreover, because so many kinds of data problems require data adjustments and because so many different data modification techniques are implemented, we are concerned that the resulting model databases may be rife with internal inconsistencies that have deleterious effects on the quality of the simulation estimates. We argue in Chapter 3 that investments in data quality are sorely needed to improve the payoffs from all social welfare policy analysis and research. We conclude even more strongly that such investments are specifically needed to improve the payoffs from microsimulation modeling. By making input data more suitable for modeling purposes, there is the opportunity both to improve the quality of model estimates and to reduce the cost and complexity of the models themselves.

Evaluating the Quality of Existing Data Sources

Survey Data

The necessary prelude to an investment in data quality is an investment in understanding the problems and limitations of the existing data. As a corollary, there must be an investment in making available what is known about data quality in a manner that both informs the agenda for long-range quality improvements and provides guidance in using the data appropriately in the near term.

Copyright National Academy of Sciences. All rights reserved.

What is termed a "quality profile"—namely, a document that brings together what is known about all the sources of error that may affect the estimates from a survey or other data collection system (see Bailar, 1983—offers a very effective way to describe the results of data evaluation and to design a continuing program to fill in knowledge gaps. Not many such profiles (sometimes termed error profiles) have been prepared by federal statistical agencies, no doubt due to the difficulty in assessing sources of error other than sampling variability and to the cost and time required to collate separate evaluation studies. The profiles that do exist have proved invaluable to agencies and data users alike.

The first quality profile—and still a landmark in the field—was that prepared in the late 1970s to describe and assess the sources of error for the employment and unemployment estimates based on the monthly CPS (Brooks and Bailar, 1978). The 1970s also saw considerable research on the quality of the earnings data in the March CPS (see Citro, in Volume II). However, no complete quality profile was ever developed for the March CPS, and very little research has been conducted, certainly not in recent years, on the quality of income and program-related data. Although the outlines of many of the problems in the March CPS, such as undercoverage and underreporting, are known, we are not aware of any thoroughgoing evaluation of their import for use in modeling income support programs.

In contrast, SIPP, as a new survey that incorporates many novel design features, has had a relatively high level of resources devoted to identifying and assessing the magnitude of its data problems and to experimenting with changes in questions and procedures that could improve data quality. Outside researchers as well as Census Bureau staff have conducted a number of analyses of SIPP data quality. The results of the evaluations to date have been summarized in the SIPP *Quality Profile,* which, now in its second edition (Jabine, King, and Petroni, 1990), sets a standard for the field.

We applaud the effort that has gone into evaluating and documenting SIPP data quality. We believe that a similar effort is required for the March CPS. Given the limitations of the current SIPP sample size and the problems of obtaining and working with the complex SIPP data files, the March CPS will undoubtedly remain a primary data source for microsimulation modeling and other kinds of policy analysis for the near term. For the longer term, the March CPS may well continue its primary role and, in any event, it will continue to be used for many policy estimates. Allocating resources for an in-depth evaluation of the quality of the March CPS data is therefore an urgent need.

Such an evaluation will not be straightforward or easy to conduct. To take just one example, reports—or imputations—of welfare receipt by households that appear to be ineligible for benefits on the basis of their annual income may not be incorrect. These households may have had one or more months of sufficiently low income to qualify for benefits in those months, but these poor spells would not be observed in the March CPS. A record-check study of the

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

type being performed for SIPP, in which reports of benefits are checked against administrative data, could perhaps also help in evaluating the quality of CPS income reporting.

We note that the Census Bureau has recently begun a program designed to obtain improved estimates of the income distribution for the U.S. population (see further discussion below). The program has two main goals: first, to evaluate and construct error profiles of the quality of the income data in the March CPS, as well as the SIPP, by using data from administrative records and other sources; second, to develop methods for adjusting the income data in these two surveys. If sufficient resources are allocated for timely implementation of this program, it will generate much of the information that we believe is required to understand the implications of data problems for policy applications and to plan corrective actions.

***Recommendation 5-1.* We recommend that the Census Bureau evaluate the Current Population Survey March income supplement in its role as a primary source of data for analysis of the income distribution and economic well-being of the population. The evaluation should be designed with input from the policy analysis agencies that are major users of the data. It should be comprehensive, covering the impact on data quality of every stage of data collection and processing. It should also compare the March CPS estimates with estimates from other sources. The results should be brought together in a quality profile that is published for users and updated periodically as further evaluations are conducted and new findings obtained.**

Administrative Data

Administrative records, as well as survey data, play an important role in microsimulation modeling. Data from program administrative records are commonly used as control totals for calibrating baseline simulations, and they also provide the databases for program benefit calculators. For example, data about the characteristics of welfare recipients from the IQCS play a crucial role in calibrating the survey databases used in models of income support programs and also serve as databases for benefit-calculator models of the AFDC and food stamp programs. Moreover, for some kinds of models, such as those for taxes and retirement income programs, administrative records provide essential data that are not available elsewhere. One heavily used administrative data source is the Statistics of Income samples of federal income tax returns (discussed in detail in Chapter 8).

Just as relatively little investment has been made in evaluating the quality of the March CPS in recent years, the quality of the information in administrative

Copyright National Academy of Sciences. All rights reserved.

records systems has not been assessed rigorously. For example, the quality of IQCS data, which are drawn from large monthly samples of AFDC, Medicaid, and food stamp case records, is believed to be problematic in some respects. (See Food and Nutrition Service, 1988, and Family Support Administration, 1988, for descriptions of the IQCS information on, respectively, food stamp and AFDC beneficiaries. The primary purpose of the IQCS is to provide measures of errors in eligibility status and payment amounts for the programs in each state.) Wide discrepancies exist between the IQCS and the March CPS in information on characteristics such as the relationship of the AFDC unit head to the household head. The nature of the discrepancies suggests that the problem may lie more with the former than the latter source. The IQCS should be reviewed **to** identify problem areas and the consequent implications for microsimulation model estimates. It would be useful in this regard for the major user agencies, including ASPE and FNS, to coordinate their evaluation efforts. However, the different ways in which they currently extract and use the IQCS data for their models—for example, ASPE works with a year's worth of data from the IQCS, while FNS works with data for selected months—may be an impediment.

Recommendation 5-2. **We recommend that the responsible agencies sponsor in-depth evaluations of the quality of administrative data that are used as primary or supplemental inputs to social welfare policy microsimulation models. Such data sets include the Integrated Quality Control System samples on the characteristics of welfare recipients and the Statistics of Income samples from federal income tax returns. The results of each evaluation should be brought together in a quality profile that is published for users and updated periodically as further evaluations are conducted and new findings obtained.**

Improving Databases for the Near and Long Term

After databases for microsimulation modeling have been evaluated, the next step in improving data quality is to seek ways to eliminate or compensate for important errors. Ideally, data quality problems would be addressed at the source, that is, through revision of survey questionnaires and data collection procedures to provide the needed variables with the required level of quality. Indeed, over the past two decades, numerous improvements were effected in the CPS March income supplement to accommodate the needs of modelers and policy analysts generally. Moreover, SIPP was launched to provide a more appropriate vehicle than the CPS for social welfare program analysis.

However, investment in data collection has greatly lagged behind the needs in the past decade, and, in many cases, there has actually been disinvestment

Copyright National Academy of Sciences. All rights reserved.

in needed data. A prime example concerns SIPP, for which the sample size was cut back repeatedly so that it cannot be considered for use as the database for modeling income support programs on a regular basis.[15] The CPS sample, after increasing by about 20,000 households from 1967 to 1981, was cut back subsequently by about 15,000 households (Levitan and Gallo, 1989:3). Still another example is the long delay in revising the public-use microdata file processing system needed for making available the enhanced income detail that was added to the March CPS questionnaire.

Additional resources for improving data quality at the source would clearly be most welcome. Given the high expenses associated with field work for original data collection, however, it is not likely that sufficient resources could ever be obtained to alleviate all of the important problems at the data collection stage, nor is it even feasible to effect all needed improvements at that stage. Considerations of respondent burden alone would preclude an attempt to build into one single data collection instrument, whether SIPP or the March CPS, all of the information required for modeling social welfare programs. Hence, it will remain incumbent and indeed cost-effective to make many needed data improvements subsequent to data collection, through such techniques as imputation and matching from other sources.

It is imperative, in our view, to keep in mind that the goal of federal statistical activities is to generate the best data possible for policy analysis and other important purposes. Given inevitable resource constraints and burden limitations, achieving this goal will almost always require looking beyond the confines of a particular survey and seeking to relate data from multiple sources, including administrative records. Hence, the challenge in allocating budget resources for enhanced data quality is to achieve the best possible balance among spending for additional data obtained through surveys; spending for additional data obtained from administrative records and other sources; spending for better measurements from surveys and administrative systems, through improvements to questionnaires and procedures; and spending for better databases, through improved techniques for combining data from multiple sources.

Near-Term Approaches

Turning to the near-term outlook for the quality of available data for modeling social welfare programs, we are very encouraged that the administration is seeking additional funding to effect a range of improvements to federal statistical

[15] However, there have been some modeling applications of SIPP. Mathematica Policy Research, Inc., recently completed development of a microsimulation model of the food stamp program (FOSTERS) that combines data from the 1984 and 1985 SIPP panels. Mathematica is also developing an updated version of the model using combined data from the 1986 and 1987 SIPP panels. The Social Security Administration is also planning to develop a model of the SSI program with the 1984 SIPP panel.

Copyright National Academy of Sciences. All rights reserved.

programs. We are concerned, however, that the funding for income and program-related statistics may not be allocated in the most effective manner. The biggest new budget request in the income area is to restore the original SIPP design by increasing the sample size for each panel to 20,000 households and reinstituting overlapping panels. Once the restoration is complete—it is scheduled to be phased in over the next few years—the annual budget for SIPP will be about $27 million, compared with current spending of $20 million. In comparison, the entire CPS costs about $28 million per year, of which perhaps $2-3 million represents the marginal cost for the March income supplement.[16] But, as detailed above, it is clear that microsimulation models will need to continue their reliance on the March CPS in the near term as a primary source of input data; SIPP cannot be competitive as a modeling database until at least the mid-1990s. Thus, we question the decision to allocate additional funding to more panels in SIPP at this time. An alternative strategy, which might have higher payoff in terms of the uses of the data, would be to retain the 1990 SIPP design and allocate the added resources to a combination of initiatives (including use of SIPP and administrative records data) designed to improve the quality and utility of the March CPS database.

Under this alternative strategy, the large 1990 SIPP panel of 21,500 households would be completed and another large panel begun in 1992 and again in 1994. (Most likely it would be necessary to discontinue the smaller 1991 panel.) If each of these panels comprises the full eight waves, the 1990-1992 and 19921994 panels would overlap for two interviews, which would be advantageous for producing a consistent time series on income and program participation. In addition, just as the 1990 panel contains selected cases from the 1989 panel in order to improve estimates for the low-income population, the 1992 and 1994 panels might include some cases from a preceding panel for the same purpose. This design for SIPP would preclude combining panels to obtain a sample size of perhaps as many as 35,000 households. However, each of the 1990, 1992, and 1994 panels, like the original 1984 panel, would be of sufficient size for many kinds of useful analyses, including studies that could allow improved imputations in the March CPS.

In place of introducing new SIPP panels on an annual basis, the Census Bureau, together with the policy analysis agencies, should consider several possible steps, such as:

- adding a low-income sample to the March CPS, which could be done readily in the same manner as the current Hispanic supplement by using cases from earlier months, thereby directly enhancing the utility of the data for microsimulation modeling and other analyses of income support programs;

[16] Cost estimates for SIPP were provided to the panel by Daniel Kasprzyk of the Census Bureau (July 1990); cost estimates for the CPS are from Levitan and Gallo (198 $1$2$3).

Copyright National Academy of Sciences. All rights reserved.

- adding a limited set of questions to the March CPS to ascertain family composition during the income reference year, which would permit assessing the impact of the current fixed measurement of family composition on the quality of the income data and facilitate cross-comparisons with SIPP information;
- adding a limited set of questions on income during the last month (or, possibly, the last December) to the March CPS, which would provide some (albeit limited) information on intrayear income variability and, more important, facilitate the use of SIPP intrayear income data to improve the allocation of the CPS annual data to monthly values;
- experimenting with changes in interviewer instructions to try to raise the dismayingly low response rates to the income questions (at present, CPS procedures accord priority to obtaining responses to the core labor force questions and to retaining households in the sample for the next month);
- exploiting the longitudinal information available in the CPS to improve the quality of the data in the March income supplement (employment information is available from the previous 1-3 months for some portion of the March sample, as is income and employment information from the previous March);
- exploring sophisticated imputations using SIPP data to improve the March CPS information on intrayear income, employment status, and other variables; and
- exploring matches of administrative data with SIPP and the March CPS in a form that can be made publicly available.

Adding a low-income sample and a few questions to the March CPS are steps that could be taken relatively quickly, as could limited experimentation with changes to interviewer instructions or other procedures to try to improve response. Given sufficient resources and priority attention, the development of SIPP-based imputations to improve the March CPS data could also be accomplished within a reasonably short time, as could making use of the longitudinal information in the CPS. Conducting matches of SIPP and March CPS data with administrative records, in our view, has perhaps the greatest potential to improve the quality and utility of the information in both surveys. For example, adding information on earnings histories from social security records and on tax returns from IRS records would greatly expand the policy relevance of the survey data, and incorporating administrative reports of earnings and property income would greatly improve the accuracy of the survey data.[17] However, the use of administrative data raises formidable problems of obtaining the records and working out the very tough problems involved in making the

[17] Since 1978, the Social Security Administration has collected information on individuals' total annual earnings, including earnings above the social security payroll tax limit. The SIPP currently includes a module that collects tax return data; however, most of the tax-related variables are not available to outside researchers (see further discussion in Chapter 8). Incorporating administrative reports **of** transfer income would also be highly beneficial; however, in addition to problems of confidentiality, obtaining such reports from all **50** states poses substantial administrative difficulties and costs.

Copyright National Academy of Sciences. All rights reserved.

information accessible to users. We urge that work begin immediately on tackling these problems, although we recognize that they are not likely to be solved in the near term.

Recommendation 5-3. **We recommend that the Census Bureau, in conjunction with policy analysis agencies, immediately evaluate alternative options for short-term improvements to the data used for microsimulation modeling, and policy analysis generally, of income support and related social welfare programs. Alternatives that should be investigated include:**

- **proceeding with the current plan to obtain added resources to restore the SIPP sample size and overlapping panels, beginning with the 1991 panel; and**
- **keeping the SIPP budget at its current level with the 1990 design of fewer, larger panels, while reallocating the added budget to some combination of initiatives, including adding a low-income sample to the March CPS; adding a limited set of questions to the March CPS to ascertain family composition during the income reference year; exploiting the longitudinal information available in the CPS; exploring sophisticated imputations that use SIPP data to improve CPS information on intrayear income, employment status, and other variables; and exploring matches of SIPP and CPS data with administrative records in a form that can be made publicly available.**

Long-Term Approaches

Looking to the longer term, we note that there are plans that could materially affect the role of the March CPS and the SIPP as databases for microsimulation modeling and other policy analyses of social welfare programs. The Bureau of Labor Statistics proposed a large expansion of the CPS sample in order to support monthly unemployment estimates for every state, to be implemented in the late 1990s after the sample is redesigned following the 1990 census (Butz and Plewes, 1990). If this expansion was applied to the supplements as well, the utility of the March income supplement for microsimulation modeling and other analyses requiring large sample sizes would increase markedly. Other planned enhancements to the CPS that could benefit microsimulation modeling include use of the longitudinal information to improve monthly labor force status reports and questionnaire experimentation.[18]

Concurrently, the Census Bureau is sponsoring several studies to evaluate

[18] The administration was not successful in obtaining fiscal 1991 funding to plan for expansion of the CPS sample; hence, work has stopped. However, the concept could be revived at a later date.

Copyright National Academy of Sciences. All rights reserved.

the SIPP design and recommend any needed changes to be implemented when the SIPP sample is redesigned on the basis of the 1990 census. Tentative decisions have already been reached with regard to the sample redesign for SIPP, namely, that the sample should disproportionately select low-income households. SIPP was originally intended to support analysis of both tax and transfer programs for the entire range of the income distribution, but the processing difficulties experienced by the Census Bureau and the overall complexity of the survey have led, in recent years, to the decision to focus the survey on data needs for lower to middle-income groups. At the Census Bureau's request, the Committee on National Statistics recently appointed a panel to assess the appropriate use of SIPP longitudinal data for policy analysis purposes and to consider the best design for the survey to serve the goal of providing improved income statistics for the nation. The Census Bureau is also conducting studies of questionnaire content and other aspects of the program.

We believe these are important efforts. We believe that all aspects of the SIPP design need to be reconsidered with a fresh perspective, including:

- whether and what type of overlapping panel design is desirable. An overlapping design may be needed to support reliable time series on income (because of the biases from attrition over the course of a single panel). However, given that sample size must likely be traded off for frequency of panels, it may be that fewer, larger panels would be better suited to the needs of microsimulation and policy analysis of income support programs, which require adequate observations on subgroups of the low-income population.
- the length of each panel. Some analysts have wanted SIPP to follow panels for more than 28-32 months. Certainly, the longitudinal information on intrayear income dynamics afforded by SIPP is invaluable for research purposes and could, in the long run, support new kinds of microsimulation models that simulate the dynamics of intrayear participation in social welfare programs. However, increased panel length usually entails decreased data quality due to attrition. It may be that not every panel needs to run even as long as 28 months and that a better use of scarce resources could be to have some shorter but larger panels.
- the way in which SIPP can best be designed to facilitate linkages with the March CPS and administrative records, particularly to improve the capability for modeling and analysis of the full range of the income distribution and of the joint impact of tax and transfer programs on the population. We cannot overstate our belief in the advantages for modeling and other kinds of policy analysis of developing integrated databases that incorporate administrative records and survey information. We see consideration of ways to carry out such linkages in a manner that permits access to the data while protecting their confidentiality as a high priority task for the Census Bureau's review of the SIPP program.

Copyright National Academy of Sciences. All rights reserved.

Recommendation 5-4. **For the longer term, we note that the Census Bureau now has studies under way to consider the future design of SIPP. We recommend that these studies focus on improving the databases for modeling and analysis of income support and related social welfare programs. We recommend that the studies review all aspects of the SIPP design (such as the sample size and length of each panel and the extent to which overlapping of panels is desirable) and consider how best to design SIPP to facilitate relating data from the SIPP, the March CPS, and administrative records.**

Within the next 2 years, the various studies that are under way of the SIPP design and the CPS improvement plans should be completed and their recommendations known. At that time, the policy analysis agencies will be in a position to plan a future course for microsimulation models of tax and transfer programs. In our expectation, it is likely that the March CPS, particularly if the sample is expanded, will remain a key source of input data for microsimulation models. Hence, continuing attention will be needed to issues of content, design, and data quality of the March CPS, instead of assuming, as in the recent past, that the CPS is familiar territory and that all of the effort on evaluating and improving income data should be directed to SIPP.

We also expect that the SIPP will be placed on a firm footing, available design adopted, and the processing problems that have plagued the survey overcome. In that case, we expect to see increased use of SIPP for policy analysis and, particularly, for improvements to model databases provided by the March CPS.

The availability of greatly enhanced computer hardware and software technology that supports added power, flexibility, and user accessibility at reduced cost (see Chapter 7) may make it possible to pursue two tracks for model development. The agencies could redesign the current CPS-based models to improve their cost-effectiveness and retain them as the workhorses for policy estimation requiring large samples for reliable analysis of detailed subgroups. At the same time, the agencies could experiment with new models that exploit the longitudinal nature of the SIPP to explore program dynamics and investigate special issues. In any event, we urge the policy analysis agencies to prepare to move forward rapidly once the outlines of the future data systems in the area of income and program participation are clear.

Recommendation 5-5. **After current studies of SIPP and the CPS are completed, we recommend that the policy analysis agencies plan to redesign their income-support program microsimulation models to make best use of the improved data on income and related subjects that should be available after 1995.**

Copyright National Academy of Sciences. All rights reserved.

Expanding the Census Bureau's Role

The discussion to this point has concerned strategies for improving the input data for microsimulation modeling of social welfare programs. We now consider the question of the locus for carrying out various improvement operations. Of course, there will continue to be data processing steps required for modeling programs such as AFDC and food stamps that only the modelers can implement, and it is important that they continually seek ways to evaluate and improve the data components of their models. Indeed, we believe that a very high priority for microsimulation modelers is to evaluate thoroughly the processes they undertake to generate input databases, including such steps as calibrating to control totals. They need to assess the quality of the resulting output and determine ways to make improvements in quality. (Chapter 9 presents the program of model validation and assessment of sensitivity and variance that we recommend, including documentation of the impact of each stage of data processing on the model estimates.) However, we believe that it is time to give serious consideration to enlarging the role of statistical agencies in generating data files that are useful for analysis.

The Census Bureau currently has responsibility for the two major surveys that provide data for modeling and analysis of social welfare programs—the March CPS and the SIPP—but it is not charged with detailed policy analyses of these data. Although it has made a number of changes in the two surveys to respond to the data needs of policy analysis agencies, the Census Bureau has not seen its job as preparing analytical databases, distinct from survey files. In other words, the Census Bureau has concentrated on such tasks as weighting and imputation for nonresponse, leaving to the users the tasks of further processing the survey data, such as correcting income reporting errors, imputing needed variables such as asset holdings or expenditures, and adjusting the data to match administrative control totals.

We believe that the policy analysis agencies as well as the Census Bureau itself could benefit from a division of labor whereby the Census Bureau performed more of the steps involved in turning survey files into usable databases. The research community would benefit as well. Having the Census Bureau perform such tasks as adjusting income amounts for underreporting could result in cost savings (when the total data processing costs of both the Census Bureau and the agencies' contractors are considered) and also promote consistency across model estimates. (The data files should, of course, contain the amounts before adjustment so that outside analysts can evaluate the adjustment procedures used and modify them if desired.) Moreover, the Census Bureau is much better placed to make adjustments to survey information because of its greater access to administrative records sources that can be used to evaluate and inform

Copyright National Academy of Sciences. All rights reserved.

such adjustments.[19] Finally, increasing the dialogue between policy analysts and survey statisticians should benefit the design of the models, the design of the surveys, and the design of the databases that relate the models and the surveys.

We recognize that it will not be easy to implement a major change in the relationship of the federal policy analysis and statistical agencies. The analysis agencies will have concerns as to whether the Census Bureau can provide enhanced databases in a manner that is timely and that addresses the needs of particular programs. On its part, the Census Bureau will have concerns about whether meeting agency requirements for enhanced databases will adversely affect its responsibilities for original data collection and processing. Nonetheless, we urge that a dialogue be started.

We are encouraged in this regard by the Census Bureau's recent decision to begin exploring ways to relate data from the March CPS, SIPP, and administrative records with the goal of producing improved statistics for the full range of the income distribution.[20] The Census Bureau has asked the new Committee on National Statistics' panel on SIPP to review its plans for implementing this concept, which moves away from the notion that the goal is to publish income (or other) statistics from a particular survey such as the March CPS or SIPP and, instead, looks to the goal of publishing the best set of income statistics from all available sources.

Briefly, the Census Bureau proposes to use administrative records and other sources to assess the extent and nature of income reporting errors in the March CPS and SIPP. The next step would be to adjust the SIPP information, perhaps through use of some kind of multivariate imputation or weighting technique. Alternatively, exactly matched administrative values, perhaps with random noise added, might replace the reported amounts in the SIPP records if the problems of data access can be worked out. Then, the adjusted SIPP data would be used to improve the quality of the income data from the March CPS through a related imputation or modeling procedure. The CPS data would retain the advantages of sample size and timeliness (if adjustments were made from an earlier SIPP file); while the later SIPP data would provide the advantage of additional subject detail.

This project (which implicitly assumes that the March CPS and SIPP will continue to coexist, rather than the latter replacing the former as the primary source of income information) has enormous implications for the quality of databases available for microsimulation modeling. Work has already begun on

---

[19] Statistics Canada has devoted considerable resources to developing microsimulation model databases using techniques such as imputation and statistical matching that permit access to the records but are informed by exact matches of administrative and survey data carried out within the agency.

[20] The Census Bureau's project to evaluate and adjust for nonsampling error in the March CPS and SIPP income reports was described by John Coder of the Census Bureau in a presentation to the Committee on National Statistics' Panel to Evaluate the Survey of Income and Program Participation (July 30, 1990).

Copyright National Academy of Sciences. All rights reserved.

the first step of developing error profiles for income reporting in the March CPS and SIPP. However, the project faces formidable obstacles in terms of obtaining necessary administrative records and devising ways to make adjusted SIPP and CPS data available to users, in addition to the inevitable problems of developing, testing, and implementing new types of adjustment procedures.[21]

***Recommendation 5-6.* We recommend that the Census Bureau assume a more active role in adding value to databases for modeling and research purposes and for generating published data series. In particular, we recommend that the Census Bureau seek to produce the best estimates of the income distribution and related variables, such as household and family composition. Steps necessary to achieve this goal include evaluating income reporting errors in the SIPP and March CPS, on the basis of administrative records and other information sources, and using data from multiple sources to develop improved estimates.**

At the present time, very few resources are available at the Census Bureau for moving the project forward. We urge the Census Bureau to seek support **to** obtain adequate resources and priority attention for this important project.

Another way in which we believe the Census Bureau could and should act to improve the quality of household survey data used for microsimulation modeling of social welfare programs concerns population undercoverage. All of the household surveys, including the CPS, SIPP, and others, are plagued by net undercounts of the population that are substantial overall and very high for some subgroups such as minorities, people not members of nuclear families, and the low-income population. The sample weighting procedures that adjust for the undercount in the surveys relative to the decennial census do so on only a few dimensions, and they do not adjust at all for the undercount in the census itself. The panel's validation experiment, which implemented a very crude correction for the census undercount to the March 1984 and 1988 CPS (see Cohen et al., in Volume II), found that AFDC participation rates dropped by about 4 percent in each year because the adjusted database generated more eligible units.

In Chapter 3 we recommend that a high priority task be to assess the implications of coverage errors in censuses and surveys for policy analysis

[21] As an example of problems in obtaining administrative records, the Census Bureau currently is able to use only a limited set of IRS tax return data. The Census Bureau has had access to IRS data on tax filing status (joint or single return) and income reported from wages and salaries, interest, and dividends for many years. (The data were originally made available to support development of small-area population and income estimates for the General Revenue Sharing Program.) Recently, after several years of negotiation, the Census Bureau obtained permission to receive IRS data on total income reported on each return. Further lengthy negotiations, with no promise of success, will be required to obtain access to additional information that the Census Bureau would like for this project, such as private pension income reported on tax returns.

Copyright National Academy of Sciences. All rights reserved.

and research purposes, using data from the 1990 census coverage evaluation program and other sources. We reinforce this recommendation here in urging that the Census Bureau work with the policy analysis agencies and their modeling contractors to carry out an evaluation of the effects of coverage errors in household surveys for microsimulation model estimates. Should important effects be determined, we urge the Census Bureau to develop ways to implement coverage error adjustments in the March CPS, SIPP, and other surveys that the models use and to improve the procedures that are currently implemented to adjust for the higher undercoverage in household surveys relative to censuses.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

6

Model Design and Development

We begin our consideration of the structure and design of the models themselves with a strong belief in the importance of following good design principles and practices in the development of policy analysis tools that are intended to be used many times in response to the changing requirements of the policy process. Skimping on resources for model design and testing in the development stage is likely to be a classic "penny-wise, pound-foolish" decision that comes back to haunt the model users in the application stage, particularly for highly complex microsimulation models. A poorly designed model may not perform all of the functions that were intended, may entail higher-than-expected costs of operation, may be inordinately difficult to modify or to understand, or may experience other difficulties that require analysts to scale back their expectations or invest resources in an effort to patch up the problems after the fact.

We thus begin this chapter by outlining the design principles and practices that are appropriate for microsimulation models as a general class, and we briefly review the history and current status of microsimulation model development in the United States from the viewpoint of good and bad design practices. We then turn to the substantive issues involved in determining the kinds of capabilities that might be added to existing microsimulation models or included in new models. We discuss the issues and trade-offs involved in the choice of aging strategies for microsimulation models, in the decision of whether and how to incorporate first-round behavioral responses, and in the decision of whether and how to model second-round responses. We do not make specific recommendations about preferred approaches to the modeling task in any of these areas because there is an inadequate body of knowledge evaluating

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

the impact of particular design choices on the quality and utility of model outputs.[1] In the absence of such evidence, it would be foolhardy and arrogant to attempt to dictate the future course of microsimulation model development. Our recommendations, instead, lay out an agenda for research and evaluation that we hope will result in the information needed to make wise choices for future model development.

One concern that is common to all the issues we address is when a microsimulation model becomes "too big": When do the benefits from additional capabilities sink under the weight of added costs for model development and use—when costs include not only staff and computer resources but also reduced capability for timely response, reduced flexibility, and reduced accessibility of the model? The issue of how wide a range of program interactions to simulate poses this concern as well.

Historically, policy analysts have asked microsimulation models, as a bedrock capability, to estimate the direct effects of proposed program changes on individual decision units, that is, the effects if one assumes no changes in people's behavior in response to the program (other than to accept or reject the program's benefits). Analysts have also typically asked models to estimate interactions among closely related programs, such as SSI, AFDC, and food stamps. At times, analysts have asked models to provide additional capabilities, such as the following:

- the ability to project, inside the model, the cost and population effects of a program change when a new policy is expected to be implemented and for some number of years into the future (referred to as the process for "aging" the data);
- the ability to estimate, inside the model, the likely extent and effects of program-induced behavioral responses, such as the likelihood that people who are ineligible for income benefits will behave so as to become eligible (e.g., by quitting their jobs in order to obtain AFDC and Medicaid benefits) or that households will change their investment patterns in response to tax incentives;
- the ability to estimate, inside the model or through links to other models, so-called second-round effects, that is, the longer run effects of program changes—for example, the effects on wage rates of labor supply responses to welfare program changes, or the effects on health care providers of changes in the demand for health care services induced by mandated changes in the proportion of costs that patients have to pay themselves; and

[1] In general, we also do not discuss the detailed approach of particular microsimulation models to specific aspects of the models' operations. However, we note considerable variability in the way in which similar functions are implemented in different models due to factors such as differences in client interests and modeling "style." Citro and Ross (in Volume II) describe some of the differences among three static models, and Ross (in Volume II) describes some of the differences between two dynamic models. The panel's validation experiment with TRIM2 (see Chapter 9 and Cohen et al., in Volume II) found important effects of different approaches to such operations as aging the data.

Copyright National Academy of Sciences. All rights reserved.

- the ability to model interactions among a wide variety of programs across several issue areas, such as income support, taxation, retirement income, and health care benefits.

MODEL DESIGN PRINCIPLES AND PRACTICES

The majority of microsimulation models today incorporate some, but far from all, of these capabilities. For example, most models analyze interactions among multiple programs (although, typically, within just one or two issue areas), but few of them analyze second-round effects or, indeed, first-round behavioral effects other than program participation choices. Yet, in simply modeling the complex provisions of the nation's income support programs or tax codes, today's microsimulation models are already very complicated and difficult for anyone other than experienced analysts to use or modify to respond to changing policy concerns.

Design Principles

The complexity that reflects the real world of policies and individual circumstances is an inherent feature of microsimulation modeling, so it is especially important to minimize unnecessary complexities. We believe models should follow design strategies that enable them to meet the three criteria that we believe are essential for their cost-effective use in the policy process:

- flexibility and timeliness in responding to emerging policy needs;
- understandable outputs that can be assessed in terms of their quality; and
- tolerable costs of model development, use, and modification.

These requirements argue strongly for a design approach that has the following characteristics: clear analytic goals, self-contained modules, facilitated documentation and evaluation, linkages, computational efficiency, and accessibility to users.

Clear Analytic Goals

It is important that the main analytic goals of a model are well thought-out and that development work gives priority to those goals, accepting limitations with respect to other possible goals. In other words, no one model can or should try to be all things to all people. Examples of models that foundered because the goals were far too ambitious, particularly given the restricted capabilities of the computer hardware and software technology available at the time, litter the history of microsimulation model development in the United States (see below).

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

We encourage broad conceptual thinking about model goals, and we support, when appropriate, model designs that are ambitious in their scope. For example, there are strong arguments that future model development in the health care policy world (whether for models of health care consumers or providers or both) ought to simulate behavioral responses and second-round effects as well as the direct effects of policy changes (see Chapter 8). Nonetheless, we argue strongly that model developers must make explicit choices about priorities. If they design a model that is broad in scope, then they should plan to develop only some of the model components in depth. To continue with the example, the developers of a behavioral health care policy model might pick one type of response to model in detail, while constructing a much cruder model of other responses. Conversely, a very detailed model should most likely restrict the range of issues or behaviors that it encompasses.

We are hopeful that current and expected future improvements in computer technology will make it possible to build more ambitious models that do not crumble under their own weight. Improvements in research knowledge and databases will also support this goal. However, we still see the need to trade off model breadth and depth. If such trade-offs are not made, the model development process is likely to abort or result in a model that is dysfunctional, wasting scarce resources.

Self-Contained Modules

Good design should be based on self-contained modules that can be readily added to (or deleted from) a model. As a corollary, the design should permit an analyst to specify all or only a subset of the existing modules to be used for an application. Furthermore, the modular principle applies not only to the components of the model that simulate programs such as AFDC or SSI, but also to the various steps that are used to generate a model database. Finally, each module should be highly "parameterized," that is, allow the user to change various features of the module by simply supplying a new value for one or more parameters instead of writing new code. For example, it should be possible to change the value of allowable deductions in the formula for calculating AFDC benefits or to change the value of the factor used to adjust income amounts for underreporting.

In our view, a well-specified modular structure is one of the most important means for enabling a model to respond in a timely and flexible manner to new policy issues, while minimizing development costs because the entire model does not need to be rebuilt. Similarly, a modular structure minimizes costs of model use because only those modules that are required for a particular application have to be used. Looking to the future, it may be that modular designs implemented with enhanced computer technology will enable microsimulation

Copyright National Academy of Sciences. All rights reserved.

models to become bigger in scope than current models without becoming "too big."[2]

Facilitated Documentation and Evaluation

A good design provides for features in each module that facilitate documentation and evaluation. Although many existing microsimulation models have been designed along modular lines, relatively few have sought ways to improve the production of model documentation, and none has paid sufficient attention to model validation.

Good documentation is essential to enable people other than a model's original developers to understand the model and modify it in a cost-effective manner and to evaluate the interactions of model components and their effects on the quality of the model output. Good model design should seek ways to reduce the costs and time of preparing complete documentation, such as providing for self-documenting program code in each module and automated audit trails that can track the effects of modifications in a module on other model components (see David, 1991; see also Chapter 10).

Good model design should also seek ways to facilitate the evaluation of model components, separately and collectively. To date, the task of model validation has been costly and time consuming and, partly for those reasons, has been generally ignored. Providing features such as the ready ability to use different procedures for constructing variables in the database could greatly facilitate the conduct of validation studies, including sensitivity analyses and estimation of variance in the model results through sample reuse techniques. Again, developments in computer technology should make possible great strides on these dimensions of good model design.

Linkages

Good design provides for entry and exit points in the model that facilitate linkages with other models. In addition to a modular structure, another way to tackle the problem of providing enhanced capability without designing an overly elaborate and cumbersome model is to build in pointers that make it possible to use more than one model. For example, with such a design, a microsimulation model could feed results to a macroeconomic model in order to obtain a reading of second-round effects of proposed policy changes and, in turn, use outputs from the macroeconomic model in running further microsimulations.

---

[2] Indeed, enhanced computer technology may well make it possible for a good model design to be "integrated," that is, to require running all instead of just some of the model's component modules without incurring excessive computational costs. Such a design requires careful specification of default values for each module to minimize the burden on the user of determining whether all modules are properly specified for the particular application.

Copyright National Academy of Sciences. All rights reserved.

Computational Efficiency

Good design strives to attain a high degree of computational efficiency of a model and its components consonant with other important design objectives. It is obvious to assert that the design of computer-based models should strive to minimize computational costs, and, indeed, developers have historically given high priority to this aspect of model design. However, declining costs of computing mean that it may be feasible, and perhaps desirable, to give up some computational efficiency for other goals, such as ease of use and ease of modification of the model. In any case, in considering ways to reduce computing costs, developers need to consider not just the costs of a single run but the combined costs of all of the runs that are necessary for a particular application.

Accessibility to Users

Good model design incorporates features that provide a high degree of accessibility of a model to analysts and other users who are not computer systems experts. At present, policy analysis agencies generally rely heavily on a few expert analysts, often affiliated with a contractor, to carry out microsimulation model applications. In turn, the analysts rely on programming staff to make changes in the model and even to set up model runs. Although, to our knowledge, the experts have met high performance standards, we believe that the cost-effectiveness of microsimulation modeling for the policy process would be improved by making the models accessible to more people.

This goal requires making the models more "user friendly" to people who are not computer buffs, as well as providing adequate documentation, training, and support services. Greater accessibility should reduce overall costs for model applications by making it possible for analysts to run the models themselves (at least for straightforward applications), should enhance the analysts' understanding of the properties of the model, and should encourage greater experimentation with the models and closer scrutiny of the quality of the model outputs.

Recommendation 6-1. **We recommend that policy analysis agencies set standards for the design of future microsimulation models that include:**

- **setting clear goals and priorities for the model;**
- **using self-contained modules that can be readily added to (or deleted from) the model and that are constructed to facilitate documentation and validation, including the assessment of uncertainty through the use of sensitivity analysis and the application of sample reuse techniques to measure variance;**

Copyright National Academy of Sciences. All rights reserved.

- **providing for entry and exit points in the model that facilitate linkages with other models;**
- **attaining a high degree of computational efficiency of the model and its components consonant with other objectives such as ease of use; and**
- **attaining a high degree of accessibility of the model to analysts and other users who are not computer systems experts.**

Practices

In addition to adopting good design principles, it is important to follow good practices in the actual implementation of a microsimulation model design. Such practices include careful development, full documentation, validation, and follow-up assessment.

The development process has to be carefully staged, with work directed toward intermediate goals or milestones, so that progress can be assessed and timely corrective action taken if there are undue delays or budget problems. In addition, the process should include the construction of prototypes of model components so that design flaws can be identified and corrected at an early stage. Prototypes or test versions of the model can also provide the sponsor agency with some analysis capability before the model is completed, an important consideration if the development process is expected to extend over several years.

Fully adequate documentation should be prepared on a timely basis for the model, its components, and its database. Not only is it important for the design to facilitate documentation, but the development process must assign sufficient priority and resources so that good documentation does, in fact, result. Ideally, documentation is developed on a flow basis, including an audit trail of all modifications, instead of being left to the end.

It is important to carry out validation studies of the model and its components, including sensitivity analyses as well as estimation of variance. It is particularly important to perform sensitivity analyses for each new module prior to full implementation, including studies of the impact on the rest of the model, in order to identify any unexpected or dysfunctional interactions or adverse effects on use. Again, the development process should assign sufficient priority and resources to this task.

In our experience, model developers typically follow good practice with regard to validation in the development of individual components of modules. For example, standard procedure in developing a regression equation for imputing variables, such as child care deductions for the AFDC program, is to look at various measures of goodness of fit before determining the final form of the equation. However, model developers are much less likely to look at the effect

Copyright National Academy of Sciences. All rights reserved.

of alternative forms of an equation on the operation of the module as a whole or of alternative modules on the operation of the entire model.[3]

Finally, once a model is in use, developers should periodically take stock of the model's capabilities and functioning. That is, they should conduct regular "sunset" or "zero-based" reviews, in which they reevaluate the structure of the model, determine components that have outlived their usefulness and need to be dropped, and identify other components that need to be rebuilt. Such zero-based reviews are important because models constantly undergo revision and adaptation to respond to users' changing requirements for estimates. Without regular reevaluations, it is all too likely that a once well-designed model will become increasingly less cost-effective over time. It is important that policy analysis agencies build into their budgets provision for reassessing and reoptimizing the microsimulation models they use for input to the policy process.

Recommendation 6-2. **We recommend that policy analysis agencies set standards of good practice for the development of future microsimulation models that require:**

- **constructing prototypes and establishing milestones throughout the development process so that design flaws can be identified at an early stage and the agency provided with some analysis capability before the entire model is completed;**
- **preparing fully adequate documentation on a timely basis for the model and its components;**
- **conducting validation studies of the model and its components, including estimates of variance and sensitivity analyses (the latter should be conducted for each new module, prior to full implementation, by examining its impact on the rest of the model in order to identify any unexpected or dysfunctional interactions or adverse effects on use); and**
- **subjecting the model to a "sunset" provision, whereby the model is periodically reevaluated, obsolete components are deleted, and other components are respecified to optimize the model's usefulness and efficiency.**

[3] An exception is a recent study of food stamp participation rates, based on comparing counts and characteristics of participants from the IQCS with simulations of eligible units from the SIPP. In it, Doyle (1990) discussed sources of error in the participation rate estimates and provided an assessment of some components of uncertainty, including the impact of two different ways of estimating countable assets for households; adjusting versus not adjusting the program administrative data for administrative errors; and comparing wave 7 of the 1984 SIPP panel with wave 3 of the 1985 SIPP panel, with the two waves combined as the database for the eligibility simulations.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

CURRENT MICROSIMULATION MODEL DESIGN

Many current microsimulation models have been constructed on the general principles that we advocate. But some modeling efforts have foundered because (in addition to whatever data and other problems occurred) the design was too grandiose, the structure did not provide for sufficient modularization, or the documentation was inadequate to permit easy updating or adaptation of model components. In addition, some models that were initially well designed became encumbered with ill thought-out additions or obsolete components and lost significant flexibility and ease of use.

In the history of microsimulation modeling, RIM (Reforms in Income Maintenance) and the KGB (Kasten, Greenberg, Betson) model represent notable examples of models that were developed under acute time pressures; asked to simulate more and more proposed policy changes; and then, having collapsed under their own weight, abandoned. Insufficient attention to documentation, modularity, and computing efficiency made them increasingly difficult for their own developers—let alone other analysts—to use. The developers of TRIM (Transfer Income Model) originally tried to rewrite RIM but gave up the attempt and instead developed an entirely new model; the KGB model wound up on the shelf after its developers left the originating agency, ASPE.

Other models have been constructed according to good design principles but, over the years, having become increasingly more cumbersome and difficult to access. Thus, a major impetus to the development of the KGB model was the perception that TRIM, as it then existed, could not readily be adapted to the requirements of simulating a combined income support and jobs program and that it would take too many resources to conduct such simulations with the MATH (Micro Analysis of Transfers to Households) model. The current TRIM2 model is the result of rewriting TRIM to achieve greater efficiency and modularity. The MATH model was also rewritten to improve computing efficiency.

Other modeling efforts have suffered from being overly ambitious, particularly given the limitations of the computing technology at the time of their development. The original DYNASIM (Dynamic Simulation of Income Model) was intended to model a wide range of human behaviors and policy issues and to be accessible to a broad research community. The goals were laudable, but the costs of running the model on more than a very small sample of households proved prohibitive. DYNASIM2 is a scaled-back version that was redesigned to maximize computational efficiency and to focus on policy issues surrounding the provision of adequate retirement income support. The MRPIS (Multi-Regional Policy Impact Simulation) model has also experienced high costs and long development times in its developers' effort to construct a complex system of fully linked models. The system includes a detailed microsimulation model of the household sector, together with input-output and cell-based models that

Copyright National Academy of Sciences. All rights reserved.

simulate the effects of changes in household disposable income on savings and consumption that, in turn, affect regional economic production and, ultimately, in a feedback loop, affect household employment. To date, MRPIS has had only limited application.

Overall, the history of microsimulation has exhibited a pattern of advances and retreats. Advances occurred as new capabilities of various kinds were added to the models to respond to policy needs. Invariably, there came a time when the added capability proved dysfunctional for the continued cost-effective operation of the model. Then came retreat. Historically, the response has been (if the model was not simply shelved) to redesign the model for maximum efficiency with the available computing technology. Another typical response has been to refocus the model on core modules: for example, current maintenance and development efforts for the MATH model are limited almost entirely to those modules that are necessary to simulate the food stamp program. Other modules have been largely shelved.

Today, the major microsimulation models are optimized for the mainframe, batch-oriented computing technology that prevailed in the late 1970s and early 1980s. Although this strategy has greatly reduced computing costs, the models are not well positioned to move toward adding new capabilities. Development as well as operation of the models can entail substantial overall costs because many applications require multiple runs and the intervention of skilled programmers to make necessary changes in the program code. Moreover, the linearity embedded in the models' hardware and software configuration imposes constraints, sometimes quite severe, on analysts' flexibility of use. For example, once a new TRIM2 or MATH baseline file has been constructed, it is very expensive to go back and alter its characteristics, say, to age the data forward in time if aging has not already been done or to construct different kinds of program filing units. This inflexibility also inhibits evaluation studies of models.

Today, the major models—particularly the cross-sectional "static" models such as TRIM2 and MATH—are focused on those capabilities that are central to simulating the direct effects of proposed program changes. The core of these models is the set of modules that mimic the detailed operation of government programs, and new model development in the past decade has generally focused on further elaborating these modules or developing modules for additional programs. For example, the federal and state income tax modules in TRIM2 were completely rewritten and expanded in the mid-1980s, as was the TRIM2 Medicaid module. This type of development increases the breadth of a microsimulation model in terms of the policy areas for which it is relevant. Another example of adding breadth, in this case to a dynamic model, is the recent development of a module for simulating long-term care financing programs in PRISM (Pension and Retirement Income Simulation Model).

In contrast, development of the kinds of capabilities that increase the depth of a microsimulation model in terms of simulating the effects of a proposed

Copyright National Academy of Sciences. All rights reserved.

policy change has either not taken place or resulted in less-than-satisfactory modules that were later mothballed. Examples of depth include functions for aging the data and simulating first-round and second-round behavioral responses inside a model. All of the major static models include aging modules; however, this capability in TRIM2 has not been used for policy purposes since the early 1980s. As we have noted, none of the static models currently includes modules for behavioral response (other than the participation decision) or for second-round effects.[4] The modules that were developed for MATH in the mid-1970s to simulate work responses to welfare program changes have been dormant for many years.

The dynamic models, such as DYNASIM2 and PRISM, that are used to model long-range policy issues involving retirement income and other areas obviously include aging capabilities. Indeed, aging is the focus of these models. They also make heavy use of research knowledge about behavior to build their databases. However, they incorporate relatively few feedback effects of future policy changes on behavior, except for the decision to retire, which is the functional equivalent to the program participation decision in the static income support models.

Many factors have resulted in the situation we just described. The need to keep their models competitive in the policy analysis market has obviously motivated the model developers (and the analysts who use the results) to strive for reduction of computer costs and sufficient streamlining so as to facilitate their use.

In turn, the predilections of policy makers (and their staffs) have exerted a strong influence in the direction of emphasizing the models' accounting elements —that is, the detailed program rules—in preference to other capabilities. Particularly in the Gramm-Rudman-Hollings era, policy makers have often been more concerned about the immediate consequences of a proposed welfare or tax policy change for next year's budget deficit, for example, than about longer term behavioral effects. They have also been interested in obtaining results for very fine-grained changes in programs. An added factor in reinforcing emphasis on the accounting modules is the concern of the staffs in the policy analysis agencies to have credible models whose baseline results closely match available administrative data on program recipients, although, as noted above, the process of calibrating model results to control totals carries its own perils (see Chapter 5).

Last, but far from least, an important factor in the underemphasis on capabilities such as modeling behavioral response is the skimpy base of research knowledge on which to develop these kinds of capabilities. Social science is far from being able to predict the interactions of individuals with government policies and other social and economic forces with the precision that would be

[4] Recently, work started on developing a labor supply response module for TRIM2.

Copyright National Academy of Sciences. All rights reserved.

desired for policy analysis purposes or in a form that readily adapts to models. Given the thin research knowledge base and the costs of implementing new model capabilities in the current microsimulation model computing framework, it is no wonder that the model developers have focused their attention on the accounting functions of their models.

STRATEGIC DIRECTIONS FOR MICROSIMULATION MODEL DEVELOPMENT

The question we address in this section is what paths the policy analysis agencies should follow for the future of microsimulation model development.[5] Clearly, a critical line of attack for building improved microsimulation models involves taking advantage of recent developments in computer hardware and software that will not only facilitate expansion of model capabilities at reduced costs, but also increase their flexibility and ease of use. (This topic merits a full discussion of its own, which we provide in the next chapter.) In this section we assume that a successful jump to a new computing environment will be made and, hence, that it is timely once again to consider expanding the capabilities of microsimulation models.

Three major types of capabilities might be incorporated or enhanced in future microsimulation models: (1) aging the data forward in time, (2) simulating behavioral response to proposed policy changes, and (3) simulating second-round effects of policy changes. It is important to note that our recommendations in each area do not prescribe a preferred path for microsimulation model development. In the absence of evaluation studies, there is simply not enough information—for example, about the merits of static versus dynamic aging techniques, of omitting or incorporating behavioral effects, or of limiting models to first-round effects or simulating second-round or later effects—to reach specific recommendations about preferred model capabilities. Even with greatly enhanced computing power at lower costs, it is important to consider carefully the addition of complex new capabilities for which the needed research knowledge may be very costly to obtain and to determine effective ways of implementing that knowledge in models. Thus, our recommendations concentrate on a research agenda with two major goals: to identify the kinds of enhancements to microsimulation models that are important because of their impact on the resulting policy estimates and to provide an improved knowledge base to support the effective implementation of key new model capabilities.

Data Aging

Microsimulation models share with other kinds of models the need for updating

[5] This section benefited greatly from a set of notes prepared by panel member Robert Moffitt.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

or projecting available data so that they can be used to develop current or future-year impacts of existing and proposed new programs.[6] The need for this process, commonly called "aging," arises for several reasons:

- Needed data inevitably become available with a time lag, so that even current-year estimates require aging.
- Programs are never adopted instantaneously so that aging of the data is required to the year when a program change is expected to take effect.
- Policy makers typically want estimates of the costs and population impacts of proposed changes for a period of several years beyond the date when the change is to take effect.
- Policy makers sometimes want to see the impacts of past or proposed changes over much longer time spans, 30 or more years for programs for retirement income or health care of the elderly.

Although the need for aging databases forward in time confronts virtually all types of policy analysis tools, microsimulation models present particularly difficult problems due to the large number of variables they include. Moreover, the choice of aging strategies, whether static or dynamic, has far-reaching implications for the design and nature of the resulting model. The complex issues presented by aging in microsimulation models have led some policy analysis agencies to forgo aging inside the model and, instead, conduct simulations on the latest available data and then extrapolate the simulated values to develop future projections outside the model.

The dimensions of a database most often of concern in aging analyses are its population proportions and its economic variables. The age-race-sex and family type proportions in a population change over time, and these changes are likely to affect the aggregate simulated values of most policies under consideration in microsimulation models. Likewise, real growth as well as nominal growth in incomes and prices are both certain to occur over time and likely to affect simulated values. Another variable often considered for aging is the unemployment rate, because a cross-sectional database will by definition capture only one point in the business cycle, a point that may not be of interest to policy makers.

Existing Aging Techniques

There are, roughly speaking, two main and very different aging strategies used in existing microsimulation models. First, many models (e.g., MATH) employ static aging of the population, a technique by which the individual weights for each record on a micro file are adjusted to meet a set of projected population

[6] The discussion in this section benefited greatly from papers by Caldwell (1989) and Ross (in Volume II) and from presentations to the panel by David Kennell and John Sheils of Lewin/lICF, Inc., and Sheila Zedlewski of the Urban Institute.

Copyright National Academy of Sciences. All rights reserved.

proportions; monetary variables are scaled upward in a separate exercise according to outside macro forecasts; and forecasted changes in unemployment are distributed across the micro units in some fashion in a third and yet separate exercise.[7] Static aging is the technique most commonly used in microsimulation models because of its relative ease of implementation.

Second, some models (e.g., DYNASIM2, PRISM) use dynamic aging, a technique by which individual units on a micro file are aged directly (see Caldwell, 1989; Ross, in Volume II). For each year that a database is projected, the age of each individual is increased by 1; economic variables are scaled upward by individual-specific magnitudes; and many other variables, if not all, are moved forward on an individual-specific basis according to outside, econometrically estimated transition probabilities and dynamic micro equations.[8] Older individuals are deleted from the sample according to estimated mortality rates, and hypothetical newborns are added to the sample according to estimated fertility rates.

However, some users of microsimulation models do not use either type of aging capability. For example, analysts in ASPE ask the TRIM2 modelers to perform AFDC simulations on the most current database, which at best is 1-2 years out of date. The analysts then apply the percentage differences estimated by TRIM2 for total costs and caseloads under current versus proposed law to projections of current-law total costs and caseloads developed through other means.[9] This method avoids any controversy over use of a particular aging method, but it shifts the controversy to the method used for the out-of-model projections. With regard to population effects of proposed AFDC changes, such as "gainers" and "losers," ASPE uses no aging technique at all. Instead, the ASPE analysts use the unadjusted TRIM2 distributions, assuming that there will be no population changes of any import. ASPE adopted its out-of-model aging strategy in the early 1980s, after deciding that the cost and effort required to

[7] The methods for obtaining the control totals for the static models vary. For example, changes in prices and unemployment may be forecasts from a macroeconomic model, while population projections are typically the output of cell-based models such as those used by the Census Bureau or the Social Security Actuary (see Grummer-Strawn and Espenshade, in Volume II). Dynamic microsimulation techniques could also be used to develop population projections, although the use of such projections in a reweighting process would still involve static aging.

[8] Most dynamic models use a hybrid approach, in which many variables are dynamically aged, but others—for reasons of convenience and lack of a reasonable set of estimated transition probabilities—are adjusted in a static fashion. We note that the panel did not review dynamic microsimulation models, such as DYNASIM and PRISM, in as great detail as it reviewed static models, such as TRIM2 and MATH. The discussion throughout this section focuses on general issues in the use of dynamic versus static techniques for aging model databases forward in time; it does not address the techniques used **in** particular dynamic models, which are quite complex and vary in a number of ways.

[9] Projected current program costs are developed by the Family Support Administration (now in the Administration for Children and Families), using a variety of information from federal and state sources.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

perform static aging inside TRIM2 and the uncertainty regarding the quality of the resulting database were unacceptably high.

Issues and Trade-Offs

In the literature on in-model aging strategies, the central trade-offs between the different methods involve cost, theoretical desirability, and reliability. Theoretically, the preferred method is dynamic aging because it fully incorporates cohort effects and therefore properly preserves all the covariances in the data between age and other variables. However, the reliability of dynamic aging methods may be low if the estimated transition functions necessary to implement the aging have high variance or suffer from serious misspecification. Furthermore, it may be difficult to implement a dynamic technique in a partial manner. If it is implemented at all, it should be implemented for all age-related variables on a database. For example, if 30-year-old individuals on a file are increased in age by 1 year, all other variables for their income, labor force status, earnings, number of children, marital status, education, and so on should also be dynamically aged. Any variable not so aged but left at its age-30 value may provide a less accurate indicator of its age-31 value than the values of that variable for the age-31 sample at the initial time period. In other words, any variable not dynamically aged may be inferior to the values that would be obtained from static aging. However, aging all variables is a time-consuming process, requiring the estimation of many sets of dynamic equations on panel data, with unknown reliability.

The major advantage of static aging is its comparative ease of implementation. Its theoretical weakness mirrors the strength of dynamic aging, namely, the failure to capture adequately the cohort effects in the variables on the file. Simply changing the weight on 30-year-olds, for example, fails to account for the differences in the values of the other variables among 30-year-olds of different cohorts.

Another issue in aging is how a short-run time horizon, as in the MATH and TRIM2 models, or a long-run time horizon, as in DYNASIM2 and PRISM, affects the relative advantages of static and dynamic aging. Although dynamic aging has traditionally been used for long-run projections and static aging for short-run projections, in general either method could be used in either instance. Dynamic aging may be theoretically superior in the short run as well as the long run (see Caldwell, 1989); however, the degree of damage done by static aging is probably less for short-run projections than for long-run projections, because cohort effects are not likely to distort seriously the age-related covariances of the variables in a database over a 3- to 5-year period.

The out-of-model aging technique adopted by users of TRIM2 avoids these controversies at the cost of shifting problems to a later stage in the development of policy estimates. If a model is run on unaged data and the results of the

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

simulation extrapolated to future periods outside the model, the assumptions related to aging are not eliminated but only moved to a different place in the simulation procedure. The method of extrapolation then must be based upon static aging, dynamic aging, or some other concept.

Only limited evaluations of microsimulation model aging techniques have been conducted to date. The panel, as part of its validation experiment using the TRIM2 model, varied the degree of static aging that was performed (see Chapter 9 and Cohen et al., in Volume II). Model runs included no aging at all (in or out of the model), aging of population characteristics only, demographic aging together with adjusting for price changes, and full aging (that is, demographic aging, price adjustments, and labor force adjustments). To eliminate uncertainties with regard to the quality of the control totals, a 1983 database was aged to 1987 by using known values. The results showed significant differences in projected 1987 estimates of AFDC costs, caseloads, and characteristics of recipients across the four aging methods and between each set of estimates and administrative records data from the IQCS for 1987.

Doyle and Trippe (1989) carried out a more limited study of the quality of demographic aging and price adjustments in a 1980 MATH model database projected to 1984 using known control totals. This study found significant differences between the aged database and the actual 1984 database, for example, in the distribution of households by income as a percent of the poverty line. Overall, however, the aging process produced a population that more closely resembled the actual 1984 population than did the unaged base-year file.

Both Doyle and Trippe (1989) and the panel's study found that, even in the short run, demographic and economic changes occur and are important for estimating program costs and caseloads. They also found that current static aging routines capture some but not all of the changes.[10] However, neither study compared alternative forms of the TRIM2 or MATH model aging procedures (both models use virtually the same aging routines), such as varying the assumptions used to distribute projected changes in unemployment rates to individual records. Nor did either study compare dynamic with static aging techniques.

Evaluations of dynamic models have examined their performance on specific dimensions, such as their ability to reproduce historical patterns of labor force participation over time, compared with actual data from panel surveys such as the Michigan Panel Study of Income Dynamics (see Cohen, Chapter 7 in Volume II). These studies in some cases led to modifications in the models' transition functions. Dynamic models are also routinely calibrated for agreement

[10] Not surprisingly, Doyle and Trippe (1989) determined that the aging process worked best for characteristics that were controlled for in the process, for other demographic characteristics that did not change much or were closely associated with characteristics that were controlled for, and for income characteristics that did not change much. The aging process did not work as well for some key measures of the economic status of the population.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

with aggregate forecasts of macroeconomic models. Again, however, there has been no systematic evaluation of alternative forms of dynamic aging.

Given the need for some method of updating and projecting the data required for future-year estimates of program impacts, policy analysis agencies should devote more attention to the question of appropriate aging strategies for microsimulation models. We suggest that systematic evaluation begin with methods for aging the databases used by models with short time horizons, such as MATH and TRIM2, because the evaluation task seems less daunting for short-term than for long-term projections. We propose an evaluation that would be much more extensive than, but similar in design to, the panel's own experiment: the aging methods would use known control totals (so as to eliminate errors due to mistaken population and economic projections) and would be evaluated against targets from program administrative records, including characteristics that are and are not related to the control variables.

A principal focus of the evaluation of aging methods should be to determine how well a method performs, on average, when it is assessed for a number of time periods. Extensions of an evaluation would be to vary the length of time over which the aging is performed, for example, from 1 to 5 (or perhaps as many as 10) years, as well as the aging method itself. The techniques to be evaluated should include the current static aging routines in TRIM2 and MATH, as well as variants of those routines (e.g., modifying the scaling factors in the economic aging routine to reflect varying patterns of income growth by socioeconomic status or using raking techniques such as iterative proportional fitting to make simultaneous adjustments to multiple sets of control totals). An evaluation should also include the out-of-model aging technique currently employed with TRIM2, comparing the administrative data with the model estimates after ASPE analysts have made extrapolation adjustments. It should also be possible to generate aged populations by using a dynamic model such as DYNASIM2, to run the TRIM2 or MATH program modules on these populations, and to evaluate how well the output matches the program administrative data. The results should be useful for charting the future development of all models oriented to the short term, whether they be models of income support programs, taxation, or health care policies.

***Recommendation 6-3.* We recommend that policy analysis agencies sponsor an evaluation program to assess the quality of estimates from current static microsimulation models as a function of the aging technique that is used and that they further support such evaluations on a periodic basis for future models.**

Copyright National Academy of Sciences. All rights reserved.

In conducting our experiment, we found that the cost and time of implementing the various aging scenarios in TRIM2 were high.[11] The aging must clearly be carried out at an early stage in the preparation of the database, which then necessitates rerunning many other database and program simulation modules each time a different aging method is used. These costs are one reason for the out-of-model aging strategy currently followed with TRIM2; they are also why new aged databases for the MATH model are developed only every 2 or 3 years (in contrast, new (unaged) TRIM2 databases are created annually from each March CPS). This situation clearly has a chilling effect on efforts to evaluate and experiment with alternative aging procedures. One goal for future model development or redevelopment should be to find ways to make it easier to add and experiment with alternative aging routines and control values.[12]

Recommendation 6-4. **We recommend that policy analysis agencies require that future static microsimulation models build in an aging capability in a manner that facilitates evaluation and use of alternative aging assumptions and procedures.**

Incorporation of Behavioral Response

Most of the policy measures simulated by the models that we reviewed have potential behavioral effects on the individuals and organizations affected by the policy.[13] Altering the structure of transfer programs for the low-income population, changing the tax code, or changing the form of health care benefit programs may all affect individual and organizational decisions in many ways:

- Altering the level of cash or in-kind (e.g., Medicaid) benefits in the AFDC program may affect the work decisions of welfare recipients.
- Changing the tax rate on capital gains may affect decisions on the part of investors as to when to realize their gains.
- Mandating employer provision of health insurance may affect the demand for labor by small businesses.

---

[11] One factor unique to our experiment was that the TRIM2 programmers had to become familiar again with the aging routines, which had been on the shelf for a number of years.

[12] We note that the most recent version of the Office of Tax Analysis's microsimulation model incorporates a major revision to the aging procedures. The aging routines are integrated within the structure of the model and allow the user to select one of three scenarios based on alternative projected growth rates of the economy (see Chapter 8). However, it does not appear that a user could readily experiment with different forms of aging, such as choosing different variables to include in the aging process.

[13] This discussion of modeling first-round behavioral responses benefited greatly from papers prepared for the panel by Burtless (1989), who examined labor force and retirement responses in the MATH, TRIM2, KGB, DYNASIM2, and PRISM models; Grannemann (1989), who reviewed issues in modeling behavioral responses to health care policy changes; and Strauss (1989), who considered issues in modeling behavioral responses to changes in federal income tax policies.

Copyright National Academy of Sciences. All rights reserved.

More generally, changes in transfer policy, tax policy, and health care policy can be expected to affect participation by individuals in transfer and health care programs, levels of work effort among program recipients or the population as a whole, levels of health care expenditures and other health outcomes, and levels of investment and other variables affected by the policies under consideration.

Conceptually, one can distinguish two separate functions of microsimulation modeling: the accounting function of aggregating the effects of a policy change across the individual units in a given population and the function of predicting the behavioral response to a policy change. The first function is the more basic and must precede the second. This section considers the desirability and feasibility of incorporating behavioral responses into microsimulation models in addition to their accounting structures. We consider initially only first-round behavioral effects and hence only partial responses to policy changes; later, we discuss the incorporation of second-round effects.

Behavioral Responses in Current Models

The documentation for some of the models that we reviewed left us in doubt as to the extent to which the model incorporated behavioral responses. Nonetheless, it is clear that in the microsimulation models currently in operation, as well as those implemented in the past, the incorporation of behavioral response has been the exception rather than the rule.

MATH and TRIM2 do not currently have operational components for behavioral response in labor supply (or other dimensions) to policy changes in transfer programs; thus, the models almost exclusively serve the accounting function. Previously, however, labor supply modules were developed in both MATH and KGB to simulate the labor supply effects of the Carter welfare reform program, and a labor supply module is currently under development in TRIM2.

The dynamic microsimulation models of retirement income programs (e.g., DYNASIM2 and PRISM) make use of many behavioral parameters in developing their longitudinal databases, for example, using probabilities of job change as a function of variables such as age and previous job experience to build individual employment and earnings histories. However, they model only a limited range of the likely behavioral *responses* to proposed changes in retirement income programs. Thus, in neither model does the level or potential level of pension and social security income affect hours of work or earnings of retirees or nonretirees. Rather, changes in retirement income programs affect labor supply only by their effects, if any, on the decision to retire at a particular age (see Burtless, 1989:20-27; Ross, in Volume II).

For tax policy, most microsimulation models are tax calculators that simulate the revenue effects of changes in the tax code. Tax models typically include a function to decide whether or not taxpayers will itemize, on the basis

Copyright National Academy of Sciences. All rights reserved.

of the assumption that they will choose to minimize their tax liability. Other behavioral responses are also frequently taken into account, but not inside the models. Rather, the agencies and congressional committees using tax simulators adjust their accounting estimates for behavioral response outside the models, for example, adjusting the revenue estimates of a change in the capital gains tax rate to reflect assumptions about the effects on the timing of sales of capital assets (see Strauss, 1989).

For health care policy, strong behavioral responses have been documented in some cases: for example, hospitals responded when Medicare implemented a prospective payment system by reducing the average length of stay (see Grannemann, 1989). However, behavioral responses are rarely or only crudely incorporated in health care policy models at the present time.

As we have noted, the one behavioral response that is incorporated in MATH, TRIM2, and other microsimulation models for income support programs is participation in such programs. Given that some proportion of eligible households voluntarily choose not to apply for benefits, the models must incorporate the determinants of the participation decision. The analogous behavioral response function in models of retirement income programs is the decision to accept retirement benefits, which entails a corresponding decision to stop work or reduce hours.[14]

Issues and Trade-Offs

There are many issues that must be considered in an assessment of the desirability and feasibility of incorporating behavioral response in microsimulation models. First is the potential magnitude of behavioral responses that are thought to occur. Clearly, the greater the potential for a behavioral response, the more seriously it merits consideration for incorporation into models. The level of the potential response is likely to differ by the type of policy considered, the magnitude of the policy stimulus, the type of individual unit whose response is under consideration, and the type of behavior being examined. A related issue comes into play when the microsimulation modeling is designed to rank-order different policy alternatives rather than to obtain the best estimate of a single alternative. In this case, the relevant question is whether the potential magnitude of behavioral response across the alternatives being ranked is likely to affect their respective rankings.

A second major consideration is whether there are generally agreed-upon and reasonably reliable statistical estimates of the magnitude of a behavioral response in order to simulate it in the first place. To some extent, this issue

[14] The participation and retirement decision components of existing microsimulation models vary widely in their form and complexity; see the comparison of the MATH, TRIM2, and HITSM participation functions in Citro and Ross (in Volume II) and the comparison of DYNASIM2 and PRISM in Ross (in Volume II).

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

precedes the first one, because judgment of the potential magnitude of behavioral response presumes the existence of relatively good estimates. Nevertheless, it is often the case that the statistical estimates available from the research literature are sufficiently large to warrant concern over behavioral response but far from sufficiently reliable to give any degree of confidence to incorporating them in models.

Unfortunately, it appears that this situation is the rule rather than the exception for social welfare policy issues. For income support issues the variances of estimates of the work-effort effects of welfare programs are often large and, more important, the estimates obtained from different studies range widely (see Burtless, 1989). The same characterization applies to tax policy estimates, such as the effect of capital gains taxation (see Strauss, 1989). Estimates of other kinds of behavioral responses with important policy implications, including charitable giving, demand for health services, and demand for health insurance, also vary widely (Strauss, 1989:Tables 4-1 to 4-10). In many cases, it appears that research designed to replicate and reconcile the different findings from different studies could narrow considerably the range of estimates in the statistical literature. However, studies that attempt such a narrowing of the range of estimates are rarely conducted.[15]

Because of the problem of the unreliability of response estimates, there is a need to develop concrete and well-defined methods for assessing the uncertainty of behavioral response parameters if they are incorporated in microsimulation models. If a single estimated parameter is introduced in a microsimulation model, the variance of that parameter must be translated into variances of the aggregates simulated from the model. If multiple parameters are introduced, their covariance must be similarly incorporated; this task is often difficult because the different parameters are drawn from different studies. In addition, the uncertainty of parameters resulting from the range of estimates in the statistical literature must be assessed through some type of sensitivity analysis and the findings reflected in the caveats that accompany the output from a model. The degree of uncertainty from the range of estimates may often be much larger than that from the variance of any single estimated parameter.

We have not found any systematic attempt in microsimulation modeling efforts to date to address the problem of measuring the degree of uncertainty in model outputs attributable to the use of particular behavioral parameters

[15] Meta-analysis techniques, although still evolving and posing a number of thorny methodological issues, may be useful in helping to narrow the range of estimates of behavioral responses that are important to include in microsimulation models. Meta-analysis involves a systematic way to aggregate all available studies on a topic and, with the aid of statistical techniques, to determine the best estimate based on all of the studies, without conducting new research or secondary data analysis; see Wachter and Straf (1990). Empirical or hierarchical Bayes techniques, which provide a means to combine information from different sources, of varying quality, through a weighting process, may also be helpful in developing the best available estimate of a parameter for inclusion in a microsimulation model (see Efron and Morris, 1973; Lindley and Smith, 1972).

Copyright National Academy of Sciences. All rights reserved.

or to alternative specifications for those parameters.[16] Indeed, there has been little attention to the development of scientifically based methods for measuring uncertainty in the accounting function estimates of the models. Yet the assessment of uncertainty with regard to the estimation of behavioral response should not proceed without a similar assessment for the no-response estimates from microsimulation models that are obtained in the first place (see further discussion of this set of issues in Chapter 9).

The third major consideration concerns the problems involved in actually incorporating behavioral parameters in microsimulation models in an appropriate manner. This task is not always straightforward (see Burtless, 1989; also Grossman and Watts, 1982). The database used to estimate a behavioral response is usually not the same as the database used by the microsimulation model. Thus, researchers have typically estimated labor supply and other responses on panel data sets such as the Michigan Panel Survey of Income Dynamics and the Retirement History Survey. However, these surveys, because of small sample size, limited populations, and other reasons, are rarely suitable as simulation databases. In contrast, large, regularly updated, nationally representative surveys that are heavily used for simulation, such as the March CPS income supplement, generally lack the richness of information desired for estimation. For example, the March CPS does not obtain direct measures of hourly wage rates, so that modelers must impute wage rates to the CPS records in order to simulate a labor supply response to a policy change. As Burtless (1989) suggests, a useful, but rarely performed, step to help assess the impact of basing the simulation on imputed values would be to compare the estimated response in the original data set using both actual and imputed variables.

A related concern is that the available behavioral specifications may reflect an inappropriate degree of aggregation in relation to the microsimulation model. Thus, a good deal of behavioral analysis is carried out in terms of a representative economic agent, although microsimulation models are designed to carry out simulations on diverse kinds of units. For example, using a specification that predicts a uniform decline in work hours in response to an increase in welfare benefits fails to take advantage of the capabilities of microsimulation models for detailed analysis of population subgroups. Moreover, use of such a specification may distort the results, if, in fact, the pattern of response is that some workers drop out of the labor force entirely while the rest maintain their work hours.

Furthermore, the interaction of a behavioral response module with other aspects of a microsimulation model may produce anomalous effects. Burtless (1989) reported a comparison of the labor supply effects of four negative income tax plans simulated by the MATH model and a variant of MATH developed

---

[16] Betson's (1988) is one of the few studies to address such questions. Betson delineates major sources of errors in the microsimulation modeling process; he then explores the issue of the variability in microsimulation model estimates of a negative income tax plan that could be expected from the variability in estimates of labor supply response parameters from experimental versus nonexperimental studies.

Copyright National Academy of Sciences. All rights reserved.

by SRI International (called TATSIM). The two models used the same labor supply functions, estimated from the Seattle and Denver Income Maintenance Experiments: MATH simulated that aggregate labor supply and earnings of single mothers would decline under all four plans; TATSIM simulated that labor supply and earnings for such women would rise. These discrepant results could not be attributed to the behavioral response function itself; rather, they must have been caused by differences in the two models' simulation of the baseline prereform policy environment.

The fourth major consideration concerns the means by which behavioral responses are presented to policy makers—provided that reasonably good estimates are available, that they can be incorporated into a microsimulation model, and that they are of sufficient magnitude to be important for policy purposes. One issue is whether simulations should be presented both with and without the behavioral response included or just including the response. In the case of cost estimates, it may be that the simulated values should be provided to policy makers with the behavioral response, because such response will generally affect costs. However, in the case of other estimates more directly related to the response variable in question, such as labor supply, health expenditures, or capital gains realizations, it may be that the simulated values should be presented both with and without response, because the magnitude of the response itself may be of interest to policy makers.

Another issue involving presentation is how to convey the uncertainty in the estimates to policy makers. Many policy analysts argue that only the best possible point estimates should be provided because policy makers will base their decisions on such point estimates and not on a range, even if it is presented. We argue throughout this report that policy makers should be made aware of the degree of uncertainty in the simulated values either directly or indirectly so that they can take uncertainty into account in their decisions.

Over the past decade there has been little investment in capabilities for simulating behavioral response inside microsimulation models. There has been even less investment in the kinds of research and evaluation studies that would permit charting a sensible course for the development of future behavioral components of such models.

Concern about the cost implications of major development work has been one important factor in creating this situation. As we noted above, the nature of the legislative process in this decade—in which policy makers have been most concerned about total costs and distributional effects only in the very short run —has limited the motivation to add behavioral responses to models. The absence of a body of research demonstrating the impact on model estimates of incorporating behavioral parameters has further dampened interest, while the absence of sufficient research to produce parameter estimates within a reasonably narrow range or in a form suitable for microsimulation modeling has made the modelers cautious about expending much effort in this direction. In short,

Copyright National Academy of Sciences. All rights reserved.

there has been a self-perpetuating cycle in which lack of investment leads to downplaying the need and also makes it more difficult to incorporate behavioral response in models, which, in turn, leads to continued lack of investment.

Yet, undoubtedly, there are important behavioral effects of proposed policy changes about which decision makers should be informed. Certainly, legislative initiatives such as the 1988 Family Support Act and the 1986 Tax Reform Act were predicated on the assumption that policy changes would in fact lead to behavioral changes—for example, that transitional Medicaid and day care benefits would make work more attractive to AFDC recipients and that lower tax rates would stimulate investment and economic growth. But these effects were not simulated, nor with current modeling capability could they have been simulated; instead, to the extent that the effects were treated at all, they were handled with ad hoc, out-of-model adjustments. Yet microsimulation modeling cannot continue to ignore these behavioral response issues and expect to remain intellectually strong or policy relevant in the future. (In Chapter 8 we note the particular importance of incorporating behavioral response effects for modeling policy changes related to health care. Such responses are also very important for modeling tax policy issues.) Thus, the goal must be to devise a strategy for development of enhanced modeling capabilities in the area of behavioral response that makes the best use of scarce investment resources.

Clearly, a balance is required between the two extremes of always and never developing in-model estimates of behavioral responses. Research is required to determine the appropriate mix. In some cases, the best method of obtaining estimates may be through behavioral response studies separate from microsimulation models, perhaps involving the application of microeconometric behavioral research results to aggregate estimates of key population subgroups. In other cases, it may be that estimates of the behavioral effects of policy changes are best obtained within the structure of microsimulation models themselves. Or it may be that a dual strategy should be followed in which behavioral response estimates are obtained in both ways. We note that, even where it appears more cost-effective to obtain estimates of behavioral responses outside microsimulation models, it is quite possible **to** use the models to generate estimates of the population groups that are most likely to have a response in order to obtain a rough assessment of the likely overall impact on program costs and caseloads.

***Recommendation 6-5.* We recommend that policy analysis agencies devote resources to studies of the relationship between behavioral research and microsimulation modeling, including studies of ways in which research and modeling can complement one another, as well as ways in which the two are alternative modes of deriving answers to policy questions.**

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

An important criterion for deciding to spend the resources necessary to imbed behavioral response capabilities in microsimulation models is the likely magnitude of the impact. Hence, policy analysis agencies need to determine the kinds of behavioral responses to policy measures that could be important to consider. If behavioral responses are potentially large for some policy issues, and if it is decided to build some or all of those responses into microsimulation models, policy analysis agencies should attempt to reduce the extent of uncertainty surrounding existing statistical estimates of such responses. For many reasons, existing incentives for research, especially in academic circles, do not lead to the replication studies, additional data analysis, and sensitivity testing that are needed to reconcile the diversity of estimates from different analyses. Policy analysis agencies will need to provide those incentives by supporting such work.

Recommendation 6-6. **We recommend that policy analysis agencies sponsor studies to determine when behavioral response effects are most likely to be important in different policy simulations and, hence, how investment in developing behavioral response capabilities in microsimulation should be concentrated. On the basis of such studies, policy analysis agencies should commission research to attempt to narrow the range of statistical estimates of behavioral parameters that may be of major importance to critical policy changes. Such research may require additional data analysis, replication studies, and multiple econometric analyses that use different data sets and analytic techniques.**

One of the major issues regarding the incorporation of behavioral response parameters into microsimulation models concerns measurement of their uncertainty and the resulting impact on the uncertainty of the model outputs. After the analyses that we recommend above have been conducted, some uncertainty will still remain, both because the different results of varying studies will prove difficult to reconcile and because all behavioral estimates have variance. How to incorporate that uncertainty in measuring the total uncertainty of the estimates from a microsimulation model is far from obvious and requires the resolution of difficult technical issues (see further discussion in Chapter 9).

Recommendation 6-7. **We recommend that policy analysis agencies commission methodological research to develop methods for systematically assessing the impact on microsimulation model estimates of the degree of uncertainty in the behavioral parameters that are used—both the uncertainty arising from the variance of specific parameters and that arising from the range of estimates from different behavioral studies. This work should be tied into the development of similar methods for assessing uncertainty of the estimates produced by microsimulation models without behavioral response.**

Copyright National Academy of Sciences. All rights reserved.

INCORPORATION OF SECOND-ROUND EFFECTS

As discussed above, the first function of a microsimulation model is the accounting, or aggregation, function; a second function is the simulation of behavioral response to a policy change. Here, we discuss a third possible function of microsimulation models: simulation of the second-round effects of a policy change.[17] By second-round effects, we refer not to the immediate responses by individual economic units directly affected, but to 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.

Examples of potential second-round effects are many and pervasive:

- 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.
- Altering the coinsurance rate in Medicare or health care benefit programs in general will affect the demand for health care and therefore its price. If care is rationed because prices are inflexible, the policy change may affect the amount of health care in the market.
- Lowering the marginal tax rate in the federal income tax code will affect the tax price of all deductions and therefore affect the demand for the goods that provide such deductions, such as housing.
- Changes in transfer, health care, or tax policy will generally affect the level of consumption of different goods, which will affect employment in different industries, which, in turn, will affect the distribution of income in the population.

Existing Models of Second-Round Effects

The incorporation of second-round effects has been studied for many years by economists and others in the field of macroeconomic policy modeling. Indeed, the development of macroeconomic models preceded the development of microsimulation models, and early microsimulation modelers such as Guy Orcutt explicitly looked to macroeconomic models as sources of input. Thus, Orcutt

[17] The discussion in this section benefited greatly from presentations to the panel by Don Fullerton on computable general equilibrium (CGE) models and by Barry Bluestone, Alan Clayton-Matthews, and John Havens on MRPIS. Useful references on CGE models are Berkovec and Fullerton (1989), Shoven and Whalley (1984), and Whalley (1988). Useful references on MRPIS are Havens and Clayton-Matthews (1989), Havens et al. (1985), and Social Welfare Research Institute (no date). Finally, Anderson (1990) and Hanson and Robinson (1988) discuss linkages of macroeconomic and microsimulation models generally.

Copyright National Academy of Sciences. All rights reserved.

proposed originally to build a macroeconomic component into DYNASIM that would interact with the microsimulation components (Zedlewski, 1990). In addition to this type of interaction, the development of input-output models has provided another avenue for linkage with microsimulation models, and input-output models have been incorporated in some microsimulation models even on a regional basis (Haveman et al., 1980). Inadequate documentation made it difficult for us to determine the detailed structure of many such models. To focus our efforts, we examined two types of second-round models: computable general equilibrium (CGE) models and the Multi-Regional Policy Impact Simulation model.

CGE models, such as those developed by Shoven and Whalley (1984), simulate the equilibration of a full set of interconnected markets in an economy and permit full long-run adjustment of prices to changes in supply and demand.[18] CGE models are calibrated by a process of setting parameter values for supply and demand elasticities drawn from the econometric literature or picked to fit the available data on market prices and quantities. Although CGE models as generally implemented are not micro in nature, it is possible to develop disaggregated CGE models or to link them with microsimulation models by iterating back and forth between them until market equilibrium is reached (see, e.g., Berkovec and Fullerton, 1989; Slemrod, 1985).

The MRPIS model simulates shorter run responses (see Havens and Clayton-Matthews, 1989). Its focus is on modeling the impact of program changes on regional product mix and therefore on regional employment. Program changes, such as welfare reform or tax reform, are simulated first on a microlevel database with a microsimulation model, and marginal propensities to consume different goods are then applied to the changes in household income on the micro file. The implied region-specific changes in consumption demand are then applied to a regional input-output matrix to obtain regional employment demands by occupation. These changes in employment are then allocated across the individuals in the original micro file, which in turns generates income changes that are used to begin the process anew. The model iterates to convergence. All prices and wages are held fixed during the iteration.

Issues and Trade-Offs

Many of the same issues and trade-offs discussed for the incorporation of behavioral response in microsimulation models apply to the incorporation of

[18] David (1991) reviews the general equilibrium modeling that was conducted by the U.S. Department of the Treasury as part of the debate on tax reform. He asserts (p. 784) that this approach is sophisticated in that it "integrates consumption, saving, and factor supply decisions within a household population. And it models the total response of the economy to a change in tax structure." However, he notes a number of problems in the ability of general equilibrium approaches thus far to model in a satisfactory manner the complex behavior of actors in the economy.

Copyright National Academy of Sciences. All rights reserved.

second-round effects. First, the potential empirical importance of second-round effects should be a primary determinant of whether attempts to incorporate such effects are undertaken. Second, the uncertainty surrounding the estimates must be an important concern. Third, the best manner to present estimates of second-round effects to policy makers—if the effects are of sufficient magnitude and the estimates of sufficient quality—is an important issue.

The second issue, that concerning uncertainty, is likely to be even more important for the modeling of second-round effects than for the modeling of first-round behavioral effects. For example, CGE models include simulation of behavioral effects in all markets on both supply and demand sides, and hence must rely on estimated elasticities for many different relationships. The range of uncertainty created by a simulation from such a large number of uncertain parameter values is likely to be larger than that of a first-round simulation, by one or more orders of magnitude, because the uncertainty is likely to increase multiplicatively rather than additively in the number of parameters. The MRPIS model also involves a large number of parameters, including detailed marginal propensities to consume, in order to project changes in consumption, and it also has a large number of parameters from a multiregional input-output matrix. The model uses estimated econometric equations to allocate changes in labor demand across the micro units as well. Although the literature on these models, particularly the CGE literature, has discussed issues of uncertainty, too little attention has been paid to the development of methods by which uncertainty can be quantified. This inattention reflects a more general lack of attention to the issue of model validation.

The addition of second-round effects to models also raises difficult issues relating to short-run versus long-run adjustment. For example, CGE models are best thought of as relatively long-run simulations, but such models rarely provide guidance about the time horizon for full adjustment or the dynamic path of the adjustment process. Policy makers will naturally be quite concerned with these issues.

A related issue is the nature of the counterfactual scenario that is being simulated by second-round models. For example, the MRPIS model simulates increases in transfer program benefits to increase national output and to lower the unemployment rate, and the simulated estimates reflect this change as well as the change in distribution of employment across different skill classes. Yet policy makers are unlikely to be directly interested in the fiscal stimulus provided by a particular change in a transfer program, because fiscal stimulus could be provided in many other ways as well. This issue has been discussed for some years in the economic theory of tax incidence; it is often proposed that programs be evaluated on the basis of their distributional effects, while total government expenditures are held constant.

Given these difficulties with second-round models, it is worth considering whether there is any middle ground between the polar extremes of completely

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

ignoring second-round effects and fully incorporating them into microsimulation models through a process of iterating with a second-round model. Such iteration is likely to be difficult to carry out in an appropriate manner with any degree of confidence in the results. The quality of the simulation depends critically on the proper allocation of outcomes from the aggregated second-round model across the individual units in the micro model.

A middle ground could be achieved by a single iteration in which the initial aggregated outcomes from a microsimulation model are put into a second-round model, from which second-round effects on prices and quantities are estimated. Such second-round estimates would provide a barometer of the importance of such effects as well as an indication of their sign (or direction). These outputs from the second-round model could then be presented in an appropriate way to policy makers as qualifiers to initial estimates.

We are less than sanguine about the cost-effectiveness of devoting investment resources to modeling second-round effects, given the difficult modeling issues involved and the high degree of uncertainty surrounding the outputs. However, it seems advisable for policy analysis agencies to support research to determine what kinds of second-round effects of policy changes are likely to be of large magnitude and hence important to understand. It also seems useful for the agencies to explore the issues involved in attempting to link microsimulation models with second-round effects models. At a minimum, the agencies should require that microsimulation models be designed to facilitate linkages with such models. As we discuss in Chapter 8, it may be particularly important to consider modeling second-round effects of policy changes for health care issues.

Recommendation 6-8. **We recommend that policy analysis agencies support research on second-round effects of policy changes that may be important to understand. We also recommend that the agencies require that future microsimulation models include entry and exit points that could facilitate linkages with second-round effects models. However, except perhaps for health care issues, we do not recommend investment at this time in building second-round effects capabilities into microsimulation models.**

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

7

Computing Technology and Microsimulation

Microsimulation models for social welfare policies depend on the use of digital computers to process the input data and generate the simulation results. Even the simplest microsimulation models (and most such models are large and complex) require significant computing power because they process large microlevel databases, mimic the features of complex government programs, and apply probabilistic techniques to simulate the behavior of individual decision units. The historical development of microsimulation models as a policy analysis tool is closely intertwined with the expansion of computing technology over the past 30 years.

The major microsimulation models in widespread use today are designed for computing environments that are rapidly becoming obsolete and that impose significant costs for use of and access to the models. Although computing costs per simulation run have been greatly reduced with the hardware and software platforms used by the models in the 1980s in comparison with those used by the models in the 1960s and 1970s, other costs, such as those for staff and for modifying models to respond to new policy initiatives, remain high.

In this chapter, we briefly review the history and characteristics of the computing environments for two widely used models, DYNASIM2 and TRIM2. We then consider the potential of new hardware and software technologies for improving microsimulation models. There are exciting developments on the horizon that offer the prospect of building a new generation of microsimulation models that are much more flexible and accessible to policy analysts. Of course, implementing new computer technology is never easy or without pitfalls. There

Copyright National Academy of Sciences. All rights reserved.

are many examples of spectacular failures—involving large sums of money—from technological initiatives in both the public and the private sectors. We recommend a strategy that we hope will minimize the risks and maximize the payoffs from investment by policy analysis agencies in new computing platforms for microsimulation modeling.

THE EVOLUTION OF MICROSIMULATION COMPUTING PLATFORMS

DYNASIM2

In the late 1960s, Guy Orcutt and his colleagues obtained funding to begin work on a full-scale dynamic microsimulation model at the Urban Institute, building on his pioneering work to develop dynamic microsimulation techniques for research and policy analysis in the late 1950s.[1] The first version of DYNASIM, completed in 1975, represented an ambitious effort to simulate all major demographic and economic life events, including birth, death, marriage, remarriage, unemployment, and migration. DYNASIM was intended not only for policy analysis, but also as a general social science research tool.

DYNASIM's computer software, the Microanalytic Simulation of Households (MASH) system, written in FORTRAN for a DEC system-10, qualified as state of the art. The software system was interactive and provided an extensive command language structured to allow researchers to work directly with the model, its database, and each other. The MASH system included an on-line attribute library, a machine-readable codebook, and on-line dictionaries for each system user. The population data for the model were stored in virtual memory rather than on magnetic tape, allowing nonsequential direct access to each person and family in the file.

Yet the system had serious shortcomings for policy analysis and research. The virtual memory simulator and its disk space requirements could not handle large sample sizes. Moreover, the cost and elapsed time required to process a large number of persons in an interactive time-sharing environment was excessive: it could take from 4 to 12 hours to complete a single year's projection.

In the late 1970s the Urban Institute obtained funding to write a new version of the model, DYNASIM2, which was designed with a narrower focus. DYNASIM2 includes elements of the original DYNASIM model, the PENSIM model developed by James Schulz to simulate private pension alternatives, and other features specifically geared to simulating long-range forecasts of earnings and family histories needed for analysis of retirement income policy issues. The

[1] The discussion in this section draws heavily on Zedlewski (1990); see also Ross (in Volume II), for background on the evolution of DYNASIM2.

Copyright National Academy of Sciences. All rights reserved.

software system for DYNASIM2 is much simpler than that for DYNASIM: it is also written in FORTRAN, but processes the population data sequentially in a binary integer format and is optimized for batch processing with a mainframe (or large mini) computer. The software includes some of the original features of the MASH system, such as an on-line time-series dictionary and a machine-readable codebook. It has a simplified command structure for the user.

The DYNASIM2 software can process large data files much more inexpensively than its predecessor. Processing costs increase during simulation runs, because DYNASIM2 adds persons to the database and records new data elements for each person in each simulation year, but the costs do not become prohibitive. For example, the number of CPU (central processing unit) minutes for a run of the family and employment history module increases by about 50 percent from projections in the early 1990s to those in 2030. DYNASIM[2], in contrast to the original DYNASIM, has also proved to be fairly portable across types of computers.

TRIM2

TRIM2 is the latest version of a cumulative, 20-year effort to model accurately and efficiently income support and taxation systems in the United States. 2 RIM (Reforms in Income Maintenance), the first model in this lineage, was designed and developed in 1969 by President Johnson's Commission on Income Maintenance. RIM was programmed in FORTRAN for CDC (Control Data Corporation) computers. It lacked any overall architecture or framework, and the virtual absence of documentation made it difficult to use and modify. The Urban Institute attempted to build a more useful system by modifying RIM, but had to abandon the effort. Instead, the Institute built an entirely new system called TRIM.[3]

TRIM was programmed in FORTRAN, with some use of assembly language, for running on an IBM mainframe computer. The software system was designed to enable TRIM to

- run with large microlevel databases that could support reporting simulation results for very disaggregated groups;
- simulate a wide variety of tax and income support programs through a modular structure;
- simulate programs that use different filing units in the same model run;
- model alternative versions of a single program or different program interactions in a single model run;
- support flexible specification of program characteristics to be simulated;

[2] The discussion in this section draws heavily on Cotton and Sadowsky (in Volume II) and Webb, Michel, and Bergsman (1990).

[3] TRIM was also the basis for the MATH model developed by Mathematica Policy Research, Inc.

Copyright National Academy of Sciences. All rights reserved.

- facilitate making changes to the computer code to simulate new programs or features of programs that could not be modeled simply by setting one or more existing parameters;
- execute runs in a timely fashion; and
- provide adequate documentation for programmers, analysts, and users.

These characteristics were valuable, but the heavy demands placed on TRIM for policy analysis in the mid-1970s left little time for following good practice with regard to structured programming and adequate documentation, with the result that TRIM increasingly came to be viewed as difficult to understand and use. Hence, when ASPE needed to produce estimates for the Carter administration's proposed Better Jobs and Income Program, the agency decided to develop a special-purpose in-house model that could simulate a linked public jobs and cash strategy, rather than asking the Urban Institute to make the necessary modifications to TRIM. The resulting KGB model was developed in approximately five weeks by using data files derived from TRIM and simplified tax and transfer program operating characteristics. Although it was used heavily for several years, KGB again exhibited the worst features of RIM. It lacked a generalized software framework, its operating efficiency was just barely adequate, and the system could be used only by its creators.

In the late 1970s the Urban Institute received funding to redesign TRIM to improve overall efficiency, improve the printed output, and develop new data structures and processing procedures in order to make the system easier to use by analysts and easier to develop and modify by programmers. The old TRIM software is still used to preprocess each year's March CPS file. The new TRIM2 software, also written in FORTRAN for an IBM mainframe computer, then takes over to complete the processing of the input data and to run program simulations.[4]

NEW DEVELOPMENTS IN COMPUTING TECHNOLOGY

Both DYNASIM2 and TRIM2—as well as other major microsimulation models that are widely used today such as MATH—are optimized for batch processing with mainframe technology (or, as for DYNASIM2, a large minicomputer). Their software system developers have focused on enhancing processing efficiency and reducing computer costs through such techniques as data compression, variable transposition, and overlaying run phases in memory. They have also assigned high priority to techniques for maintaining control over the simulation process. For example, all parameter definitions and their values are predefined in the central TRIM2 directory.

[4] See Figure 1 in Citro (in Volume II) for an overview of the steps involved in creating a new TRIM2 baseline file from the March CPS, by using the TRIM and TRIM2 software. The MATH model software was also rewritten at about the same time to optimize processing efficiency.

Copyright National Academy of Sciences. All rights reserved.

But, the effort to achieve processing efficiency and centralized control has resulted in less than optimum performance on other dimensions. Total turnaround time can be slow. For example, only two or three TRIM2 production jobs can generally be run during a day, and, indeed, large jobs are typically run overnight both to take advantage of cost discounts and because the scheduling algorithm of the computer center would be unlikely to allow them to be run during the day. Although the software redesign for models such as TRIM2 and DYNASIM2 made them easier to use than their predecessors, analysts, whether on the staff of the sponsor agency or the modeling contractor, still almost always rely on programmers to prepare runs for submission. This is true for runs that involve simply setting parameter switches as well as runs that involve modifying the model code to handle new program options. The investment in learning how to work with the models directly is prohibitive for most analysts.

Just about 10 years ago, a new technology based on individual desktop personal computing burst on the scene.[5] Although actual implementation has lagged somewhat behind the promises and perception, it is arguable that the widespread adoption of microcomputer technology has revolutionized the way in which people engaged in basic or applied research and development carry out their work. Analysts have been able to interact in real time with their data and analysis systems without leaving their desks and without the need for programmers or systems analysts to serve as intermediaries. Linking individual microcomputers through local and remote electronic networks has preserved flexibility for individual researchers while enabling groups of researchers to work collaboratively and share ideas and information.

Some agencies and researchers have already begun making use of microcomputers as platforms for microsimulation modeling. The CORSIM (Cornell Simulation) model, developed by Steven Caldwell, is a dynamic model based on DYNASIM2 written in the C language that is currently being used on an IBM mainframe and on MS-DOS microcomputers. In a related development, Mathematica Policy Research has developed a PC-based benefit-calculator model of the food stamp program for use by the Food and Nutrition Service (FNS), while the Urban Institute has developed a PC-based benefit-calculator model of the AFDC program for use by ASPE. Statistics Canada several years ago instituted a major microsimulation modeling effort that led to the Social Policy Simulation Database/Model (SPSD/M)—a static model of Canadian household tax and transfer programs, written in C for an MS-DOS microcomputer, that is available to the general public (Statistics Canada, 1989).

---

[5] By "personal," "desktop," and "micro" computing, we mean computers that not only are physically located on the analyst's desk (or close by in the work area) but also incorporate high-bandwidth communications, which enable the user to interact directly with the data or text files in real time. We frequently use the term "workstation" for this technology. We do not include remote "dumb" terminals, which simply let a user enter commands locally, to be sent to a distant computer (or receive output locally), but do not enable the user to manipulate the information.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

For policy analysis generally, microcomputing technology has come into widespread use. Policy analysts have become accustomed to placing heavy reliance on personal computer hardware and such software as spreadsheets to develop ad hoc, special-purpose models to respond to particular policy issues. As described in Chapter 2, CBO and ASPE analysts relied heavily on personal computer-based ad hoc models to generate many of the estimates for the debate on the Family Support Act (FSA). Analysts value highly the flexibility to develop their own models and the facility that personal computer-based models afford them for timely response to changing policy issues. In contrast, some of the changes required to enable TRIM2 to handle particular features of the FSA could not be implemented in time to provide input to the policy debate.

Many aspects of the computer implementation of microsimulation models could usefully be reviewed to determine approaches that might improve the models' cost-effectiveness. For example, the capabilities for data management, retrieval, and documentation afforded by relational database management systems (see David, 1991) may have applicability to microsimulation models. Similarly, high-level languages developed especially for simulation may have application to microsimulation. Our review of computing issues for microsimulation, however, has focused on the potential for innovations in microcomputer-based hardware and software to improve the next generation of microsimulation models.

We believe that microcomputer platforms have the potential to enable microsimulation models to provide more flexible, timely responses to policy questions; to facilitate validation of models and their outputs; and, generally, to make the models more accessible to users. Although microsimulation models (along with other formal models) offer important benefits to the policy process, we fear that they will not be competitive with special-purpose ad hoc models for many policy needs so long as they remain relatively inaccessible to the analyst. For the same reason, we fear that microsimulation models will remain less well utilized and studied by academic researchers and, hence, that there will be less useful feedback for the improvement of policy models (see Chapter 11).

Alternative Computing Platforms: SPSD/M Versus TRIM2

The panel commissioned a comparative evaluation of a microcomputer-based tax and transfer simulation model, SPSD/M, and a mainframe-based model, TRIM2 (reported in Cotton and Sadowsky, in Volume II). This study was not intended to be a representative sample but rather to give a concrete comparison of major options. The study found advantages and disadvantages to both models' design features and computing platforms; however, the clear weight of the evaluation was in favor of the microcomputer-based platform of SPSD/M:

Copyright National Academy of Sciences. All rights reserved.

- The interactive SPSD/M system permits users to specify interactively what a model run is to do and to interrupt the execution of the model to check intermediate results, a capability that the TRIM2 system does not support.
- An analyst or programmer would probably be able to complete a research project at lower cost and in less elapsed time using SPSD/M rather than TRIM2, because of the interactive features in SPSD/M and its reasonably efficient processing (about 20 minutes of wall-clock time on a Compaq Deskpro 386/20 to run a complete simulation) coupled with the ability to run at any and all times of the day. In contrast, production runs for TRIM2, although requiring only 5 to 10 minutes of IBM System 3090 CPU time, must typically wait to be scheduled during the day or run overnight.[6]
- The C language in which SPSD/M is written permits it to be moved to the next generation of personal workstations: for example, the current system could be ported to a UNIX workstation with less than 1 week of development effort. In contrast, moving TRIM2 away from its IBM MVS environment would involve a major software development project. Also, the SPSD/M user interface could be adapted relatively easily to a graphic user interface; however, such an adaptation is not even appropriate for the TRIM2 mainframe environment.
- SPSD/M is attractive to new users because of its ability to interface with other MS-DOS software packages, such as Lotus 1-2-3 and PC SAS. TRIM2 can produce SAS input code, and there has been some experience in downloading TRIM2 data to a personal computer for use in Lotus 1-2-3; however, the interface with personal computer software packages is inconvenient to carry out.
- The SPSD/M user community has grown as more individuals and organizations became interested in using the model to participate in the tax reform debate in Canada. Most of the new SPSD/M users are using the model in the "black box" mode, that is, a mode in which they do not seek to understand or modify the program code. These users receive no formal training from Statistics Canada, but they appear to be using the model successfully with the SPSD/M documentation and the model's interactive mode of operation. Several outside researchers and consultants are modifying the model's internal structure (i.e., they are using the model in the "glass box" mode) with minimal assistance from Statistics Canada. TRIM2, in contrast, is currently used almost exclusively by the technical programming staff at the Urban Institute, and no outside researchers are attempting to enhance the model by modifying the internal FORTRAN source code.

[6] Cotton and Sadowsky did not conduct a rigorous cost comparison of TRIM2 and SPSD/M. Hence, their conclusion regarding the comparative cost-effectiveness of the two models should be regarded **as** provisional.

Copyright National Academy of Sciences. All rights reserved.

New Computing Capabilities Versus Microsimulation Requirements

Computer systems for microsimulation models must effectively perform several key functions, including: storing the microdata in an efficient and cost-effective manner, managing the database and associated documentation, ensuring logical consistency among individual modules and operating characteristics within the model, providing a means to link modules for particular applications, and providing an effective user interface. Good microcomputer-based solutions do not yet exist for supporting all of the necessary functions; however, current developments appear to promise considerably more effective computing environments than are now used for microsimulation modeling activities.[7]

The computing platform must be able to run large and complex models. By 1995, desktop workstations should be available with the capability to support microsimulation models of considerable size. Single processors are expected to range in speed between 40 and 100 millions of instructions per second (MIPS); hence, the speed of executing microsimulation model runs will no longer be a significant issue. To the extent that parallel processor architecture becomes available for workstations in such a way that tasks can be easily divided to run in parallel, the increase in speed will be multiples of what a single processor will be able to achieve.

For efficient processing, mainframe microsimulation models have compressed the input data in various ways. The desktop workstation environments of 1995 are likely also to require some type of compression. However, primary RAM (random access) memories of 32-64 megabytes are expected to be common, so that compressed files of the size currently processed by TRIM2 may well fit within primary memory during the course of a simulation. With respect to secondary storage devices (e.g., hard disks), current developments indicate that secondary storage on workstations in 1995 will be at least as good as secondary storage on current medium-size mainframes.

The history of electronic computing has been characterized by hardware developments leading software developments. However, the workstation software environment in 1995 is likely to feature several elements of importance to microsimulation modeling activities. For example, powerful graphical user interfaces should greatly facilitate direct experimentation by an analyst in simulating alternative policy proposals, extending model capabilities, and performing sensitivity analyses and estimates of variance. These interfaces are characterized by icons, windows, and the use of "point and click" tools that enable users to work more effectively and easily with complex models and data. Similarly,

[7] The discussion in this section draws heavily on a detailed assessment of the likely trends in microcomputer capabilities in relation to the computing requirements of microsimulation models; that assessment was part of the study conducted for the panel by Cotton and Sadowsky (in Volume II). We note that Cotton and Sadowsky's review was carried out in 1989; the development of cost-effective microcomputing capabilities may be proceeding at an even faster pace than they forecast.

Copyright National Academy of Sciences. All rights reserved.

substantial advances in computer-assisted software engineering (CASE) tools—which are almost always based on a graphical user interface and embody the notion of a construction kit approach to model design—should magnify the productivity of software designers as well as users.[8]

In sum, there appears to be no reason that applying microcomputer technology to the needs of microsimulation modeling on a large scale should not be feasible in the near future and should not bring the same kinds of important benefits that such technology has brought in other contexts to researchers and analysts. Chief among these benefits is the ability for users, as well as programmers, to interact directly with models and data in a manner that encourages experimentation and use. Arguments against moving to microcomputer technology because of the limited storage capacity of microcomputers (both immediate access and secondary storage) and slow processing times are rapidly losing relevance with the pace of change in the microcomputer world.

Another argument questions the utility for policy analysis of encouraging many people to engage in modifying and using models and thereby generating possibly inconsistent or erroneous estimates. We believe that this danger can be minimized by building extensive documentation features into new models. Moreover, we believe that the benefits from broader access to models equipped with the capabilities for ready evaluation of model estimates and ready adaptation to new policy needs outweigh the risks from a proliferation of estimates.

Although we have focused our discussion on microcomputer technology (specifically, powerful workstations), we are mindful that such platforms do not represent the only possible future for microsimulation models. The pace of change in hardware and software technology is not only rapid, but also multidirectional. It may well be that some other type of platform would be feasible and, indeed, optimal for the next generation of microsimulation models. For example, one could envision a computing environment that used graphical software tools operating on a workstation to develop and test model applications, coupled with links to a supercomputer for making production runs using large samples. Such an environment might facilitate experimentation and direct access by an analyst for many applications and also permit systems staff to run large, complex applications involving a range of capabilities, such as behavioral response or links to other models. Such an environment could also make it possible to keep track of variants of the model, thereby minimizing the problems from multiple estimates.

In our view, the overriding consideration is not the specific hardware or software technology that is adopted, but the need to design a new generation

[8] Cotton and Sadowsky (in Volume II) also note that so-called object-oriented programming systems (OOPS)—which conceptualize computer programs in terms of discrete objects that receive messages as inputs; produce other messages as outputs; and provide the basis for building other, more complex objects—may have applicability to microsimulation. For an example of an object-oriented microsimulation design effort, see Hayes (1989); see also Dabrowski, Fong, and Yang (1990).

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

of microsimulation models that meet the design criteria we have identified (see Chapter 6). As we recommend, new models need to provide enhanced flexibility; enhanced accessibility; the ability to generate clear and complete documentation; the ability to evaluate model components as well as the overall model; and, finally, acceptable cost and time for development and use. To achieve these goals, we believe, requires soon leaving behind the current computing environment for microsimulation modeling and moving toward new technology.

FUTURE DIRECTIONS FOR COMPUTING IN MICROSIMULATION

We are excited about the prospect that powerful new hardware and software technology will make possible a new generation of microsimulation models that support increased modeling capabilities and flexibility for analysts and thereby attract a new, broader user community. In considering what microcomputer hardware is likely to be available by 1995, it seems probable that the computing power currently required for a TRIM2-equivalent simulation (including CPU, memory, and secondary storage) will be available on a workstation costing no more than $10,000. Linked microcomputer and mainframe systems that provide useful functionality may also be available.

Considering software developments that could apply to workstations (and possibly other platforms), there are promising innovations noted above such as graphical user interfaces. However, these software environments are not yet settled. In particular, no market leader or industry standard has yet emerged. It may be several years before it would be prudent to choose the software environment in which to make substantial investments in microsimulation model development.

Given the need for improved functionality of microsimulation models, while recognizing the risks that always accompany an investment in new technology, we urge policy analysis agencies to proceed both resolutely and cautiously to explore the potential of new computing technology. It is probably a mistake to plan immediately to port an entire existing model (such as TRIM2) to a new computing platform (such as that used by SPSD/M), in part because there are aspects of the SPSD/M computational design that could be improved. More over, some of the potentially most fruitful software developments, such as graphical user interfaces, are still in an early stage of development. Finally, to port TRIM2 to a radically different environment could entail substantial costs with little offsetting benefits. As Cotton and Sadowsky concluded (in Volume II):

> The basic difficulty in extracting benefits from a desktop version of TRIM2 goes back to the . . . [fact that TRIM2 and other models like it] have all been optimized to run on a central mainframe computing system that relies primarily on batch processing. Moving these systems into a very different

Copyright National Academy of Sciences. All rights reserved.

> environment minimizes their operational strengths and exposes their lack of ability to exploit the new environment. The benefits of porting TRIM2 in its present form . . . are moderate at best and justify neither the real costs involved nor the opportunity costs of preempting investment resources that could better be used elsewhere.

In other words, if substantial resources are put into creating a personal workstation version of TRIM2 (or MATH), there is the real danger that the opportunity for making a breakthrough in microsimulation model technology will be lost and the pattern of the past 20 years—with present needs crowding out medium-term investments, whether they be in adding capability, assessing uncertainty, or improving the software design—will repeat itself once again.[9]

We believe that the best way to proceed at this time is by taking a series of relatively small steps, particularly until the directions of new software (and hardware) developments become clearer. First, policy analysis agencies could well make some modest investments in developing workstation-based front-end tools for helping programmers modify and manage the existing mainframe-oriented models in a more cost-effective manner. Second, it is not too soon for the agencies to consider how to translate the large volume of code in models such as TRIM2 and MATH, which embody the various social welfare program accounting schemes and rules of operation, to a new computing environment with minimum error.

Third, and most important, the agencies should invest in developing prototypes of both static and dynamic models using the best available new hardware and software technology. The prototype development could seek to mimic, in a very skeletonized form, the basic functions of a model such as MATH or DYNASIM2 (or of a component module, such as the simulation of SSI or AFDC), and then to add to these functions one or another important new capability. For example, a worthwhile prototype to develop and test would be a model that makes it possible for analysts to conduct validation studies, such as sensitivity analyses or estimates of variance, without programmer intervention. Another prototype might be one to make it possible for analysts to readily alter the aging or imputation routines that are used to construct the database. The object would be not to produce realistic simulations of policy changes, but to test new ways to enhance functionality such as increased flexibility and accessibility.

Recommendation 7-1. **We recommend that policy analysis agencies invest resources in developing prototypes of static and dynamic microsimulation models that use new computer technologies to**

[9] We understand that ASPE recently investigated the ability of TRIM2 to run in a personal computer version of FORTRAN and was pleased with the results. However, we still believe that what is called for is not simply moving a model such as TRIM2 in its entirety to a microcomputer but, rather, designing a new version of TRIM2 or component modules to take maximum advantage of hardware and software features for increasing flexibility and accessibility to users.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

provide enhanced capabilities, such as the ability for a wider group of analysts to apply the models; conduct timely and cost-effective validation studies, including variance estimation and sensitivity analyses; and alter major components, such as the aging routines, without requiring programmer intervention.

At the same time, the agencies need to keep abreast of ongoing technological developments, particularly the kinds of software, such as graphical user interfaces, that are emerging with the potential to provide substantial benefits for microsimulation. In this regard, we urge policy analysis agencies to obtain input from computer science researchers who know in what directions software technology is moving and what advantages new technology could offer for microsimulation. We also urge the agencies to keep in touch with developments abroad. Although the details of government policies and programs vary, many countries are involved in microsimulation for a range of social welfare programs and share an interest in developing cost-effective computing platforms that support increased model functionality. Recently, the United Kingdom, under the auspices of the Working Party on Social Policy of the Organization for Economic Cooperation and Development (OECD), organized a panel of interested countries to look at the current state of the art in the use of microsimulation methodology for analysis of social policies and taxation. The activities we recommend that U.S. agencies undertake to investigate and invest in new technology for microsimulation could well be coordinated with the activities of this OECD panel.[10]

Once experience has been gained with prototypes (we urge the agencies **to** move rapidly in their development) and the software environment has settled down so that reasonable choices can be made, agencies can move ahead with efforts to develop new microsimulation models based on next-generation computing technology. We urge agencies in the United States to form a broad consortium for this work and to consider involving interested agencies from other countries that have strong commitments to microsimulation modeling as a policy analysis tool.

Recommendation 7-2. **We recommend that policy analysis agencies, after experience with prototypes and reviews of developments in computer hardware and software technologies, make plans to invest in a new generation of microsimulation models that facilitate such design criteria as user accessibility and adequate documentation and evaluation of model components, as well as computational efficiency.**

---

[10] ASPE participated in a recent conference of this group.

Copyright National Academy of Sciences. All rights reserved.

8

Microsimulation Modeling of Health Care, Retirement Income, and Tax Policies

Our discussion to this point has focused largely on the problems of modeling income support programs such as AFDC and food stamps. Microsimulation techniques are also appropriate and have been used extensively in other social welfare policy areas. Many of the issues that we raise with regard to data inputs, model design, and computing technology are common across policy areas, although each area presents some unique features and problems. In this chapter we briefly discuss special issues in microsimulation modeling of health care, retirement income, and tax policies. Because we did not review the models and data in these areas in as great depth as those in the income support area we make few specific recommendations; instead, we do raise issues that we believe are particularly important to address.

One general question that arises is the relative weight to give to investments in microsimulation models for different policy areas. Because we do not pretend to have any particular expertise in foreseeing the future mix of policy issues, we cannot offer unequivocal advice on this question. We note, however, that health care policy is an area of growing importance because of the escalating costs of providing health services and the evidence of glaring gaps in the health care system, such as the large population not covered by private or public health insurance. Moreover, as we indicate below, available data, research, and models for health care policy analysis exhibit many inadequacies relative to the information needs.

Yet we believe it would be unwise to concentrate investment resources on any one set of issues. Welfare policy provides a cautionary example in this regard. After the collapse of the Carter administration's push for the Better Jobs

Copyright National Academy of Sciences. All rights reserved.

and Income Program in 1977-1978 and the subsequent focus of the Reagan administration on restricting welfare benefits, one might well have concluded that capabilities for modeling significant welfare reform initiatives would be of little importance. That conclusion would have been wrong, as evidenced in our review of the policy debate that led to the Family Support Act (FSA) of 1988 (see Chapter 2). Indeed, failure to invest in improvements to models and data handicapped the ability of policy analysts to use microsimulation techniques to develop estimates for many of the proposed welfare policy changes considered in the FSA debate. We do not expect that the FSA represents the last word in welfare policy, either. Hence, we see a continued need to scrutinize income support models and data and to determine ways in which they can be improved.

Similarly, we see the need to scrutinize models and data for retirement income and tax policies, as well as for health care policies. Indeed, the problem is not so much to pick policy areas for investment—all of them are and will continue to be important—but to discern particular aspects of each broad area that are likely to assume priority in the policy debate. For example, the need in the FSA debate was for models that could link the AFDC program with new initiatives such as child support enforcement, job training, and transitional assistance programs, tasks for which the existing models and data were not well suited.

Although there are many sources of information that can help agencies anticipate future policy proposals, there is no crystal ball that will furnish them with infallible forecasts for guiding their investments in policy analysis tools. The difficulties of predicting the policy agenda underscore the importance of investments that are aimed at improving the overall capabilities of microsimulation models (and other policy analysis tools) for flexible, timely, and cost-effective responses to changing policy concerns. To achieve this goal, whether for income support, health care, or any other policy area, databases need to be broad in scope, models need to follow good design principles and practices, and agencies need to find ways to further fruitful interactions between policy research and modeling.

HEALTH CARE POLICIES

Some of the reasons that health care policy issues are of continuing and increasing concern to decision makers are evident from the following selected indicators:

- Total public and private spending for health care in the United States, which currently amounts to more than $600 billion, increased from 7.3 percent of the gross national product in 1970 to 11.6 percent in 1989. Over this same period, the proportionate share of national health care costs assumed by the public sector increased from 33 to 40 percent; the costs of the federal Medicare program rose by 300 percent (in real terms) to $102 billion, and the costs of

Copyright National Academy of Sciences. All rights reserved.

federal-state public assistance programs for health care (chiefly, Medicaid) rose by 215 percent (in real terms) to $67 billion.[1]

- While the consumer price index (CPI) increased overall by 200 percent from 1970 to 1988, the medical care component of the CPI increased by 300 percent over the same period (Bureau of the Census, 1990b:Table 762).
- In the fourth quarter of 1988, an estimated 31.5 million people, or 13 percent of the total population, were not covered by health insurance of any kind, either private insurance, Medicare, or Medicaid (Nelson and Short, 199$1$2$3).
- Over the period 1984-1989, total diagnosed AIDS cases increased from 4,000 to 78,000; federal spending for AIDS rose from $0.06 to $1.3 billion; and state spending for AIDS rose from $0.01 to $0.2 billion (Bureau of the Census, 1990b:Table 188).
- A recent microsimulation study projected that, over the next 30 years (1986-1990 to 2016-2020), the elderly (people age 65 and older) will increase from 31 to 50 million; those elderly receiving long-term nursing home or home care services will increase from 6.3 to 10.4 percent of the total; total public and private expenditures for nursing home services will increase by 197 percent, **to** $98 billion; and total expenditures for home health care services will increase by 154 percent, to $22 billion (Rivlin and Wiener, 198$1$2$30-11) (all dollar amounts in constant 1987 dollars).

These indicators and others underscore the policy interest in the health care area.[2]One pressing set of issues revolves around how to manage and contain what appear to be runaway costs for medical services. These costs are driven by many factors—ranging from the development of expensive new medical technology and treatments, to demographic and socioeconomic changes in the population, to the demand incentives for medical care that result from public and private health insurance programs. Another equally pressing set of issues revolves around how to ensure that people who need health care services will have access to them at a reasonable cost.

Before proceeding, we should make clear that our discussion of health-related policy modeling is limited to cost and coverage issues pertaining to the provision of health care services. Another topic that we did not take up but that deserves serious consideration is the use of policy analysis tools for modeling the relationships of health care interventions and other determinants of health status to health outcomes in the population and for estimating the cost-effectiveness

---

[1] All figures are from Bureau of the Census (1991 :Table 136). Total national health care expenditures include costs of medical research and construction of medical facilities, in addition to direct health care costs. Percentage increases in real terms were determined by using the GNP implicit price deflator for government purchases of goods and services (calculated from Bureau of the Census, 1991 :Table 767).

[2] The discussion in this section benefited greatly from papers prepared for the panel by Chollet (1990), who reviewed some of the existing health care policy models, and Grannemann (1989), who reviewed issues in modeling behavioral responses to health care policy changes.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

of alternative treatments and the feedback effects on overall costs of the health care system. It is clear that expenditures on health care may not always translate into improvements in health. (Thus, although the United States spends a higher proportion of its GNP on health care than other industrialized countries, it ranks lower than many countries in such health indicators as infant mortality and life expectancy; see Bureau of the Census, 1990b:Tables 1440,1444.) Hence, it is important in assessing alternative health care policies to consider not only the costs in terms of payments for services, but also the costs and benefits in terms of the effects on health outcomes. This topic presents challenging modeling issues that we have not investigated, but it seems likely that microsimulation techniques, with their ability to model complex relationships and individual circumstances, could make potentially useful contributions.

We note in this regard that demographers have recently developed complex stochastic models of disease processes and disability states, using hazard techniques with longitudinal data such as the Framingham Heart Study and the 1982-1984 National Long-Term Care Survey (see, e.g., Manton and Stallard, 1990; Manton, Woodbury, and Stallard, 1990, 1991). (The National Institute on Aging has supported much of this work.) They have used the resulting equations to analyze a number of policy-relevant issues. For example, they have analyzed the implications for the age structure of the population and total versus active life expectancy (i.e., years free of disability) of alternative assumptions about the elimination of particular diseases or risk factors (e.g., smoking) in the population. They have also examined the savings in nursing home and home health care costs that might be achieved by eliminating such diseases as Alzheimer's. To date, researchers working in this field have largely applied their estimated parameters to cell-based models, such as life tables, to analyze alternative scenarios. It may well be that putting this type of risk-factor analysis in a microsimulation framework and, further, effecting a linkage with microsimulation models of health care financing and coverage issues could have potential payoffs for analysis of health policy issues.[3]

Microsimulation Modeling for Health Care Policy

Microsimulation has played a role in analysis of health care cost and coverage issues since the technique was first introduced to the political process. The RIM model was used in the late 1960s to estimate the costs and distributional effects

[3] Wolfson (1989b, 1991), writing from a Canadian perspective, asserts that the health statistics system has given too much attention, relatively, to information about inputs to the health care system, including financial costs, and too little attention to "the supposed outputs of the system, namely the health status of the Canadian population." He argues for microsimulation models that would estimate the impacts of alternative health care policies and programs on health status as well as models of costs of and access to medical services. At present, Statistics Canada is developing a Population Health Model that includes microsimulation components for disease processes and health care interventions.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

of alternative national health insurance programs (Orcutt et al., 1980:85). A Medicaid module was developed for the MATH model in the late 1970s (Pappas, 1980). A microlevel database of households from the 1976 Survey of Income and Education and the 1976 Survey of Institutionalized Persons formed the population component of the ASPE Health Financing Model that was used to estimate alternative national health insurance programs during the Carter administration (Office of the Assistant Secretary for Planning and Evaluation, 1981).

However, throughout the 1970s and early 1980s, microsimulation techniques were applied less frequently to health care policy analysis than they were to other social welfare policy areas. None of the microsimulation models developed for health care policy gained the kind of widespread use enjoyed by models such as MATH and TRIM2 for analysis of tax and transfer programs or models such as DYNASIM2 and PRISM for analysis of public and private pension systems.

Recently, the use of microsimulation techniques for health care policy analysis has shown signs of new life. A module in TRIM2 to simulate the costs and distributional effects of expanding Medicaid coverage was recently redesigned and updated (see Holahan and Zedlewski, 1989); in addition, work is under way, sponsored by the Department of Labor, to add capabilities to TRIM2 to model employer-provided health insurance benefits. A major expansion of the PRISM model was effected to simulate alternative financing programs for long-term care of the elderly (Kennell and Sheils, 1986; Kennell et al., 1988; Rivlin and Wiener, 1988). DYNASIM2 has also been used to look at longterm care issues. The developers of PRISM recently built the Health Benefits Simulation Model, a comprehensive model for the household sector designed to simulate health insurance coverage, health services use, total health care spending, and sources of payment among the noninstitutionalized population (see Chollet, 1990). CBO has developed a microsimulation model for simulating changes in Medicare benefits, based largely on Medicare administrative records, that was used to estimate the costs and distributional effects of alternative ways to insure against catastrophic health care costs under the Medicare program (Congressional Budget Office, 1988).[4] CBO has also developed models of Medicare and Medicaid eligibility. The Health Care Financing Administration (HCFA) has sponsored work by the Actuarial Research Corporation to apply static aging techniques to update the 1977 National Medical Care Expenditure Survey (NMCES) and to use the resulting database to simulate policy issues,

[4] Catastrophic coverage for Medicare beneficiaries was enacted in 1988 but repealed the next year. Decision makers in Congress had underestimated the antipathy of relatively well-off elderly people toward paying income tax surcharges to finance the program. In addition, revised CBO estimates of costs to the government for reimbursing catastrophic prescription drug charges, produced from its Medicare microsimulation model by using newly available data on growth in prescription drug usage, were more than double the original estimates (see further discussion in text).

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

such as the effects of extending Medicaid coverage. (Wilensky, 1982, originally proposed basing a comprehensive microsimulation model for health care issues on the 1977 NMCES.)

All of these health care policy microsimulation modeling efforts have dealt with the household sector and primarily with issues of expanding health insurance coverage and the associated costs to the federal and state governments for reimbursing medical care charges. Microsimulation-based models have also been developed to examine issues that affect the supply side of the health care market. For example, CBO developed a model, based on Medicare administrative records of payments to physicians, augmented with county-level data, that has been used to estimate the effects of changing fee schedules for different types of medical specialties and geographic areas (Congressional Budget Office, 1990). (Legislation to alter fee schedules, which is to take effect in 1992, was recently passed.) CBO has also developed a model, based on data for each of the hospitals in the United States, that is designed to estimate the cost and distributional effects of changing various provisions of the Medicare prospective payment scheme used to reimburse hospital costs. For example, the CBO model could determine effects by geographic area, type of ownership of the hospital (private, nonprofit, etc.), and hospital size (number of beds). The Health Care Financing Administration has a similar prospective payment scheme model.

In looking back over the past 20 years, however, it is clear that microsimulation models have played a distinctly subordinate role in health care policy analysis. Moreover, many of the microsimulation models that exist were developed on an ad hoc basis for special purposes and are neither well documented nor used outside the agency that developed them.[5]

Cell-based models, often with links to macroeconomic models, have played a much more prominent role in health care policy analysis. The Health Resources Administration supported development of a cell-based health care sector simulation model in the 1970s (Yett et al., 1980). The model, which never went beyond the prototype stage, represented an ambitious effort to relate demand and supply relationships in the health care market. It included submodels for projecting the population of consumers, classified by demographic and income categories; the supply of physicians, classified by age, specialty, and a few other characteristics; quantities and prices of physician services; quantities and prices of hospital services; and the supply of nonphysician health personnel.

The ASPE Health Financing Model was essentially a large cell-based

---

[5] In this regard, Chollet (1990:2) notes: "Other microsimulation efforts [in addition to CBO's in the health care area] have been mounted by the Prospective Payment Assessment Commission (ProPAC), the Physician Payment Reimbursement Commission (PPRC), the Health Care Financing Administration (HCFA), and by the Employee Benefit Research Institute (EBRI). Like CBO's microsimulation models, each of these efforts was undertaken for specific, in-house analyses; none [is] documented for external use."

Copyright National Academy of Sciences. All rights reserved.

model. The microdatabase, developed from merging the Survey of Income and Education and the Survey of Institutionalized Persons, was projected forward in time by using static aging procedures and then aggregated into several thousand population cells, classified by family income; employment of adult members of the family; primary insurance coverage; family size; and age, sex, and disability status of the person. Data on health care utilization from the National Health Interview Survey were matched to the population data on a cell-by-cell basis. Finally, data on health care expenditures by type of medical service, source of payment, and population group were assembled from a variety of administrative sources and, in turn, matched with the population cells to generate the medical expenditure profile in the simulation year under current law. Simulations of alternative profiles altered the cell values to reflect a number of direct and indirect effects of policy changes, such as changes in the level of utilization of services in response to changes in cost-sharing and other patient payment provisions. Often, the population microdata records were retabulated to support simulation of alternative policies that affected different subgroups, but the core of the model operated on a cell basis.

The Macroeconomic-Demographic Model of Health Care Expenditures, developed for the National Institute on Aging by Lewin/ICF, Inc., is a large cellbased model for projecting health-care costs over the long term (Anderson, 1990; Cartwright, 1989).[6] The model includes a macroeconomic growth model with two goods—investment and consumer goods—and two factor inputs—labor and capital services. Interacting with the macro model are large cell-based models of population growth, the labor market, pension benefits, family formation, consumer expenditures, housing demand, health care expenditures, and health care benefits. As an example of the model's size, health care expenditures are estimated for 3,136 family groups, classified by family size, sex of head, age of head, race of head, geographic region, urban or rural residence, and whether or not covered by private health insurance. The equations used to estimate expenditure shares, labor supply, and other parameters in the model were developed by using microdata, but the model itself operates on a cell basis.

Issues in Modeling Health Care Policy Alternatives

The limited application of microsimulation modeling in the health care policy area and, indeed, the failure of any particular model(s), regardless of type, to gain widespread use for health care policy analysis result from the complexity

[6] The President's Commission on Pension Policy and the National Institute on Aging initiated development of this model in 1979. At that time, the model focused on the retirement income system and the interactions of social security and private pensions. After the 1983 Social Security Act Amendments, resources were concentrated on adding components to the model to simulate the impacts of population aging on the health care system.

Copyright National Academy of Sciences. All rights reserved.

of the issues involved and the magnitude of the modeling task. Moreover, the complexities and broad scope of health care policy issues have led to fragmented efforts to develop needed databases and research knowledge, which, in turn, have handicapped model development efforts.

Why is health care policy so daunting? For one thing, many actors are involved in health care, including:

- patients and prospective patients, in both the household and the institutionalized population;
- informal caregivers, such as relatives;
- doctors, hospitals, nursing homes, and other service providers;
- state and federal insurers and regulators; and
- private profit and nonprofit insurers of several types, ranging from traditional insurers that reimburse on a fee-for-service basis to health maintenance organizations and other prepaid medical service plans.

The interconnections of the various actors exhibit a bewildering variety that makes it difficult to assemble relevant information or even determine an appropriate unit of analysis. For example, an elderly patient, during the course of one illness, may be treated by several different specialists in one or more hospitals or other service centers and may obtain reimbursement from both Medicare and a private Medigap policy and also pay some costs directly. For some health care policy questions, it may be important to have sufficient information to analyze the spell of illness (or another broad measure) instead of working with a narrower unit of analysis such as doctor visits or hospital days. Yet, in our example, none of the service providers or insurers is likely to have complete information about the patient or about the procedures or costs involved in treating the illness (e.g., Medicare reimburses doctors and hospitals separately).

Mirroring the complexity of the health care sector, there are many federal agencies involved in health care data collection, research, and policy analysis, often with overlapping mandates and interests that do not always make it easy for coordination or progress toward a common set of goals. Agencies of the Department of Health and Human Services (HHS) with important roles in this regard include ASPE, the Health Care Financing Administration, the Agency for Health Care Policy and Research, the National Institute on Aging, and the National Center for Health Statistics, among others.

Research Knowledge

It is certainly possible to model a particular class of actors and consider policy questions that directly affect that class—for example, simulating the extension of Medicaid coverage to a broader population or the effects of higher coinsurance rates on Medicare beneficiaries. However, unlike the case with a program

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

such as AFDC, it is clearly perilous to ignore first-round behavioral responses to health care policy changes. For example, research has shown that people alter their demand for medical services in response to changes in coverage, coinsurance rates, and other provisions of health insurance plans that affect the price of services.[7] Moreover, demand responses can involve large numbers of people and thus have large effects on program costs. In contrast, although features of income support programs can certainly affect aspects of behavior such as labor supply, research has demonstrated relatively small effects for eligible groups, and these groups represent relatively small proportions of the total population.[8]

It appears likely to be perilous as well to ignore second-round effects in modeling health care policy changes. For example, changes in insurance coverage may ultimately lead to changes in hospital pricing policies, such as increasing prices for services that tend to be used more heavily by insured patients (see Grannemann, 1989:8-9). As another example, changes in public health insurance benefits are likely to affect the market for private health insurance and thereby influence total and public-sector health costs in unforeseen ways.[9] As still another example, it is widely acknowledged that physicians have a large influence over levels and costs of medical services because of such factors as the limited information available to patients on prices and services. Hence, it is important to consider physician behavior in evaluating alternative health care financing and payment policies.

The research knowledge that could support modeling the complex health care sector has many inadequacies. There are many gaps and deficiencies in understanding the behavior of consumers and providers of health care, including evidence on which to base estimates of physician response to changes in payment schedules. There are some studies showing that the decisions of physicians to participate in Medicaid are responsive to the level of Medicaid reimbursement rates. Other studies suggest that state restrictions on Medicaid payments may reduce the number of visits to physicians in private offices but increase visits to other, more expensive, ambulatory care providers, such

[7] For example, the National Health Insurance Experiment conducted by the Rand Corporation found differences in health care expenditures of up to 45 percent between groups with different coinsurance rates. The National Long-Term Care Channeling Demonstration showed greatly increased use of community-based long-term care services among impaired elderly people who received coverage for this type of care; both studies are cited in Grannemann (1989:4).

[8] In an analysis of effects on employment, using data for female-headed families from the 1984 SIPP panel, Moffitt and Wolfe (1990) found weak negative effects of the AFDC guarantee, stronger negative effects of both the food stamp guarantee and an index of the expected benefits from Medicaid, and very strong positive effects of an index of the expected benefits from private health insurance.

[9] Cartwright (1989:11) notes that a reduction in Medicare benefits is likely to reduce Medicare expenditures by a proportionately greater amount because of the impact on the demand for Medicaid and private Medigap insurance plans. Cartwright does not speculate about the combined effects on total health care costs.

Copyright National Academy of Sciences. All rights reserved.

as hospital emergency rooms. However, the evidence for these conclusions is very limited (see Grannemann, 1989:15). In general, there is insufficient evidence to choose among models of physician behavior, for example, that physicians tend to act according to the "target-income hypothesis," whereby they alter prices or quantity of services in order to achieve a target level of income; or that physicians in metropolitan areas are in a position to raise prices more freely compared with physicians in rural areas, because of differences in consumer access to information (Grannemann: 1989:13). Yet in simulating the effects of the most recent changes in Medicare physician payment schedules, it would have been very desirable to have good estimates of the extent to which physicians would shift their activities in favor of more generously reimbursed procedures; of how their responses, in turn, would affect usage and costs under Medicare; and of what the impact would be, in the long run, on physician specialty choices.[10]

Similarly, there are competing models of hospital behavior—including models based on profit maximization, utility maximization, or physician control (see Grannemann, 1989:7-ll)—but the available evidence has not established the superiority of any particular model of hospital behavior. For yet another area, very little is known about the potentially great impact that current research on the effectiveness of alternative medical treatments could have on delivery of health care services. Such research may well lead public and private insurers to deny payment for services that are deemed ineffective or not worth their cost. Such actions would presumably lead to responses on the part of providers and patients that, in turn, would affect usage and costs of medical services (Grannemann, 1989:36).

Databases

There is no lack of information sources pertinent to health care, including many surveys and administrative records.[11] However, there is no database, such as the March Income Supplement to the Current Population Survey (CPS) or the

[10] The new payment system, instead of reimbursing doctors' "customary, prevailing, and reasonable" charges, assigns specific amounts to specific procedures based on assumptions about their relative value. The result of the system chosen is expected to raise receipts for internists and general practitioners and lower those for surgeons (Congressional Budget Office, 1990). The CBO analysts who produced estimates for the reimbursement changes did not attempt to include long-range responses in their microsimulation model. They did include a first-round behavioral response, with the assumption that physicians would strive to meet a target income. Specifically, they assumed that 50 percent of a physician's potential income loss due to the new Medicare fee schedule would be offset by an increase in volume of services, and 35 percent of a physician's potential gains would be offset by a decline in volume (Congressional Budget Office, 1990:Appendixes; see also Chollet, 1990).

[11] See Gilford (1988: Appendix C) for descriptions of some of the major health data sets; see Panel on Statistics for an Aging Population (1986) for an inventory of data sets related to the health of the elderly.

Copyright National Academy of Sciences. All rights reserved.

Survey of Income and Program Participation (SIPP), that provides, on a regular basis, the majority of the variables needed for evaluating alternative health care policies (although the CPS and SIPP, particularly the latter, do provide relevant information). This is true even in models confined to particular sectors and types of issues, such as insurance coverage of the population. Hence, although microsimulation models typically face the task of generating a suitable database by combining variables from multiple sources through techniques such as imputation or statistical matching, models for health care (including cell-based models) face an unusually daunting database construction task.[12]

There are several examples of the database problem. The current TRIM2 Medicaid module is designed to answer fairly narrow questions about the subgroups of the noninstitutionalized population that will benefit from extending coverage of the Medicaid program to more low-income Americans and the government budgetary implications of such an extension.[13] Yet it requires the use of several data sources. The principal sources include the March CPS (supplemented with other variables, such as deductible expense imputations based on Consumer Expenditure Survey data) and the Medicaid "tape-to-tape" data from state administrative records.[14] The March CPS serves as the primary database for simulating Medicaid eligibility on the basis of AFDC and SSI, while the Medicaid data provide a basis for estimating enrollment (participation) rates and medical care utilization and expenses.

The ICF/Brookings Long-Term Care Financing Model (LTCFM), which is built on top of the PRISM model, is designed to answer questions about costs and coverage of alternative ways of financing long-term nursing home and home health care for the elderly. The LTCFM submodel uses the basic PRISM model, with some modifications, to project the numbers and characteristics of the elderly with regard to family structure, employment, income, assets, and

---

[12] Tax policy modeling also requires extensive matching and imputation to develop a suitable simulation database (see below). However, tax policy analysts do have samples of administrative records from tax returns that regularly provide many of the needed variables.

[13] Modeling the institutionalized would require a database on this population (which the CPS does not provide) and somewhat more extensive simulation of the "medically needy" eligibility provisions of Medicaid than the current TRIM2 module contains. Under these provisions, people who would not ordinarily be eligible on the basis of their income and assets, but who incur high medical care costs that result in their "spending down" their resources, can become eligible for Medicaid. Most medically needy eligible people are the elderly and disabled who need but cannot pay for nursing home care.

The cost implications for federal and state governments are not the same as the net cost to society. Currently, the expense of covering low-income people who are not enrolled in Medicaid is being borne by patients (through out-of-pocket payments), private insurers, and providers (uncompensated care). Moreover, extended Medicaid coverage might well increase use of services and thereby increase total health care costs, as well as redistribute them among payment sources.

[14] The Medicaid tape-to-tape data are commonly formatted files created by the Health Care Financing Administration from the administrative records of five participating states that have computerized Medicaid management information systems. There are separate files for Medicaid enrollees, claims, and providers.

Copyright National Academy of Sciences. All rights reserved.

private health insurance coverage. In addition, the submodel then simulates disability status of the elderly, their use of and expenditures for nursing home and home care services, and their accumulation and spending down of assets to gain Medicaid eligibility. Major databases used for the LTCFM submodel include the 1983 Survey of Consumer Finances, the 1977 National Nursing Home Survey, the 1982 National Long-Term Care Survey, and the March 1982 CPS. (The submodel was recently revised to include data from the 1984 SIPP panel, the 1982-1984 National Long-Term Care Survey, and the 1985 National Nursing Home Survey.) Many other data sources and studies were consulted in determining such critical assumptions of the model as the relationship of mortality and morbidity (the model assumes that people will enjoy longer but not necessarily healthier lives) and the projected rate of increase in nursing home prices.

Finally, the Health Benefits Simulation Model (HBSM) is designed as a comprehensive model of health care policies as they affect the noninstitutionalized population. The primary database for HBSM is the 1980 National Medical Care Utilization and Expenditure Survey (NMCUES), aged to approximate current levels of income, other socioeconomic characteristics, employer health insurance coverage, health services use, and average length of hospital stays. The aging process uses data from the March CPS, the National Health Interview Survey, and a Lewin/ICF survey of employer health insurance plans. Control totals from the National Health Accounts are used to calibrate the total health expenditures by service type and payment source estimated from the NMCUES.[15] Data about characteristics of employer-provided insurance plans from the Lewin/ICF survey are appended to the NMCUES records through a statistical matching process. Other data sources that are used to impute needed variables include the 1986 Employee Benefit Survey of the Bureau of Labor Statistics, the 1977 National Medical Care Expenditure Survey (NMCES), and the March CPS.

As this brief review highlights, there is no single, regularly repeated survey or administrative records system that provides data on most of the needed variables for health care microsimulation models. The series of surveys of medical care utilization and expenditures—the 1977 NMCES, the 1980 NMCUES, and the 1987 National Medical Expenditure Survey (NMES)—come closest to filling the need. The 1977 NMCES obtained data over an 18-month period from about 14,000 households that were queried about income, employment, family structure, sources of health insurance, disability status, health services use and expenditures, and amounts for health care paid out-of-pocket and by third parties. Additional data were obtained for these households from surveys of their physicians and other health care providers, their employers, and their

[15] The National Health Accounts are compiled by HCFA, from data from IRS income tax returns and business and provider reports of health expenditures by service type and payment source.

Copyright National Academy of Sciences. All rights reserved.

health insurance carriers. The 1980 NMCUES was a similar but more limited survey, conducted over 15 months for a sample of 10,000 households. The sample comprised a nationally representative component of 6,000 households supplemented with samples drawn from Medicaid records in four states; see Subcommittee on Federal Longitudinal Surveys (1986). The 1987 NMES was similar to the original NMCES, with a sample of 14,000 households with oversampling of blacks, Hispanics, elderly people, low-income people, and people with functional limitations. Like NMCES, the NMES also included ancillary surveys of employer health insurance plans and medical providers listed by NMES households. Finally, the 1987 NMES included an institutionalized component, comprising 13,000 residents of nursing and personal care homes, psychiatric facilities, and facilities for the mentally retarded.

Yet the NMCES-NMCUES-NMES family of surveys does not satisfy the data needs for health care modeling: the sample sizes are small and, with the exception of NMES, cover just the household sector, excluding institutionalized people; the surveys have been conducted infrequently; and there have been long lags in making the data available for public use—data have just recently become available from the 1987 round.[16] Other useful surveys have also been conducted infrequently. For example, the two most recent rounds of the National Nursing Home Survey were in 1977 and 1985; the next round is scheduled for 1991 or 1992.

Other surveys are limited in other ways. For example, the National Health Interview Survey is a large household survey that is conducted on an annual basis; however, it focuses on self-reported health conditions and use of health care services, and obtains only limited data on other variables that are needed for modeling such as income and health care expenditures.

Outdated information can present serious problems for health policy modeling, given rapid technological advances and changes in treatment that affect demand and costs. For example, the cost estimates originally developed by CBO for the prescription drug component of the 1988 Medicare Catastrophic Coverage Act, by using data from the 1977 NMCES and 1980 NMCUES, were much lower than those prepared after passage of the act by using data from the 1987 NMCES (Congressional Budget Office, 1989c). Given the uncertainties in the original estimates, Congress mandated that CBO reestimate the costs of the prescription drug component when the 1987 data became available. Comparison of the 1980 and 1987 data showed that prescription drug usage by the elderly and the average price paid had increased much faster than CBO had originally projected. CBO's original estimates had projected $6 billion in government outlays for covering prescription drugs and $8 billion for insurance premiums

[16] Indeed, the developers of HBSM, which is based on the 1980 NMCUES, note that the necessity of using so many control totals from other data sources to age the data forward in time makes the model more like a cell-based than a microsimulation model (personal communication, John Sheils, Lewin/ICF, Inc., March 1991).

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

paid by the elderly for the period 1990-1994; the revised estimates were $12 billion for the government and $9 billion for the elderly.

In terms of administrative records, there are a plethora of sources, including records of Medicare claims and payments maintained by HCFA; records of Medicaid eligibility, claims, and payments maintained by the states; records maintained by private insurers; hospital admission and discharge records; and so on. There are a number of useful files from these sources, including the HCFA Continuous Medicare History Sample, which is a continuously updated 5 percent sample of all Medicare claims records, beginning with 1974. However, users face many problems in gaining access to administrative records, in working with their large volume, and in developing integrated files of needed data from multiple sources. The task is particularly daunting in cases for which state data are needed, for example, modeling changes to the Medicaid program.[17]

There are a number of heartening indications of improvements in needed data for health care modeling. For example, HCFA is planning to inaugurate an ongoing Current Beneficiary Survey (CBS) of the Medicare population, which will consist of continuously updated samples of about 12,000 (mainly elderly) people, each of whom will be interviewed three times a year over several years. The CBS is intended to provide larger samples of the elderly than do most other surveys and to obtain additional information about beneficiary characteristics that the Medicare claims records cannot provide (Health Care Financing Administration, 1990).[18] As another example, the National Center for Health Statistics is planning to coordinate its surveys of providers, including hospitals and nursing homes, and conduct them on a more frequent basis. As yet other examples, the March CPS, which added questions on health insurance coverage in 1980, added questions in 1988 to improve measures of health insurance coverage for children and for nonworking adults (such as retirees) covered by a former employer, while the Health Interview Survey recently added more detailed questions on household income.

---

[17] As noted above, the Medicaid tape-to-tape files that provide commonly formatted records are available for only five states. The Medicaid expenditure equations in TRIM2 are based on only 1 year of data from these files for just two states; the equations are then adjusted for state differences through use of aggregate Medicaid expenditure data provided by the states to HCFA (Holahan and Zedlewski, 1989:Appendix B). For a discussion of problems and suggested solutions related to linking data for research purposes from the multiple Medicare files, see Fisher et al. (1990).

[18] The need for the CBS was underscored by the Health Care Financing Administration (1990): "Medicare has entered a period of rapid change. . . . Ten years ago, program coverage and the beneficiary population were relatively stable. In that environment, program data, even with large availability lags, were sufficient to make many of the estimates that were required. We are now operating a program whose most basic features are continually altered by legislative changes, as well as changes in regulations, technology, and consumer and provider behavior."

Copyright National Academy of Sciences. All rights reserved.

Suggestions and Recommendations

There is clearly strong and growing concern about the need for improved models and data for evaluating alternative health care policies. And there has been visible progress on some fronts, notably in initiating programs to obtain better data sets for modeling and research purposes. At the same time, modeling efforts remain, for the most part, fragmented and limited in scope. There also appears to be a good deal of confusion about priorities and directions for health care policy modeling. This confusion is not surprising, given the wide range of health care policy issues and the diverse actors whose behavior must be taken into account. At present, various interagency coordination bodies are concerned with health data needs, including committees on data for the elderly and the disabled. However, there is no coordinating body that is focused specifically on data and research requirements for developing improved models for health care policy analysis.[19] We do not pretend to have answers regarding priority directions for microsimulation modeling of health care areas. We do have some general views, and we propose an approach to help in the task of setting priorities.

Model Development

We believe that microsimulation techniques have a useful role to play in answering many detailed questions about the effects of alternative health care policies. Yet we are also aware that microsimulation models can be costly and time consuming to develop and apply, particularly when first-round and even second-round behavioral responses have to be modeled in addition to direct effects. In our view, it is imperative for HHS to set up a department-wide coordinating and steering body to determine priorities both for microsimulation model development in the health care area and for needed data collection and research studies that will lead to improved models. This steering group should be established at a level that will give weight to its recommendations and should be led by an agency (or agencies) with a mission that is oriented primarily to policy analysis (rather than data collection or research). Given the sizeable needs of microsimulation models for data and research results, we believe that

[19] The history of the NMCES-NMCUES-NMES family of surveys is instructive in this regard. The primary sponsor of the 1977 NMCES was the National Center for Health Services Research (NCHSR), now the Agency for Health Care Policy and Research (AHCPR). Data files were very slow in being released for use by other agencies or outside researchers, at least in part because NCHSR was oriented to research use of the data internally and not to serving the larger user community. The primary sponsor of the 1980 NMCUES was the National Center for Health Statistics (NCHS), which is a major collector and distributor of health data with a primary focus on assessing health conditions among the population. NCHS has relatively little experience in considering data needs from the perspective of assessing the impact of alternative health care policies on costs or access to medical care. The 1987 NMES was again sponsored by AHCPR, with design input from other HHS agencies, including ASPE, HCFA, and NCHS.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

the efforts of such a coordinating body will be fruitful for many other types of health care policy research and analysis as well.

Recommendation 8-1. **We recommend that the U. S. Department of Health and Human Services establish a high-level, department-wide coordinating and steering body to set priorities for development of microsimulation models and related data collection and research needed for improved analysis of alternative government policies and programs for health care.**

Overall, we believe that it would be unwise, at this juncture, for HHS to plunge forward with the development of large-scale microsimulation models for health care policies. There are too many data and research gaps to be filled in first. Moreover, the complexity of the modeling task in health care argues strongly that new model development (or significant expansion of existing models) should be based on the kinds of emerging innovative computer hardware and software technologies (described in Chapter 7) that promise to facilitate model validation and experimentation.

When the decision is made to proceed with model development, the general principles of model design and implementation presented in Chapter 6 will need to be rigorously applied. In particular, microsimulation models for health care policy should be developed with superior capabilities to facilitate validation. Long-term care and other health-related policy issues involve long time horizons, for which it is important to conduct extensive sensitivity analyses, as well as to prepare variance estimates, in order to establish a sense of the range of reasonable projections.[20] Modifying physician or hospital reimbursement schemes, as well as many other health care benefit changes, can be expected to have first- and second-round behavioral effects, again necessitating careful and thorough validation.

For health care models, especially, model development strategies must not become overly grand. It is obviously important to have models that can take account of multiple actors and indirect as well as direct effects of policy

[20] Jacobs (1989) criticizes the Brookings/ICF Long-Term Care Financing Model from the validation perspective. He notes that the policy debate focused on the model's central forecast, namely, that a majority of the elderly in the future could not be covered by private long-term care insurance at a price they could afford, but the debate ignored other simulations that were much less pessimistic in this regard. He notes further that the model's results are highly sensitive to key assumptions, specifically, about changes in disability rates over time and the relative rate of inflation of nursing home costs. The initial simulations, reported in Rivlin and Wiener (1988), tested some alternative assumptions. Later work performed for ASPE tested additional assumptions, specifically, concerning the impact of private-sector initiatives to provide long-term care insurance under different assumptions about inflation and disability rates. Jacobs argues that additional refinement and sensitivity analysis of the model should still be done, for example, to add some assumptions about the level of induced demand that might result from more extensive insurance coverage for long-term care. He also argues (p. 17) that model results should be presented to policy makers "with variations to illustrate the impact of key assumptions."

Copyright National Academy of Sciences. All rights reserved.

changes. However, we do not believe it is possible for any one model to achieve sufficient breadth and depth to serve the full range of health care policy analysis needs. As argued in Chapter 6, model developers must consciously trade off the scope of a model versus the detail of model components, and this principle is particularly important to apply for health care models. Clearly, the HHS coordinating body, and others involved in health care model development, must think broadly in conceptualizing the requirements for improved microsimulation models and must develop a comprehensive plan to guide their specifications for needed data and research. However, the actual modeling capabilities that ultimately emerge do not have to be—probably should not be—tied to a single model. It is likely to be more cost-effective to develop separate models for different sets of actors (such as patients or physicians), designed in such a way that they can interchange outputs and be used in a coordinated way.

Given, however, that investments in actual model development on a large scale are deferred until developments in data, research, and computing technology bear fruit, the question for the coordinating body becomes one of how to choose priorities for new research and collection of new or modified data. These tasks require careful deliberation, given their high costs in comparison with those for secondary data analysis.

Data

There are many gains to be made from incremental strategies that seek to take advantage of existing data sources.[21] For example, there is a wealth of administrative data on health care that could be linked with survey data. We understand that HCFA's plans for the Current Beneficiary Survey of the Medicare population include exact matches of the survey data with Medicare claims records. We urge that such linked data sets be made available for policy analysis and research purposes. We also note the utility of including questions for key variables, such as income, employment status, and employer firm size, in multiple surveys to facilitate relating them for modeling purposes.[22] Recently, ASPE encouraged the National Center for Health Statistics (NCHS) to improve the income questions in the National Health Interview Survey so that those data can be more readily used in conjunction with other data sets. The Interagency Forum on Aging-Related Statistics has also drafted guidelines for including income items in surveys about the elderly (Income Working Group, 1990).

---

[21] The Committee on National Statistics' Panel on Statistics for an Aging Population made a large number of recommendations for improving data sets related to the health of the elderly in ways that would build on existing data and minimize additional costs (Gilford, 1988). The HHS coordinating body for health care microsimulation models that we recommend should certainly review those recommendations.

[22] As Chollet (1990) notes, variables such as employer firm size are important for modeling private alternatives to public financing of medical care costs, such as mandated private employer health insurance.

Copyright National Academy of Sciences. All rights reserved.

Although more frequent data collection entails costs, we believe it is important to consider increasing the periodicity of key data sets. For example, given the important role of private as well as public health insurance and rapid changes in medical technology and treatments that have major effects on health care usage and costs, we believe that conducting surveys of the NMCES-NMCUES-NMES type more frequently should be given serious consideration. Provider-based surveys, such as the National Nursing Home Survey, are also important to conduct on a reasonably frequent basis (as NCHS is currently planning to do).

Finally, just as a strategy of designing one grand health care policy model is not likely to be cost-effective, neither is a strategy of planning for one comprehensive health care survey, given the many different data needs and respondent burden. Hence, health care policy modelers will continue to confront the necessity to use multiple data sources and to relate them through the use of techniques such as imputation and statistical matching. Given this situation, we urge that HHS consider, periodically, sponsoring very comprehensive small-scale surveys that can be used to validate the quality of the imputations and matches performed by health care policy models (see Chollet, 1990, on this point).

Research

It is clear that additional research is required to develop robust estimates of both direct and indirect effects of alternative health care policies. (Again, our discussion pertains to cost and coverage issues but not to the important topic of effects on health outcomes.) The question is where to start, given the range of actors and responses involved. Grannemann (1989) proposed several criteria for ranking research topics, including:

- the estimated magnitude of the associated costs, in terms of the proportion of total health care costs accounted for by the actor whose behavior will be analyzed;
- the estimated magnitude of the behavioral response, in terms of the proportion of medical care services that will be affected;
- the extent to which new information is likely to improve prior assumptions about outcomes, that is, to reduce the uncertainty in the estimates of behavioral parameters; and
- the congruence of the assumed behavioral effect with priority policy concerns.

Grannemann developed a matrix of scores on each of these dimensions for five types of actors—hospitals, physicians, nursing homes, insurers, and patients—and several dimensions for each actor (such as price and volume of

Copyright National Academy of Sciences. All rights reserved.

services for physicians, and usage, insurance purchase, and program participation for patients). For example, hospital services scored higher on the cost dimension than physician or nursing home services; patient use decisions and utilization review by insurers, which can apply to all services, scored even higher. On the dimension of magnitude of the impact, physician and hospital pricing scored high, while community-based alternatives to nursing homes scored low, reflecting the limited ability such programs have shown to affect patient and family choices regarding nursing home use. On the dimension of policy relevance, physician payment and utilization review (potentially resulting from effectiveness research) scored high, while hospital price responses, which would apply mainly to private-sector payers, scored low.

Overall, Grannemann identified six areas as priority targets for behavioral research: physician pricing responses, insurer responses to utilization review, patient use decisions, patient program participation, determinants of volume of physician services, and hospital admission decisions. Of course, one need not agree with Grannemann's list of factors or his assigned scores. Nonetheless, the approach of laying out formal criteria for selecting priority areas for research on behavioral responses to health care policies has a good deal of merit, and we recommend it as a high-priority task for the HHS coordinating group charged with planning the development of improved microsimulation models for policy analysis related to health care.

RETIREMENT INCOME POLICIES

Provision of adequate retirement income has been a concern of the federal government going back at least to the days of the New Deal. Over the years, the policy debate in this area has addressed both the benefits and the financing of the federal social security system. On the benefit side, the debate has considered coverage or entitlement (e.g., coverage was extended to federal workers just a few years ago); the level of the benefit replacement ratio (i.e., the percentage of earnings prior to retirement replaced by social security benefits); the retirement age at which benefits begin; the extent to which earnings after retirement reduce benefits; and many other issues. On the financing side, the debate has considered the social security tax rate, the level of income subject to tax, and the extent of income taxation of benefits. The policy debate has also addressed issues of private pension coverage and benefits, and how the private pension system meshes (or does not mesh) with social security and civil service retirement.[23]

[23] The discussion in this section benefited greatly from a paper reviewing the two major retirement income microsimulation models, DYNASIM2 and PRISM, which was prepared for the panel by Ross (in Volume II). For documentation of DYNASIM2, see Johnson and Zedlewski (1982); Johnson, Wertheimer, and Zedlewski (1983); and Zedlewski (1990); for documentation of PRISM, see Kennell and Sheils (1986, 1990). The panel also benefited from presentations by Zedlewski and by Kennell

Copyright National Academy of Sciences. All rights reserved.

The stakes in this area are high. To cite some figures: in fiscal 1988, payments to retirees under the federal social security system totaled $157 billion, and federal civil service and military pensions totaled another $36 billion (Bureau of the Census, 1991:Table 588). The sum of payments under social security (to retirees, survivors, and the disabled) and other federal payments to retirees and the disabled totaled $271 billion in fiscal 1988 and had climbed to $306 billion in fiscal 1990—almost one-quarter of total federal outlays (Congressional Budget Office, 1991:150, 152).

Microsimulation Modeling for Retirement Policy

The Social Security Actuary has for many years maintained cell-based models to project the people expected to be eligible for social security benefits and those expected to pay social security taxes and to estimate the balances between expenditures and receipts in the social security trust fund.[24] These projections are typically made for as many as 75 years into the future because of the interest in assessing likely changes in the trust fund balances due to such factors as increased or decreased fertility and labor force participation rates.

Another largely cell-based retirement policy model is the Macroeconomic-Demographic Model (MDM), which combines a neoclassical macroeconomic growth model with cell-based models for population growth, family formation, labor markets, and pension benefits. The MDM model, which was developed by Lewin/ICF, Inc., for the President's Commission on Pension Policy and the National Institute on Aging, has been used to analyze the impact of the aging of the U.S. population on retirement income systems over the long term and the interactions of private pensions with social security. More recently, health expenditure modeling components were added to the MDM model (see Anderson, 1990).

Microsimulation models have also played an important role in evaluating alternative pension policies, particularly when the questions raised involve complex issues affecting particular population groups or the intersection of public and private pension systems. Because of the need to see how changes in retirement income policies interact with demographic and employment changes over the long term, simulations of pension programs have almost always used dynamic models that can project detailed individual and family histories over periods of 20-40 or more years. The two most heavily used models in this area are DYNASIM2 and PRISM.

and Sheils and from a paper by Burtless (1989) that reviewed issues in modeling behavioral response to retirement income policies.

[24] See Grummer-Strawn and Espenshade (in Volume II) for a review of studies evaluating the quality of the Actuary's population projections, which are used not only in the Actuary's trust fund models but also to control the population projections of dynamic microsimulation models.

Copyright National Academy of Sciences. All rights reserved.

The original DYNASIM model was completed in 1975, but it proved too complex for regular use. DYNASIM2 represents a revision, completed in the early 1980s, that was designed to be more streamlined and cost-effective. The redesign also focused DYNASIM2 on retirement issues, although the model has been used for other applications as well, such as an analysis of the implications of alternative rates of teenage pregnancy for government transfer program costs and an analysis of the demand for long-term care services over the period 19902030. The DNYASIM2 model has been applied in many different retirement-related analyses: for example, it was used to project the impact of earnings sharing proposals, which the Actuary's model could not handle.[25] Agencies using DYNASIM2 for analysis of retirement income programs include the Congressional Budget Office, Department of Labor, and Department of Health and Human Services.

The PRISM model, which was first developed in 1980 to evaluate alternative national retirement income policies combining public and private pension coverage for the President's Commission on Pension Policy, has been used extensively since that time. It was developed specifically as a retirement income model and consequently does not have some of the features of DYNASIM2. For example, PRISM models childbirth, but it does not create records for newborns because they are not needed for retirement simulations. Likewise, PRISM does not include an educational attainment module, because it works with the population age 20 and older as of 1979, which has already completed sufficient education to support a simulation of the retirement age population through the year 2025. In the mid-1980s, PRISM was expanded to include a subsystem for modeling financing options for long-term care of the elderly (see Kennell and Sheils, 1986; Rivlin and Wiener, 1988).

Dynamic strategies for aging the initial database are at the heart of the DYNASIM2 and PRISM models. As noted in Chapter 6, dynamic aging is theoretically superior to static aging because it more fully maintains the covariances between age and other variables by applying transition probabilities for a large number of variables.[26] In practice, dynamic aging is a much more complex process than static aging, and the quality of the resulting database depends heavily on the quality of the many transition probabilities that are used. (To keep the time paths for major variables in line with other accepted projections of these variables over time, dynamic models usually include a step to calibrate each year's results to assumptions used in other forecasts.) Although static aging is less complex, it is not likely to become a viable strategy for

[25] Under earnings sharing, the covered earnings of married couples are split between both spouses. Hence, in the case of divorce of a couple with only one covered worker, the nonworking spouse would still have an earnings history and be entitled to benefits.

[26] Not all variables are dynamically aged in dynamic models: for example, DYNASIM2 includes cross-sectional routines to impute assets, disability, and SSI variables to each record.

Copyright National Academy of Sciences. All rights reserved.

modeling public and private pension systems because of the need for rich detail throughout the simulation period and not just at the end year. Detailed income, earnings, and employment *histories* are needed to evaluate alternative retirement income proposals, such as changing the replacement ratio or requiring portability of private pensions after a specified time period.[27] Dynamic models also afford the capability to change assumptions about transition probabilities, as well as independent control totals, so that different scenarios can be played out over time.

The high degree of complexity of dynamic models means that they tend to be resource-intensive to develop, update, and apply. They are also subject to many sources of error because of their large numbers of parameters and, hence, difficult and resource-intensive to validate. Moreover, in applications that require long projection periods, as is the case for most retirement income issues, dynamic microsimulation models confront the inescapable problem that the quality of their projections deteriorates over time. Not only are errors in multiple sources likely to compound, but people are likely to change their behavior in ways that cannot be foreseen. DYNASIM2 and PRISM typically use aggregate population and economic growth assumptions from projections of cell-based or macroeconomic models to control their own projections, but, of course, errors inevitably affect the quality of the outside projections as well, particularly as the projection period lengthens.

Clearly, dynamic microsimulation models need to be reassessed periodically to determine if there are ways to take advantage of new computing technologies and alternative designs to enhance their cost-effectiveness and, even more important, to facilitate the task of validation and communication of uncertainty in the projections to decision makers. The kinds of validation studies that are needed include external validity studies that compare model estimates with measures of truth, sensitivity analyses, and studies to estimate variance from sampling error in the primary database and other sources (see Chapters 3 and 9). For a review of studies of the quality of the labor force projections from DYNASIM2, see Cohen (Chapter 7 in Volume II). Wertheimer et al. (1986) compared DYNASIM2 projections with those produced by the cell-based MDM. In that study, MDM outputs were used to control the DYNASIM2 projections, and the comparisons looked at variables not explicitly linked in the control process. Overgeneralizing from a very complex analysis, the models exhibited minor differences in projections of social security beneficiaries, somewhat greater differences in projections of social security benefits, and substantial

[27] To simply calculate social security benefit entitlements under current law requires information for each person about covered quarters of employment and highest wages attained. Theoretically, a database that already contained historical information for each record could be statically reweighted to some future year. However, it is unlikely that retirement income analysts would be satisfied with having to assume that only gross demographic and price changes were likely to take place over the projection period.

Copyright National Academy of Sciences. All rights reserved.

differences in projections of private pension recipients and benefits. However, few such studies have been done.

Suggestions and Recommendations

We have not reviewed in detail the components of the DYNASIM2 and PRISM models and therefore do not have specific recommendations for functions that may need improvement. Burtless (1989) comments that DYNASIM2 and PRISM simulate the effects of retirement policy changes on labor supply in terms of the effect on the decision to retire, but not in terms of the effects on work hours or earnings before or after retirement. For example, DYNASIM2 assumes that labor force participation ends once a worker retires. PRISM provides for partial retirement, but the level of pension income does not affect work hours or earnings. Burtless (1989) makes a more general observation that dynamic models, although they draw heavily on behavioral research for their transition probabilities, include very few feedback effects of behavioral changes in response to changes in government programs and policies. Because these kinds of effects may also be of policy interest, it could be useful to consider expanding the models' capabilities in this direction.

We see two fundamental requirements for dynamic models of retirement income policies—first, the need for linked survey and administrative data that are periodically updated to provide the initial database; and second, the need for continual modification from updated research results of the transition probabilities that are used to project the individual data records in terms of marital, childbearing, labor force, and other behaviors.

Linked Data

Models such as DYNASIM2 and PRISM rely on social security administrative records of earnings histories linked with the cross-sectional demographic and socioeconomic information in the March CPS as the foundation of their longitudinal databases for simulating retirement income policies. (DYNASIM2 uses the March 1973 CPS-SSA exact-match file; PRISM uses the March 1978 CPS-SSA exact-match file, which in turn, is matched with the March 1979 CPS and the May 1979 CPS.) Because confidentiality concerns and resource constraints have forestalled development and public release of any CPS-SSA exact-match files in the 1980s, models such as DYNASIM2 and PRISM are working with increasingly outdated initial databases. They have to simulate more and more years of historical data before they can even begin projecting their databases into the future.

We believe that it is vitally important for resources to be found to produce updated exact-match files of social security and survey data and for ways to

Copyright National Academy of Sciences. All rights reserved.

be devised to make such files available for modeling and analysis of retirement security options—a policy area that will be of continuing importance as the population ages. Performing matches with both the March CPS and the SIPP should be considered.[28] SIPP has demonstrated excellent performance in obtaining social security numbers from respondents, and, indeed, the survey was designed with the expectation that administrative data would be used in conjunction with the interview data.[29]

SIPP also regularly includes needed data on investments in individual retirement accounts (IRAs) and private and public pension plan coverage. (The CPS has also periodically included pension supplements.) SIPP has also experimented with obtaining detailed information about employer-provided benefits, including pensions and others. This information is needed to simulate provisions such as vesting, comparisons of defined benefit and defined contribution types of plans, etc. The SIPP experiment entailed getting a signed release from a sample of respondents to allow the Census Bureau to query their employers. The response rate for the releases was only 44 percent, but the rate for queried employers was very high, 92 percent (Jabine, King, and Petroni, 1990:40-41). Currently, DYNASIM2 imputes pension plan variables, and PRISM matches data from a sample of pension plan sponsors.

Ideally, matched survey and administrative data would be publicly available, as in the past. If legal constraints prohibit public release, other ways should be devised to provide research access to the data. A very limiting option would be to allow analysts to be sworn in as special census employees to use the data at the Census Bureau. A preferable option might be to release the data files to the policy analysis agencies under special security provisions and allow agency analysts and their contractors to use the data at a secure facility.

Recommendation 8-2. **We recommend that the Census Bureau perform a new exact match of social security earnings histories with the March CPS as soon as possible. The Census Bureau should develop a program for periodically conducting matches of social security earnings histories with both the March CPS and SIPP records. Ways should be found to make the matched data files available for research and modeling use.**

Behavioral Research

In addition to good initial databases, models that are used for simulating

[28] The administration recently endorsed efforts to develop linked files of SIPP and administrative records data that could be available for research use (see Council of Economic Advisers, 1990).

[29] The Census Bureau recently matched the 1984 SIPP panel with SSA data, but this file is available for a limited time only to researchers at SSA who are sworn in as special census employees. Moreover, the smaller sample size of the SIPP vis-a-vis the CPS limits the analysis that the matched file will support.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

retirement income policies need good estimates of a wide range of behaviors to use in projecting the database forward in time. For example, DYNASIM2 simulates changes in labor force status annually as a function of previous labor force status, age, race, sex, education, region, disability status, marital status, presence of children, and spouse's earnings by using 16 profit equations (for different demographic groups) estimated from pooled 1968-1981 data from the Panel Study of Income Dynamics.

The need for the kind of research that we recommend in Chapter 6 to narrow the range of estimates of behavioral parameters—such as the propensity to move in and out of the labor force—is clearly critical to the quality of dynamic models for retirement income policies. Also critical is that the parameters be reestimated periodically, to reflect any changes in relationships that may have occurred. Alternative specifications that take advantage of improved methodology should also be investigated periodically. Although DYNASIM2 and PRISM can implement updated coefficients for parameters such as fertility rates or female labor force participation rates, it is harder to implement alternative modules with different functional forms, such as alternative labor force participation equations. Yet many of the basic routines in these models are based on research that was conducted some years ago and, hence, may not reflect methodological advances, let alone behavioral changes, since then. Greater flexibility in this regard would also greatly enhance the ability to conduct sensitivity analyses of model components.

Of course, up-to-date and better-fitting estimates do not solve the problem that relationships may change in unforeseen ways during the course of a projection period.[30] Detailed sensitivity analysis involving a number of socioeconomic and demographic assumptions for each policy alternative may be the best way to indicate the likely uncertainty due to a range of plausible scenarios.[31] However, not having updated parameters available when it is known, or believed highly likely, that important behavioral relationships have altered—due to some combination of policy and population changes—places an extra burden on a model. As just one example, the retirement probabilities incorporated in the DYNASIM2 model were estimated by Burkhauser and Quinn (1981) from the Retirement History Survey for people who were working at the beginning of the survey in 1969. To the extent that program and population changes have altered the labor force distribution of people of this age subsequent to 1969,

[30] Burtless (1989:23) gives the following example: "Changes in average health and economic circumstances cannot entirely explain the trend in older men's labor force participation over the past five decades. We must infer that some unobserved factor—probably taste for retirement—has varied over time. If average tastes change, we cannot confidently predict future behavior on the basis of observed behavior. . . ."

[31] Grummer-Strawn and Espenshade (in Volume II) discuss the problem of interpreting the range that is obtained when key assumptions are varied in projection models, specifically, the Actuary's and the Census Bureau's population projections.

Copyright National Academy of Sciences. All rights reserved.

the transition probabilities estimated by Burkhauser and Quinn may no longer apply (Burtless, 1989).[32]

Academic research of the type that can best support dynamic microsimulation modeling of retirement income and other policies depends critically on the availability of regularly updated information from panel surveys that provide repeated measurements on the same individuals over time. It is important that the federal statistical and policy agencies support panel surveys, which represent the most appropriate mechanism to provide the information needed to study individual responses to changing events and to calculate transition probabilities.[33] We believe that the costs of continued investment in panel surveys will be more than repaid by the benefits from improved understanding of behavior. In turn, a better research knowledge base for modeling and other types of policy analysis should make it possible to improve the information available for public policy debate in many critical areas.

TAX POLICIES

Our review of the use of microsimulation techniques for public policy analysis has focused thus far on assistance programs, that is, the expenditure side of the government ledger. However, the earliest, most continuous, and perhaps most widespread use of microsimulation modeling has been in the area of tax policy analysis, the revenue side of the ledger. The complexities of the federal individual income tax code, coupled with the diverse economic circumstances of the U.S. population, virtually require that tax policy analysis be conducted at the micro level.[34]

Microsimulation Modeling for Tax Policies

Historically, there has been a need to apply microsimulation techniques to calculate the revenue effects of proposed changes in tax policies, and to answer questions about the fairness of tax policies for different population subgroups. In the 1980s, the constraints of the growing federal budget deficit and the Gramm-Rudman-Hollings Act increased the demand for estimates of the impact of changes to the tax code. Typically, the only way to produce the needed revenue

[32] The Retirement History Survey followed 12,500 people—men, and women with no husband in the household—aged 58-63 at the start of the survey for 10 years, interviewing them every 2 years; see Subcommittee on Federal Longitudinal Surveys (1986:111-114).

[33] The National Institute on Aging recently announced that the University of Michigan Institute for Social Research will begin in 1992 a long-term panel survey designed to obtain updated information on health and retirement patterns. This survey should provide a rich resource for behavioral research.

[34] The discussion in this section benefited greatly from a set of notes prepared by Harvey Galper, a member of the panel, and from a background paper prepared for the panel by Strauss (1989). Bristol (1988) also provides a useful discussion of issues in tax policy modeling.

Copyright National Academy of Sciences. All rights reserved.

and distributional estimates is by means of microsimulation models applied to detailed microlevel databases. Most recently, the debate that engulfed the Congress and the President in considering the fiscal 1991 federal budget—the debate over which kinds of taxes to raise (or lower) and by how much, and which income groups ought to bear the tax burden—required microsimulation modelers on the staff of the Joint Committee on Taxation to work virtually around the clock for days on end (Russakoff, 1990).

The tax policy microsimulation models that are in use today have a long history, beginning in work conducted in the early 1960s by researchers at the Brookings Institution. Joseph Pechman and his colleagues at Brookings, with the encouragement of analysts in the Treasury Department, developed microlevel databases using tax return information. Subsequently, they developed even more elaborate databases for tax policy modeling called MERGE files. The first MERGE file contained statistically matched records from the Statistics of Income (SOI) public-use sample of individual tax returns and the 1967 Survey of Economic Opportunity. Subsequent MERGE files matched the March CPS with tax return records (see Pechman, 1965, 1985; see also Minarik, 1980).

The Treasury Department in the mid-1960s brought tax modeling capability in-house, and the Office of Tax Analysis (OTA) still maintains a comprehensive tax policy microsimulation model that is based on a statistical match of the detailed SOI sample with the March CPS. The full OTA model, including both SOI and March CPS data, is updated every 2 or 3 years; OTA analysts also use more recent SOI data for tabulations and ad hoc simulations. OTA is currently running version 11 of the model, which includes data from the 1985 SOI and the 1986 March CPS (Cilke and Wyscarver, 1990). The next full OTA model, which will use the 1989 SOI and the March 1990 CPS, is scheduled for completion in 1992.

The Joint Committee on Taxation (JCT), which is responsible for preparing revenue estimates of all congressionally proposed tax law changes, initially relied on analyses conducted on its behalf by OTA. Subsequently, the JCT staff ran the OTA model on the Treasury Department's computer. In the mid-1970s, the JCT brought the OTA model in-house and, since then, has modified some of the aging and imputation routines.

Many other public and private agencies maintain tax policy microsimulation models, including more than half the states; a number of private research organizations concerned with the effects of federal and state taxation policies, such as the Brookings Institution and the National Bureau of Economic Research; several private economic consulting firms; and the governments of most Western nations. And, in addition to OTA and JCT, other U.S. agencies with tax models include:

Copyright National Academy of Sciences. All rights reserved.

- the Congressional Budget Office, which has conducted several studies of the effects of tax policies on the distribution of family and personal income using microsimulation techniques (see, e.g., Kasten and Sammartino, 1990);
- ASPE, which invested substantial resources in the early and mid-1980s to enhance the federal and state income tax modules in TRIM2 (see Weinberg, 1987, for an analysis of the distributional effects of the 1986 Tax Reform Act based on running the TRIM2 model) and also recently made extensive use of the TRIM2 federal tax module to simulate proposed changes to the child care tax credit; and
- the Census Bureau, which uses its model to estimate the distribution of after-tax income rather than for policy analysis purposes (see Bureau of the Census, 1988).

Although U.S. tax policy models share features in common with income support models, they also have several distinguishing characteristics. Currently used tax policy models fall into the class of cross-sectional microsimulation models that simulate policy effects on a population database at a given time and use static aging techniques to project the database forward in time.[35] In general, the models are very elaborate tax liability calculators that incorporate very few explicit behavioral responses. Most models assume that taxpayers will choose whether or not to itemize deductions on the basis of which alternative reduces their tax liability, but other types of behavioral response, to the extent that they are considered, are usually handled outside the model.[36]

One distinguishing feature of the tax models is their very heavy reliance on information from administrative records, specifically, from the SOI samples of tax returns. Each year, the Statistics of Income Division of the Internal Revenue Service (IRS) obtains a random, stratified sample of individual income tax returns, which it processes to produce published statistical series and analyses (see Coleman, 1988). (The division also provides tabulations and analyses of corporate tax returns, partnership returns, and estate and gift tax returns, as well as other specialized tabulations based on IRS records.) In odd years, the division obtains a much larger sample of individual returns and also keys in additional variables. It is these biennial detailed files that are most heavily used in tax modeling and tax policy research.

Tax models are also distinguished from income support models, at least in degree, if not in kind, by their very heavy reliance on imputations, statistical

[35] In the future, the Statistics of Income samples that provide input data for the tax models (see the description of the SOI in the text) will include an embedded panel: that is, tax returns of some individuals will be sampled each year on a continuing basis. The Office of Tax Analysis hopes to develop a dynamic model to use these panel data.

[36] From time to time, important behavioral responses, such as changes in handling assets in response to changes in capital gains tax rates and changes in charitable deductions in response to changes in marginal tax rates, have been simulated within rather than outside of tax policy models.

Copyright National Academy of Sciences. All rights reserved.

matching, and complex static aging techniques. More extensive use of imputation and aging techniques is needed for tax modeling because the SOI data generally become available with a lag of 2-3 years and lack information, such as family characteristics, that is important for evaluating alternative tax proposals (see further discussion, below). The March CPS provides some of the missing information and is available on a more timely basis, but it is not suited to stand on its own as a tax policy analysis database. It lacks information on deductions, some components of income (such as capital gains), and other variables (such as net operating loss carryovers) that are critical to calculating tax liability. The March CPS also "topcodes" income amounts for high-income people (the top income category shown is $100,000 or more) and has other problems from a tax modeling perspective.

One can divide the family of U.S. tax models into at least three main categories on the basis of their principal purpose and data source. The first category includes the very detailed tax model used by the OTA and the JCT. The primary purpose of this model is to prepare revenue estimates for proposed legislative changes to the federal individual income tax. The model is also frequently used to examine the distributional effects of tax law changes. The OTA/JCT model is built on information from the detailed SOI files that have not been modified for public use, with additional variables imputed or matched from the March CPS and other sources.

The second category includes tax models that start with the public-use SOI files and impute or match additional information from household surveys such as the March CPS. The primary purpose of these models is to examine broad issues surrounding the distributional effects of government tax policies. The models maintained by the Brookings Institution and the National Bureau of Economic Research are in this category.

The third category includes tax simulation routines that are embedded in models of income support programs and that, consequently, start with a March CPS database to which information from the public-use SOI files is imputed. Models such as TRIM2 fall into this category. The tax routines of this class of models are also typically used to study broad distributional effects of tax policy changes, sometimes in conjunction with the transfer program routines.

Because OTA and JCT have a need for, and also have access to, the detailed SOI files, the OTA/JCT model is more elaborate than any of the other tax policy simulation models. With regard to the differences between the detailed and public-use SOI files (see Statistics of Income Division, no date), the public-use files are processed so that names, addresses, and social security numbers are deleted from all records; records sampled at 100 percent from the full IRS file of tax returns (which are returns with very high incomes or very high losses) are subsampled at a 33 percent rate in the public-use SOI files; state codes and other geographic indicators are deleted from records with adjusted gross income of $200,000 or more and, in addition, other codes (such

Copyright National Academy of Sciences. All rights reserved.

as exemptions and alimony paid or received) are removed or modified for these records; wage and salary income, state and local income tax deductions, and real estate tax deductions are blurred for high-income records; state and local tax deductions and alimony paid or received are blurred for all other records; and many detailed items reported by small sets of taxpayers are deleted (e.g., preferences under the minimum tax).

Developing a Tax Model Database: The Case of OTA

To illustrate the array of problems and the extensive procedures involved in developing a suitable database for tax modeling, particularly at the level of detail required for estimating federal revenues, we describe the process of generating the database for version 11 of the OTA model (see Cilke and Wyscarver, 1990). This version of the model implemented several major improvements over earlier versions, including new aging routines. Converting the model from a UNIVAC mainframe to a VAX minicomputer provided greater computational power and speed that made possible enhancements to the model's capabilities.

To develop version 11, OTA began with the 1985 SOI sample, containing about 121,000 individual income tax returns, each of which had more than 1,000 data fields. OTA deleted unnecessary fields, created necessary recodes, packed the data to minimize computer storage requirements, and verified that the taxpayer's liability calculated by using the OTA model agreed with the liability reported in the SOI file. For the 5 percent of returns for which the reported and simulated tax liabilities did not agree, OTA examined the records and edited one or more variables as needed.

OTA next imputed additional variables. For example, because not all taxpayers itemize deductions, but some proposed changes in tax law might induce some current nonitemizers to itemize, OTA imputes deductible expenses for nonitemizing taxpayers. (Such expenses include charitable contributions, real estate taxes, home mortgage and other categories of interest, medical expenditures, and others.) Imputations were performed in some cases from regression equations estimated from the Survey of Consumer Finances and, in other cases, from probabilities estimated for itemizers in the SOI file itself.

As another example, OTA imputed the earnings attributable to husbands and wives, for married couples filing jointly, based on tax return information provided in support of the two-earner deduction for working couples. Such information is needed to assess tax provisions that impose a so-called marriage penalty and other analyses, but it was not available for many working couples except for the 1982-1986 period when the two-earner deduction was in effect. OTA then exactly matched a file containing each taxpayer's month and year of birth (from social security records) with the SOI tax return sample.

The next step was to reduce the size of the SOI sample in order to improve processing efficiency. This procedure involved an optimization model designed

Copyright National Academy of Sciences. All rights reserved.

to select about 89,000 observations from the full file of 121,000 tax returns in such a manner as to minimize the total loss of information when measured by an explicit loss function.

Yet (Cilke and Wyscarver, 1990:2-8),

> In spite of vast improvements . . . to the individual tax model over the years, one shortcoming still remains: the data base relies almost entirely on tax return data. Analyzing proposals that could radically alter the tax base, the tax unit, and the tax rates requires information that is not tied to a particular tax law or limited to what is reported on tax returns filed under that tax regime.

Additional information that OTA needed includes: (1) sources of income and types of expenditures not subject to taxation under current law; (2) links between taxpayers and family or household units, so that examination of tax burdens can be based on logical economic units; (3) information on people whose incomes are too low to require filing a return under current law but who might, under some proposals such as a consumption tax scheme, become tax filers; and (4) information necessary to construct a comprehensive income measure, such as consumption plus change in net worth, that permits an analysis of the distributional effects of tax reform proposals for a broader concept than that of taxable income.

OTA's approach to obtaining much of the additional information needed for version 11 of its tax model was to carry out a statistical match of the March 1986 CPS and SOI records. The first step in this process was to apply tax filing unit provisions to each CPS person aged 16 and over in order to convert CPS households into one or more potential tax filing units or potential nonfiling units with low incomes. Corrections for underreporting of certain forms of income in the March CPS were also made by using data from the National Income and Product Accounts and other sources. The SOI file of taxpayers age 16 and older and the March CPS file of potential filers were then statistically matched on the basis of a common set of core variables, including marital status, age, number of children, housing tenure (owner or renter), and the presence or absence and amount of various sources of income. The CPS nonfilers were then appended to the matched file, and the latter file was sorted by CPS family number. Finally, taxpayers under age 16 from the SOI file were appended to families judged to have appropriate income and demographic characteristics. The resulting file size expanded to about 213,000 records, given the addition of nonfilers from the March CPS, taxpayers under age 16 from the SOI, and the need to match some CPS records to more than one SOI record and vice versa in order to preserve the properties of each of the two samples. In particular, the March CPS sample is very thin in the number of cases of high-income tax filing units in comparison with the SOI.

The last stage in the process of creating the OTA tax model database

Copyright National Academy of Sciences. All rights reserved.

involved several series of imputations. One series was undertaken to provide information needed to simulate the 1986 Tax Reform Act (TRA). Imputations required for this purpose included tax-exempt interest, health insurance expenses for self-employed people, meal and entertainment expenses, adjustments to pension income, contributions to 401(k) retirement plans, net operating losses carried over to future tax years, employer contributions to pension plans, and active and passive income by source. A variety of data sources were used to develop the TRA imputations. A new set of parameters to simulate the TRA provisions also had to be incorporated into the tax calculator portion of the model.

Another set of imputations was performed to generate information needed to simulate tax liabilities for catastrophic health insurance coverage under Medicare. These imputations illustrate the additional database enhancement that is frequently required to model new provisions of the tax code. Required imputations in this instance included Medicare eligibility for the tax unit head and spouse, employer contributions to Medigap insurance, reduced medical expense deductions because of provided catastrophic health insurance, pension income by source for the unit head and spouse, and social security income for the unit head and spouse. Repeal of the catastrophic health insurance program in 1989 obviated the need for such information at the present time, although Congress may in the future consider a similar type of plan.

Finally, a series of imputations was undertaken to provide information to construct the broad concept of family economic income used by OTA in analyzing the distributional effects of proposed tax law changes on income classes of the population. For example, untaxed income sources (such as AFDC and SSI), employer-provided benefits, and exclusions from taxable income (such as 401(k) contributions) had to be imputed and added to adjusted gross income for filers and added to estimated adjusted gross income for nonfilers.

The result of all of these operations was a greatly enhanced database for modeling tax policies as of 1985. However, because OTA must develop projections for the current year and a minimum of 5 years into the future, it has developed and regularly employs static aging procedures. The aging procedure for version 11 of the OTA model, which reflects a major effort to improve this part of the model, generates a new database for each year from 1986 through 1995 and is implemented in two stages (see Gillette, 1989).[37] The procedure is performed each time a simulation is carried out and is under the control of the user, who can select one of three aging alternatives developed by OTA to represent different expected rates of economic growth.

The first stage in the aging process applies growth factors on each dollar amount in the database to reflect actual and projected per capita real growth

[37] Prior versions of the model aged the database only to the current year; projections to future years were developed through an out-of-model extrapolation technique.

Copyright National Academy of Sciences. All rights reserved.

and inflation. The second stage adjusts the weights of each family head in the file to hit aggregate targets for 33 different variables, such as adjusted gross income by income class and type of return, for which the targets are chosen to be consistent with forecasts of national income, population, and inflation. The weight adjustment is carried out in a series of iterations designed to hit the targets while minimizing a loss function.

Finally, OTA has included in the aging procedures an optional special adjustment to simulate the behavioral response of taxpayers to the increase in the top tax rate on long-term capital gains resulting from the 1986 Tax Reform Act. The simulation changes the level of realizations of long-term capital gains as a function of the difference in marginal tax rates before and after tax reform and the prereform level of realized gains. Other behavioral responses to increased tax rates on capital gains, such as changing holding periods, shifting mixes of capital assets, and converting ordinary income to (or from) capital gains, are generally estimated outside the model.

Issues in Modeling Tax Policies

Clearly, the use of SOI tax return information in tax policy microsimulation models is beneficial and, indeed, essential for detailed simulation of the revenue consequences of proposed changes to the tax code. The SOI files provide documented sets of income amounts, deductions, and other tax-related variables for large samples of actual filing units and thereby portray all of the detail of the current tax code. Just as clearly, however, the SOI files are inadequate in many ways. They lack variables necessary to characterize fully the socioeconomic status of households and families for use in analyses of the distributional effects of tax policy changes. They also lack variables needed to simulate proposed tax reforms that broaden the base of taxable income, expand the types of allowable deductions and credits, change the definition of a tax filing unit, or in other ways significantly alter the tax code. Users of the SOI data alone for tax policy modeling are at the mercy of changes in the tax laws that not only may require new imputations but also add, or delete, useful information. For example, the information required to support two-earner deductions for working couples from 1982 through 1986 provided a good basis for determining the share of wages attributable to each spouse; however, this information is no longer available.[38]

To obtain the range of needed variables, tax policy modelers must confront a heavy task of extensive data creation using imputation and matching techniques. Imputations for missing variables are also required for simulations of current and proposed income support programs, but not to the degree that is usually required for detailed simulation of proposed changes to the tax code.

[38] Other administrative records, such as social security earnings records, could provide information on spousal wages through an exact matching procedure, but such information is not available on public use tapes.

Copyright National Academy of Sciences. All rights reserved.

A factor that complicates the task of creating a tax policy simulation database is that tax return data differ in important ways from the household survey data in the March CPS and other files that are often used in matching and imputation. For example, there is a difference between tax filing units and households or families. Data on filing units are required to estimate tax liabilities, but data on households are most often required for distributional analyses. Recent changes in the tax law that require social security numbers for most dependents should make it easier to reconstruct family units from tax returns when dependents file separately and, thereby, improve the match with household survey records. However, the tax return data cannot, by definition, provide information on families that are not currently required to file tax returns. Survey data are needed to supply information about nonfilers, and one or more potential tax filing units must be constructed for these cases.

Another important difference that affects the quality of matches of household survey data with tax return files is the imbalance in the distribution of records by income class in the two samples. Household surveys such as the March CPS contain many more low-income and many fewer high-income records in comparison with the SOI data. Moreover, survey data for high-income people are typically top coded to protect confidentiality. These differences constrain the ability of the matching algorithm to find good matches.

Still another important difference between SOI and household survey data concerns underreporting (and misreporting) of income. Underreporting affects both types of data, but the opportunities to underreport (or misreport) differ for different types of income and, thus, differentially affect tax returns and household surveys. For example, W-2 forms must support reports of wages and salaries to the IRS, but this is not the case for household surveys.[39] The growing use of 1099 forms by providers of interest, dividends, and other types of nonwage income presumably encourages more complete reporting to the IRS, while these kinds of income sources are of notoriously poor quality in surveys. However, reports of income from odd jobs and cash transactions may be better, or at least no worse, in surveys than on tax returns.

As noted above, OTA, as part of creating its tax model database, adjusts the March CPS income amounts to national targets. One could also use tax compliance data to correct the SOI figures for the entire population of filers. However, if one wants to model income tax *collections,* one should only implement corrections for the small number of people who are actually audited. In any case, the differences in mechanisms underlying income reporting in administrative and survey data bear examination to determine their effects on the quality of the resulting merged database.

[39] Although wages and salaries are presumably reported more accurately in tax returns than in surveys, definitional problems plague comparisons for this as well as other types of income. Thus, contributions to so-called 401(k) pension plans are legitimately excluded from taxable wage and salary income and, hence, from W-2 forms; they may or may not be reported in surveys.

Copyright National Academy of Sciences. All rights reserved.

One of the original goals of SIPP was to support improved modeling of tax as well as transfer policies. Each SIPP panel to date has annually included a module that obtains information about tax-related variables, including yearly income amounts from employment and assets, type of tax return filed, number and type of exemptions, type and amount of deductions and credits, and taxes paid. Clearly, such information, together with the improved reporting of income in SIPP, would be valuable for tax policy models that currently start with the March CPS and for models, such as OTA's, that match CPS with SOI data. However, small sample sizes and confidentiality restrictions have severely limited the utility of the SIPP tax information to date. Only a few variables from the tax module are included in files that are available for research and policy analysis use (no information on amounts is made available), and these files are obtainable only under special access arrangements.

One way to assess and improve the quality of tax modeling databases would be to conduct exact matches, based on social security numbers, of household survey records from the March CPS or SIPP with tax return data from the full IRS files. Confidentiality restrictions may well preclude access to such a matched file, but the Statistics of Income Division and the Census Bureau could explore the possibility of conducting exact matches for use solely in evaluating and enhancing the quality of the statistical matches and imputations that are currently done to construct tax policy analysis databases.

Confidentiality concerns generally are a major and growing problem for access to tax return data for research and analysis purposes. More and more data on the public-use versions of the SOI files are being blurred. For example, deductions for state taxes are blurred because some states disclose information such as the amount of property taxes paid. (No state codes are provided at all for returns with adjusted gross income of $200,000 or more.)

Even if access problems are solved, tax modeling still requires substantial data generation. For example, to support comprehensive modeling of the incidence of taxes, one needs to impute consumption so that one can model excise and sales taxes. One also needs to impute all data for nonfilers. These are big imputation jobs. Much more attention needs to be paid to imputation procedures and their quality, particularly for models that try to assess tax incidence broadly. More sensitivity analysis of imputation techniques is needed. OTA has commissioned some work on this topic (see, e.g., Kadane, 1978), but more is needed, for example, on the impact of imputations on joint distributions of variables.

The need for updating and projecting the information required for tax policy modeling further increases the difficulties of data generation and adds uncertainties, which could be substantial, to the estimates. Just as imputation procedures generally need much more extensive evaluation, so do the procedures used by tax models to age their databases.

Finally, the issue of behavioral response is serious for tax models because

Copyright National Academy of Sciences. All rights reserved.

behavioral effects of tax policy changes affect almost the entire population. One is not dealing with a subset of the population, such as welfare recipients, for which behavioral effects may be relatively unimportant. Changes in marginal tax rates undoubtedly affect behavior across the population and also have important second-round effects. To model behavioral responses in a satisfactory manner, one would need to link tax models with other models of wage rate and factor price determination. Data from panel surveys that follow people over time are also needed for development of good estimates of behavioral responses to such changes as the tax treatment of capital gains. Another important area of research concerns ways to narrow the range of parameter estimates and address other technical considerations in developing and implementing behavioral response functions in microsimulation models. An added issue to consider concerns presentation of estimates of first-round and second-round effects of tax policy changes. One can make a real case for using models for short-term estimates and then discussing possible longer-run behavioral responses in qualitative terms. This was the approach followed by the Treasury Department in the 1984-1985 tax reform debate leading to the 1986 Tax Reform Act.

Given that our review concentrated principally on models of government expenditure programs in the social welfare area, we do not offer explicit recommendations for tax policy models. However, we believe that there are at least three priority needs for attention. One important need is to address confidentiality and data access issues for tax return information and, specifically, to develop ways to take advantage of exact matches of survey and administrative records information to improve the quality of tax model databases. The second important need is to improve the capability for estimating the effects of behavioral responses to tax policy changes, whether these estimates are developed inside or outside the tax policy simulation models themselves.

The third, and perhaps the most important, need is to conduct sensitivity analyses and other validation studies of the extensive array of imputations and matches that are performed in creating databases for tax modeling and of the procedures used to age these databases. We are concerned that the imputation and aging procedures, regardless of how much care is taken in implementing them, may distort important covariances for individual filing units. Under the current methodology, each imputation is performed individually to meet aggregate control totals and other known information about the distribution of the item by income class. When several imputations are done in this fashion, however, the effect may be to cause variations among individual units within income classes that reflect the imputations more than the real world. The imputation procedures may not have adverse effects on the quality of comparisons involving vertical equity, that is, assessment of the effects of tax policy changes on different income classes; however, they may well adversely affect the quality of comparisons involving horizontal equity, that is,

Copyright National Academy of Sciences. All rights reserved.

assessment of the impact of tax policy changes on taxpayers in similar economic circumstances.

These needs are important not only for microsimulation models of tax policies, but also for models in other policy areas. We encourage a broad range of policy analysis agencies to work together to make the needed improvements in microsimulation models.

Copyright National Academy of Sciences. All rights reserved.

9

Validation

As far as I can see, in all uses of models for policy purposes . . . there is no confidence or error band . . . . But we know that no model is correct. It couldn't be; there are so many factors omitted, even under ideal circumstances . . . . Whether or not these remarks are fair, they lead to the questions of validation. How do we validate these relationships? . . . Forecasting or hindcasting is a way of validating a whole system. One would like, really, to be able to validate individual relations as well because, if the whole system doesn't work well, it is necessary to know where the repair job is needed. This problem requires a methodological discussion which I have not yet seen.

Kenneth Arrow (1980:260,263)

The purpose of microsimulation modeling, as of any policy analysis tool, is to produce useful estimates for the legislative process. "Useful" connotes many things, but of prime importance in the panel's view is that the estimates be reasonably accurate. It seems obvious to say that better estimates are preferable to worse estimates, particularly when millions, often billions, of dollars ride on decisions to implement one or another program change. Yet agencies have typically underinvested in model validation, and microsimulation models have not been exempt from this pattern of underinvestment. Indeed, it is arguable that microsimulation models have received even less attention in terms of validation than have other model types—not so much due to deliberate oversight as to the particularly daunting nature of the task. Unlike, for example, cell-based population projection models, which have relatively

Copyright National Academy of Sciences. All rights reserved.

few components, microsimulation models are typically very large and complex, which greatly complicates the job of validation. Unlike macroeconomic models, which produce unconditional forecasts that can be and are frequently checked against reality, microsimulation models produce conditional estimates that are more difficult to measure against reality.

On the positive side, however, validation of microsimulation models is simplified because in many cases their focus is on estimating differences among policy alternatives. The typical application of microsimulation models is to estimate the total costs, caseloads, and caseload characteristics for the current program and for one or more alternative programs. Therefore, the focus is not on levels of costs or caseloads for an alternative program but on differences between the alternative and the current program. This focus means that validation exercises can perhaps safely ignore some of the factors on which the estimates would ordinarily be conditional: a misspecified module or parameter that wrongly affects both a current program and alternative programs in the same way may have little effect on the estimates of differences. Thus, fewer factors are probably "active" for any given problem, facilitating model validation.

Still, an analyst cannot assume that particular factors, such as the macroeconomic or demographic forecasts used in the simulations, have little or no impact on the estimates of differences between a current and alternative programs. Such a determination has to be carefully considered. And microsimulation models are also asked to estimate costs and caseloads of entirely new programs. In these instances, the interest is not in a difference but in an estimated level, which will be affected by all of the operative factors in the simulation. Thus, in general, one should take estimates of levels from microsimulation models to be less certain than estimates of differences.

Because microsimulation models typically provide a variety of outputs, including aggregate estimates of costs and caseloads in terms of both levels and percentage changes, in comparison with current programs, and detailed estimates of the effects on population subgroups, the question of an appropriate loss function for an evaluation is a difficult one. For example, should one use as the evaluation criterion the minimization of error in estimates of levels of total costs or caseloads, percentage changes for a key population characteristic, or some other criterion?

In this chapter we first note the kinds of corroborative activities, such as assessing model outputs against the analyst's view of the world or against outputs from other similar models, that, to date, have constituted the bulk of the effort devoted to microsimulation model validation. Then, we briefly review three basic types of validation studies—external validation, sensitivity analysis, and variance estimation—and note some issues specific to their use in validating microsimulation models. We next discuss briefly the use of loss functions in conjunction with microsimulation output. We then review the handful of studies that have attempted to carry out validations of microsimulation models and draw

Copyright National Academy of Sciences. All rights reserved.

what limited conclusions we can from these studies about the quality of current models.

Next, we describe a validation study of the TRIM2 model that the panel commissioned as part of its work. This study, our "validation experiment," started out as a sensitivity analysis of three components of the TRIM2 model: the procedures used for aging the database, the procedures used for generating monthly income and employment variables from the annual values on the March CPS, and the use of the standard TRIM2 March CPS database versus the use of a database corrected for undercoverage of the population. In addition, we turned the experiment into an external validation study, by having the TRIM2 model use its 1983 database to simulate AFDC program costs and caseloads according to the rules in effect for these programs in 1987. This approach enabled us to compare the TRIM2 results with actual administrative data for 1987. Our experiment was very limited because we examined only three components of one model for only one time period. However, we learned something about elements of TRIM2 that merit further evaluation, and, more important, our experience offers an example of the kind of analysis that should be a regular feature of the microsimulation modeling enterprise.[1]

In the final section of this chapter we discuss strategies for microsimulation model validation and present our recommendations for policy analysis agencies.

CORROBORATION AS A STAND-IN FOR VALIDATION

We do not want to convey the impression that policy analysis agencies and modeling contractors have been insensitive to the need for validation of model output; indeed, in a number of areas, they have worked hard to ensure the accuracy of the models. In particular, they have typically devoted considerable resources to "debugging" activities designed to verify the accuracy of the computer code and a model's representation of the detailed accounting rules for the various programs within its purview. One debugging technique that is used is to identify a test sample of households, with complex structures or characteristics that relate to special programmatic features, and to check the records for these households—variable by variable—before and after simulations are run to ensure that the program specifications have been properly implemented. A related technique that could be used is to print the records for a relatively small number of cases that are outliers for benefits or some other characteristic of interest so that the computations for those cases can be checked. Debugging activities are important to continue and to improve, when possible, given the large volume

[1] We did not single out TRIM2 because of any belief that it was much worse—or better—than other models. The primary sponsor for our study, ASPE, expressed interest in such an evaluation of TRIM2 and made available contractor and computing resources for the work. We would have liked to include other models as well but lacked sufficient resources to overcome various problems, such as different initial databases.

Copyright National Academy of Sciences. All rights reserved.

of complex code in microsimulation models and the correspondingly increased probability of coding errors.

Microsimulation model analysts also typically engage in other activities that are designed to flag errors and problems in the model estimates. These activities, which amount to seeking to corroborate model results rather than to validate them, include:

- assessing model output against the analyst's view of the world, which is often informed by many years of experience (as a simple example, a simulated increase in benefits that produced smaller rather than larger numbers of beneficiaries compared with current law would act as a red flag to the analyst for further investigation);
- comparing model output with the analyst's "back-of-the-envelope" estimates (e.g., comparing TRIM2 projections for mandating the unemployed-parent program with the result of using for all states a simple ratio of existing unemployed-parent to total AFDC caseloads; alternatively, TRIM2 projections might be compared with those from a simple time-series model relating AFDC caseloads to a few key variables such as the unemployment rate); and
- comparing model output with output from other similar models or from the same model run by another agency.

The last corroboration activity is an important mechanism for identifying and correcting problems with model estimates during the course of policy debates. Typically, two or more agencies—such as ASPE and CBO—are preparing estimates of the cost and distributional effects of proposed changes. Communication channels among analysts in the various agencies are usually fairly open, and analysts will compare their estimates. If there are large discrepancies, there will be an effort to determine the source, which often stems from different assumptions but may also stem from problems with one or another of the models.

One example of the role of this type of corroborative activity from the history of the Family Support Act concerns the estimates for extending the unemployed-parent program to all 50 states (mentioned above). Analysts at CBO noted that the TRIM2 estimates used by ASPE for the states not previously covering unemployed parents also showed increases in the basic caseload in these states; the CBO estimates for the basic caseload, which were derived from an AFDC benefit-calculator model, did not show comparable increases. Investigation determined that the participation algorithm used in TRIM2 contains a parameter for the generosity of the state in which the eligible program unit resides. The existence of an unemployed-parent program serves as the proxy to distinguish between more and less generous states. Hence, mandating coverage of unemployed parents in all 50 states resulted in raising participation rates for other kinds of eligible units in the states not previously covering unemployed parents. But this is not a plausible outcome if one assumes that mandated

Copyright National Academy of Sciences. All rights reserved.

coverage does not indicate program generosity in the way that elective coverage does.

Useful as these kinds of activities are, they are not a substitute for rigorous external validation studies and sensitivity analyses of model outputs. Indeed, there are grave dangers in relying on corroboration alone, even when the policy analysts are very knowledgeable and careful in debugging their code and checking their results with others. It is possible that the collective wisdom of the entire policy analysis community is simply in error. One example of an erroneous assumption related to AFDC policy concerns participation rates. The conventional wisdom for years held that the basic AFDC program, after experiencing enormous growth in the caseload in the 1960s and early 1970s, had saturated the eligible population. Overall participation rates simulated by the major income support program models regularly exceeded 90 percent. However, a marked drop in the simulated participation rate—to about 80 percent—occurred after 1980. Investigation determined that the primary cause was what appeared to be a minor change in coding family relationships, which the Census Bureau implemented beginning with the March 1981 CPS. The result was to add about 1 million subfamilies who were eligible for the AFDC program but exhibited lower-than-average participation rates (see Ruggles and Michel, 1987).[2] As a result of this evaluation, analysts had to revise their views of the participation behavior of the population eligible for AFDC and hence their expectations of "reasonable" participation rates that were simulated by the models. This is but one example of the need for rigorous validation, not just corroboration, of microsimulation models.

TECHNIQUES OF MODEL VALIDATION

External Validation

External validation of a model is a comparison of the estimates provided by the model against "the truth"—that is, against values furnished by administrative records or other sources that are considered to represent a standard for comparison. Several factors complicate, although they do not preclude, the task of externally validating the output of microsimulation models.

First, as noted above, these models produce conditional estimates. If the social or policy environment changes and thus is different from the assumptions on which the estimates are based, it will not be surprising that the estimates differ from the reality. If a model simulated the actual program that was enacted, but other factors, such as the economic environment, changed, it may

[2] As another example (but with the opposite substantive effect), the use of more detailed data from SIPP on asset holdings and other variables to calculate food stamp participation rates has resulted in higher participation estimates compared with estimates based on the March CPS (see Doyle, 1990).

Copyright National Academy of Sciences. All rights reserved.

require considerable effort to respecify an appropriate set of model runs that are conditional on the correct factors for comparison. More frequently, a model did *not* simulate the particular policy alternative that was actually enacted, and so direct comparisons are not possible. Instead, ex post forecasting or backcasting techniques must be utilized. Ex post forecasting is the use of an archived data set from several years ago with an archived (or current) model to predict the program in use a short time ago. In backcasting, one takes the current model and database and simulates program provisions that were operative at some period in the past.

Another factor complicating external validation studies of microsimulation models concerns the appropriate time period for comparison, which may not always be clear. Because most models produce estimates of direct effects, the comparison period needs to be after the program changes are fully implemented but before any feedback effects could be expected to show up.

Finally, the measures of "truth" that are used for the model comparisons are likely to present several problems. Like the microsimulation estimates themselves, they may be subject to both sampling and nonsampling errors. For example, detailed information on the AFDC caseload is based on a sample rather than on the complete set of administrative records. The available comparison values may also lack the cross-tabular detail needed for a full evaluation: such detail is critical given that a raison d'etre for microsimulation is its ability to provide estimates of distributions as well as aggregates.

Despite these difficulties, external validation of microsimulation model estimates is crucial to carry out on a regular, systematic basis. It remains the best method for measuring the total uncertainty in a model, especially when the modeling environment is fairly stable and one can conduct several studies assessing the performance of a model for different time periods and policies. Of course, the information so obtained will not alone suffice to guide policy makers or funding agencies. Because external validation is an ex post operation, by definition it cannot give policy makers the real-time information they need about the quality of the estimates being produced for a current legislative debate. However, backcasting techniques may help policy analysts assess the likely quality of current estimates, as may experience with a model based on a long track record of external validation (both forecasting and backcasting). Another limitation of the results of external validation as a direct guide for policy makers is the fact that differences between model estimates and comparison values will include chance variation. Therefore, providing a complete picture of the model's performance requires generating a set of results that allows distinguishing among underlying sources of error. Further, because of likely practical limitations on the character and extent of external validation studies, they are unlikely to provide specific information about areas of needed improvement. For this type of information, one must turn to sensitivity analysis.

Copyright National Academy of Sciences. All rights reserved.

Sensitivity Analysis

Sensitivity analysis is a technique that measures the effects on model outputs of alternative choices about model structure by replacing one or more existing modules (or the entire model) with reasonable alternative modules. The variability of the resulting estimates helps gauge the model's susceptibility to bias[3] in the estimates due to model misspecification—one of four sources of variability.[4] The resulting observed variability is not directly useful as a variance estimate because there is no indication of the true value. However, if the various alternatives used are equally plausible (or nearly so, given the current state of knowledge), the resulting range of the output estimates will provide some information on the variability in the output that could be attributed to misspecification of that component. Importantly, if there is an indication that the model is performing poorly, a sensitivity analysis should help to identify those component alternatives that make a difference and therefore should help direct the search for components that need to be improved (possibly by being replaced by one of the tested alternatives). This is one way in which a feedback loop can be constructed to identify and remedy model defects.

Two key requirements for the use of sensitivity analysis are that alternative specifications of components exist and that the algorithms and software to implement them be easy to obtain and use. The first requirement of the existence of alternative modules is not a serious problem. Within the framework of microsimulation, there appear to be many candidates for experimentation, such as alternative functional forms for participation equations or alternative methods of calibration. A more important problem is that many alternatives may be difficult to implement. The panel's validation experiment involved a simple sensitivity analysis of TRIM2. Of the three modules investigated—distributing incomes on a monthly basis, treating undercoverage, and making use of various procedures for aging—competing methodologies for the first two modules were easily programmed, but the third, alternative aging techniques, required a large investment of resources. This component of TRIM2 was not specifically designed to accommodate a sensitivity analysis—a situation not, of course, peculiar to TRIM2, or more generally to microsimulation models. In addition, the proprietary nature of some microsimulation models restricts the free exchange of model components.

Interpreting the results of a sensitivity analysis is clearly easier when the alternative components have little interaction. Sensitivity analysis is simply

---

[3] The term bias is difficult to define precisely in this context: the general meaning here is the difference between the central tendency and the truth.

[4] The four sources are sampling variability in the primary database; sampling variability in other input sources; errors in the primary database and other input sources; and errors from model misspecification. The concept of mean square error properly includes all four sources, but in practice almost always ignores the fourth source and often the third source as well (see the Appendix to Part I).

Copyright National Academy of Sciences. All rights reserved.

more feasible if the effect of joint modifications of the model's components on the outputs can be decomposed into essentially additive effects from changes to individual components. Because one is often interested in the effect of several factors on model estimates and it is expensive to investigate one component at a time, one usually modifies several factors simultaneously. If there is little interaction between components, it is easier to identify the factors that had the greatest impact on different output estimates. Whether or not it is possible to separate out different components depends, of course, on the characteristics of the underlying model. The panel's validation experiment found some interaction among the effects of adjusting the three modules of TRIM2 we investigated, which complicates the interpretation of the findings. It remains to be seen whether this degree of interaction is common to microsimulation models generally.

Sensitivity analysis and external validation have different advantages and disadvantages. An advantage of sensitivity analysis over external validation is that it can be done during model development and use, that is, during the decision-making process; external validation, by definition, must be conducted after the fact. However, a sensitivity analysis cannot by itself identify which components are working well, because there are no comparison values. On the other hand, external validation cannot usually identify specific weaknesses in individual model components because, when the entire model is tested against the comparison values, sufficient information on the components is difficult to generate.

Both sensitivity analysis and external validation can benefit from simultaneous application. By testing a variety of model alternatives through use of various component alternatives, and by making use of comparison values to identify models that produce estimates closer and further from comparison values, one can identify superior combinations of components. If the components have little interaction, it becomes easier to identify the combination of alternatives that is performing best (for that situation) and more feasible to study a larger set of factors in a limited number of model runs. If the components do interact, however, simultaneous modification is the only way of identifying the factors that need to be jointly improved, even though the identification process is complicated by the presence of interaction.

Another way of learning about deficiencies in a microsimulation model, related to sensitivity analysis, is to make use of completely different modeling approaches to the entire problem, rather than exchanging individual components. It is clear that, for specific outputs, analysts using macroeconomic models, cell-based models, or other approaches, could produce estimates with error properties that were competitive with the estimates produced by microsimulation models. Many of these models would be relatively inexpensive to implement, and they could be very effective for helping to determine the likely variability in the estimates produced by microsimulation and diagnosing areas of weakness.

Copyright National Academy of Sciences. All rights reserved.

When replacing an entire model, it is harder to determine precisely which component is causing a discrepancy; however, experienced analysts can often pinpoint the likely source.

Variance Estimation

Because the input data set is one sample out of a family of equally likely input data sets, the output of a microsimulation model, potentially estimated on each member of this family, has an associated variance. In simpler modeling contexts, such as multiple regression analysis, statistical theory has produced methods of directly estimating the variance in estimates arising from sampling. In the complicated world of microsimulation modeling, however, no such theory is available. In this discussion of the measurement of this variance, we focus on one specific technique, the bootstrap (see Efron, 1982). However, it is important to point out that most of the techniques in the general literature on sample reuse or nonparametric variance estimation are applicable; for a full discussion of these techniques, see Cohen (Chapter 6, in Volume II).

The idea behind the bootstrap is as follows. The variance of an estimator is a function of the average squared distance between an estimate (a function of the sample) and the true value (a function of the population). Because data are not available for the entire population, the relative frequencies of different estimates, needed to evaluate the average squared distance, are not known. However, if there is a relatively large sample, the observed relative frequencies of different estimates obtained by evaluating the estimator on pseudosamples drawn from the observed population will approximate the unknown relative frequencies obtained by sampling from the entire population and evaluating the estimator for those samples. Hence, possibly a good estimate, even for small samples, of the average squared distance between an estimate and the truth is the average squared distance between estimates derived from repeated sampling, with replacement, from the original sample and the estimate from the original sample, which then serves as a proxy for the truth.

The actual application of bootstrap techniques to microsimulation modeling is complicated and rests on many choices that are difficult to make in the abstract. The underlying theoretical and practical aspects are discussed in Cohen (Chapter 6, in Volume II). We are convinced at this stage that the bootstrap or another of the currently available sample reuse techniques can be used to estimate sampling variances for outputs of microsimulation models. In fact, Wolfson and Rowe (1990) provide an example in which they used the bootstrap to estimate the variance of estimates from the SPSD/M model.

These bootstrap variance estimates, however, will only measure directly the first component of variability of model estimates, that is, the sampling variability resulting from using one rather than another database. This component may well be the least important source of error in model outputs (although the relative

Copyright National Academy of Sciences. All rights reserved.

magnitude of the various sources of uncertainty is currently unknown). The concept of the bootstrap could be extended to measure the second component as well, that is, the variability in the model outputs resulting from sampling variability in the many other database inputs, such as imputations for variables not contained on the primary database, control totals from macroeconomic forecasts or population projection models, and behavioral response equations. Each of these inputs contains variability from being based on one rather than another database.

To extend the bootstrap in this respect would be a complicated process. One would need to have pseudosamples for the primary database before modification or augmentation with other sources. One would also need to have distributions for each input—such as a series of child care expense imputation equations with varying coefficients—developed by applying the bootstrap (or possibly a less exacting procedure) to the Consumer Expenditure Survey or other data on which the equations are based.

It may be possible to use sensitivity analysis in conjunction with bootstrap resampling as a more feasible way to develop estimates of total model uncertainty. Roughly speaking, one could use a bootstrap to measure the variance and a sensitivity analysis to weakly measure bias. Thus, one could use sensitivity analysis alone or together with the bootstrap to evaluate the contribution of errors in the primary database or in other sources, such as undercoverage of the target population, inappropriate imputation methodologies to overcome nonresponse, and misreporting of key variables. In Chapter 5 we recommend that originating agencies, such as the Census Bureau, play a more active role in preparing databases that are suitable for microsimulation modeling and other forms of policy analysis. This role should include more vigorous investigation and correction of data quality problems. In our view, it is not appropriate or cost-effective to ask microsimulation modelers to take on this burden. However, the modelers can make a contribution by using their expertise to identify the data quality problems that may have most import for modeling and conducting some limited evaluations to provide feedback to the originating agencies.

Loss Functions

The estimates produced by microsimulation models are in reality quite complicated, providing a wealth of detailed information about the distribution of outcomes from any policy alteration. This in itself presents challenges in evaluating the validity of any modeling effort. It is certain to be the case that a modeling effort—characterized by choices of modules, data sources, and the like—will do better in producing estimates for certain outcomes than it will do for others. Moreover, alterations in a model to improve the estimates in one area, such as AFDC participation, could actually worsen the estimates in

Copyright National Academy of Sciences. All rights reserved.

another area, such as the estimates of benefits transferred to different types of families.

The full analysis of models, and the feedback to further development, requires that the producers and users of estimates specify the relative importance of accuracy with respect to the different estimates produced in any given effort. This specification can be characterized by a "loss function." The loss function provides a quantitative measure of the importance of errors of different types. For example, if the model were invoked solely to estimate total costs of an AFDC option and if policy makers were increasingly unhappy the farther the estimate was from the true cost, the loss function might be posed simply as the square of the distance between the estimate and the true answer.

There are many possible alternative specifications of loss functions that could be employed. In the previous example, if the policy makers were more distressed by underestimates of the budget implications than by overestimates, a nonsymmetric loss function might be appropriate. Alternatively, other loss functions could be defined in terms of the deviations of the estimates from certain aspects of the overall distribution of program effects.

Past work has seldom employed explicit loss functions. This is unfortunate, because it leads to considerable ambiguity in evaluating model outcomes. In the context of formal estimation of the variance of model outputs, lack of a well-defined loss function presents severe problems. Therefore, one aspect of the future work on model validation should involve developing more explicit notions of an appropriate loss function. This work will require direct interaction between users and producers of policy analysis. An extra benefit from being explicit about the loss function is the considerable guidance gained thereby for any model development process. The loss function provides direct information for analysts about which aspects of a modeling endeavor are most in need of attention and possible modification.

REVIEW OF VALIDATION STUDIES

As part of our examination of the state of microsimulation model validation, we searched for and collected validation studies for two purposes: to obtain information on the performance of the models currently in use and to gather examples of the methods that others have used for microsimulation validation. Although the literature review that we commissioned was not fully comprehensive, we are reasonably certain that we have not missed any major validation studies of microsimulation models.[5]

We found only 13 validation studies of microsimulation models. When one considers that microsimulation modeling techniques have been used for policy analysis for over 20 years, and that during this time at least 6 major

[5] The discussion here summarizes the material in Cohen (Chapter 7, in Volume II).

Copyright National Academy of Sciences. All rights reserved.

and 10-20 minor microsimulation models have been developed (see Chapter 4), it is surprising that so little effort has been devoted to determining the uncertainty present in these models. We are not the first to come to this conclusion. Betson (1988:12) notes, "Efforts to explore the statistical properties of the estimates derived from microsimulation models have been scant." Doyle and Trippe (1989:v) agree: "Despite the fact that microsimulation models have been used extensively to set public policy, little effort had been invested in ascertaining the quality of the simulation models until very recently." Burtless (1989:49) adds, "No behavioral predictions from microsimulation models have been compared with actual historical experience from a different period than the one used to derive the original behavioral estimates. We can increase the public's confidence in microsimulation results (and probably improve the reliability of behavioral routines) if we periodically compare the model predictions against actual experience." Finally, there is the quote that led off this section by Kenneth Arrow.

Table 9-1 lists the studies that we reviewed and some of their characteristics. The second column of the table indicates the model(s) covered; the third and fourth columns indicate whether the study included a sensitivity analysis or external validation of the model; the fifth column indicates whether the study stated that the results of the analysis subsequently led directly to changes in the model; and the sixth column indicates the number of replications that were involved in each study. The entries for each of these characteristics were not always obvious, and many are subject to debate. For example, it was not always clear whether a finding was used to modify the model, and it was not always clear whether an analysis of a dynamic model over several years was a single replication, because some factors undoubtedly changed during the period of analysis.

These 13 studies demonstrate, first, the lack of any (at least formal) sensitivity analysis for many of the models; our review found only five models that have had a formal sensitivity analysis. This lack greatly hinders the feedback process of model improvement. Our review also found only nine studies involving an external validation, and then only one of them clearly with more than one replication. These studies do not allow us to develop direct measures of model performance, except for individual situations that are unlikely to apply generally and that have great uncertainty associated with them. Lastly, although not covered in the table, to our knowledge there has been only one recent effort to make use of sample reuse methods to determine the variance of the output from microsimulation models (Wolfson and Rowe, 1990).

As shown in Table 9-1, model validation can lead to model improvement. This is all the more reason to be concerned that very few microsimulation models have been validated. Because there have been so few attempts to develop error estimates for the output of any microsimulation model, there is

Copyright National Academy of Sciences. All rights reserved.

TABLE 9-1 Summary of Literature Review: Microsimulation Model Validation

Study	Model	Feature: Sensitivity Analysis	Feature: External Validation	Feature: Feedback Improvement	Replications
Hendricks and Holden (1976a)	DYNASIM	Yes	Yes	Yes	1
Hendricks and Holden (1976b)	DYNASIM	Yes	No	No	N.A.
General Accounting Office (1977)	TRIM	Yes	Yes	?	1
Holden (1977)	DYNASIM	Yes	No	Yes	1
Hayes (1982)	MICROSIM	Yes	Yes	Yes	1
Jefferson (1983)	DYNASIM	No	Yes	No	1
Haveman and Lacker (1984)	DYNASIM and PRISM	No	No	?	N.A.
ICF (1987)	TRIM2 and HITSM	No	Yes	Yes	1
Kormendi and Meguire (1988)	TRIM2	Yes	Yes	No	2
Betson (1988)	KGB	Yes	No	No	1
Doyle and Trippe (1989)	MATH	Yes	Yes	Yes	1
Beebout and Haworth (1989)	MATH	No	Yes	No	1
Burtless (1989)	MATH and TATSIM	No	Yes	No	1

NOTE: N.A., not available
SOURCE: Data from Cohen (Chapter 7, in Volume II).

Copyright National Academy of Sciences. All rights reserved.

no answer today for questions such as the relative contribution of sampling variance to total uncertainty for microsimulation models.

Our review does give us some reason for optimism. Much of the work that was done is of very high quality, indicating that there is an appetite for such analysis and that this type of work is currently feasible. We discuss below three of these studies to give a sense of some of the methods used and the results of the application of those methods.

Doyle and Trippe

Doyle and Trippe (1989) conducted a two-phase validation of the MATH model forecasts of the parameters of the food stamp program in 1984. In the first phase, they compared administrative data for August 1984 with simulated values from the model implemented with a March 1985 CPS database, using program parameters for that time. This procedure removed the contribution of forecasting error, which is usually present in MATH simulations based on databases that have been aged forward from an earlier year. They also compared the CPS-based results with simulations based on the 1984 SIPP.

In the second phase, Doyle and Trippe directly evaluated the aging module by projecting an unaged 1980 database (from the March 1981 CPS) to 1984, making use of historical data from the March 1985 CPS to generate the control totals, again eliminating the contribution of forecasting error due to incorrect control totals. They compared the distribution of household characteristics for the low-income population—including characteristics controlled for and those not controlled for in the aging process—across the unaged 1980 database, the aged 1984 database, and the actual 1984 database (from the March 1985 CPS).

Doyle and Trippe began their analysis with the following a priori assumptions about possible causes of problems originating in the data inputs for MATH: data limitations in understanding behavioral decisions; weak macroeconomic projections; nonsampling errors in the March CPS, such as underreporting of income; undercoverage of selected population groups in the March CPS; and the omission, in the March CPS, of variables such as assets that are necessary to determine program eligibility and benefit level. For comparison values in the first-phase study, Doyle and Trippe used the administrative data on the food stamp caseload, while recognizing that these data are also subject to error. To account for the sampling error present in the administrative estimates, they used confidence intervals about the administrative estimates for comparison, rather than only the estimates themselves. In addition, they used tolerance levels that attempted to represent differences that were not important in a subject-matter sense. They defined the tolerance levels to be the greater of two values: 5 percent of the value of the administrative estimate or twice the sampling standard error.

Doyle and Trippe's first-phase analysis demonstrated that MATH was

Copyright National Academy of Sciences. All rights reserved.

effective in estimating program costs and caseloads given by administrative data and in providing estimates of the distribution of the caseload by important characteristics, including household size and gross monthly income. However, they showed that, in other respects, MATH was less effective. A major problem was that MATH simulated too few food stamp households that receive public assistance and have school-age children and too many food stamp households that include elderly people and have earnings. Doyle and Trippe thought that these problems likely stemmed from errors in the March CPS, and so they examined whether using SIPP data to perform the simulations in place of the March CPS would alleviate this problem, but the discrepancies persisted. The analysis also determined that the MATH model's procedure for imputing asset values, which are not reported in the March CPS, is inadequate and needs to be improved.

The second-phase analysis conducted by Doyle and Trippe examined the benefits of using the aging module rather than working with an unaged file, in the usual modeling situation in which one must develop estimates for a future year. As would be expected, aging was beneficial for variables that are used as controls in the aging process. However, aging was either not beneficial or somewhat harmful for variables that are not included as controls, such as the distribution of households by income as a percentage of the poverty threshold.

Haveman and Lacker

Haveman and Lacker (1984) analyzed the differences between DYNASIM and PRISM in their estimation of future public and private pension benefits. The initial database for DYNASIM was the March 1973 CPS-SSA exact-match file; the initial database for PRISM was an exact match of the 1978 CPS-SSA exact-match file with the May 1979 and March 1979 CPS. Haveman and Lacker (1984:3) comment: "Although the microdata simulation procedures on which these models rest are marked improvements over previous methodologies, the ability to project retirement income with accuracy has yet to be demonstrated . . . . For the two models under consideration here, the baseline projections do diverge." For example, for the year 2000, DYNASIM projected the average private pension benefits for 65-year-old males to be $3,509; PRISM projected a value of $6,160.

Haveman and Lacker put forward five possible sources of these discrepancies: (1) different initial samples, (2) use of different specifications for the endogenous relationships, (3) relationship estimates from different data sets, (4) different judgments for situations in which no data existed, and (5) use of different exogenous parameter values. Because of budget constraints, Haveman and Lacker could not carry out a sensitivity analysis; instead, they did a qualitative assessment of the likely source of the differences. They admit that this approach does not permit a clear answer as to which model's projections are

Copyright National Academy of Sciences. All rights reserved.

more reliable. For example, they note that DYNASIM takes race, education, and marital status into consideration in its mortality module; PRISM does not. On the other hand, PRISM makes use of disability. However, without exchanging these modules in the two models, it is not clear how much difference this makes.

Haveman and Lacker do point out that it is often possible to perform the equivalent of a sensitivity analysis by simply examining the competing algorithms. For example, they argue that the PRISM estimates of outlays in private defined-benefit plans would be nearly 50 percent higher than the estimates from DYNASIM as a result of indexing nominal-valued constants to different series. However, Haveman and Lacker admit to the limitations of their approach (1984:109):

> Overall, we found both models to be impressive and highly innovative pieces of model-building research. Their impressiveness, however, is simultaneously their weakness. The enormous complexity which they embody makes them, effectively, black boxes. The input into them can be seen, understood, and judged. The projections which they yield can be understood and are illuminating. Yet how the assumptions and other inputs came to yield the printed-out projections cannot be seen, understood, or judged. The interaction of the complex relationships, transition matrices, time-triggered status changes, random drawings from unknown pools, and constraints to insure comparability is so complicated that little intuition or 'feel' is possible for why the resulting projections are what they are. The evidence which would lead a reviewer to believe the predictions of one model more than that of another is slim, indeed.

They recommend three approaches for investigating microsimulation models: backcasting, sensitivity analyses in which certain modules are switched between the models, and the analysis they performed for PRISM and DYNASIM.

The kind of careful scrutiny of DYNASIM and PRISM that Haveman and Lacker gave in their paper, in conjunction with an external validation, would greatly expedite determination of the reasons for any discrepancies between the models and the truth. In addition, work of this sort provides modelers with obvious examples of alternative methodologies that can be used in a sensitivity analysis. Therefore, although the Haveman and Lacker analysis is not a validation per se, it represents an important component of validation.

Kormendi and Meguire

Kormendi and Meguire (1988) examined the performance of TRIM2 in estimating the number of households that participate in some form of welfare assistance program and the benefits they receive. They took two different approaches to this validation of TRIM2. The first approach was randomly to perturb parameters in the AFDC participation module of TRIM2, according to reasonable

Copyright National Academy of Sciences. All rights reserved.

estimates of the variability of these estimated parameters, and then to examine the resulting variability in the estimates of participation and level of benefits. This exercise represents an interesting form of sensitivity analysis.

The results of this first part of the validation of TRIM2 indicated that the variability of simulated benefits was less than that of participating units. In nearly all cases, the coefficients of variation (i.e., the standard errors as percents of the estimates) of simulated benefits were less than 3 percent. At the same time, the coefficients of variation of simulated units ranged from 4 to 18 percent.

In the second part of the analysis, which Kormendi and Meguire label dynamic validation, they used TRIM2 to simulate the changes in welfare participation and benefits that occurred from 1979 to 1985 and from 1983 to 1985, comparing the results to those from the March 1986 CPS. The first time period spanned the enactment of the 1981 Omnibus Budget Reconciliation Act, which made a number of cutbacks in welfare programs. The second period represented a time when no major legislation was enacted.

Kormendi and Meguire considered the 50 states and the District of Columbia as mini-TRIM2 models, which allowed them to make use of regression analysis in their validation of TRIM2. One example of this was to regress, for several outputs, the forecasts derived from the 1979 and 1983 baseline files on the "true" values from the March 1986 CPS and compare the estimated regression coefficients with 1, which they defined as one type of unbiasedness. They also used reverse regression to examine whether the forecasts might be improved through correction of the bias.

Another application of regression used by Kormendi and Meguire was in the context of a sensitivity analysis. They attempted to attribute the changes from 1979 to 1985 to either administrative or economic-demographic changes. To do this, they ran TRIM2 based on the March 1986 CPS on the 1979 law, and they ran TRIM2 based on the March 1980 CPS on the 1985 law, in addition to the usual runs with the March 1980 CPS on 1979 law and the March 1986 CPS on 1985 law. With this approach, they were able to measure the percentage change in forecasts conditional on the law remaining constant, adding this measure as a covariate in the regression described above. Hence, they were able to control for variation due to changes in economic and demographic conditions.

Their results showed that the forecast errors of TRIM2 for the period 1979-1985 were much smaller for benefits than for units: the absolute value of the errors for units generally ranged from 5 to 30 percent; those for benefits ranged from 1 to 6 percent. These results are somewhat surprising in view of the fact that TRIM2 baseline files are routinely calibrated to administrative values for participating AFDC units by state and not to the values for total benefits.

Copyright National Academy of Sciences. All rights reserved.

A VALIDATION STUDY OF TRIM2: THE PANEL'S EXPERIMENT

The panel, in conjunction with the Urban Institute, decided to perform an illustrative validation of TRIM2—our "experiment." There were four primary goals of the experiment: (1) to determine the resources required to conduct a sensitivity analysis and an external validation and, more generally, to examine the feasibility of microsimulation model validation; (2) to identify those modules that, when altered, have an appreciable effect on model outputs; (3) to obtain an admittedly limited measure of model validity against comparison values from administrative records (limited because only one time period was examined); and (4) to illustrate some of the data analytic techniques that analysts can use to help answer these questions. In addition, we were interested in attempting to identify component alternatives that were superior to those currently in TRIM2, although this was not a major objective of the study.[6]

Description

The first task in planning the experiment was to identify several modules currently in use in TRIM2 that could be replaced by alternative modules that were, a priori, reasonable substitutes. Some of the alternative modules either had been used previously or had been considered as substitutes in place of the current algorithms. TRIM2 would then be run with various combinations of alternatives for these modules to determine which choices resulted in the greatest variability in the model's outputs. In addition, using administrative quality control estimates as surrogates for the truth, we could determine which alternatives were more successful in approximating the "true" values.

We decided to use TRIM2 with 1983 data (based on the March 1984 CPS) to estimate the costs of the 1987 law for the AFDC program in 1987, as well as a number of other distributional characteristics of the 1987 program in 1987. We chose the years 1983 and 1987 because the March 1988 CPS (which we needed to generate known population control totals) was the latest file available when the panel began its experiment; we wanted at least a 3-year forecasting window; and definitional and other comparability problems began cropping up for CPS data as the forecasting horizon grew appreciably longer. Although we avoided some comparability problems, the period we examined and the March 1984 and 1988 CPS files exhibited some unique features that limit the generalizability of the results. Of course, every time period and every database are unique in some respects, and therefore no single experiment can be used to make general inferences about the efficacy of a model.

[6] Cohen et al., in Volume II, provide a full discussion of the experiment. The analysis data set is available from the Committee on National Statistics upon request.

Copyright National Academy of Sciences. All rights reserved.

We based our choice of modules for the experiment on considerations of ready availability of alternatives and substantive interest. Because of expected difficulties in reprogramming, we excluded some modules of strong interest on grounds of excessive cost. Our criteria resulted in the choice of three modules: adjustment for population undercoverage, imputation of monthly employment and earnings, and aging. The following descriptions of our modules also show our "name" for each module alternative.

Adjustment for population undercoverage ("adjustment"). TRIM2 currently does not attempt any correction for undercoverage of certain population groups in the March CPS. By modifying models in Fein (1989), we derived a logistic regression model specifying rates of undercoverage for households with various characteristics. The effectiveness of this logistic regression model was limited by the variables that existed in the TRIM2 database and by the necessity of using the same weight for all individuals in a household to adjust for undercoverage. Nevertheless, the model provided an opportunity to see if weighting for undercoverage would make a difference in the model's outputs.

Imputation of monthly employment and earnings ("months"). Because the March CPS provides information only on annual employment and earnings, while the AFDC program operates on a monthly accounting basis, TRIM2 includes a module "MONTHS" that endeavors to capture monthly variation in employment, unemployment, and earnings ("current"). A simpler procedure employed in an earlier version of TRIM2, which we refer to as "old MONTHS," simulates a maximum of two spells during the year, one working and one not working ("old").

Aging. Although static aging modules are available for TRIM2, the model as currently applied does not invoke them. We specified three static aging alternatives and not aging ("none") (see details in Giannarelli, 1989). The first aging alternative was to invoke the demographic aging routine to reweight the records in the March 1984 CPS to match target values for population totals by age, sex, and race generated from the March 1988 CPS ("demo"). The second aging alternative, unemployment aging coupled with demographic aging, additionally invoked the routine to adjust labor force activity on the demographically aged file to meet targets from the March 1988 CPS for unemployment during the week of the survey and the preceding calendar year ("unemp"). The third aging alternative, full aging, additionally invoked the routine to adjust income amounts for price changes and economic growth between 1983 and 1987 ("full"). In every case, known control totals were used from the March 1988 CPS to eliminate the source of variability due to erroneous demographic and macroeconomic forecasts.

We therefore specified 16 different "alternative" versions of TRIM2 (2 x 2 x 4) for three modules. It is important to note that we selected the alternatives for each of the three modules either because, a priori, they were not clearly

Copyright National Academy of Sciences. All rights reserved.

worse than the existing module in TRIM2—notably the aging alternatives and undercount adjustment—or because they were substantially simpler and, we had reason to believe, not substantially worse than the existing TRIM2 module—notably old MONTHS.

In addition to the 16 variations simulating 1987 law with the 1983 database, we had available: (1) the 1983 baseline files, both adjusted ("A.Base") and unadjusted ("Base") for undercoverage, which simulated 1983 law with the March 1984 CPS; (2) the 1987 baseline files, both adjusted and unadjusted for undercoverage, which simulated 1987 law with the March 1988 CPS; and (3) tabulations from the Integrated Quality Control System ("IQCS") of characteristics of the AFDC recipient population in both 1983 and 1987. Throughout the remainder of this section, we treat the 1983 and 1987 IQCS data as the "truth," that is, the comparison values. We have remarked elsewhere about the dangers in this assumption and have noted methods that might be used to deal with the problem of sampling error in the comparison values. In addition, the quality control data are believed to be subject to nonsampling errors in measurement of characteristics such as the composition of the AFDC unit's household. Table 9-2 shows the features of all 22 runs in the experiment.

Results

The first question that occurs is what differences in the outputs resulted from the alterations in the three modules. In other words, do the alternatives make a difference?

For estimates of change in the total number of AFDC participants from 1983 to 1987, the estimates from the 16 versions of TRIM2 range (in units of 1,000) from -213 to +293; the comparison value from the IQCS is 98 (2.7% of total AFDC participants in 1983). For estimates of change in total benefits (not adjusted for inflation), the estimates from the 16 versions of TRIM2 range (in millions of dollars) from $1,578 to $3,742; the comparison value is $2,499 (18.2% of total AFDC benefits in 1983). That is, in using alternatives that were a priori thought to have similar success in modeling AFDC, the alternatives disagree about whether the number of participants is going up or down, and by at least twice the size of the observed change of 100,000 people. The alternatives also provide estimates of increased costs in total benefits that range from $1.6 billion to $3.7 billion. Moreover, these estimates of change are with a forecast horizon of only 4 years.

Table 9-3 provides ranges for estimates of both level and change for other statistics of interest, many of which are percentages of dichotomized variables. (The latter are analyzed in their original, undichotomized form below.) Note the instances when a model version incorrectly estimated the direction of change. Also note that, for estimates of change for three variables—race of head, earnings, age of youngest child—the range of values does not include

Copyright National Academy of Sciences. All rights reserved.

the comparison value. In general, the results in Table 9-3 show that there are important differences in most of the estimates when alternate modules are used.

TABLE 9-2 Description of TRIM2 Experimental Runs

Run Identification	AFDC Law	CPS Year	Adjustment	Months	Aging
$Base_{83,83}$	1983	1983	No	Current	—
$Base_{87,87}$	1987	1987	No	Current	—
1	1987	1983	No	Current	None
2	1987	1983	No	Current	Demo
3	1987	1983	No	Current	Unemp
4	1987	1983	No	Current	Full
5	1987	1983	No	Old	None
6	1987	1983	No	Old	Demo
7	1987	1983	No	Old	Unemp
8	1987	1983	No	Old	Full
$A.Base_{83,83}$	1983	1983	Yes	Current	—
$A.Base_{87,87}$	1987	1987	Yes	Current	—
9	1987	1983	Yes	Current	None
10	1987	1983	Yes	Current	Demo
11	1987	1983	Yes	Current	Unemp
12	1987	1983	Yes	Current	Full
13	1987	1983	Yes	Old	None
14	1987	1983	Yes	Old	Demo
15	1987	1983	Yes	Old	Unemp
16	1987	1983	Yes	Old	Full
IQCS83		IQCS data for 1983			
IQCS87		IQCS data for 1987			

NOTE: See text for definitions and descriptions of terms.

These observations are not meant to imply any weakness of TRIM2 relative to other models, microsimulation or otherwise, that have the goal of providing estimates of characteristics. No such conclusions can be drawn because this analysis does not include other modeling approaches. What our experiment does demonstrate is the large amount of variability that results from reasonable changes to the basic model.[7]

The observed variability has a great deal of structure because underlying the 16 observations for each response lies a 2 x 2 x 4 factorial design with one replication per cell. To help understand the variability, we used analysis of variance to measure the relative sizes of the main effects, to test for significance of the main effects, and to test for significance of the interactions. (See the report of this analysis in Cohen et al., in Volume II.)

[7] Note that these ranges are not intended to be interpreted as any type of confidence interval.

Copyright National Academy of Sciences. All rights reserved.

TABLE 9-3 Observed Differences for Various Estimates from TRIM2 Validation Experiment

Variable	IQCS87 Level	Range of Level, Runs 1-16	IQCS87 Minus IQCS83	Range of Change, Runs 1-16 Minus 83 Base
Type of AFDC unit (% basic)	90.40	87.99, 91.90	1.20	–0.45, 3.52
No. of adults (% 2)	9.58	6.32, 10.43	–0.65	–3.41, 0.42
No. of children (% > 2)	25.98	26.36, 28.21	0.58	0.29, 1.79
Total no. in unit (% > 3)	28.20	28.72, 31.10	0.13	–0.27, 1.46
Age of youngest child (% < 5)	53.58	51.96, 55.25	–0.09	–2.76, –0.64
Race of head (% black non-Hispanic)	41.30	39.50, 42.75	–2.84	–1.85, 0.01
Earnings of adults (% none)	92.31	91.22, 94.88	–2.26	–0.91, 1.67
Marital status of head (% no spouse)	86.82	86.01, 90.42	1.61	–0.45, 3.60
Sex of head (% female)	86.77	82.96, 86.89	1.30	–0.50, 3.53
Age of head (% < 20)	5.73	6.51, 7.30	–1.18	–0.62, –1.51

NOTE: Differences among model runs and between a model run and the IQCS may not be statistically significant.

The principal finding from our analysis of estimates of change was that changing from unadjusted to adjusted data and changing the type of aging used make a substantial difference in most of the output estimates. Changing from MONTHS to old MONTHS makes a much smaller difference. In addition, some interaction effects were observed, which complicates the identification of better and worse modules. We note that no interactions were observed for the parallel analysis of estimates of levels.

Because we also had comparison values, we were able to compare outputs of the different model versions with the comparison values. There are many ways to examine whether some models are "better" than others with respect to closeness to the comparison values (see Cohen et al., in Volume II). First, in conjunction with the analysis of variance, looking at the comparison values indicates which main effects are bringing the overall mean for the 16 model versions closer to or further away from the comparison value; this analysis provides an indication of which alternatives are more promising. Second, the errors themselves are easy to display in tabular form, which provides a visual impression of a superior version, if one exists. A difficulty with this type of presentation is the vast difference in size of the variables for which estimates

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

are given. A third possibility is to replace the errors by their ranks over the 16 model versions; one could use this procedure as a descriptive tool or use available nonparametric statistics to base inferences on these ranks. The general impression we gained from these analyses is that the results show no clear pattern, no clearly superior version of TRIM2. For various outputs, different versions show alternately strong and weak results, but there are no general patterns.

We also examined, for some of the frequency table outputs, how close the various versions of TRIM2 corresponded to the same frequencies from the 1987 IQCS. Note that here we are not examining estimates of change because there is no simple way to examine changes in a distribution. (We dichotomized some of these frequencies in the previous analysis, which is one way of focusing on a characteristic of a frequency table that makes analysis of change feasible.)

Table 9-4 presents the $?^2$ test of independence for a 2 x r contingency table, in which one row contains the frequencies from one model version and the second row the frequencies from the IQCS data for 1987. Notationally, in which n_{ij} is the number of persons in category j estimated by run i; $n_{i.}$ is the sum over categories—the total number of people "produced" by run i; $r_{j.}$ is the sum over runs—the total number of people in a certain category; and N is the total number of people "produced" by the two runs (in which one run, in this case, is the IQCS data). The statistic Q has a chi-square distribution with r - 1 degrees of freedom. (This test can also be used for more than two sets of estimates, if desired.) This statistic, along with the associated degrees of freedom, is presented in Table 9-4 for all 16 model versions, for a variety of model outputs.

$$Q = N[\{\sum(1/n_{i.})\sum(n_{ij}^2/n_{.j})\} - 1],$$

One difficulty in the interpretation of the $_x{}^2$ statistics shown is that the statistics for different rows are not comparable because they are associated with different degrees of freedom. However, one can compare model versions within each row and form an assessment of whether some model versions consistently outperform others. (Table 9-4 provides the $_x{}^2$ values for degrees of freedom from 1 through 8 at the 99 percent confidence limit. A higher value in the table for a model version indicates that it differs from the 1987 IQCS by an amount greater than one could expect by chance.) For some variables, such as gross income of unit, all model versions obviously differ substantially from the IQCS.

As in the analysis of change, we found that no model version appears to have any noticeable advantage over other versions. For example, model 1 (the current version of TRIM2) has a low $_x{}^2$, relatively speaking, for the dichotomized variables unit size, number of adults, number of children, age of head of household, marital status of head of household, and size of benefit.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

TABLE 9-4 X2 Goodness-of-Fit Statistics for Distributions, from TRIM2 Validation Experiment

	Run Identification							
Variable	1	2	3	4	5	6	7	8
Total no. in unit	26	37	28	28	27	36	28	27
No. of adults	10	27	39	45	11	27	41	45
No. of children	2	5	6	6	2	4	5	5
Age of youngest child	11	14	12	12	11	16	13	12
Gross income of unit	568	508	461	396	506	456	423	362
Earnings of adults	15	19	18	8	14	26	23	17
Type of AFDC unit	15	3	13	15	17	3	12	14
Race of head	4	9	2	2	6	10	2	2
Sex of head	28	19	1	0	31	21	1	0
Age of head	30	47	45	48	29	49	45	46
Relationship of unit head to household head	1.2	0.9	0.9	0.9	1.2	0.9	0.9	0.9
Marital status of head	I	1	15	20	1	2	16	19
Size of benefit	85	114	101	106	87	116	108	109

D.F. = 1, ^{2} = 6.635	D.F. =5, ^{2} = 15.086
D.F. = 2, ^{2}= 9.210	D.F. = 6, ^{2} = 16.812
D.F. = 3, ^{2} = 11.341	D.F. = 7, ^{2} = 18.475
D.F. = 4, ^{2} = 13.277	D.F. = 8, ^{2} = 20.090

* D.F. indicates degrees of freedom. The x^2 values at the 99 percent confidence limit are as follows (a higher value in the table indicates that a model version differs from the 1987 IQCS by an amount greater than one could expect by chance):

Therefore, for these variables, model 1 was one of the more successful models for approximating the IQCS data. At the same time, model 1 has a relatively high ?^{2} for the variables gross income of unit, type of unit, and sex of head of household. This lack of general superiority or inferiority is true for all 16 model versions.

The last column of Table 9-4 displays the results from using the 1983 IQCS, which assumes that the characteristics of the caseload in 1987 remain unchanged from those in 1983. Under some circumstances, the comparison values for the beginning of a period provide an interesting challenge to the model versions. If a study examines a situation in which a substantial policy change has occurred, the performance of the old IQCS data provides a standard

Copyright National Academy of Sciences. All rights reserved.

that a reasonable model should exceed, namely, the model should do better than an estimate based on the assumption of no change over the period. Our experiment (as discussed further, below) examined a situation in which the policy change was modest, but economic changes in the period were relatively large. Under these circumstances, the comparison is less important because the noise is in some sense too large a fraction of the signal. Nevertheless, the 1983 IQCS data set does not compete well with the 16 versions of TRIM2 in the analysis (shown in Table 9-4), but it does outperform many TRIM2 versions in estimating total participants and other aggregates (see Cohen et al., in Volume II). In many situations, this type of comparison is extremely informative in providing a naive estimate of how well one can do with a very simple model. Also, this comparison provides an estimate of how much variability is natural to the problem, which can be compared with the variability left unexplained by the model versions.

9	10	11	12	13	14	15	16	IQCS83	D.F.[a]
48	19	40	40	49	49	42	41	73	4
33	8	45	48	35	39	48	51	438	2
4	6	6	6	6	5	6	6	41	3
12	10	5	5	10	7	6	6	77	4
436	508	362	310	419	412	338	287	259	8
16	38	8	13	20	24	20	29	426	8
9	0	5	6	12	7	5	6	75	2
7	16	6	8	5	5	5	6	493	3
14	3	0	0	20	19	0	0	48	1
115	101	40	40	32	38	39	40	272	7
1.3	1.1	0.9	0.9	1.2	0.9	1.0	1.0	0	4
1	4	20	23	0	1	21	24	57	1
50	85	120	119	99	127	121	119	2015	7

Limitations of the Experiment

Our experiment was designed both to illustrate the types of methods that can

Copyright National Academy of Sciences. All rights reserved.

be used to validate a model's outputs and to provide some indication of the current performance of TRIM2. With respect to the second of these goals, the experiment is limited in a variety of ways.[8] First, the experiment studied TRIM2 during only one time period. With respect to estimating error properties, even in a replicable situation, one replication is very limiting: a model can perform better than it would on average, or worse. Moreover, forecasting situations should not be considered replicates without further investigation. Different time periods will typically present different challenges to a model. In particular, any characteristics of either the 1984 or 1988 March CPS data, or of the period under study, that are peculiar to those data sets or to that time period reduce the opportunity for generalizing from our results.

With regard to peculiarities of the data, an analysis conducted as part of the experiment showed that simulations of 1987 law using the 1983 baseline file sometimes outperformed simulations of 1987 law using the 1987 baseline file. This finding triggered a more extensive analysis of the quality of the March 1988 and 1984 CPS data (see Giannarelli, 1990). That analysis documented that the March CPS files typically include some states that have insufficient simulated units eligible for AFDC compared with administrative counts of participants—a phenomenon that complicates the calibration effort. It turned out that the March 1988 CPS had an unusually large number of such states (8), which made the calibration in that year less successful than usual. The number of simulated eligible units dropped by large percentages in some states from the previous year (e.g., by 29% in Connecticut, 22% in Michigan, and 32% in New Mexico). Although sampling variability in the CPS appears to explain some of these changes, Michigan's drop remains something of a mystery.

It is important to point out that, although microsimulation modelers are aware of quality problems with the March CPS data, they do not regularly investigate changes in quality from year to year. Hence, we have an example of how validation can lead to more pointed validation and to identification of problems for further investigation and possible correction.

The period from 1983 to 1987 was also special because of the large drop in the unemployment rate, from 8 percent in 1983 to 5 percent in 1987, and the differential impact of the change in unemployment on different subpopulations. We note that the changes in welfare program regulations between 1983 and

[8] Indeed, the limitations of the experiment, including that only one time period was examined, constrained our analysis to emphasize primarily descriptive rather than inferential techniques and interpretations of the data. However, inferential analysis, including hypothesis testing, with the goal of identifying differences and patterns that are statistically significant versus those that are not (which is not often possible with the descriptive approaches that we used) has a great deal to offer when the number of replications increases. We certainly encourage the use of inferential techniques, such as nonparametric analysis of variance (see, e.g., Lehmann, 1975) and anticipate that expertise with respect to which models and techniques are most applicable will follow as experience is gained with these types of validation studies.

Copyright National Academy of Sciences. All rights reserved.

1987 were relatively minor, which limits our findings to periods when there are few changes in law.

The limitation imposed by having only one time replication is somewhat offset by having outputs for several characteristics that focus on fairly distinct portions of the model and by having outputs for states that can serve as mini models (for which we have not done an extensive analysis). Clearly though, there are circumstances that would cause a model to perform poorly for all states for a single time period but would not be indicative of the model's overall efficacy, due to the likely correlation among states for some response variables.

The limitations resulting from examining TRIM2 at only one time help demonstrate that validation should not be an occasional examination of a model. Rather, model validation should be a continuous process that accumulates knowledge about potential model weaknesses, the size of errors, and situations in which a model is less reliable.

Another limitation of our experiment was that, in addition to simulating just one time period, we simulated just one change in law. Had time and resources permitted, we would have liked to simulate an alternative policy— for example, a mandated AFDC minimum benefit nationwide—that would have represented a major policy change. Of course, we could not have conducted an external validation of such an alternative policy because it was never enacted. However, we could have conducted sensitivity analyses of both policies and obtained information relevant to the question of whether, in fact, microsimulation model estimates of differences between two policies are less variable than estimates for a particular policy, because of sources of error affecting both policies to about the same extent.

Yet another limitation of our experiment (noted above) is that the surrogate for the truth that we used throughout may have weaknesses. First, the IQCS data are from a sample survey and therefore subject to sampling error. They are also subject to bias from a number of sources. For example, different states have different collection procedures for their quality control data, which may lead to different kinds of biases across states. In our experiment, we ignored the problems raised by the use of an imperfect surrogate for the truth. Whenever feasible, analysts should search for the sources of major discrepancies between model estimates and the comparison values in both data systems. At the same time, analysts should ignore discrepancies between model estimates and the comparison values that are smaller than what would be explained by ordinary sampling variability.

Another limitation of the experiment was that we examined only one model, due to the time and resource constraints under which we, the sponsoring agencies, and the agencies' contractors operated. Ideally, it would have been desirable to expand the experiment to include other major models in use today, such as MATH, HITSM, DYNASIM2, PRISM, and MRPIS. However, we should note that all of these models are not directly comparable because they

Copyright National Academy of Sciences. All rights reserved.

are often used for different purposes and, by design, cannot necessarily all produce estimates of the same variables. Therefore, adding the model class as another factor in an analysis of variance, for example, may not generally be feasible. However, in cases in which models are directly comparable, one could certainly expand our basic experiment in that direction.

If one could overcome the problems of comparability, there are major advantages to be gained through comparing the effectiveness of different broad modeling strategies. For example, the question of the relative advantages of static versus dynamic aging could be addressed in this way. Comparing several models with the truth (or a reasonable surrogate), along with a comprehensive analysis of model differences exemplified by Haveman and Lacker (1984), will often yield great insight into the strengths and weaknesses of the various models.

Major Conclusions

The panel's experiment was successful in demonstrating that microsimulation model validation is feasible. With current methods, analysts can measure the degree of variability attributable to the use of alternative components. Such information helps indicate overall model uncertainty, as well as which components to examine further to make improvements to a model. Thus, sensitivity analysis methods, especially when augmented with comparison values in an external validation, provide a great deal of data with which to direct efforts at model development as well as to measure model uncertainty.

Our experiment demonstrated that there is considerable uncertainty due to changes in the three modules we studied in TRIM2. Therefore, the choice of which model version to use makes a difference. Yet it is not clear that any of the 16 versions has any advantage over the others. Certainly, for individual responses, particular versions fared better. However, given that the experiment is only one replication, it would be foolish to assume that our results provide confirmation of any real modeling advantage.

Because the experiment did not attempt to measure the variance of any of the versions of TRIM2, we have no idea of the relative sizes of various sources of uncertainty in relation to variance. Therefore, it is difficult to assign a priority to development of variance estimates vis-á-vis use of sensitivity analysis. We do believe that it is important to investigate both.

We stress that our experiment was purely illustrative. The benefits from a continued process of validation are rarely evidenced through study of a single situation. There is an important question about the degree to which different studies of this sort of the same model in different modeling situations would represent replications in any sense. However, even if the studies are not replications, use of these methods will provide evidence of general trends in model performance. Their use will generate a great deal of information as

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

to the situations under which a model performs well and provides accurate information.

While we have made a convincing case for the feasibility of sensitivity analyses and external validation, the experiment was not cheap. The Urban Institute estimated staff costs (including overhead) to conduct the experiment of about $60,000 for 1,400 person-hours of effort or, roughly, 35 person-weeks. These estimates are probably low because it was difficult for the Urban Institute to separate activities that were needed for the experiment from their own day-to-day efforts; in addition, they do not take into account the time taken to specify the experiment and analyze the data. Moreover, we do not have estimates of computer costs. Overall, it is clear that the way in which TRIM2 (and most other microsimulation models) is currently configured can make a sensitivity analysis very costly.

The costs were dramatically affected by our interest in trying out different forms of aging. The overall cost would have been substantially reduced, possibly by a factor of 2 or 3, had an easier module been selected for the experiment. But there were other modules that we did not investigate because the costs of working with them would have been higher still. It is obvious that, for model validation to become a routine part of the model development and policy analysis process, the structure of the next generation of models must facilitate the type of module substitution that is used in sensitivity analysis.

In summary, our experiment gave mixed signals on the effectiveness of TRIM2. That TRIM2 is sensitive to the inclusion or exclusion of various factors is apparent. Our results suggest that, in some instances, nothing was gained by implementing TRIM2 rather than available IQCS data. However, in other instances, TRIM2 indeed provided very valid estimates. Our main goal was to show how one might undertake sensitivity and validity studies for microsimulation models. It is quite reasonable to speculate that similar studies on other microsimulation models will produce comparably mixed results, namely, that the model under study will prove to be useful for some variables but not as good as had been believed for others. This knowledge can only be valuable to the analysts using models to inform policy makers as well as to those involved in making improvements to the models. We have made a small start toward this end.

STRATEGIES FOR VALIDATING MICROSIMULATION MODELS: RECOMMENDATIONS

Our validation study of TRIM2 illustrates both the benefits and the costs of serious attempts to investigate the quality of estimates produced by microsimulation models. The benefits, even in our very limited study, seem clear to us. We determined that TRIM2 estimates are sensitive to alternative choices for model components. We also observed weaknesses in the March CPS database for

Copyright National Academy of Sciences. All rights reserved.

modeling income support programs. These weaknesses in the March CPS were more or less well known, but our findings underscore the need to investigate them further and to take corrective action of some kind.

The costs of validation are also evident from our experiment, in terms of both time and resources. Indeed, the kinds of external validation studies, sensitivity analyses, and variance estimation procedures that we outline for microsimulation models may well appear to involve a dismayingly high expenditure of staff and budget resources, particularly in light of the limited resources that have been allocated to these activities in the past. Clearly, in the context of an ongoing policy debate, there is no possibility of applying such evaluations to even a fraction of the estimates and proposals that are modeled.

However, we believe that the cost-benefit ratio for microsimulation model validation can be improved substantially through several mechanisms. First, implementation of the next generation of models with new computer technology, as recommended in Chapter 7, should dramatically reduce the costs and increase the scope of feasible validation studies, particularly if the modeling software is designed—as it should be—with validation in mind. Second, improvements in model documentation and archiving that we recommend (see Chapter 10) should make it easier to carry out validation studies, particularly external validations. Third, academic researchers should find model validation questions of considerable interest, particularly when they are able to access the models directly through new technology, and their work on validation methodology and applications should prove fruitful. Finally, we believe that, as policy analysis agencies and their contractors gain more experience with validation, the task will become easier and more rewarding, particularly when validation results prove helpful in making decisions about priorities for investment in models.

Although the greatest improvements in microsimulation modelers' ability to carry out validation studies will come with the implementation of new technology, we believe that more validation can and should be accomplished in the short term with the current models. We outline a set of institutional arrangements that we believe will facilitate cost-effective model validation in the near term. We also recommend agency support of research on model validation methods and agency adoption of the "quality profile" concept as a way of communicating information about model strengths and weaknesses to a broad user community and a way of organizing a continuing program of model validation targeted to priority areas for improvement. We urge policy analysis agencies to allocate the necessary resources and make the needed commitment so that validation becomes a regular part of the microsimulation modeling enterprise.

Institutional Arrangements for Model Validation

In formulating our recommendations for model validation, we took cognizance

Copyright National Academy of Sciences. All rights reserved.

of the severe time pressures for producing estimates that characterize the policy analysis process. We also took cognizance of the relatively limited capabilities of current models for cost-effective validation. Hence, we do not suggest that each and every set of policy estimates be evaluated, either in real time or after the fact—such a recommendation would deservedly be ignored. We present instead a plan for a reasonable approach to the validation task.

In our view, major contracts for the development, maintenance, and application of microsimulation models should target a percentage of funds sufficient to carry out validation studies that can provide useful information to analysts and policy makers who are engaged in shaping legislation on a real-time basis. For those agencies that maintain and apply their own models in-house, rather than contracting for these services, the agency should allocate its own modeling budget in this way.[9] These contracts should also include an allocation of funds earmarked to implement model revisions on the basis of the results of validation studies conducted by the modeling contractor and others.

With regard to specific types of validation, the contractor would be expected to provide estimates of variability (such as a bootstrap estimate) and the results of sensitivity analyses for key sets of model estimates, where an example of a "key set" might be the first set of estimates prepared for the initially proposed version of the Family Support Act The contractor would not hold up delivery of these estimates until the validation was finished, but would endeavor to complete the validation as soon as possible. The results of the validation for one set of estimates would help interpret the quality of the estimates for alternative proposals (unless major new provisions were added). The sensitivity analysis would focus on those model components the analysts believe are most likely to have an impact on the particular set of estimates. In other words, the validation performed by the contractor would be "rough and ready," focused on helping to inform the policy debate. (See Chapter 3 for a discussion of the issues involved in communicating validation results to decision makers.)

In addition to the validation efforts performed by the modeling contractor, we believe it is essential for policy analysis agencies to commission independent validation studies that include external validation as well as sensitivity analysis. In principle, independent evaluation is preferable to evaluation performed by the developer and user of a model (just as academic journals appoint independent reviewers for articles submitted for publication). In practice, independent evaluation is preferable as well, given the pressures confronting a modeling contractor to respond to insistent and frequently changing policy demands for large volumes of estimates prepared within short time frames. Agency staff

[9] About 10-15 percent of funds might suffice for validation activities on an ongoing basis. However, given the relative lack of investment in microsimulation model validation to date, it may be that the percentage of funds earmarked in major contracts for validation purposes should be higher until sufficient experience is gained with validation techniques.

Copyright National Academy of Sciences. All rights reserved.

could carry out independent validation studies, but they, too, are usually under severe pressures from the demands of current policy issues.

Hence, our recommendation is that, for every major microsimulation model development, maintenance, and application contract that policy analysis agencies let, they also let another contract to a separate organization to carry out longer term, more comprehensive validation studies of the particular model(s). The validation contractor would be expected to carry out external validation studies of selected model estimates and also to conduct extensive sensitivity analyses in order to identify areas for needed model improvement or revision.

To implement a program of independent evaluations will clearly entail working out a number of practical matters. There will need to be ways to guard against conflicts of interest, such as a validation contractor deliberately downgrading a model in order to boost the chances for that contractor's own model winning the next bidding round. Most important, there will need to be cost-effective ways of providing validation contractors with access to the models they are evaluating and to modeling experts, without impairing the ability of the modeling contractors to respond to agency needs for real-time policy analysis. A possible approach is for a knowledgeable programmer to bring a second copy of the model to work on-site with the validation contractor; alternatively, staff of the validation contractor could work on-site with the modeling contractor. We are confident that workable arrangements can be devised. Looking ahead, we note that implementation of a new generation of models with computer technology that facilitates their use should make it much easier to deal with these problems and to expand the scope of the validation studies that are feasible to perform. Enhanced documentation will also facilitate independent validation of the type that we describe.

Recommendation 9-1. **We recommend that policy analysis agencies commit sufficient resources and accord high priority to studies validating the outputs of microsimulation models. Specifically, we recommend:**

- Agencies, in letting major contracts for development, maintenance, and application of microsimulation models, should allocate a percentage of resources for model validation and revisions based on validation results. The types of validation studies to be carried out by the modeling contractor should include estimates of variance and focused sensitivity analyses of key sets of model outputs. The goal of these efforts should be to provide timely, rough-and-ready assessments of selected estimates that are important for informing current policy debates.
- In addition, agencies, when practical, should let separate microsimulation model validation contracts to independent organizations or in other ways arrange to carry out comprehensive,

Copyright National Academy of Sciences. All rights reserved.

in-depth evaluations. The types of studies to be performed by a validation contractor should include external validation studies that compare model outputs with other values and detailed sensitivity analyses. The goal of these longer range efforts should be to identify priority areas for model improvement.

Research on Model Validation Methods

In addition to letting the kinds of validation contracts that we describe above, it would be useful for policy analysis agencies to support research specifically designed to develop improved methods for microsimulation model validation. For example, a useful topic for investigation could be ways to increase the cost-effectiveness of techniques such as the bootstrap for estimating the variance in model outputs. For this kind of work, the agencies could let separate methodological research grants to academic researchers. The agencies could also take steps to interest the National Science Foundation and perhaps the National Institute of Standards and Technology in supporting this type of research, which could well have application for validation of complex models in other fields.

Such work should be attractive to researchers, although the difficulties in providing them with access to current microsimulation models is an impediment. In the short term, perhaps an effective strategy would be to support fellowships for researchers to carry out methodological work on-site with the agencies' modeling contractors. The fellowships could be similar to those currently offered by several federal statistical agencies, including the Bureau of Labor Statistics, the Census Bureau, and the National Center for Education Statistics. (These programs are supported by a combination of National Science Foundation and agency funds, and are administered through the American Statistical Association.) Over the longer term, on the assumption that the next generation of models is successfully implemented with new computer technology, the agencies should find it quite easy to attract academic interest in the kinds of methodological work needed to improve model validation methods. Indeed, academic researchers would be able as well to conduct model validations.

Recommendation 9-2. **We recommend that policy analysis agencies provide support, through such mechanisms as grants and fellowships, for research on improved methods for validating microsimulation model output.**

Quality Profiles

Finally, as a way of organizing an ongoing, comprehensive program of validating microsimulation models and communicating the results of validation studies to users, we urge policy analysis agencies to adopt a concept that is gaining

Copyright National Academy of Sciences. All rights reserved.

ground in statistical agencies, namely, that of developing "quality profiles." A quality profile is a document that brings together all of the available information about sources of error that may affect estimates from a survey or other data collection effort (see Bailar, 1983, and the discussion in Chapter 5). The profile identifies measures and procedures for monitoring errors; assembles what is currently known about each source of error and its impact on the estimates; provides comparisons with estimates from other data sources; and outlines needed research and experimentation designed to gain better understanding of sources of error and to lead to the development of techniques to reduce their magnitude.

Clearly, developing a profile for a microsimulation model is a much bigger task than developing one for a single survey; however, the effort to conceptualize the sources of error and to bring together what is known about them in a document can be very helpful. Analysts who make use of the model output can benefit from the knowledge and caveats provided in a quality profile; model developers can use a profile to guide methodological work on understanding and reducing sources of error and to build a cumulative body of knowledge about their models' strengths and weaknesses.

Recommendation 9-3. **We recommend that policy analysis agencies support the development of quality profiles for the major microsimulation models that they use. The profiles should list and describe sources of uncertainty and identify priorities for validation work.**

Copyright National Academy of Sciences. All rights reserved.

10

Documentation and Archiving

Microsimulation models, like other models that are implemented with electronic computing technology, require documentation: of their input data, their software specifications and program code, their mathematical formulas and equations, their interface with the computer's operating system, the way in which analysts and programmers can interact with them, and other aspects of their design and use. Without adequate documentation, users cannot run a model or modify its software. Equally important, users cannot understand a model's design or operation and hence cannot properly specify changes to the inputs or interpret the outputs.[1]

The principles and practices that we propose for cost-effective design and implementation of microsimulation models in Chapter 6 stress the need to incorporate model features that facilitate good documentation and to provide sufficient resources and priority attention to complete it on a timely basis. In this chapter we discuss at greater length some of the qualities that we believe characterize "good" documentation for microsimulation models. We also consider the related issue of archiving of models, databases, and model runs. Archiving is an important part of model documentation, broadly conceived, because models and associated databases frequently change. Without good archiving, it is very difficult to undertake activities such as validation studies that examine past model performance.

[1] This section benefited greatly from a paper reviewing the documentation of the TRIM2, MATH, and HITSM models, prepared by panel member Kevin Hollenbeck (in Volume II).

Copyright National Academy of Sciences. All rights reserved.

STANDARDS FOR MODEL DOCUMENTATION

As all developers of complex computer models know, the preparation of documentation is a seemingly thankless task that can entail considerable time, expense, and drudgery. Consequently, documentation is often deferred until so late in the development process that model developers and users either forget to include important points or leave the job half-finished. Often, critical aspects of a model are known only to its developers, and if they are not available, the model may become impossible to update or use. In these cases, the investment in the model is essentially lost: the RIM and KGB models are object lessons in this regard. In other cases, the documentation may be adequate for the programmers and analysts who work with the model regularly, but inadequate for others who may use the model only occasionally or for analysts who work only with the model outputs.

There are strategies that can help facilitate good documentation, such as requiring programmers to include adequate comments in their code, building features into the models such as automatic links to a central variable directory whenever new or modified variables are specified, and taking full advantage of features of word processing systems that permit ready updating of all parts of the documentation affected by changes in particular model components.[2] However, an essential element of obtaining adequate documentation has to remain the commitment of the model sponsors to the importance of the task.

Although we run the risk of stating the obvious, a way to underscore what is involved in a commitment to good documentation and to begin to define what we mean by "good" is to list all of the functions related to a microsimulation model that are dependent on documentation:

- Use. Documentation must be adequate to permit people to set up and run a model correctly and efficiently and to detect and correct errors in a timely fashion.
- Development. Documentation must be adequate to permit people to change a model—in both minor and major ways—in a cost-effective manner that minimizes the introduction of errors into other parts of the model.
- Comprehension. Documentation must be adequate to permit people to understand the assumptions and operation of a model, supply usable specifications for model inputs, and appropriately interpret the model outputs.
- Validation. Documentation must be adequate to allow people, including those who are and are not directly associated with a model, to conduct studies of the quality of the model outputs and otherwise contribute to model improvement.

[2] TRIM2 includes a central directory system that produces much of the documentation for the individual simulation modules on an automated basis.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

- Access. Documentation must be adequate to support portability of a model to different sites and to allow would-be users to develop facility in working with a model in a reasonable amount of time.

These various functions give rise to the need for several distinctly different kinds of documentation. For example, people who set up model runs need technical information, such as the software instructions that are necessary for the model to perform in the operating environment of the particular computer being used. People who interpret model outputs need understandable descriptions of the model's design, how it works, and the assumptions underlying various model components. They also need information about the particular specifications that were used in model runs, such as the parameters of the proposed policy change(s) that were simulated. People who are not currently familiar with the model but who want to begin using it need tutorial material.

With reference to documentation of a model per se (in contrast to documentation of particular model applications), one can identify three major types: informational, that is, documentation that provides general information about the design and operation of the model; instructional, that is, documentation that provides instructions on how to operate (or interpret the operation of) the model; and reference, that is, documentation to which users can refer easily and at random to answer specific questions about any and all aspects of the model. Each of these types of documentation, in turn, can be targeted to different audiences, such as analysts with some expertise who specify model runs, programmers who implement model runs, and nontechnical users who interpret model outputs.

Although particular documents may serve a particular purpose for a specific audience, several general criteria apply to all model documentation. These criteria have to do with content and format: documentation should be accurate, clear, and complete; and documentation should be designed for ease of use, with all components formatted consistently.

With respect to the various elements that ought to make up a documentation package, the Institute of Electrical and Electronics Engineers (IEEE) (1988) has published an industry-wide standard for either instructional or reference documentation of software. According to the standard, both types of documents should include the following kinds of front matter: title page, restrictions (e.g., copying restrictions), warranties and contractual obligations, table of contents, and list of illustrations. Both instructional and reference documents should include an introduction that provides: a description of the audience, a statement of applicability (e.g., model version, required hardware, and required operating system), a statement of purpose, a document usage description, a list of related documents, conventions (such as use of symbols), and instructions for how users can report problems.

The IEEE standard for the body of the document differs for instructional

Copyright National Academy of Sciences. All rights reserved.

and reference materials. For instructional documents, the body should include: a statement of scope, descriptions of materials the user will need and the preparations the user must make to complete the tutorial, cautions and warnings to help learners avoid major problems, and the actual tutorial informing the user how to invoke functions and what results or possible errors to expect. Finally, instructional documents should include related information such as constraints or limitations.

For reference documents, the body pertains to the components of the software or model itself: each component should begin with a statement of purpose, describe needed materials and preparations, describe all the needed input data for the particular module or function, provide cautions and warnings, describe how to invoke the module or function and how to interrupt and recognize normal or abnormal terminations, and describe the expected outputs.

The IEEE standard also calls for a separate document on error messages, known problems, and error recovery, along with cross-references to these topics in other documents that make up the documentation package. The standard also suggests that appendixes be included to provide detailed descriptions of input and output data, file structures, and global processing constraints, along with sample inputs and outputs. Finally, the standard requires a bibliography, glossary, and index. We believe the IEEE standard is both applicable and appropriate for documentation of microsimulation models.

A DOCUMENTATION CASE STUDY

Using the criteria and standards outlined above, the panel undertook a review of the available documentation for three representative static microsimulation models of income support programs—TRIM2, MATH, and HITSM.[3] Our findings from this review were discouraging with regard to the current state of model documentation:

- The TRIM2 and MATH documentation serve primarily a reference function for both experienced analysts and programmers, although TRIM2 provides some useful instructional material as well. The HITSM documentation serves primarily an informational purpose for a nonexpert audience and is wholly inadequate as a guide for model users.
- The technical writing in these documents is of highly uneven quality. Formatting is inconsistent, and the text is filled with jargon and mnemonics. There are inaccuracies and typographical errors.
- Many of the documents refer to different versions or releases of model components, but the archiving system is not explained, and obsolete cross-references suggest that updates are not carefully integrated into the documents.

[3] Hollenbeck carried out this review on behalf of the panel (see Volume II, Chapter 10, which lists the specific documents that were evaluated).

Copyright National Academy of Sciences. All rights reserved.

- All of the documents lack some components of the IEEE standard. A key component missing from virtually all of them is an index. The body of the reference documentation for particular components of the models is generally complete with regard to descriptions of purpose, inputs, and outputs. The TRIM2 and MATH documents also provide satisfactory information about how to invoke each component. Weaknesses are in the lack of cautions and warnings and a general lack of discussion of error conditions.
- The copies of the documentation that were provided for review have many publication flaws, such as missing and upside-down pages, suggesting that the documentation function has low priority.
- Finally, the documentation of these models is formidable in content, format, and sheer size. The lack of user-friendly documentation has probably contributed to the lack of use of these models by a broad community.

Some of the features of the documentation for these three models, such as the lack of instructional materials and insufficient attention to such matters as consistent formatting, have to do with the way in which microsimulation models have been developed and applied.

Historically, the models have been marketed as a service rather than a product. Typically, policy analysis agencies contract for the services of experienced model analysts who, in turn, provide specifications for programmers to implement. In the process, the models are constantly being updated and changed. Until now, the effective audience for model documentation has been the purveyor of the service who, presumably, can cope with such problems as lack of cross-references or tutorial materials. The purchaser of the service—namely, the policy analysis agency—has generally been expected to have primary interest only in the outputs, not in the technical details.

This picture of the microsimulation model industry (also see Chapter 11) is somewhat exaggerated. Some agencies use models such as TRIM2 in-house, and other agencies have analysts who are very knowledgeable in the workings of the models. Moreover, the quality of the documentation for the models included in our review is much better than that for many models that agency analysts have developed for their own use (indeed, the latter often do not have any documentation). Nonetheless, we conclude that the documentation for the major microsimulation models that are used heavily for policy analysis exhibits many weaknesses. These weaknesses militate against user understanding of model outputs and the ability of people not closely involved with the models to evaluate them or to become members of the user community.[4]

[4] The panel's own review was constrained in this regard. We could not, for example, readily evaluate the accuracy of the technical descriptions of model components in any comprehensive way.

Copyright National Academy of Sciences. All rights reserved.

RECOMMENDATIONS

Clearly, we see a need for increased investment in the quality and scope of microsimulation model documentation. We believe it is imperative that policy analysis agencies set high standards for documentation and provide the resources to make it possible to achieve those standards. The IEEE standards are an obvious set to adopt. However, given the complexity of most microsimulation models, it may be advisable to investigate whether and what kinds of added standards are required to achieve the goal of high-quality documentation that meets users' needs.

We believe that policy analysis agencies will readily acknowledge the value of investing in the quality of the models' reference documentation (with regard to formatting, accuracy of cross-references, etc.) and perhaps in additional informational material about the models. These kinds of investments not only help the people who run the models (whether agency or contractor technical staff), but also help analysts who work with the outputs. However, the agencies may question the value of investing in such materials as tutorials, which would primarily help to enlarge the user community rather than give direct support to the agencies' policy analysis needs.

We argue, however (as we have in other parts of the report), that the continued ability of microsimulation models to provide high-quality, relevant, and cost-effective service to the policy analysis function depends critically on making the models more accessible to a broader community of users, including the agencies' own staffs. (We develop the latter point further in Chapter 11.) Expanded access is needed to enlarge the pool of ideas and perspectives for model development, to facilitate experimentation with alternative model structures and uses, to support model validation, and to encourage the highest level of understanding and the appropriate use of model outputs in the policy debate. Expanded and improved documentation is one of the most important means toward a broader user community.

Recommendation 10-1. **We recommend that policy analysis agencies set high standards for documentation of microsimulation models and their inputs and outputs. Agencies should investigate existing standards, such as those published by the Institute of Electrical and Electronics Engineering, for relevance to microsimulation models and determine what additional standards are needed. The kinds of documentation that agencies should require to be developed for analysts and programmers who use, or expect to use, the models include general informational materials; tutorials; and detailed reference documents for model components that describe their theoretical basis, assumptions, operation, inputs, and outputs.**

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

To this point, we have discussed primarily documentation for models and associated databases in their current form, abstracted from any particular applications. However, models and databases are frequently updated and modified, and the policy issues they are asked to address also change over time. As discussed in Chapter 9, this situation complicates the task of model validation, particularly of the type that involves comparing projections with actual outcomes. Often, such external validation studies will be carried out with a current model, simulating current law in comparison with the law in effect at some earlier period. These studies will require access to earlier versions of the database. In other cases, in which a legislative change corresponded to an actual simulation made at the time, a complete validation will require access not only to the original database, but also to the original model and specifications.

For these kinds of uses, it is essential to establish a workable system for documenting major model applications by archiving the model outputs and the versions of the models, databases, other inputs (such as control totals), documentation, and specifications that were invoked at the time. In addition, databases (such as successive March CPS files) should be regularly archived. Finally, the archiving system must facilitate ready retrieval of materials, including appropriate comparison values, when they are needed for particular validation purposes.

Recommendation 10-2. **In order to facilitate model validation, we recommend that policy analysis agencies require archiving of microsimulation model databases on a regular basis. In addition, we recommend that the agencies require full documentation and archiving of major applications of microsimulation models. The archived materials should include the model itself, the documentation of the model, the database and other inputs, the analyst's specifications, and the outputs.**

We recommend in Chapter 9 that policy analysis agencies let separate contracts for validation of microsimulation models and their outputs. The agencies may also want to consider letting separate contracts for model documentation and archiving. The work involved in preparing and updating the documentation for large, complex models, particularly to meet the needs of different audiences, can be arduous and time-consuming. Similarly, developing and maintaining an archiving system, even one limited to databases and major applications, is a substantial undertaking. The people involved in working actively with the models must necessarily give priority to the exigencies of the policy debate. They are rarely in a position to forecast when they will have slack time that can be used for background activities such as documentation, and the all-too-likely outcome is that the documentation effort will be stinted. Of course, the model developers and users must be involved in documentation and archiving—they are the ones with first-hand knowledge—but having a separate group assume the

Copyright National Academy of Sciences. All rights reserved.

major responsibility for preparing comprehensive documentation on a regular and timely basis and for managing an archiving system may be beneficial. Such an arrangement would also bring fresh perspectives to the models themselves.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

11

The Microsimulation Modeling Community

Throughout our report we single out aspects of the structure of the microsimulation modeling community that bear on the quality, relevance, and cost-effectiveness of the use of the models to generate estimates for the policy process. In this chapter, we bring together those scattered references and discuss other aspects of the community that merit attention.

By "structure of the microsimulation modeling community," we mean the relationships among all of the organizational entities that are involved in developing microsimulation models and databases, applying them for policy analysis purposes, and using the resulting estimates to inform policy making. Hence, we are concerned about the relationships among many members of the community: federal policy analysis agencies that support model development and applications, and prepare estimates for decision makers; federal statistical agencies that produce needed input data; research and modeling contractors that conduct studies and analyses for policy analysis agencies; and the staffs of executive and legislative branch decision makers who receive the outputs from the models and interpret and use—or do not use—them in making policy. We are also concerned about the use of microsimulation techniques by the community of basic and applied social science researchers that, in turn, can benefit the development of improved microsimulation models for policy analysis purposes.

Model estimates do not spring forth unbidden as wisdom from the head of Zeus. They are developed, interpreted, and applied by people in diverse settings, including government agencies, contracting firms, congressional committee staffs, university departments, and others. Organizational characteristics

Copyright National Academy of Sciences. All rights reserved.

and modes of operation of those settings affect the ways in which the people in them act and interact. In turn, these interactions influence how well microsimulation modeling supports the policy analysis function and, ultimately, the legislative process. Sometimes, these interactions have positive results on the cost-effective use of microsimulation modeling; sometimes, their effects are detrimental. We necessarily focus on the negative aspects in order to indicate opportunities for improving the structure of the community from the viewpoint of facilitating the production and use of high-quality, relevant policy estimates from microsimulation models.

RELATIONSHIPS AMONG FEDERAL AGENCIES

Policy Analysis Agencies

One of the themes throughout our report concerns the fragmented and decentralized character of data production and policy analysis in the federal government. In some ways, this decentralization is obviously appropriate and functional. For example, there is no need for the Food and Nutrition Service (FNS) to be particularly concerned about tax policy or the Office of Tax Analysis (OTA) to be concerned about the food stamp program. Each agency can focus on the data, research, and modeling that are most relevant to the policy mission of its department. Similarly, the various federal statistical agencies, such as the Bureau of Labor Statistics and the Statistics of Income Division of IRS, are well located to be responsive to the data requirements for the policy roles of their departments. But policy interests are not always easily separable. Many policies interact, and many policy analysis agencies have overlapping responsibilities and interests. One example is the number of executive and legislative agencies that are involved in tax policy analysis using microsimulation models: Office of Tax Analysis, Joint Committee on Taxation, CBO, ASPE, and others. Each of these agencies has a somewhat different perspective and takes a somewhat different approach to database construction and modeling of tax policies.

Another example of more than one agency sharing policy interests concerns income support. FNS in the Department of Agriculture is responsible for monitoring the food stamp program, and its model of the program is more fully developed than the ASPE model. The U.S. Department of Health and Human Services (HHS) is responsible for overseeing the AFDC program, and ASPE, on behalf of HHS, has supported AFDC modeling capabilities that are more fully developed than those of FNS. This division of labor makes sense from the viewpoint of each department's primary policy interest; however, given program interactions, both ASPE and FNS need good models of all the major income support programs. To date, only limited efforts have been made to explore ways for each agency to borrow strengths from the other's model.

Health care policy provides still another example of multiple agencies

Copyright National Academy of Sciences. All rights reserved.

with overlapping concerns and responsibilities. For example, the Health Care Financing Administration has responsibility for administering the Medicare and Medicaid programs. The Agency for Health Care Policy and Research has a broad responsibility for policy research in health and is specifically mandated to coordinate data and analysis programs related to research on the effectiveness of medical treatments. The National Institute on Aging has a mandate to assess health trends and policies from the perspective of the elderly population. ASPE has historically played and continues to play a role in health care policy analysis, for instance, in the recurring initiatives for some sort of national health insurance program or for mandated employer-provided health insurance.[1] The fact that these agencies are lodged in the same department, HHS, has not in our assessment made the task of developing good interagency working relationships any easier. Indeed, for health care policy, it appears to us that communication and turf problems have hampered effective coordination of modeling, database construction, and behavioral research needed to support policy analysis.

In general, the existence of multiple models and analyses in policy areas such as taxation and income support serves the positive function of providing a means of cross-checking to identify errors in computer codes and to highlight differences in key assumptions underlying model estimates produced by different agencies. Moreover, there is considerable potential for the differing perspectives of analysts in different agencies to contribute valuable insights for improvements in models and databases that could benefit a broad range of users. On the negative side, however, the involvement of multiple agencies adds costs because of duplication of effort. And differences in analytical approaches, models, and databases among the agencies make it difficult to determine the reasons for different outputs and, hence, difficult to pinpoint aspects of particular models or data sets that need improvement. Finally, barriers to interagency communication, including time pressures and the orientation of the analysts to the needs and viewpoints of their own agency and department, can all too easily nullify the potential gains from different perspectives.[2]

As we have noted repeatedly, microsimulation models and databases are usually large and complex, and thus likely to be large consumers of resources for development and use. This fact makes it particularly important for federal agencies to seek ways to improve cross-agency communication and coordination, both within and among departments and branches of government. Because there are frictional costs to any interagency collaboration, however, coordination efforts should not be allowed to become counterproductive. In particular, we are wary of joint model-building efforts that could result in a model that serves

[1] The Department of Labor also has an interest in employer-provided health insurance and has funded some modeling efforts on this topic.

[2] It appears to us that the barriers to effective communication are often particularly high between the policy analysis agencies in the executive and legislative branches.

Copyright National Academy of Sciences. All rights reserved.

none of its sponsor agencies' interests very well. Yet, there are many ways in which enhanced collaboration would be beneficial to the development of improved models and databases. For example, agencies could usefully work together to:

- specify common data elements for inclusion in surveys;
- consider mechanisms to facilitate data record linkages and access to linked files;
- develop model design standards to facilitate more ready interchange of model components, inputs, and outputs;
- sponsor evaluations of alternative model aging techniques and specifications;
- sponsor broad behavioral research;
- sponsor investigations of promising new computer technologies; and
- sponsor research on model validation methods.

***Recommendation 11-1.* We recommend that executive and legislative branch policy analysis agencies expand their communications and undertake cooperative efforts to improve the quality of microsimulation models and associated databases through such means as cosponsoring research on model validation methods and other initiatives.**

(In Chapter 8, we specifically recommend that the Department of Health and Human Services establish a high-level steering group to coordinate development of microsimulation models, data, and research needed to improve the scope and quality of policy analysis in the vital health care policy area.)

Statistical and Policy Analysis Agencies

The fragmentation of authority and the overlapping responsibilities and interests among executive and legislative branch policy analysis agencies also characterize federal statistical agencies. Again, the area of health care is a prime example: not only the National Center for Health Statistics and the Census Bureau, but also the Agency for Health Care Policy and Research, the Health Care Financing Administration, and ASPE, have played a role in sponsoring important data collections that are needed for health care policy analysis. This fragmentation has had negative consequences. For example, health-related surveys have not always included variables that are important from the perspective of modeling and policy use, nor have they always become available in a timely fashion to users in other agencies.

Another problem with the decentralized nature of federal statistical activities is that collection of administrative data is typically lodged in one set of agencies (for example, the IRS or the Social Security Administration), while other agencies, such as the Census Bureau or the Bureau of Labor Statistics,

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

have primary responsibility for data from household and other types of surveys. This situation exacerbates the problems in trying to link administrative and survey data; it is one reason that relatively few linked data files exist or are widely available, although, potentially, such files represent a cost-effective means of obtaining valuable data for statistical and research uses.

The relationships between the statistical agencies, on one hand, and the policy analysis agencies, on the other, present some troubling aspects for cost-effective production of useful, high-quality databases for microsimulation modeling and other types of policy analysis. Because statistical agencies perceive a limited responsibility to add value to the data they collect, the users in the policy analysis agencies must spend considerable resources, often for duplicative work, to make various corrections and enhancements in order to have suitable databases for modeling. Moreover, differing perspectives often impair effective communication between statistical and policy analysis agencies. As a consequence, policy analysis needs are not always well reflected in statistical agency decisions about the design and content of surveys, and, conversely, user agencies do not always properly use or interpret the available data.

In other chapters of the report we make several recommendations that bear directly or indirectly on the general relationships among statistical agencies and between those agencies and the user agencies. In summary, we recommend:

- Federal statistical agencies, generally, and the Census Bureau, particularly, should assume a more active role in adding value to databases for modeling and research purposes, through such means as adjusting responses for reporting errors.
- Federal data collection strategies should emphasize breadth of use and ability to respond to changing policy needs, including duplication of selected questions across surveys to the extent that such duplication enhances utility and facilitates evaluation of data quality.
- The federal statistical community should develop mechanisms to improve access, under appropriate circumstances, to administrative and survey microdata for statistical research and analysis purposes: in particular, the Census Bureau should conduct updated exact matches of social security earnings records with household survey data and develop ways to make them available for research use.
- The Census Bureau and the policy analysis agencies should work together to evaluate alternatives for short-term improvements to the data used for microsimulation modeling of income support and related social welfare programs.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

POLICY ANALYSIS AGENCIES AND THEIR SUPPLIERS

The extent to which policy analysis agencies are directly involved with the development and use of microsimulation models varies substantially. Some policy analysis agencies, such as the Office of Tax Analysis, develop, maintain, and operate their own models in-house. Other agencies, such as the Congressional Budget Office, do some modeling in-house but also rely on outside contractors. Still other agencies rely almost exclusively on contractors. Historically, ASPE has developed and used models in-house; currently, ASPE staff use the AFDC benefit-calculator model in-house but rely exclusively on contractors for model development and maintenance and for application of complex models, such as TRIM2. FNS has historically contracted for all microsimulation model services (although FNS staff now use a food stamp benefit-calculator model in-house).

The past decade has seen increasing reliance on contractors. There are many reasons for that reliance. Cutbacks in agency staff and personnel ceilings have made it difficult for many agencies to recruit in-house modeling staff, particularly given competing needs for other kinds of staff capabilities. Furthermore, some agencies have preferred to develop long-term relationships with contractors whose staffs have developed in-depth expertise in the "care and feeding" of the models and databases and in working closely with the agency analysts.

Microsimulation modeling contractors typically sell a package of services, not an off-the-shelf product. These services include routine maintenance, such as updating program parameters in the models and creating new baseline files as new survey data are released; new model development; and rapid response to agency requests for model runs. Underlying the attractiveness of the service package offered by contractors is the complexity of the microsimulation models, which places a premium on the knowledge and experience of staff. That same complexity limits the number of suppliers by putting high barriers in the way of new firms entering the field or competing successfully with established contractors. In fact, looking at microsimulation suppliers, a small handful of contracting firms are involved in providing the majority of modeling services to the policy analysis agencies.

We have every reason to believe that microsimulation modeling contractors have ably fulfilled the terms of their contracts and, indeed, have often provided outstanding service to their agency clients. We also recognize that there are real costs to the agencies to try to increase the contractor pool, given that heightened competition may disrupt agency-contractor relationships that are beneficial for obtaining the most experienced and knowledgeable assistance in carrying out the policy analysis function. Yet we are disturbed that the small number of suppliers may well result in missed opportunities for new ideas and perspectives that could lead to improvements in models and model estimates.

Copyright National Academy of Sciences. All rights reserved.

Whether an agency runs microsimulation models in-house or uses contractors, we are disturbed about another aspect of the industry as currently structured, namely, that analysts almost never have hands-on experience with the models. Given the complexity of most current models and their reliance on batch-oriented mainframe technology, even straightforward model applications require intervention by skilled programming staff.[3] This situation prevails not only at the agencies themselves but also at the contractors. Again, in our view it is likely that there are missed opportunities for new ideas and experimentation with alternative policy proposals and alternative model structures. Moreover, the distance of analysts from a model, which is particularly great for the staff of an agency that contracts for modeling services, is not conducive to the analysts' fully understanding the model or the resulting output. Such distance increases the opportunities for miscommunication, erroneous specifications, and erroneous interpretations of model results.[4]

We believe that it is critically important for agencies to expand the community of knowledgeable microsimulation model developers and users and, in particular, to bring the models closer to the analysts who have the ultimate responsibility for preparing estimates for the policy debate. There will undoubtedly be short-run costs for these efforts, but the long-run payoffs are vital for the continued ability of microsimulation models to support the policy analysis function in a cost-effective manner.

As concrete means to these ends, following the design principles that we present in Chapter 6 should produce models that are more accessible to a broader range of users, not only for applications, but also for experimentation and validation efforts that, in turn, may well lead to improved models and model estimates. Similarly, implementing models with new computer hardware and software technologies, as discussed in Chapter 7, should be very effective in this regard. Providing more extensive model documentation, which includes both reference and instructional and informational materials (see Chapter 10), is another way of encouraging more extensive use and better understanding of the models by a broader community of users.

Our recommendation that agencies let separate contracts for model validation and our suggestion that this approach be considered for documenting and archiving models (see Chapters 9 and 10) also offer important benefits from the perspective of model development and use. Such contracts can be expected to provide direct incentives for other firms and academic researchers to become involved with microsimulation modeling and, at least, to broaden the range of perspectives that address issues of model quality and utility. In this regard, it

[3] The exceptions are the personal computer-based benefit-calculator models, which agency analysts often run themselves.

[4] In Chapter 9, we discuss an example from the Family Support Act-for the unemployed-parent component of AFDC—that illustrates the problems for analysts' understanding of model outputs because of the models' complex nature.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

is critically important that agencies lend their support only to those models that can readily be made available to all interested users.

Recommendation 11-2. **We recommend that policy analysis agencies have a strict policy that only public-use, nonproprietary microsimulation models—for which documentation, inputs, outputs, and programming code can be freely exchanged—will be considered for agency use.**

Finally, a goal that we encourage policy analysis agencies to adopt is to increase the in-house use of microsimulation models by their own analysts. Such a goal will confirm the agencies' commitment to improving models and databases through expanding the pool of knowledgeable users. Such a goal will also help motivate the agencies to take other needed actions. Specifically, effective use of models by agency analysts will necessarily require better design, computer implementation, and documentation of models, all of which are important for improving the flexibility of modeling tools and the quality and understandability of model estimates.

Recommendation 11-3. **We recommend that the policy analysis agencies set a goal of increasing the in-house use of microsimulation models by agency analysts, who have the ultimate responsibility of interpreting model results for policy makers.**

POLICY ANALYSIS AGENCIES AND DECISION MAKERS' STAFFS

To this point we have considered relationships among organizational entities that are involved in the production of policy estimates. We now very briefly consider the relationships among organizational entities that are involved in the use of those estimates. Specifically, we are concerned with the interactions of the policy analysis agencies, which are responsible for supplying estimates to the policy debate, and the staffs of decision makers, who play a key role in interpreting the estimates and mediating their influence on policy choices.

The single point that we want to stress is the need for policy analysis agencies to work with decision makers' staffs to improve understanding of the quality of the estimates and the uncertainty that inevitably surrounds them. We realize that, ultimately, decision makers will use point estimates and not the confidence intervals or other measures of uncertainty that accompany the estimates. Indeed, this is appropriate for such purposes as adding up cost estimates for various components of a legislative proposal and making the best assessment of the distributional effects of a proposed policy change.

However, it is critical that decision makers' staffs receive, understand, and take cognizance of measures of uncertainty and other quality assessments

Copyright National Academy of Sciences. All rights reserved.

for the estimates produced by policy analysis agencies. The information that a particular cost estimate has an uncertainty range of +$50 billion to -$50 billion, for example, and another estimate has an uncertainty range of +$25 to +$35 billion, is vital input in assessing the weight that decision makers should accord to each of the two estimates in their deliberations. Furthermore, such information will help decision makers and their staffs make judgments about needed improvements in databases and analytical tools that promise to improve the quality of the estimates for future deliberations. We urge both policy analysis agencies and decision makers' staffs to give serious heed to our recommendation that quality measures be developed for policy estimates and used in the policy debate.

THE ROLE OF RESEARCH

Contributions of Research to Policy Analysis

Another factor in the limited extent of the community involved in the use of microsimulation modeling for policy purposes has been the disassociation of much of academia from such analysis. Although the concepts of microsimulation originated in academia, relatively few academic researchers have been actively involved in the development and use of microsimulation models for policy applications over the past decade. In some areas of basic and applied social science research, notably the study of human populations by demographers and others (see further discussion below), academic researchers have a substantial tradition of development and use of microsimulation models. In other areas, however, such as economics, microsimulation techniques have been relatively little used. In our view, more direct involvement of academic researchers with microsimulation techniques would benefit the development of improved models for policy analysis as well as research purposes.

Academic contributions to microsimulation modeling as a tool for policy analysis come not only from researchers' direct involvement with models, but also from their development of knowledge that provides needed input to model functions. Policy models can only be as good as the underlying base of knowledge about human behavior. In microsimulation, even those models that emphasize the accounting aspects of government programs must take into account decisions on the part of families and individuals to accept benefits, itemize deductions, and so on. Should the models attempt to simulate other kinds of behavioral responses, such as a work response to welfare reform, they need more information about behavioral relationships. Models that look at changing characteristics over time, such as the dynamic microsimulation models used to simulate retirement income programs, depend very heavily on basic knowledge about the factors that influence marriage, divorce, educational attainment, labor force participation, and other modeled characteristics.

Copyright National Academy of Sciences. All rights reserved.

Model developers rely on social science researchers for their knowledge base. Yet the interests and incentives of researchers, even those in applied research, are not necessarily directed to the topics of most concern for policy modeling. As we noted in Chapter 6, even ostensibly relevant research may not be carried out in a form that is readily usable for models. And, of course, all too often research results vary widely and hence leave model developers in a quandary about which estimate to use.

Generally, social science researchers are heavily influenced by the requirements of academic publishing, which emphasize innovative work in highly specialized areas. In contrast, microsimulation modeling and other forms of policy analysis require well-established results from cross-disciplinary studies that incorporate replication or reanalysis of previous work. The results must be in a form that the models can readily implement and must, of course, relate to topics of policy concern. The issue for policy analysis agencies is how to obtain a higher payoff from social science research. An obvious answer is to increase contract research and grant funds for needed studies. Although additional funding may well be part of the answer, agency budgets may not be able to provide it. The funds that are available need to be carefully targeted to focus researchers' attention on the priority needs of the agencies. In other chapters of the report, we make several recommendations for priority research areas for agencies to support. In summary, we recommend that policy analysis agencies sponsor

- studies of the relationship between behavioral research and microsimulation modeling, including studies of ways in which research and modeling can complement one another, as well as ways in which the two are alternative modes of deriving answers to policy questions;
- studies to determine when behavioral response effects are most likely to be important in different policy simulations and to attempt to narrow the range of statistical estimates of those behavioral parameters that may be of major importance to critical policy changes; also, studies on second-round effects of policy changes that may be important to understand;
- studies to develop methods for systematically assessing the impact on microsimulation model estimates of the degree of uncertainty in the behavioral parameters that are used—both the uncertainty arising from the variance of specific parameters and that arising from the range of estimates from different behavioral studies; and
- studies to develop improved methods for validating microsimulation model output.

We also believe that agencies may benefit from relatively low-cost activities designed to encourage and reward researchers' interest generally in policy analysis concerns that could benefit from a microanalytic perspective. Specifically, we suggest that the agencies inaugurate regular programs of conferences

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

that feature research results and methods that are pertinent to microsimulation modeling and other forms of microlevel policy analysis and highlight emerging policy issues on which research is needed. The resulting conference proceedings should be published and disseminated to a broad audience. We note that the Food and Nutrition Service has sponsored two such conferences in recent years (Food and Nutrition Service, 1986, 1990). The Statistics of Income Division of IRS has also sponsored conferences about challenges confronting microsimulation models related to tax policy, of both the household and the corporate sectors, that attracted academic interest (Internal Revenue Service, 1989; U.S. Department of the Treasury, 1988a, 1988b). We encourage other agencies to follow their lead.

Recommendation 11-4. **We recommend that policy analysis agencies encourage and support the involvement of social science researchers in work that is relevant to microsimulation modeling, and other microlevel policy analysis, through sponsoring regular research conferences. The conferences should highlight pertinent research results that can be used for models, with an emphasis on the synthesis of research findings and the reconciliation of conflicting results. These conferences should also work to develop research agendas to address emerging policy needs. The agencies should prepare and disseminate proceedings from all such conferences.**

The Promise of Microsimulation for Social Science Research

Our report focuses on the use of microsimulation models for policy analysis purposes.[5] Indeed, a primary motivation of Guy Orcutt, the economist who pioneered the development of microsimulation modeling more than three decades ago, was to find a technique that would bring social science research knowledge to bear on policy questions. Looking back 45 years later, Orcutt (1986:26) said, "[I] shifted out of the study of physics into economics with the hope that my research might help to avoid a repeat of the international catastroph[y] referred to as the 'great depression'." He initially believed that macroeconomic time-series analysis held the key to understanding economic behavior and developing public policies to combat economic and social problems. However (p. 26), "By the middle of the 1950's it seemed evident that macroeconometrics, despite its many attractions, was a house built upon sand.… I then came to think of microanalytic simulation modeling as a potentially promising alternative.…I am still convinced of the correctness of this view." Orcutt believed that microsimulation techniques were necessary to trace through the potential

[5] The discussion in this section benefited greatly from a set of notes prepared by Michael Wolfson, a member of the panel.

Copyright National Academy of Sciences. All rights reserved.

effects of alternative public policies on individual decision units and generate individual behavioral responses that, in turn, would have feedback effects on other sectors of the economy.

Orcutt also had a vision that microsimulation models would make important contributions to social science research knowledge. He argued that the complexity of human behavior makes it very difficult for researchers to determine how interactions among social processes will play out over time. Dynamic microsimulation models, which can include multiple processes—such as the decisions to enter the labor force, marry, and have a child—and model them with a detailed population sample, provide the computational means for achieving this goal. Running a model may produce unexpected results that, in turn, lead a researcher to look for ways in which the underlying behavioral equations need to be improved. Stated another way, microsimulation models are vehicles for integrating the results of different strands of socioeconomic research within a framework that forces consistency and alerts a researcher to important gaps in data and knowledge about human behavior. Hence, Orcutt saw microsimulation modeling playing a prominent role in the development of knowledge about socioeconomic behavior.

Orcutt's dream has never been fully realized. For many subjects of inquiry, microsimulation models have remained largely a tool of policy analysts and not of social science research. Moreover, the most heavily used policy models have tended to be static models that contain complex modules to determine program eligibility and benefits but incorporate very few of the dynamic behavioral processes that Orcutt saw as essential to the technique. Dynamic models such as DYNASIM2 and PRISM have had a much more limited analytic scope than Orcutt envisioned.

There are many reasons for the failure of Orcutt's dream to excite the attention of academics in disciplines such as economics or to stimulate serious involvement on their part over the past two decades: most important, in our view, are the complexities and limited access to many current models. There is a high entry cost for academic researchers (as well as other would-be users) to learn enough about large-scale microsimulation models such as DYNASIM2 or TRIM2 to be able to use them effectively within the confines of a limited-duration research grant. Inadequate documentation and instructional tools compound this problem. The highly complex nature of microsimulation models has led some in the academic community to express outright distrust of what appear to them to be impenetrable "black boxes" (see, e.g., Taussig, 1980). Similarly, academic researchers have been disturbed that such models appear to be overly sensitive to the choices of their developers and generally to lack ready mechanisms with which to judge the reliability and generalizability of their results. In addition, the relatively high cost of using mainframe-oriented models strains research budgets, and the necessity of programmer intervention

Copyright National Academy of Sciences. All rights reserved.

runs counter to prevailing academic work habits and limits the experimentation that a researcher can attempt.

The interests of and incentive structures for academic social scientists have also operated to limit their involvement with microsimulation modeling. Academic researchers generally have been uninterested in policy applications of microsimulation models, such as estimating the costs and distributional effects of proposed program changes. Policy analysis results are likely to be of little or no theoretical relevance and unlikely to be accepted for publication in academic journals that tend to emphasize new findings and theoretical advances. For the kinds of basic socioeconomic or policy research studies that more naturally fit their agendas, academic researchers in many social science disciplines have seen microsimulation models as costly and of little relevance. Thus, it is not surprising that many social scientists have remained aloof from microsimulation techniques. However, we believe that this situation will change and, indeed, is already changing, and that microsimulation modeling will prove increasingly useful and cost-effective for a growing number of analytical situations.

The cost structure of microsimulation modeling is poised for a fundamental change in the very near future. Improvements in both hardware and software technologies should make it possible to operate very large, complex models in a manner that facilitates direct user interaction and lowers costs of access and use. Improved computing environments should facilitate the ability of users to run alternative models, modify important model components, and validate model outputs—all activities that are critical to researchers' acceptance of microsimulation modeling as an important social science analytical tool.

Other factors should also contribute to an improved climate for social science research applications of microsimulation models. Rich longitudinal microdata sets that provide the wealth of variables and repeated observations of individuals that are needed to estimate dynamic behavioral functions are available. These data sets include the Panel Study of Income Dynamics, the National Longitudinal Surveys of Labor Market Experience, SIPP, and others. Some of these surveys provide information for 20 years or more, have been extensively analyzed by many researchers, and are maintained by central facilities. These centers, in turn, provide a broad range of services to the users, such as documentation, training, and bibliographic reference services. Not only do these data sets furnish important inputs for modeling, but the central facilities that support them may serve as prototypes for the support of microsimulation models for academic research purposes.

As noted above, some of our recommendations to policy analysis agencies are directed specifically toward facilitating academic involvement with microsimulation models that are used for policy analysis. We suggest ways to attract statisticians and other academics to the issues involved in validating microsimulation model results. For example, data sets, such as the results generated from our TRIM2 validation experiment, can readily be made available to

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

researchers for more extensive analysis. Policy analysis agencies can also actively encourage research on improved methods for validating microsimulation model output through such mechanisms as fellowships that allow researchers to work on-site with agencies or their contractors. We also suggest ways for the agencies to obtain more focused research from academic social scientists that would make possible improved modeling capabilities in such areas as incorporating behavioral response. Such involvement on the part of academic researchers is important to maintain and replenish the intellectual capital that underlies the use of microsimulation models for policy analysis purposes. In turn, such involvement may help to stimulate academic interest in microsimulation techniques for more basic social science research.

We note that demographers, particularly those working in the sub discipline of family demography, have a long and rich tradition of the development and use of microsimulation models for research purposes. (Anthropologists and geneticists have also been involved in microsimulation-based research studies related to human populations.) The earliest studies in this tradition were initiated in the mid-1960s. By 1972, at a conference sponsored by the Social Science Research Council (Dyke and MacCluer, 1973), researchers reported on the use of dynamic microsimulation techniques to carry out such diverse studies as the survival probabilities of small, closed populations over long time periods under different rules regarding marriage among relatives and clans and the probabilities of pregnancy outcomes (e.g., still versus live births) as a function of consanguinity and genetic inheritance. These models, like cell-based population projection models, typically modeled marriage, fertility, and mortality; however, they applied transition probabilities to individual persons rather than to age-sex groups and hence could accommodate a much greater level of detail: for example, simulating complicated societal marriage rules and other aspects of "marriage markets," such as age preferences for spouses, and often including added functions such as divorce, remarriage, and birth outcomes. One early demographic microsimulation model, POPSIM (Population Simulation Model), was used for policy analysis of alternative family planning methods (Rao et al., 1973; see also Horvitz et al., 1971).

In the early 1970s, the National Science Foundation supported the development of the SOCSIM (Social Simulation) model, which simulated demographic processes with the added dimension of household structure (Hammel et al., 1976). One innovative application of SOCSIM, in the field of historical demography, was to model the household composition of preindustrial English villages over a 150-year period under different assumptions about demographic rates and societal preferences for living arrangements (Wachter, with Hammel and Laslett, 1978). The model was invoked to shed light on a troubling discrepancy between theory and evidence, namely, the much higher-than-expected percentages of nuclear families in listings of household membership in preindustrial English settlements. The simulation, which included 15 scenarios, each replicated 64

Copyright National Academy of Sciences. All rights reserved.

times with different random numbers, demonstrated that the empirical data were unlikely to result from demographic constraints.

SOCSIM has continued to be used and expanded. One policy-relevant set of simulations looked at the likely kin structures of the U.S. population in the year 2000 under alternative sets of assumptions about fertility rates (Hammel, Wachter, and McDaniel, 1981; see also Reeves, 1987). The study addressed such questions as the proportion of middle-aged people who could be expected to have both older and younger dependents and the proportion of the elderly who could be expected to have kin on whom to call for support, including siblings, children, and grandchildren. (Smith, 1987, reports on the use of a similar model, CANSIM, to simulate kin sets and obtain counts of various types of kin at various stages of the life cycle.)

In a paper for a volume on methods of family demography, Wachter (1987:217) comments that "simulation studies function in some respects as the social scientist's substitute for the natural scientist's controlled experiments under laboratory conditions." He summarizes the contributions of microsimulation methods for studies of complex family and household structures (also pointing to their use for simulating fertility processes); notes their advantages and disadvantages vis-a-vis analytic solutions and macrosimulations (what we term cell-based models); and indicates areas for further work. He comments that analytic, macro, and micro approaches are often all pertinent to a research problem in household structure; however, microsimulation has particular advantages when (p. 219) "the theory to be simulated specifies choices for individuals depending on their detailed circumstances…[when] the outputs desired include measurements of random dispersion as well as central tendency…[and when] complex configurations of kin…are of interest."

We also note that, in basic research fields outside social science, microsimulation modeling has become quite commonplace. For example, in the 1960s, it was an ordinary part of "atom smashing" research with cyclotrons to use Monte Carlo microsimulation of subatomic particles to assess the impact of phenomena that were not directly observable (for example, pi to mu meson decay between the time the particle left the cyclotron and smashed into the target in scattering experiments). Modern cosmology is a very active area of academic research that, for many types of studies, such as the dynamics of star clusters, relies fundamentally on forms of microsimulation. It is widely accepted that contemporary research in fields as diverse as the archaeology of settlement patterns, the management of inventories for large manufacturing plants, transportion and other queuing problems, and simulations in neural biology of neural networks, to name some examples, may require the development of large-scale microdata-based simulation models.

For social science research, we accept the premise of Orcutt's dream, namely, that microsimulation techniques afford the opportunity to address a number of complex analytical problems and thereby advance the state of

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

knowledge. As just described, the field of demography provides examples of the utility of microsimulation to draw out the implications of complicated interactions that conventional methods of demographic analysis, such as life tables and their extensions, are unable to handle. Another application of microsimulation modeling that has been closely linked to academic research—namely, the simulation of systems of behavior in urban areas—also illustrates the utility of the technique. Because of complex locational factors interacting with individual behavioral decisions, researchers have found microsimulation modeling to be useful to integrate research about locational choices and consumer demand with housing and transportation policies. See, for example, de Leeuw and Struyk (1975); Ingram, Kain, and Ginn (1972).

We believe that there are opportunities to expand the role of microsimulation for studies that combine demographic, sociological, and economic concerns and that may have policy relevance as well. For example, it has been shown that labor force experience on the part of young women increases their likelihood of getting divorced (Rowe, 1990). But researchers are generally interested in learning more than this: What effect does this factor have on the amount of time children can expect to spend growing up in a single-parent household? As another example, if similar longitudinal microdata analyses have examined the determinants of labor force entry and exit, it would be interesting to know how these two processes interact. Microsimulation models have the potential to address these and other research questions involving complex interactions.

Bergmann (1990), in describing a highly stylized microsimulation model of job search and hiring behavior on the part of workers and firms, argues strongly for the research utility of microsimulation techniques. (Her model finds, contrary to conventional analysis, that the availability of unemployment insurance does not always raise the unemployment rate.) Bergmann (1990:100-101) comments:

> Microsimulation models have a number of advantages over conventional methods used by economists. One advantage is the explicitness with which the interactions on the micro level can be depicted. Conventional methods of economic analysis often provide cogent derivations of individuals' contingency plans, but tend to be weak in describing the process that allows all the contingency plans to mesh together…Simulation can be especially helpful in dealing with problems of information and expectations…Simulation models… can be set up to be completely recursive, as real life is…Nonlinearities, switches of regime, and other complexities are easily accommodated.

In a similar vein, microsimulation has the potential to address a number of problems that confront the conventional economic theory and analysis of the firm. Anderson's (1989) analysis suggests that individual firms' investment behavior responds nonlinearly and, indeed, discontinuously to tax status. Arthur (1990) argues cogently that increasing returns to scale and chance events are

Copyright National Academy of Sciences. All rights reserved.

important factors in the dynamics of firms, and that these factors have a dominant impact on the kinds of market structures that emerge.

Microsimulation models can be developed to take account of all these factors. There is no need to impose simplified patterns of behavior on individual decision units in order to satisfy aggregation conditions. It is natural in a microsimulation framework to represent all the diversity and heterogeneity of the individuals, families, or firms that available microdata can support. There is no need to assume away major aspects of the real world in the theory of the firm, such as increasing returns to scale and "satisfying" behavior (see, e.g., Nelson and Winter, 1982, or Eliasson, 1985, for a Swedish example). Nor is there a need to ignore important determinants of key demographic processes, such as divorce (see Wolfson, 1989a), or to confine analysis to steady-state population structures, as illustrated by the multigeneration population dynamics models developed by Wachter and others (described above). Population heterogeneity, complex nonmaximizing patterns of behavior, increasing returns to scale, and disequilibrium dynamics can all be encompassed within microsimulation modeling.

The time is ripe for a new look at the utility of microsimulation methods to advance the state of social science knowledge. This objective is important in and of itself, and it can lead to advances in the state of the art of public policy research and analysis. We are hopeful that the confluence of factors identified above—including reduced costs of using microsimulation models, the availability of richer data sets, and the existence of a large and growing community of researchers who have hands-on experience in working with complex microdata—will lead not only to better policy making but also to the rebirth of a role for microsimulation modeling in social science research and, ultimately, to the full realization of Orcutt's dream.

Copyright National Academy of Sciences. All rights reserved.

Appendix

Microsimulation Models, Databases, and Modeling Terms

This appendix details the characteristics of selected social welfare policy microsimulation models, including DYNASIM2, HITSM, MATH, MRPIS, PRISM, the long-term health care PRISM submodel, the model of the Office of Tax Analysis, and TRIM2; the major databases that the models use; and the technical terms used in microsimulation modeling. Entries include references for additional information. Terms are defined with reference to their application in social welfare policy modeling.

MODELS

DYNASIM2 (Dynamic Simulation of Income Model 2)

Type Public-use dynamic model for retirement income and other social processes and policies.

Supplier The Urban Institute; original DYNASIM developed in early 1970s, and DYNASIM2 developed in early 1980s.

Major Users Congressional Budget Office; U.S. Department of Labor; in the U.S. Department of Health and Human Services, ASPE, SSA, Administration on Aging, National Institute of Child Health and Human Development.

Programs Simulated SSI, social security, employer pensions, individual retirement accounts, social security payroll tax, federal income tax.

Copyright National Academy of Sciences. All rights reserved.

Main Database Exact match of March 1973 CPS with social security earnings records for 1937-1972 (quarters of coverage) and 1951-1972 (taxable earnings).

Major Database Enhancements Cross-sectional imputation of home ownership and income from financial assets for older families, based on 1984 Survey of Consumer Finance, and of health status and institutionalization for older people, based on 1984 Supplement on Aging and 1984 Long-Term Care Survey.

Projection Strategy Uses dynamic aging to develop longitudinal histories for all sample individuals to any future year. Events simulated include birth, death, first marriage, divorce, remarriage, work disability, education, migration (by region and size of place), wage rate, labor force participation, hours of work, unemployment, job change, industry movement, social security coverage, pension coverage. Uses many data sources to estimate transition probabilities, including vital statistics, decennial census, BLS Surveys of Defined Benefit Plans, May CPS (selected years), Panel Study of Income Dynamics.

Behavioral Responses Simulated Basic participation decision (for SSI) and decision to retire and accept public or private retirement benefit, but no feedback effects of simulated future program changes (e.g., on hours of work or savings behavior).

Calibration of Yearly Histories Generally uses most recent alternative II-B assumptions from the OASDI trustees report to calibrate demographic and economic aggregates and BLS projections to calibrate employment aggregates.

Computer Implementation Hardware is IBM mainframe, VAX minicomputer, software is FORTRAN.

References Johnson, Wertheimer, and Zedlewski (1983); Johnson and Zedlewski (1982); Orcutt et al. (1980); Ross (in Volume II); Zedlewski (1990).

HITSM (Household Income and Tax Simulation Model)

Type Proprietary static income-support and tax program model.

Supplier Lewin/IlCF, Inc. (formerly ICF, Inc.); developed in mid-1980s.

Major Users U.S. Department of Health and Human Services.

Programs Simulated AFDC, SSI, food stamps, energy assistance, unemployment insurance, in-kind benefits from Medicare, Medicaid, housing assistance, and school lunch programs, federal income tax, social security payroll tax, state income tax, sales tax.

Main Database CPS March income supplement.

Major Database Enhancements Statistical match of the 1987 March CPS,

Copyright National Academy of Sciences. All rights reserved.

the 1984 SOI Individual Income Tax Return Sample, and the 1984 Consumer Expenditure Survey (CES); allocation of yearly income and employment to weeks and months (see Citro and Ross, in Volume II); adjustment for hours worked and earnings to match BLS and Census Bureau control totals; adjustment for underreporting of unearned income (except for AFDC and SSI, which are simulated by the model) by a hot deck to select additional recipients followed by scaling of income amounts to match control totals; imputation of value of employer-provided health insurance.

Projection Strategy Static aging of demographic characteristics, employment, and income amounts is integrated into the model and always carried out prior to program simulations.

Behavioral Responses Simulated Basic program participation decision (see Citro and Ross, in Volume II).

Calibration of Baseline Simulations AFDC, SSI, and food stamp participants are controlled to administrative data on national characteristics and to state caseloads for SSI and, optionally, for AFDC.

Computer Implementation Hardware is mainframe IBM; software is FORTRAN.

References ICF, Inc. (1987); Lewin/ICF, Inc. (1988).

MATH (Micro Analysis of Transfers to Households)

Type Public-use static income-support program model.

Supplier Mathematica Policy Research, Inc.; developed in mid-1970s, based on an early version of TRIM, and redesigned for processing efficiency in late 1970s and again in mid-1980s.

Major Users Food and Nutrition Service, U.S. Department of Agriculture; applications also performed for Congressional Research Service, U.S. Department of Labor, U.S. Department of Health and Human Services.

Programs Simulated AFDC, SSI, general assistance, food stamps, federal income tax, social security payroll tax; the food stamp module is the most well-developed (Medicaid and state tax modules were developed in the 1970s but are no longer used).

Main Database CPS March income supplement.

Major Database Enhancements Allocation of yearly income and employment to months based on the 1979 ISDP (see Citro and Ross, in Volume II); imputation of child care expenses based on the 1985 SIPP panel, shelter expenses based on the 1983 AHS, out-of-pocket medical expenses based on the 1980-1981 CES,

Copyright National Academy of Sciences. All rights reserved.

asset balances based on the 1985 SIPP panel, federal income tax deductions. There is an optional procedure to adjust for income underreporting.

Projection Strategy Usual practice is to age a recent March CPS file forward about 4 years, using static procedures to adjust demographic, economic, and employment characteristics; process is repeated every 2 or 3 years.

Behavioral Responses Simulated Basic participation decision in AFDC, SSI, general assistance, and food stamp program and participation response to change in food stamp benefits (see Citro and Ross, in Volume II); modules to simulate labor supply response to income benefits and take-up rates for public service jobs programs were developed in the late 1970s but are no longer used.

Calibration of Baseline Simulations Food stamp participants are controlled to administrative data on national characteristics. AFDC and SSI participants are controlled on the basis of regional characteristics, with optional state controls.

Computer Implementation Hardware is mainframe IBM; software is FORTRAN, Assembly (can output an analysis file to a personal computer and create tables using any statistical package).

References Beebout (1980); Doyle and Trippe (1989); Doyle et al. (1990).

MRPIS (Multi-Regional Policy Impact Simulation)

Type Public-use hybrid second-round effects model for income support and tax programs and other public- and private-sector economic changes.

Supplier Social Welfare Research Institute, Boston College; developed in early and mid-1980s.

Major Users U.S. Department of Health and Human Services; state of Massachusetts; private firms and nonprofit associations.

Programs Simulated AFDC, unemployment insurance, federal income tax, social security payroll tax, state income tax.

Main Databases For the household sector, March CPS; for the product market sector, CES data calibrated with NIPA consumption data; for the industrial sector, the 1977 multiregional input-output accounts; for the labor market sector, matched CPS files.

Structure The model consists of four interrelated sectors, household, product market, industrial, and labor market; it does not include a capital market, investment behavior, or a detailed financial sector. It is a short- or intermediate-term partial equilibrium model that simulates alternative economic states, but does not simulate the time path from the baseline to any alternative. It assumes

Copyright National Academy of Sciences. All rights reserved.

elastic supplies of commodities, labor, and capital, but fixed prices and wage rates.

The household sector is microsimulation based. It simulates tax and transfer programs and then calculates the changes for gross and disposable family income. The change values are aggregated into 20 gross family income categories by region—50 states and the District of Columbia.

The product market sector is cell based. It uses marginal propensities to consume to apportion income changes by income class and region into current savings and consumption and then uses marginal budget shares to allocate the changes in consumption to 73 categories.

The industrial sector is input-output based. It translates changes in demand from the household sector and from a simulation of government demand into total change in output—direct and indirect—needed to satisfy that demand for 124 industrial categories and 51 regions.

The labor market sector is cell based. It translates changes in industry output into labor demand and then allocates hours of labor demand by industry and occupation—11 and 8 categories, respectively—and region to the individuals in the household sector, using transition probabilities for the propensity of members of various demographic groups to change hours worked or industry or occupation. Changes in hours worked are then translated into changes in wage income that produce the basic input for subsequent (multiplier) iterations of the model.

Computer Implementation Hardware is mainframe IBM; software is FORTRAN.

References Havens et al. (1985); Havens and Clayton-Matthews (1989); Social Welfare Research Institute (no date).

PRISM (Pension and Retirement Income Simulation Model)

Type Public-use dynamic retirement income program model.

Supplier Lewin/ICF, Inc. (formerly ICF, Inc.); developed as proprietary model in 1980 for the President's Commission on Pension Reform.

Major Users U.S. Department of Labor; U.S. Department of Health and Human Services; Social Security Advisory Council; Congressional Budget Office.

Programs Simulated SSI, social security, employer pensions, individual retirement accounts, federal income tax, social security payroll tax, state income tax.

Main Database Exact match of March 1978 CPS with social security earnings records for 1937-1977 (quarters of coverage) and 1951-1977 (taxable earnings),

Copyright National Academy of Sciences. All rights reserved.

augmented with pension and other data from exact matches to the March and May 1979 CPS; second major database contains detailed information on a sample of public and private retirement plan sponsors.

Major Database Enhancements Imputation, based on regression equations and statistical matching, for missing pension, labor force, or earnings variables on 8,000 of 28,000 adult records in the CPS-SSA database.

Projection Strategy Uses dynamic aging to develop longitudinal histories to the year 2025, for adults. Events simulated include death, birth of child, marriage, divorce, disability, nursing home use, annual hours of work, hourly wage, job change, industry, pension coverage, pension plan assignment, pension acceptance, social security acceptance, IRA adoption and contributions. Uses many data sources to estimate transition probabilities, including vital statistics, matched March CPS files, and January, March, and May CPS (selected years).

Behavioral Responses Simulated Basic participation decision for SSI and decision to retire—partial as well as full retirement—and accept public or private retirement benefit, but no feedback effects of simulated future program changes (e.g., on hours of work or savings behavior).

Calibration of Yearly Histories Uses alternative II-B assumptions from the OASDI trustees report to calibrate demographic aggregates, average wage, unemployment, interest rates, inflation rates, and wage growth; uses BLS projections to control employment levels by industry and labor force participation; also uses output from the ICF Macroeconomic-Demographic Model (or other macroeconomic models) to control labor force participation and earnings.

Computer Implementation Hardware is mainframe IBM; software is FORTRAN.

References Kennell and Sheils (1986, 1990); Ross (in Volume II).

Long-Term Care Financing Model (Submodel of PRISM)

Type Public-use dynamic model of long-term care utilization and financing for the elderly.

Supplier Lewin/ICF, Inc. (formerly ICF, Inc.), and the Brookings Institution; submodel was added to PRISM in 1986; originally model was proprietary.

Major Users U.S. Department of Health and Human Services; state of Hawaii; foundations.

Programs Simulated Alternative public and private insurance and other programs for financing long-term care for the elderly.

Main Database PRISM database (see description of PRISM).

Copyright National Academy of Sciences. All rights reserved.

Major Database Enhancements Statistical match with the 1984 SIPP panel, adjusted for inflation, to assign asset records to all elderly people.

Projection Strategy Uses dynamic aging to develop longitudinal histories to the year 2025 for adults age 65 and over. The PRISM component handles aging of demographic, employment, income, and assets variables; the submodel adds aging of disability status, use of nursing home and home health care services, and public and private sources of financing long-term care. Data sources for transition probabilities for the submodel include the 1984 SIPP panel, the 1982-1984 National Long-Term Care Survey, the 1985 National Nursing Home Survey, and Medicare and Medicaid administrative data from HCFA.

Behavioral Responses Simulated See description of PRISM; the submodel simulates behavioral responses to changes in health care financing programs (such as increased utilization in response to more generous benefits) through user-specified parameters.

Calibration of Yearly Histories See description of PRISM.

Computer Implementation Hardware is mainframe IBM or VAX minicomputer; software is FORTRAN.

References Kennell and Sheils (1986); Kennell et al. (1988); Rivlin and Wiener (1988).

Treasury Individual Income Tax Simulation Model (OTA Model)

Type In-house static tax policy model.

Supplier Office of Tax Analysis, U.S. Department of the Treasury; developed in early 1970s, based on work at the Brookings Institution extending back to the early 1960s.

Major Users Office of Tax Analysis; Joint Committee on Taxation, U.S. Congress.

Programs Simulated Federal income tax, social security payroll tax.

Main Database Statistics of Income samples of individual income tax returns.

Major Database Enhancements Imputations of deductible expenses for nonitemizing taxpayers and earnings attributable to husbands and wives; statistical match with the CES to obtain information on consumption; statistical match with the CPS March income supplement to obtain sources of income not currently subject to taxation, links between taxpayers and family or household units, and information on low-income people not required to file a return under current law; imputations to simulate the 1986 Tax Reform Act and the 1989 Omnibus Budget Reconciliation Act, simulate catastrophic health insurance

Copyright National Academy of Sciences. All rights reserved.

coverage under Medicare, and construct a broad concept of family economic income.

Projection Strategy Includes a two-stage static routine to update and project the database for a total of 10 years, with three alternatives representing different expected rates of economic growth. The first stage applies growth factors on each dollar amount to reflect actual and projected per capita real growth and inflation; the second stage adjusts weights of each family head to hit aggregate targets for 33 different variables, such as adjusted gross income class and type of return.

Behavioral Responses Simulated The decision to itemize is simulated. There is also an optional special adjustment in the aging procedures to simulate a behavioral response of taxpayers to the increase in the top tax rate on long-term capital gains resulting from the 1986 Tax Reform Act. Other behavioral responses are generally estimated outside the model.

Calibration of Baseline Simulations Not needed, as the database is a sample of actual tax returns filed. The model's simulation of tax liability is checked for agreement with reported liability, and edits are performed for the small percentage of returns with discrepancies.

Computer Implementation Hardware is VAX minicomputer; software is FORTRAN.

References Cilke and Wyscarver (1990); Gillette (1989).

TRIM2 (Transfer Income Model 2)

Type Public-use static income-support and tax program model.

Supplier The Urban Institute; original TRIM developed in early 1970s, and TRIM2 developed in the late 1970s.

Major Users Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services; Congressional Budget Office; applications also performed for U.S. Department of Labor.

Programs Simulated AFDC, SSI, food stamps, school nutrition programs, Medicare, Medicaid (for the noninstitutionalized population), employer-sponsored health insurance, federal income tax, social security payroll tax, state income tax. The AFDC, SSI, and tax modules are the most well developed.

Main Database CPS March income supplement.

Major Database Enhancements Allocation of yearly income and employment to months (see Citro and Ross, in Volume II); imputation of child care and work-related deductions for AFDC from CES data; imputation of deductible expenses

Copyright National Academy of Sciences. All rights reserved.

for the food stamp program (dependent care, shelter, and medical) based on the 1976 and 1978 Surveys of Characteristics of Food Stamp Households; imputation of deductions and other variables for simulating income taxes; estimation of Medicaid costs based on HCFA tape-to-tape and other data; match of various surveys of employment-based health insurance plans to employees on the March CPS.

Projection Strategy Contains static aging routines; however, usual practice is to simulate current programs and alternatives on the latest available database and apply percentage differences to independently developed projections for current program law.

Behavioral Responses Simulated Basic program participation decision (see Citro and Ross, in Volume II) and also the decision to itemize taxes; a module to simulate labor supply response to income benefits is under development.

Calibration of Baseline Simulations AFDC participants are controlled to administrative data on state caseloads and a few national characteristics; SSI and food stamp participants are controlled to administrative data on national characteristics and some state-level targets, including number of units; federal income tax deductions and capital gains are calibrated to IRS totals by adjusted gross income class; Medicaid caseloads are calibrated to state-level enrollment and cost data.

Computer Implementation Hardware is mainframe IBM; software is FORTRAN, Assembly (creates output files for standard packages such as SAS).

References Webb et al. (1982, 1986); Webb, Michel, and Bergsman (1990); see also Cohen et al. (in Volume II).

DATABASES

Current Population Survey (CPS) March Income Supplement The CPS is a continuing monthly cross-sectional survey of a large sample of U.S. households. The survey's primary purpose is to collect data on labor force status for people aged 15 and older to permit determining the monthly unemployment rate for the nation and large states; it also provides annual average unemployment rates for all states. In most months, the survey includes supplemental questions on other topics; for over 40 years, the March CPS has included an extensive supplement of questions on income and employment status during the previous calendar year. The Bureau of Labor Statistics is the major sponsor of the CPS; the Census Bureau sponsors the March income and some other supplements, and other agencies occasionally provide funding for special supplements.

The current CPS sample size is about 60,000 households, and the March supplement includes another 2,500 households with at least one adult of Hispanic

Copyright National Academy of Sciences. All rights reserved.

origin as of the previous November interview, plus a small number of households of Armed Forces members. Each household (more precisely, each address) is in the sample for 4 months, out for 8 months, and in for another 4 months. Information is obtained for all residents found at the interview; out-movers are not followed. Data collected in the regular interview include demographic characteristics; labor force participation, hours worked, reason for part-time work, reason for temporary absence from job, industry and occupation in the prior week; job search behavior in the previous four weeks if not working and when last worked; usual hours and usual earnings, union membership, reasons left last job, reasons not looking for work (for selected rotation groups). Data collected in the March supplement include labor force participation and job history in the prior calendar year; annual income for the prior year by detailed source, including earnings, self-employment, public and private transfers, and assets; participation in noncash benefit programs, including energy assistance, food stamps, public housing, school lunch; and health insurance coverage. See Chapter 5 and Citro (in Volume II) for uses of March CPS data for income support program and other policy modeling; see also Bureau of the Census (1990a).

Current Population Survey-Social Security Administration (CPS-SSA) Exact-Match Files CPS-SSA exact-match files, prepared by the Census Bureau and SSA, are matches of records in the March CPS with social security administrative records of quarters of coverage and taxable earnings under the OASDI program. The records are linked for the same individuals on the basis of social security number and other information. Two such files, the 1973 and 1978 CPS-SSA exact matches, have been used for policy analysis and research. (The 1973 file also included a match to IRS tax return data.) No CPS-SSA exact matches have been carried out since 1978, largely because of concerns about confidentiality. An exact match of the 1984 SIPP panel and SSA records was recently completed for research use by employees of SSA who were sworn in as Census Bureau agents. See Chapter 8 and Ross (in Volume II) for uses of CPS-SSA exact-match files for retirement income modeling; see also Subcommittee on Matching Techniques (1980).

Integrated Quality Control System (IQCS) The IQCS includes samples of administrative case records for the AFDC, food stamp, and Medicaid programs that are drawn each month by the states for use in evaluating the accuracy of the determination of eligibility and benefits for these programs. Sample sizes vary by state; total average monthly sample sizes are about 6,000 cases for AFDC and 7,500 cases for food stamps. Data abstracted from the records include case information (e.g., most recent opening, number of case members, gross countable income, net countable income); demographic characteristics for each person (e.g., relationship to head of household, age, sex, race, employment status); total household income by household member and type and amount of

Copyright National Academy of Sciences. All rights reserved.

income. See Chapters 5 and 9, Citro (in Volume II), and Citro and Ross (in Volume II) for uses of IQCS data for calibrating and evaluating income-support program models; see also Family Support Administration (1988); Food and Nutrition Service (1988).

1977 National Medicare Care Expenditures Survey (NMCES), 1980 National Medical Care Utilization and Expenditure Survey (NMCUES), 1987 National Medical Expenditure Survey (NMES) The NMCES, sponsored by the National Center for Health Services Research with the National Center for Health Statistics, consisted of six rounds of data collection covering an 18-month period in 1977 and part of 1978 for a sample of 14,000 households. Surveys were also conducted of physicians and health care facilities providing care to members of the household sample during 1977 and of employers and insurance companies responsible for their insurance coverage. Data collected included expenditures and sources of payment for all major forms of medical care, demographic and socioeconomic characteristics of respondents, insurance coverage of respondents, information from medical providers about respondents, and access to medical care.

The NMCUES, sponsored by the National Center for Health Statistics with the Health Care Financing Administration, consisted of five rounds of data collection over a 15-month period for a national sample of 6,000 households and samples of 1,000 Medicaid cases each in New York, California, Texas, and Michigan. Medicare and Medicaid records were checked for the state samples to verify eligibility and obtain claims information. Data collected included health insurance coverage, episodes of illness, number of bed days, restricted activity days, hospital admissions, physician and dental visits, other medical care encounters, prescription purchases, access to medical care services, limitations of activities, income, demographic, and socioeconomic characteristics. For each contact with the medical care system, data were obtained on health conditions, characteristics of the provider, services provided, charges, sources, and amounts of payments.

The NMES, sponsored by the Agency for Health Care Policy and Research with the Health Care Financing Administration, consisted of five rounds of data collection between February 1987 and July 1988 for a sample of 14,000 households, including oversamples of blacks, Hispanics, people aged 65 and older, low-income people, and people with functional limitations. Surveys were also conducted of physicians and health care facilities providing care to members of the household sample during 1977 and of employers and insurance companies responsible for their insurance coverage. The NMES also included an institutional survey of 13,000 residents of nursing and personal care homes, psychiatric hospitals, and facilities for the mentally retarded. The data collected included utilization, expenditures, and sources of payment for all major forms of medical care, demographic and socioeconomic characteristics of respondents,

Copyright National Academy of Sciences. All rights reserved.

insurance coverage of respondents, information from medical providers about respondents, and access to medical care.

See Chapter 8 for uses of NMCES-NMCUES-NMES data for health care policy modeling.

Statistics of Income (SOI) Individual Income Tax Returns The SOI individual income tax files are samples of tax returns and supporting schedules abstracted each year by the Statistics of Income Division of the IRS from approximately 100 million returns. Sample sizes are about 80,000 returns in even years and 120,000 returns in odd years. The sample is based on such criteria as principal business activity, presence or absence of a schedule, state from which filed, size of adjusted gross income (or loss) or largest specific income (or loss) items, and total assets or size of business and farm receipts. Recently, the design was altered to include an embedded longitudinal sample, that is, to draw a portion of the returns for the same taxpayers from year to year. Data abstracted pertain to taxpayers' income, exemptions, deductions, credits, and taxes owed (due to changes in tax laws, items vary from tax year to tax year). See Chapter 8 for use of the SOI data for tax policy modeling; see also Coleman (1988); Minarik (1980), Statistics of Income Division (no date).

Survey of Income and Program Participation (SIPP) SIPP is an ongoing panel survey of adults aged 15 and older in the civilian, noninstitutionalized population, sponsored by the Bureau of the Census. The survey follows all adults in originally interviewed households and includes children and other adults who reside with original sample members. The first panel began in fall 1983 and completed nine interviews (waves) at 4-month intervals with an initial sample of about 20,000 households. Subsequent panels began in February of each year with initial household sample sizes of about 13,500 (1985); 12,000 (1986-1989); 21,500 (1990); 14,000 (1991). These panels were completed or are planned for eight waves (1985); seven waves (1986, 1987); six waves (1988); three waves (1989); eight waves (1990, 1991).

The data collected for each interview include demographic characteristics; monthly information on labor force participation, job characteristics, and earnings; monthly information on detailed sources and amounts of income from public and private transfer payments, noncash benefits (such as food stamps, Medicaid, Medicare, and health insurance coverage); and information for the 4-month period on income from assets. Data are also collected in topical modules, which are asked once or twice in one or more panels, that cover a wide range of subjects, including: annual income and income taxes; child care and child support; educational financing and enrollment; eligibility for selected programs; employee benefits (1984 panel only); health and disability; housing costs and finance; individual retirement accounts; personal history (fertility, marital status, migration, welfare recipiency, and other topics); and wealth (property, retirement expectations and pension plan coverage, assets and liabilities). In

Copyright National Academy of Sciences. All rights reserved.

addition, each panel includes a topical module with variable content designed to respond to the needs of policy analysis agencies. See Chapter 5 and Citro (in Volume II) for uses of SIPP data for income support program and other policy modeling; see also Allin and Doyle (1990); Committee on National Statistics (1989); Jabine, King, and Petroni (1990); Vaughan (1988).

SIPP was preceded by the Income Survey Development Program (ISDP), sponsored by ASPE and SSA. The ISDP conducted research on the design of a new income survey and sponsored several data collection efforts, including the 1979 ISDP Research Panel. The 1979 ISDP obtained data similar to SIPP for an initial sample of about 9,500 households (including oversamples of low-income and high-income households), who were interviewed six times at 3-month intervals; see David (1983).

MODELING TERMS

Aging This is the process of updating a database to represent current conditions or of projecting a database for 1 or more years to represent expected future conditions. See Chapter 6; Caldwell (1989); Cohen et al. (in Volume II); Ross (in Volume II).

Dynamic aging generally 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. 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. 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 example, DYNASIM2 dynamically ages the following characteristics of the records in the database: birth, death, first marriage, remarriage, divorce, work disability, education, migration, wage rate, labor force participation, hours of work, 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 usually used to generate histories over 30 years or more for modeling retirement income programs and other policy issues with long time horizons; however, the method can be used to generate histories for any period, short or long.

Static aging generally involves adjusting the weights and selected variables in a cross-sectional database to represent the population in a future year

Copyright National Academy of Sciences. All rights reserved.

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 (typically using the projections produced by the Census Bureau or the Social Security Actuary from cell-based models); adjust the income variables to match projections of inflation and real income growth (typically using projections from macroeconomic models); 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, on the basis of transition probabilities estimated using panel data. Static aging is typically carried out for a short period, 2-5 years; however, the method can 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 This term refers to 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 benefits (e.g., Medicaid) 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 a 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 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 this example, 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. See Chapter 6; Burtless (1989); Grannemann (1989); Shoven and Whalley (1984); Strauss (1989); see also description of the MRPIS model, above.

Calibration Calibration is 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 (see Ross, in Volume II). 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

Copyright National Academy of Sciences. All rights reserved.

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 profit equation to select AFDC participants from among eligible units on the basis of such characteristics 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. See Citro and Ross (in Volume II).

Cell-Based Model This type of model produces estimates of the effects of a program change for subgroups or cells that make up the population of interest. In a welfare or tax policy model, the cells might be socioeconomic classes, such as poor families headed by single mothers aged 25 to 34. In a model of the supply of health care, the cells might be classes of health care providers, such as hospitals in large central cities. The models use aggregate data corresponding to the cells and apply appropriate parameters, for example, the average tax rate by income class.

Cell-based models range from very limited models with a handful of cells to large systems with thousands of cells. In every case, in contrast to microsimulation models, they make the critical assumption that all elements within a subgroup behave in the same way. Examples of largely cell-based social welfare policy models are the ASPE Health Financing Model (Office of the Assistant Secretary for Planning and Evaluation, 1981) and the Lewin/ICF, Inc. Macroeconomic-Demographic Model (Anderson, 1990; Cartwright, 1989). The Social Security Actuary uses cell-based models to project the people expected to be eligible for social security benefits and those expected to pay social security taxes and to estimate the balances between expenditures and receipts in the social security trust fund. The Census Bureau uses a cell-based model to project the future size and composition of the population, taking into account expected rates of fertility, mortality, and net immigration. See Grummer-Strawn and Espenshade (in Volume II) for a review of studies evaluating the quality of the population projections produced by the Actuary and the Census Bureau, which are used in aging microsimulation model databases.

Computable General Equilibrium (CGE) Model This type of model simulates second-round behavioral effects of proposed policy changes (see *behavioral response),* specifically, the effects of a policy change on prices and quantities in the various markets of an economy, taking into account feedback effects between supply and demand. For example, an increase in transfer benefits may increase product demand in the economy, which in turn has an effect on employment and wages. Or, a decrease in marginal tax rates in the personal income tax may increase work effort, thereby leading to increased labor supply, lower wage rates, and higher employment in the labor market. CGE models are calibrated

Copyright National Academy of Sciences. All rights reserved.

by a process of setting parameter values for supply and demand elasticities, drawn from the econometric literature or picked to fit the available data on market prices and quantities. Although CGE models simulate the equilibration of a full set of interconnected markets in an economy, which permits full long-run adjustment of prices to changes in supply and demand, they rarely provide guidance about the time horizon for full adjustment or the dynamic path of the adjustment process. CGE models as generally implemented are macro rather than micro in nature; however, it is possible to develop disaggregated CGE models or to link them with microsimulation models by iterating back and forth between them until market equilibrium is reached. See Berkovec and Fullerton (1989); Shoven and Whally (1984); Slemrod (1985); Whalley (1988).

Distributional Analysis This is 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 and which would lose by a policy change as 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 in comparison 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. See, for example, Beebout (1980:Table 2.8).

Dynamic Model This term refers to 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 follow the effects of demographic and economic processes and previous and proposed policy changes (e.g., raising the retirement age for social security). They draw heavily on behavioral research for their many transition probabilities, although current models include relatively few feedback effects of behavioral changes in response to simulated changes in government programs and policies. Dynamic models typically incorporate a hybrid of static and dynamic aging techniques, using dynamic techniques for most but not all variables; DYNASIM2, for example, includes static routines to supply values for assets, disability, and SSI variables to each record. See Kennell and Sheils (1990); Zedlewski (1990); see also *aging.)*

Filing Unit This term is given to the unit of analysis in microsimulation models of income support and tax programs; that is, the unit entitled to apply for benefits or obligated to file a tax return. Filing units differ within and among

Copyright National Academy of Sciences. All rights reserved.

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 *imputation*.

Imputation Imputation is the process of assigning values to variables in a database record that are missing because of item nonresponse (nonresponse to a survey question or nontranscription of an item in an administrative records system) or because the variable was never included in the survey or administrative records system. See Chapter 5; Citro (in Volume II); and Madow, Olkin, and Rubin (1983).

Imputation procedures for item nonresponse range from very simple to 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 method 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 and processed sequentially, and reported values are 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 (e.g., 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 simple to 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. These equations are then run with the estimated coefficients and the values of the independent variables in the records requiring imputation in order to estimate values of the dependent variable to impute to the records with missing data.

Copyright National Academy of Sciences. All rights reserved.

Macroeconomic Model This type of model is used to produce forecasts of the state of the economy, including such aggregates as GNP, inflation, and unemployment. These forecasts play a role in policy formation in both the public and private sectors. Macroeconomic models are also used to simulate the effects of proposed government fiscal and budgetary policy changes on economic aggregates. Macroeconomic models, which have been developed by private firms such as DRI, academic researchers, and government economists, are composed of systems of simultaneous equations estimated with aggregate time-series data. Analysts supply expected values for exogenous factors, including policy variables such as tax rates and government spending. The models then use these values as input to their systems of equations to determine the impact on economic aggregates such as GNP and personal consumption expenditures.

Macroeconomic models vary in size and complexity. They have been developed for personal computers as well as for large mainframes, and they can include relatively few to thousands of equations based on relatively few to many thousands of time series. They can also produce a small number of outputs for the national economy or many outputs, disaggregated by such attributes as industry and region. Their distinguishing characteristic is that they model the behavior of aggregates, such as all consumers in the nation or in a region or state, not individual decision units. See Klein (1986); Kraemer et al. (1987); McNees (1989, 1990).

Matching Matching is 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. See Chapters 5 and 8; Cohen (Chapter 2 in Volume II); and Subcommittee on Matching Techniques (1980).

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 a 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. Typically, 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

Copyright National Academy of Sciences. All rights reserved.

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 This type of social welfare policy model operates at the individual decision unit level, which can be an individual or family for an income-support program model, 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 (see Beebout, 1986). In very general terms, the first step, which serves the same function as the control group for an experiment, is to prepare a baseline database representing the current situation, that is, the situation in the absence of a program change. The second step is to simulate the program change and its impact. 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 to mimic the rules of government programs (accounting functions), 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 demographic or economic behavior (such as marriage or job change) in aging 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 draws a probability for each decision unit at random, compares that probability to an estimated participation probability for the particular type of unit, and, if the former probability is less (more) than the latter, designates the unit (not) to participate. See *dynamic model* and *static model.*

Second-Round Effects See *behavioral response.*

Static Model This term refers to microsimulation models that operate on a database representing a cross-section of the population at a given time. Such models typically simulate the direct effects of policy changes, 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 can also be used to generate estimates for future years, for which they use 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 Beebout, 1980, 1986; Webb, Michel, and Bergsman, 1990; see also *aging*.

Weighting This is the process of assigning weights (factors) to observations in a sample survey so that the weighted count of all observations will approximate

Copyright National Academy of Sciences. All rights reserved.

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. The last set of adjustment factors is developed from the previous decennial census updated by vital records. See, for example, Bureau of the Census (1989a); Jabine, King, and Petroni (1990).

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Glossary of Acronyms

The list below includes legislation, programs, agencies, surveys, and models; for a survey, program, or model, the agency that developed or is responsible for it is given in parentheses.

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, 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
CES	Consumer Expenditure Survey (BLS)
CGE	Computable general equilibrium [models]
CORSIM	Cornell Simulation Model
CPI	Consumer Price Index (BLS)
CPS	Current Population Survey (BLS and Census Bureau)
DOL	[U.S.] Department of Labor
DRI	Data Resources, Inc.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

DYNASIM2	Dynamic Simulation of Income Model 2 (Urban Institute)
FNS	Food and Nutrition Service (U.S. Department of Agriculture)
FOSTERS	Food Stamp Eligibility Routines (Mathematica Policy Research, Inc.)
FSA	Family Support Act [of 1988]
GAO	General Accounting Office
GNP	Gross National Product (BEA)
HANES	Health and Nutrition Examination Survey (NCHS)
HBSM	Health Benefits Simulation Model (Lewin/ICF, Inc.)
HCFA	Health Care Financing Administration (U.S. Department of Health and Human Services)
HEW	[U.S. Department of] Health, Education, and Welfare
HHS	[U.S. Department of] Health and Human Services
HIS	Health Interview Survey (NCHS)
HITSM	Household Income and Tax Simulation Model (Lewin/ICF, Inc.)
IEEE	Institute of Electrical and Electronics Engineers
IQCS	Integrated Quality Control System (HHS and FNS)
IRA	Individual retirement account
IRS	Internal Revenue Service (U.S. Department of the Treasury)
ISDP	Income Survey Development Program (ASPE and SSA)
JCT	Joint Committee on Taxation (U.S. Congress)
KGB	Kasten, Greenberg, and Betson [model] (ASPE)
LTCFM	Long-Term Care Financing [Sub]Model (Lewin/ICF, Inc.)
MATH	Micro Analysis of Transfers to Households [model] (Mathematica Policy Research, Inc.)
MDM	Macroeconomic-Demographic Model (Lewin/ICF, Inc.)
MRPIS	Multi-Regional Policy Impact Simulation [model] (Social Welfare Research Institute, Boston College)
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, 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 Expenditure Survey (NCHSR with NCHS)
NMCUES	[1980] National Medical Care Utilization and Expenditure Survey (NCHS with HCFA)

Copyright National Academy of Sciences. All rights reserved.

NMES	[1987] National Medical Expenditure Survey (AHCPR with HCFA)
OASDI	Old Age, Survivors, and Disability Insurance [social security] (SSA)
OBRA	[1981] Omnibus Budget 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]
PRISM	Pension and Retirement Income Simulation Model (Lewin/ ICF, Inc.)
PSID	Panel Study of Income Dynamics (Institute for Social Research, University of Michigan)
RIM	Reforms in Income Maintenance [model] (President's Commission on Income Maintenance Programs)
SIPP	Survey of Income and Program Participation (Census Bureau)
SOCSIM	[Demographic]-Social Simulation [model]
SOI	Statistics of Income [Division] (Internal Revenue Service, U.S. Department of the Treasury); Statistics of Income [samples of tax returns]
SPSD/M	Social Policy Simulation Database/Model (Statistics Canada)
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)
STATS	Simulated Tax and Transfer System [model] (SSA)
TRA	[1986] Tax Reform Act
TRIM2	Transfer Income Model 2 (Urban Institute)

Copyright National Academy of Sciences. All rights reserved.

References

Allin, Susan, and Doyle, Pat 1990 *Analysis of the Quality of SIPP as It Pertains to the Food Stamp Program.* Final report to the Food and Nutrition Service, U.S. Department of Agriculture. Washington, D.C.: Mathematica Policy Research, Inc.

Anderson, Gordon 1989 Capital Stock Levels, Tax Capacity Exhaustion and Cost of Capital Endogeneity. University of Toronto, Canada.

Anderson, Joseph M. 1990 Micro-macro linkages in economic models. Pp. 187-220 in Gordon H. Lewis and Richard C. Michel, eds., *Microsimulation Techniques for Tax and Transfer Analysis.* Washington, D.C.: The Urban Institute Press.

Andrews, Richard W., Birdsal L., William C., Gentner, Frederick J., and Spivey, W. Allen 1987 Validation Methods for Microeconomic Simulation. Report prepared for the Social Security Administration, U.S. Department of Health and Human Services, Washington, D.C.

Arrow, Kenneth J. 1980 Microdata simulation: Current status, problems, prospects. Pp. 253-265 in Robert H. Haveman and Kevin Hollenbeck, eds., *Microeconomic Simulation Models for Public Policy Analysis. Vol. II, Sectoral, Regional, and General Equilibrium Models.* New York: Academic Press.

Arthur, W. Brian 1990 Positive feedbacks in the economy. *Scientific American* 262(2):92-99.

Atrostic, B. K., and Nunns, James R. 1988 Measuring Tax Burden: A Historical Perspective. Paper presented to the 50th Anniversary Conference of the Conference on Research in Income and Wealth, Washington, D.C.

Copyright National Academy of Sciences. All rights reserved.

Bailar, Barbara 1983 Error profiles: Uses and abuses. Pp. 117-130 in Tommy Wright, ed., *Statistical Methods and the Improvement of Data Quality*. New York: Academic Press.

Baseline Data Corporation 1984 *The Federal Statistical System, 1980 to 1985*. Report prepared for the Congressional Research Service in response to the Committee on Government Operations, U.S. House of Representatives. Washington, D.C.: U.S. Government Printing Office.

Beebout, Harold 1980 Food stamp policy modeling: An application of MATH. Pp. 45-72 in Robert H. Haveman and Kevin Hollenbeck, eds., *Microeconomic Simulation Models for Public Policy Analysis. Vol. 1, Distributional Impacts.* New York: Academic Press.

1986 Evaluating Reagan Administration social program changes: Two applications of MATH. Pp. 83-97 in G. H. Orcutt, J. Merz, and H. Quinke, eds., *Microanalytic Simulation Models to Support Social and Financial Policy.* Amsterdam:Elsevier Science Publishers.

Beebout, Harold, and Grossman, Jean Baldwin 1985 *A Forecasting System for AFDC Caseloads and Costs: Executive Summary.* Prepared by Mathematica Policy Research, Inc. AFDC and Related Income Maintenance Studies, Office of Family Assistance, Social Security Administration. Washington, D.C.: U.S. Department of Health and Human Services.

Beebout, Harold, and Haworth, Lauren 1989 Microsimulation Estimates Versus Reality: A Case Study. Unpublished memorandum. Mathematica Policy Research, Inc., Washington, D.C.

Bergmann, Barbara R. 1990 Micro-to-macro simulation: A primer with a labor market example. *The Journal of Economic Perspectives* 4(1):99-116.

Berkovec, James, and Fullerton, Don 1989 The general equilibrium effects of inflation on housing consumption and investment . *The American Economic Review* 79(2):277-282.

Betson, David M. 1988 How Reliable Are Conclusions Derived from Microsimulation Models? Paper prepared for the Workshop on Prospects and Limits of Simulation Models in Tax and Transfer Policy, Giessen, Federal Republic of Germany.

Betson, David, Greenberg, David, and Kasten, Richard 1980 A microsimulation model for analyzing alternative welfare reform proposals: An application to the Program for Better Jobs and Income. Pp. 153-188 in Robert H. Haveman and Kevin Hollenbeck, eds., *Microeconomic Simulation Models for Public Policy Analysis. Vol. I, Distributional Impacts.* New York: Academic Press.

Birnbaum, Jeffrey H., and Murray, Alan S. 1987 *Showdown at Gucci Gulch-Lawmakers, Lobbyists, and the Unlikely Triumph of Tax Reform.* New York: Vintage Books.

Boskin, Michael J. 1990 Improving the Quality of Economic Statistics. Testimony before the Joint Economic Committee Hearing on the Quality of Federal Statistical Programs, March 1, 1990, Washington, D.C.

Copyright National Academy of Sciences. All rights reserved.

Brieman, L. 1988 Submodel Selection and Evaluation in Regression: The Conditional Case and Little Bootstrap. Technical Report No. 169. Department of Statistics, University of California, Berkeley.

Bristol, Ralph B., Jr. 1988 Tax modeling and the policy environment of the 1990s. Pp. 115-122 in *Statistics of Income Bulletin.* 75th Anniversary Issue. Statistics of Income Division, Internal Revenue Service. Washington, D.C.: U.S. Department of the Treasury.

Brooks, C. A., and Bailar, B. A. 1978 *An Error Profile: Employment as Measured by the Current Population Survey.* Statistical Policy Working Paper No. 3. Federal Committee on Statistical Methodology, Office of Federal Statistical Policy and Standards. Washington, D.C.: U.S. Department of Commerce.

Bulmer, Martin 1986 *Social Science and Social Policy.* Winchester, Mass.: Allen & Unwin.

Bulmer, Martin, ed. 1987 *Social Science Research and Government: Comparative Essays on Britain and the United States.* New York:Cambridge University Press.

Bureau of the Census 1970 *1970 Census Users' Guide.* Part 1. Washington, D.C.: U.S. Department of Commerce.

1988 *Measuring the Effect of Benefits and Taxes on Income and Poverty: 1986.* Current Population Reports, Consumer Income, Series P-60, No. 164-RD-1. Washington, D.C.: U.S. Department of Commerce.

1989a *Money Income and Poverty Status in the United States: 1988.* Current Population Reports, Consumer Income, Series P-60, No. 166. Washington, D.C.: U.S. Department of Commerce.

1989b *Poverty in the United States: 1987.* Current Population Reports, Consumer Income, Series P-60, No. 163. Washington, D.C.: U.S. Department of Commerce.

1990a *Current Population Survey, March 1989 Tape Technical Documentation.* Washington, D.C.: U.S. Department of Commerce.

1990b *Statistical Abstract of the United States-1990.* 110th Edition. Washington, D.C.: U.S. Department of Commerce.

1991 *Statistical Abstract of the United States-1991.* 111th Edition. Washington, D.C.: U.S. Department of Commerce.

Burkhauser, R., and Quinn, J. 1981 The Relationship Between Mandatory Retirement Age Limits and Pension Rules in the Retirement Decision. Working paper 1348-03. The Urban Institute, Washington, D.C.

Burtless, Gary 1989 Labor Supply Response to Microsimulation Models. Paper prepared for the Panel to Evaluate Microsimulation Models for Social Welfare Programs. The Brookings Institution, Washington, D.C.

Butz, William P. 1984 Data Confidentiality and Public Perceptions: The Case of the European Censuses. Paper presented to the Population Association of America, Minneapolis, Minn.

Copyright National Academy of Sciences. All rights reserved.

Butz, William P., and Plewes, Thomas J. 1990 Moving the Current Population Survey into the 21st century. In *Proceedings of the Section on Survey Research Methods.* Alexandria, Va.: American Statistical Association.

Caldwell, Steven B. 1989 Strategies for Aging a Population. Paper prepared for the Panel to Evaluate Microsimulation Models for Social Welfare Programs. Cornell University, Ithaca, N.Y.

Caplan, N., Morrison, A., and Stambaugh, R.J. 1976 *The Use of Social Science Knowledge in Policy Decisions at the National Level.* Institute for Social Research. Ann Arbor, Mich.: University of Michigan .

Carlson, Steven 1989 FNS Microsimulation Case Study. Paper prepared for the Panel to Evaluate Microsimulation Models for Social Welfare Programs. Food and Nutrition Service, U.S. Department of Agriculture, Washington, D.C.

Cartwright, William S. 1989 Projections of Future Health Care Expenditures. Paper prepared for a policy forum sponsored by the Employee Benefit Research Institute-Education and Research Fund, Washington, D.C.

Chollet, Deborah J. 1990 Microsimulation Models in Health Services Research. Paper prepared for the Panel to Evaluate Microsimulation Models for Social Welfare Programs. Georgia State University, Atlanta, Ga.

Cilke, James, and Wyscarver, Roy A. 1990 *The Treasury Individual Income Tax Simulation Model.* Office of Tax Analysis. Washington, D.C.: U.S. Department of the Treasury.

Citro, Constance F. 1989 The Role of Static Microsimulation in Policy Analysis: The Case of DHHS/ASPE in the 1980s; the Case of USDA/FNS in the Late 1980s. Paper prepared for the Panel to Evaluate Microsimulation Models for Social Welfare Programs, Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council, Washington, D.C.

Citro, Constance F., and Beebout, Harold 1984 *Effects of the 1981 OBRA Welfare Changes and the Recession on Poverty.* Paper prepared for the Congressional Research Service. Washington, D.C.: Mathematica Policy Research, Inc.

Citro, Constance F., and Cohen, Michael L., eds. 1985 *The Bicentennial Census: New Directions for Methodology in 1990.* Panel on Decennial Census Methodology, Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, D.C.: National Academy Press.

Clogg, Clifford, Massagli, Michael, and Eliason, Scott 1986 Population undercount as an issue in social research. Pp. 335-343 in *Proceedings of the Bureau of the Census Second Annual Research Conference.* Washington, D.C.: U.S. Department of Commerce.

Coder, John F. No date The March Current Population Survey Earnings Imputation System: An Explanation and Evaluation. Bureau of the Census, Washington, D.C.

Copyright National Academy of Sciences. All rights reserved.

Coleman, Michael J. 1988 Statistics of income studies of individual income and taxes. Pp. 63-80 in *Statistics of Income Bulletin.* 75th Anniversary Issue. Statistics of Income Division, Internal Revenue Service. Washington, D.C.: U.S. Department of the Treasury.

Committee on National Statistics 1989 *The Survey of Income and Program Participation: An Interim Assessment.* Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, D.C.: National Academy Press.

Congressional Budget Office 1988 *The Medicare Catastrophic Coverage Act of 1988.* Staff working paper, prepared by Sandra Christensen and Rick Kasten. Washington, D.C.: U.S. Government Printing Office.

1989a *The Economic and Budget Outlook: An Update.* Washington, D.C.: U.S. Government Printing Office.

1989b *Fiscal Federalism and the Partnership Act of 1987.* Special study prepared by Janice Peskin. Washington, D.C.: U.S. Government Printing Office.

1989c *Updated Estimates of Medicare's Catastrophic Drug Insurance Program.* Special study prepared by Stephen H. Long and Nancy Gordon. Washington, D.C.: U.S. Government Printing Office.

1989d *Work and Welfare: The Family Support Act of 1988.* Staff working paper, prepared by Janice Peskin. Washington, D.C.: U.S. Government Printing Office.

1990 *Physician Payment Reform Under Medicare.* Prepared by Sandra Christensen and Scott Harrison. Washington, D.C.: U.S. Government Printing Office.

1991 *The Economic and Budget Outlook: Fiscal Years 1992-1996.* A report to the Senate and House Committees on the Budget. Washington, D.C.: U.S. Government Printing Office.

Cottingham, Phoebe H., and Ellwood, David T., eds. 1989 *Welfare Policy for the 1990s.* Cambridge, Mass.: Harvard University Press.

Council of Economic Advisers 1990 Improving the Quality of Economic Statistics. Press release. Executive Office of the President, Washington, D.C.

1991 FY 1992 Economics Statistics Initiative: Improving the Quality of Economic[s] Statistics. Press release. Executive Office of the President, Washington, D.C.

Dabrowski, Christopher E., Fong, Elizabeth N., and Yang, Deyuan. 1990 *Object Database Management Systems: Concepts and Features.* NIST special publication 500-179. National Computer Systems Laboratory, National Institute of Standards and Technology. Washington, D.C.: U.S. Department of Commerce .

Darby, Michael R. 1990 Prepared statement. Testimony before the Joint Economic Committee Hearing on the Quality of Federal Statistical Programs, March 1, 1990, Washington, D.C.

David, Martin H. 1990 Data-Driven Social Policy, Data-Driven Social Research: A Reality in the 1990's? Paper presented to the Association for Public Policy Analysis and Management.

1991 The science of data sharing: Documentation. In Joan E. Sieber, ed., *Sharing*

Copyright National Academy of Sciences. All rights reserved.

Social Science Data: Advantages and Challenges. Newbury Park, Calif.: Sage Publications.

David, Martin H., ed. 1983 *Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP): Papers Presented at a Conference.* New York: Social Science Research Council.

David, Martin, and Robbin, Alice 1989 Database design for large-scale complex data. In *Computing Science and Statistics: Proceedings of the 21st Symposium on the Interface.* Alexandria, Va.: American Statistical Association.

1990 Computation using information systems for complex data. In *Proceedings of the Conference on Advanced Computing for the Social Sciences.* Oak Ridge, Tenn.: Oak Ridge National Laboratories.

David, Martin, Little, Roderick J. A., Samuhel, Michael E., and Triest, Robert K. 1986 Alternative methods for CPS income imputation. *Journal of the American Statistical Association* 81 (393):28-41.

Davidson, Roger H. 1976 Congressional committees: The toughest customers. *Policy Analysis* Spring: 299-323.

de Leeuw, Frank, and Struyk, Raymond J. 1975 *The Web of Urban Housing—Analyzing Policy with a Market Simulation Model.* Washington, D.C.: The Urban Institute Press.

Doyle, Pat 1990 *Food Stamp Program Participation Rates: August 1985.* Prepared by Mathematica Policy Research, Inc., for the Food and Nutrition Service. Washington, D.C.: U.S. Department of Agriculture.

Doyle, Pat, and Trippe, Carole 1989 *Validation of the Food Stamp Microsimulation Model.* Final report to the Food and Nutrition Service, U. S. Department of Agriculture. Washington, D.C.: Mathematica Policy Research, Inc.

Doyle, Pat, Trippe, Carole, Huff, Ann, and Coffin, Kirsten, eds. 1990 *MATH Technical Description: Current Services Files.* Washington, D.C.: Mathematica Policy Research, Inc.

Dyke, Bennett, and MacCluer, Jean Walters, eds. 1973 *Computer Simulation in Human Population Studies.* New York: Academic Press.

Economic Policy Council 1987 Report of the Working Group on the Quality of Economic Statistics to the Economic Policy Council. Executive Office of the President, Washington, D.C.

Efron, Bradley 1979 Bootstrap methods: Another look at the jackknife. *Annals of Statistics* 7:1-26.

1982 The jackknife, the bootstrap, and other resampling plans. *SIAM-CBMS* No. 38.

Efron, Bradley, and Morris, Carl 1973 Stein's estimation rule and its competitors—An empirical Bayes approach. *Journal of the American Statistical Association* 67:130-139.

Eliasson, Gunnar 1985 The Firm and Financial Markets in the Swedish Micro to Macro Model:

Copyright National Academy of Sciences. All rights reserved.

Theory, Model and Verification. Industrial Institute for Economic and Social Research, Stockholm, Sweden.

Family Support Administration 1988 *Characteristics and Financial Circumstances of AFDC Recipients—FY 1987.* Social Security Administration. Washington, D.C.: U.S. Department of Health and Human Services.

Fay, Robert E., Passel, Jeffrey S., and Robinson, Gregory J. 1988 *The Coverage of the Population in the 1980 Census.* 1980 Census of Population and Housing Evaluation Research Report No. PHC80-E4. Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Fein, David 1989 The Social Sources of Census Omission: Racial and Ethnic Differences in Omission Rates in Recent U.S. Censuses. Dissertation submitted to Sociology Department, Princeton University.

Fisher, Elliot, Baron, John A., Malenka, David J., et al. 1990 Overcoming potential pitfalls in the use of Medicare data for epidemiological research. *American Journal of Public Health* 80 (12):1487-1490.

Food and Nutrition Service 1986 *Food Stamp Research: Results from the Income Survey Development Program and the Promise of the Survey on Income and Program Participation.* Prepared by Mathematica Policy Research, Inc., for the Office of Analysis and Evaluation. Washington, D.C.: U.S. Department of Agriculture.

1988 *Characteristics of Food Stamp Households.* Office of Analysis and Evaluation. Washington, D.C.: U.S. Department of Agriculture.

1990 *Food Stamp Policy Issues: Results from Recent Research.* Papers presented at the Food and Nutrition Service Research Conference. Office of Analysis and Evaluation. Washington, D.C.: U.S. Department of Agriculture.

Fuerbringer, Jonathan 1990 A rusty statistical compass for U.S. policy mappers. *Chance: New Directions for Statistics and Computing* 3(1):36-38,56.

General Accounting Office 1977 *An Evaluation of the Use of the Transfer Income Model—TRIM—to Analyze Welfare Programs.* Report to the Congress. Washington, D.C.: U.S. Government Printing Office.

1979 Guidelines for Model Evaluation. Exposure Draft. Washington, D.C.

1984 *Status of the Statistical Community After Sustaining Budget Reductions.* Staff study. Washington, D.C.: U.S. Government Printing Office.

Giannarelli, Linda 1989 Aging the 1983 CPS File to 1987. Unpublished memorandum. Urban Institute, Washington, D.C.

1990 Looking at the Quality of the 1987 Baseline AFDC Run, and the Quality of the March 1988 CPS. Unpublished memorandum. Urban Institute, Washington, D.C.

Gilford, Dorothy M., ed. 1988 *The Aging Population in the Twenty-First Century: Statistics for Health Policy.* Panel on Statistics for an Aging Population, Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, D.C.: National Academy Press.

Gillette, Robert 1989 Measures of Goodness of Fit for Extrapolations: Initial Results Using the

Copyright National Academy of Sciences. All rights reserved.

Individual Tax Model Data Base. OTA paper no. 62. Office of Tax Analysis, U.S. Department of the Treasury, Washington, D.C.

Grannemann, Thomas W. 1989 Modeling Behavioral Responses to Health Policy Changes. Paper prepared for the Panel to Evaluate Microsimulation Models for Social Welfare Programs. University of Colorado, Denver, Colo.

Grossman, Jean B. 1985 *The Technical Report for the AFDC Forecasting Project for the Social Security Administration/Office of Family Assistance.* Prepared by Mathematica Policy Research, Inc. AFDC and Related Income Maintenance Studies, Office of Family Assistance, Social Security Administration. Washington, D.C.: U.S. Department of Health and Human Services.

Grossman, Jean B., and Watts, Harold W. 1982 *The Structural Needs of Microsimulation Modeling of Labor Supply and Retirement Behavior.* A Report to the Economic Studies Program, Program for Research on Retirement and Aging. Washington, D.C.: Brookings Institution.

Haas, Lawrence J. 1990 Tax report: Protection squad. *National Journal* May 26:1270-1274.

Hammel, E. A., Wachter, K. W., and McDaniel, C. K. 1981 The kin of the aged in A.D. 2000: The chickens come home to roost. Pp. 1139 in Sara B. Kiesler, James N. Morgan, and Valerie Kincade Oppenheimer, eds., *Aging: Social Change.* New York: Academic Press.

Hammel, E. A., Hutchinson, D. W., Wachter, K. W., Lundy, R. T., and Deuel, R. Z. 1976 The SOCSIM Demographic-Sociological Microsimulation Program Operating Manual. Institute of International Studies Research Series, No. 27. University of California, Berkeley.

Hanson, Kenneth A., and Robinson, Sherman 1988 Equilibrium and Disequilibrium in Economy wide Microsimulation Models. Economic Research Service, U.S. Department of Agriculture, Washington, D.C.

Hanushek, Eric A. 1990 The policy research market. *Journal of Policy Analysis and Management* 9 (2): 146-154.

Haveman, Robert H. 1987 *Poverty Policy and Poverty Research: The Great Society and the Social Sciences.* Madison: University of Wisconsin Press.

Haveman, Robert H., and Lacker, Jeffrey 1984 *Discrepancies in Projecting Future Public and Private Pension Benefits: A Comparison and Critique of Two Micro-Data Simulation Models.* Institute for Research on Poverty Special Report 36. Madison: University of Wisconsin Press.

Haveman, Robert H., Hollenbeck, Kevin, Betson, David, and Holmer Martin 1980 A microeconomic simulation model for analyzing the regional and distributional effects of tax-transfer policy: An analysis of the Program for Better Jobs and Income. Pp. 137-162 in Robert H. Haveman and Kevin Hollenbeck, eds., *Microeconomic Simulation Models for Public Policy Analysis. Vol. II, Sectoral, Regional, and General Equilibrium Models.* New York: Academic Press.

Havens, John, and Clayton-Matthews, Alan 1989 The MRPIS Model: Notes for Presentation to Committee on National

Copyright National Academy of Sciences. All rights reserved.

Statistics. Prepared for the Panel to Evaluate Microsimulation Models for Social Welfare Programs. Social Welfare Research Institute, Boston College, Chestnut Hill, Mass.

Havens, John, Bluestone, Barry, Fournier, Stephen, and Pepin, Carol 1985 The Microeconomic Impacts of Macroeconomic Fiscal Policy, 1981-1985. MRPIS Project, Social Welfare Research Institute, Boston College, Chestnut Hill, Mass.

Hayes, Brian M. 1989 An Object Oriented Programming Implementation of a Taxation and Benefit Program. Discussion paper. London School of Economics.

Hayes, Richard L. 1982 An Evaluation of Forecasting Accuracy in the Microsim Model. Task 1 Interim Report. The Policy Research Group, Washington, D.C.

Health Care Financing Administration 1990 Current Beneficiary Survey: Facts. U.S. Department of Health and Human Services, Washington, D.C.

Hellwig, Otto 1989a A Chronology of Microsimulation Models and Software. Augsburg University, Federal Republic of Germany.

1989b Microsimulation Software: State of the Art and Outlook. Paper presented at Softstat '89, Mannheim, Federal Republic of Germany.

Hendricks, Gary, and Holden, Russell 1976a The role of microanalytic simulation models in projecting OASDI costs. In *Proceedings of the Business and Economic Statistics Section.* Washington, D.C.: American Statistical Association.

1976b A Report on the Variance in Simulated Earnings Histories Produced by DYNASIM. Urban Institute, Washington, D.C.

Hoagland, William 1989 Congress's Use of Statistics. Presentation to the Washington Statistical Society, May 1989.

Holahan, John, and Zedlewski, Sheila 1989 Insuring Low-Income Americans Through Medicaid Expansion. Draft report to the Robert Wood Johnson Foundation. Urban Institute, Washington, D.C.

Holden, Russell 1977 Components of Variation in Longitudinal Earnings History. Urban Institute, Washington, D.C.

Horvitz, D. G., Giesbrecht, F. G., Shah, B. V., and Lachenbruch, P.A. 1971 POPSIM, A Demographic Simulation Program. Monograph 12. Carolina Population Center, University of North Carolina, Chapel Hill, N.C.

ICF, Inc. 1987 *Microsimulation Models of Public Assistance: Similarities and Differences.* Report to the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services . Washington, D.C.: ICF, Inc.

Income Working Group 1990 Income Data for the Elderly: Guidelines. Draft report. Interagency Forum on Aging-Related Statistics, Washington, D.C.

Ingram, Gregory K., Kain, John F., and Ginn, J. Royce 1972 *The Detroit Prototype of the NBER Urban Simulation Model.* National Bureau of Economic Research. New York: Columbia University Press.

Copyright National Academy of Sciences. All rights reserved.

Institute of Electrical and Electronics Engineers 1988 *IEEE Standard for Software User Documentation.* Software Engineering Technical Committee of the IEEE Computer Society. New York: Institute of Electrical and Electronics Engineers, Inc.

Internal Revenue Service 1989 Selected Background Papers. Volume 1. Workshop on Tax Microsimulation Modeling for Business, May 11-12, 1989. Statistics of Income Division, U.S. Department of the Treasury, Washington, D.C.

Jabine, Thomas B., King, Karen E., and Petroni, Rita J. 1990 *Survey of Income and Program Participation: Quality Profile.* Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Jacobs, Bruce 1989 Assessing the Affordability of Private Long-Term Care Insurance: Some Cautionary Thoughts. Testimony before the U.S. Bipartisan Commission on Comprehensive Health Care, Washington, D.C.

Jefferson, Linda G. 1983 Validation of Mean Earnings and Earnings Distributions Simulated by DYNASIM for 1983. Urban Institute, Washington, D.C.

Johnson, Jon, and Zedlewski, Sheila R. 1982 *The Dynamic Simulation of Income Model (DYNASIM). Vol. II, The Jobs and Benefits History Model.* Washington, D.C.: The Urban Institute Press.

Johnson, Jon, Wertheimer, Richard, and Zedlewski, Sheila R. 1983 *The Dynamic Simulation of Income Model (DYNASIM). Vol. I, The Family and Earnings History Model.* Washington, D.C.: The Urban Institute Press.

Joint Economic Committee 1990 *Economic Indicators: December 1990.* Prepared by the Council of Economic Advisers. Washington, D.C.: U.S. Government Printing Office.

Juster, F. Thomas 1988 The State of U.S. Economic Statistics: Current and Prospective Quality, Policy Needs, and Resources. Paper presented to the 50th Anniversary Conference of the Conference on Research in Income and Wealth, Washington, D.C.

Kadane, Joseph B. 1978 Statistical problems of merged data files. Paper 6 in *Compilation of OTA Papers. Vol. I.* Washington, D.C.: U.S. Department of the Treasury.

Kalleberg, Arne L., ed. 1986 *America at Work: National Surveys of Employees and Employers.* Report prepared for the U.S. Department of Labor. Advisory Group on a 1986 Quality of Employment Survey. New York: Social Science Research Council.

Kasten, Richard A., and Sammartino, Frank J. 1990 A method for simulating the distribution of combined federal taxes using census, tax return, and expenditure microdata. Pp. 173-186 in Gordon H. Lewis and Richard C. Michel, eds., *Microsimulation Techniques for Tax and Transfer Analysis.* Washington, D.C.: The Urban Institute Press.

Kennell, David L., and Sheils, John F. 1986 The ICF Pension and Retirement Income Simulation Model (PRISM) with the ICF/Brookings Long-Term Care Financing Model. Draft technical documentation, ICF, Inc., Washington, D.C.

1990 PRISM: Dynamic simulation of pension and retirement income. Pp. 137-172 in Gordon H. Lewis and Richard C. Michel, eds., *Microsimulation*

Copyright National Academy of Sciences. All rights reserved.

Techniques for Tax and Transfer Analysis. Washington, D.C.: The Urban Institute Press.

Kennell, David, et al. 1988 *Financing of Long-Term Care.* Final report submitted to the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. Washington, D.C.: Lewin/ICF, Inc.

Klein, Lawrence R. 1986 Macroeconomic modeling and forecasting. Pp. 95-110 in Neil J. Smelser and Dean R. Gerstein, eds., *Behavioral and Social Science: Fifty Years of Discovery.* Committee on Basic Research in the Behavioral and Social Sciences, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, D.C.: National Academy Press.

Kormendi, Roger C., and Meguire, Philip 1988 Dynamic Validation of the TRIM Welfare Simulation Model. Draft report to the Urban Institute, Washington, D.C.

Kraemer, Kenneth L., Dickhoven, Siegfried, Tierney, Susan Fallows, and King, John Leslie 1987 *Datawars—The Politics of Modeling in Federal Policy making.* New York: Columbia University Press.

Lehmann, E. L. 1975 *Nonparametrics: Statistical Methods Based on Rank.* San Francisco: Holden-Day.

Levitan, Sar A., and Gallo, Frank 1989 *Workforce Statistics: Do We Know What We Think We Know —And What Should We Know?* Prepared for the Joint Economic Committee. Washington, D.C.: U.S. Government Printing Office.

Lewin/ICF, Inc. 1988 *The Household Income and Tax Simulation Model (HITSM): Methodology and Documentation.* Washington, D.C.: Lewin/ICF, Inc.

Lillard, Lee, Smith, James P., and Welch, Finis 1986 What do we really know about wages? The importance of nonreporting and census imputation. *Journal of Political Economy* 94 (31):489-506.

Lindley, D.V., and Smith, A.F.M. 1972 Bayes estimates for the linear model. *Journal of the Royal Statistical Society* Series B, 34:1-19.

Little, RJ.A., and Rubin, D.B. 1987 *Statistical Analysis with Missing Data.* New York: John Wiley & Sons.

Lubitz, Irene, and Doyle, Pat 1986 Food Stamp Monthly Income—Allocating Annual Income in the MATH Model. Unpublished memorandum . Mathematica Policy Research, Inc., Washington, D.C.

Lynn, Laurence E., Jr., ed. 1978 *Knowledge and Policy: The Uncertain Connection.* Study Project on Social Research and Development, Assembly of Behavioral and Social Sciences, National Research Council. Washington, D.C.: National Academy of Sciences.

McClung, Nelson 1970 Estimates of income transfer program direct effects. In *The President's Commission on Income Maintenance Programs: Technical Studies.* Washington, D.C.: U.S. Government Printing Office.

Copyright National Academy of Sciences. All rights reserved.

McNees, Stephen K. 1989 Why do forecasts differ? *New England Economic Review* January/ February: 42-54.

1990 Man vs. model? The role of judgment in forecasting. *New England Economic Review* July/ August:41-52.

Madow, William G., Olkin, Ingram, and Rubin, Donald B., eds. 1983 *Incomplete Data in Sample Surveys.* 3 vols. Panel on Incomplete Data, Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council. New York: Academic Press.

Magnuson, Ed 1990 Back to the other war. *TIME* September 10:42.

Manton, Kenneth G., and Stallard, Eric 1990 Incorporating risk factors into projections of the size and health status of the U.S. elderly population . In K. G. Manton, B. Singer, and R. Suzman, eds., *Forecasting the Health of the Old.* New York: Oxford University Press.

Manton, Kenneth G., Woodbury, Max A., and Stallard, Eric 1990 Longitudinal Models of Disability Changes and Active Life Expectancy in Elderly Populations: The Interaction of Sex, Age and Marital Status. Paper presented at the Consultation on Modeling of Noncommunicable Disease Prevention Programs, World Health Organization, Heidelburg, Germany.

1991 Statistical and Measurement Issues in Assessing the Welfare Status of Aged Individuals and Populations. Center for Demographic Studies, Duke University, Durham, N.C.

Maxfield, Myles, Jr., and Rucci, Mark 1985 *A Simulation Model of Targeted Employment and Training Programs for Long Term Welfare Recipients: Technical Documentation.* Prepared for the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. Washington, D.C.: Mathematica Policy Research, Inc.

Minarik, Joseph J. 1980 The MERGE 1973 data file. Pp. 3-28 in Robert H. Haveman and Kevin Hollenbeck, eds., *Microeconomic Simulation Models for Public Policy Analysis. Vol. 1, Distributional Impacts.* New York: Academic Press.

Moffitt, Robert, and Wolfe, Barbara 1990 *The Effect of the Medicaid Program on Welfare Participation and Labor Supply.* Discussion paper no. 909-90. Institute for Research on Poverty. Madison: University of Wisconsin Press.

Morris, Richard B. 1953 *Encyclopedia of American History.* New York: Harper & Brothers.

Nelson, Charles, and Short, Kathleen 1990 *Health Insurance Coverage: 1986-88.* Current Population Reports, Household Economic Studies, Series P-70, No. 17. Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Nelson, Richard R., and Winter, Sidney G. 1982 *An Evolutionary Theory of Economic Change.* Cambridge, Mass.: Harvard University Press.

Norwood, Janet L. 1990 ASA sesquicentennial. *The American Statistician* 44(2):67.

Office of the Assistant Secretary for Planning and Evaluation 1981 *A Description of the Health Financing Model: A Tool for Estimation.* Cost

Copyright National Academy of Sciences. All rights reserved.

Estimating Group, Office of Planning and Evaluation/Health. Washington, D.C.: U.S. Department of Health and Human Services.

Office of Management and Budget 1986 *Federal Statistics: A Special Report on the Statistical Programs and Activities of the United States Government, Fiscal Year 1987.* Office of Information and Regulatory Affairs. Washington, D.C.: Executive Office of the President.

1990 *Statistical Programs of the United States Government—Fiscal Year 1990.* Office of Information and Regulatory Affairs. Washington, D.C.: Executive Office of the President.

Office of Technology Assessment 1989a *A Comparison of OTA's and the Economic Policy Council's Recommendations for Improving Economic Statistics.* Staff paper. Energy and Materials Program. Washington, D.C.: U.S. Government Printing Office.

1989b *Statistical Needs for a Changing US. Economy.* Background paper. Washington, D.C.: U.S. Government Printing Office.

Orcutt, Guy H. 1957 A new type of socioeconomic system. *Review of Economics and Statistics* 80:1081-1100.

1986 Views on microanalytic simulation modeling. Pp. 9-29 in G. H. Orcutt, J. Merz, and H. Quinke, eds., *Microanalytic Simulation Models to Support Social and Financial Policy.* Amsterdam: Elsevier Science Publishers.

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

Orcutt, Guy H., Glazer, Amitai, Harris, Robert, and Wertheimer, Richard, II 1980 Microanalytic modeling and the analysis of public transfer policies. Pp. 81-106 in Robert H. Haveman and Kevin Hollenbeck, eds., *Microeconomic Simulation Models for Public Policy Analysis. Vol. I, Distributional Impacts.* New York: Academic Press.

Panel on Statistics for an Aging Population 1986 Inventory of data sets related to the health of the elderly. Pp. 241-661 in *Statistical Policy for an Aging America.* Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, D.C.: U.S. Government Printing Office.

Pappas, Norma Gavin 1980 A Microsimulation Model of the Medicaid Program. Draft documentation. Mathematica Policy Research, Inc., Washington, D.C.

Peat Marwick 1989 State Revenue Estimating Practices: Fall 1989 Survey Results. Peat Marwick, Washington, D.C.

Pechman, Joseph A. 1965 A new tax model for revenue estimating. In *Proceedings of a Symposium on Federal Taxation.* Washington, D.C.: American Bankers' Association.

1985 *Who Paid the Taxes, 1966-1985.* Washington, D.C.:Brookings Institution.

Policy Users Panel 1988 Discussion by panelists Rudolph G. Penner, Roger B. Porter, Charles L. Schultze, and Ian A. Stewart at 50th Anniversary Conference of the Conference on Research in Income and Wealth, Washington, D.C.

Copyright National Academy of Sciences. All rights reserved.

Primus, Wendell E. 1989 Children in poverty: A committee prepares for an informed debate. *Journal of Policy Analysis and Management* 8(1):23-34.

Rao, A. V., Lindsey, Q. W., Bhavsar, R. C., Shah, B. V., Horvitz, D. G., and Batts, J. R. 1973 The evaluation of four alternative family planning programs for POPLAND, a less developed country. Pp. 261-304 in Bennett Dyke and Jean Walters MacCluer, eds., *Computer Simulation in Human Population Studies.* New York:Academic Press.

Redfern, P. 1987 *A Study of the Future of the Census of Population: Alternative Approaches.* Luxembourg: Office for Official Publications of the European Communities.

Reeves, Jaxk H. 1987 Projection of number of kin. Pp. 228-248 in John Bongaarts, Thomas K. Burch, and Kenneth W. Wachter, eds., *Family Demography: Methods and Their Application.* Oxford: Clarendon Press.

Rivlin, Alice M., and Wiener, Joshua M., with Hanley, Raymond J., and Spence, Denise A. 1988 *Caring for the Disabled Elderly: Who Will Pay?.* Washington, D.C.: Brookings Institution.

Robinson, Gregory 1990 Preliminary Estimates of the United States Resident Population on April 1, 1990 Based on Demographic Analysis. Paper presented to the Census Advisory Committee of the American Statistical Association and the Census Advisory Committee on Population Statistics. Bureau of the Census, Washington, D.C.

Robinson, William H. 1989 Policy analysis for Congress: Lengthening the time horizon. *Journal of Policy Analysis and Management* 8(1):1-9.

Rovner, Julie 1988 Human services: Congress clears overhaul of welfare system. *Congressional Quarterly* October:2699-2701.

Rowe, G. 1990 Event history analysis of marriage and divorce in Canada. In *Proceedings of Statistics Canada Symposium on Analysis of Data in Time.* Ottawa: Statistics Canada.

Ruggles, Patricia, and Michel, Richard C. 1987 *Participation Rates in the Aid to Families with Dependent Children Program: Trends for 1967 Through 1984.* Washington, D.C.:The Urban Institute Press.

Russakoff, Dale 1990 Numbers people, in the crunch. *The Washington Post* October 4: A25.

Samuelson, Robert J. 1990 We just lost $200 billion. *The Washington Post* July 26: A27.

Shipp, P. Royal 1980 Discussion. Pp. 75-80 in Robert H. Haveman and Kevin Hollenbeck, eds., *Microeconomic Simulation Models for Public Policy Analysis. Vol. 1, Distributional Impacts.* New York: Academic Press.

1989 Anticipating future congressional action: Designing a new retirement income system. *Journal of Policy Analysis and Management* 8(1):10-22.

Shoven, John B., and Whalley, John 1984 Applied general-equilibrium models of taxation and international trade: An

Copyright National Academy of Sciences. All rights reserved.

introduction and survey. *Journal of Economic Literature* 13(September): 1007-1051.

Slemrod, Joel 1985 A general equilibrium model of taxation that uses micro-unit data: With an application to the impact of instituting a flat-rate income tax. Pp. 221-252 in John Piggott and John Whalley, eds., *New Developments in Applied General Equilibrium Analysis.* Cambridge, England: Cambridge University Press.

Smith, James E. 1987 The computer simulation of kin sets and kin counts. Pp. 249-266 in John Bongaarts, Thomas K. Burch, and Kenneth W. Wachter, eds., *Family Demography: Methods and Their Application.* Oxford: Clarendon Press.

Snipper, Reuben 1989 Estimating the Budgetary and Caseload Impacts of Mandating the AFDC-UP Program Nationwide: A Case Study of Microsimulation. Paper prepared for the Panel to Evaluate Microsimulation Models for Social Welfare Programs. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, Washington, D.C.

Social Welfare Research Institute No date Multi-Regional Policy Impact Simulation (MRPIS) Model, Level 3.0: Analyst's Guide. Boston College, Chestnut Hill, Mass.

Statistics Canada 1989 *SPSDIM Introductory Manual.* Ottawa: Statistics Canada.

Statistics of Income Division No date General Description Booklet for the 1987 Individual Public-Use Tax File. Internal Revenue Service, U.S. Department of the Treasury, Washington, D.C.

Strauss, Robert P. 1989 Micro-simulation Models and Taxpayer Behavior: Understanding Current Law and Its Alternatives. Paper prepared for the Panel to Evaluate Microsimulation Models for Social Welfare Programs. Carnegie Mellon University, Pittsburgh, Pa.

Subcommittee on Federal Longitudinal Surveys 1986 *Federal Longitudinal Surveys.* Statistical Policy Working Paper 13. Federal Committee on Statistical Methodology, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget. Washington, D.C.: Executive Office of the President.

Subcommittee on Matching Techniques 1980 *Report on Exact and Statistical Matching Techniques.* Statistical Policy Working Paper 5. Federal Committee on Statistical Methodology, Office of Federal Statistical Policy and Standards. Washington, D.C.: U.S. Department of Commerce.

Sy, Karen J., and Robbin, Alice 1990 Federal statistical policies and programs: How good are the numbers? *Annual Review of Information Science and Technology 25* (November).

Taussig, Michael K. 1980 Discussion. Pp. 111-113 in Robert H. Haveman and Kevin Hollenbeck, eds., *Microeconomic Simulation Models for Public Policy Analysis. Vol. 1, Distributional Impacts.* New York: Academic Press.

Triplett, Jack E. 1990 The Federal Statistical System's Response to Emerging Data Needs. Bureau of Economic Analysis, Washington, D.C.

Turek-Brezina, Joan 1988 Proposal for an Assessment of Nearly Two Decades of Social Welfare Microsimulation Modeling. Office of the Assistant Secretary for Planning and

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Evaluation, U.S. Department of Health and Human Services, Washington, D.C.

U.S. Department of the Treasury 1988a Afterthoughts and Future Directions. Workshop on Microsimulation Modeling, May 10, 1988. Statistics of Income Division,Internal Revenue Service, Washington, D.C.

1988b Selected Statistical Background Papers. Workshop on Microsimulation Modeling, May 10, 1988. Statistics of Income Division, Internal Revenue Service, Washington, D.C.

U.S. House of Representatives 1985 *Children in Poverty.* Prepared for the Ways and Means Committee by the Congressional Research Service and the Congressional Budget Office. Washington, D.C.: U.S. Government Printing Office.

1990 *Overview of Entitlement Programs: 1990 Green Book.* Background Material and Data on Programs Within the Jurisdiction of the Committee on Ways and Means. Washington, D.C.: U.S. Government Printing Office.

Vaughan, Denton R. 1988 Reflections on the income estimates from the initial panel of the Survey of Income and Program Participation. Pp. 333-414 in *Individuals and Families in Transition: Understanding Change Through Longitudinal Data.* Papers presented at the Social Science Research Council Conference in Annapolis, Md. Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Wachter, Kenneth W. 1987 Microsimulation of household cycles. Pp. 215-227 in John Bongaarts, Thomas K. Burch, and Kenneth W. Wachter, eds., *Family Demography: Methods and Their Application.* Oxford: Clarendon Press.

Wachter, Kenneth W., with Hammel, Eugene A., and Laslett, Peter 1978 *Statistical Studies of Historical Social Structure.* New York: Academic Press .

Wachter, Kenneth W., and Straf, Miron L., eds. 1990 *The Future of Meta-Analysis.* Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council. New York: Russell Sage Foundation.

Wallman, Katherine K. 1986 Federal statistics in the FY 1987 budget. Pp. 257-266 in Intersociety Working Group, *AAAS Report XI: Research & Development, FY 1987.* Washington, D.C.: American Association for the Advancement of Science.

1988 The Statistical System Under Stress: Framing an Agenda for Success. Council of Professional Associations on Federal Statistics, Alexandria,Va.

Webb, Randall L., Michel, Richard C., and Bergsman, Anne B. 1990 The historical development of the Transfer Income Model (TRIM2). Pp. 33-76 in Gordon H. Lewis and Richard C. Michel, eds., *Microsimulation Techniques for Tax and Transfer Analysis.* Washington, D.C.: The Urban Institute Press.

Webb, Randall L., Hager, Clara, Murray, Douglas, and Simon, Eric 1982 *TRIM2 Simulation Modules.* 2 vols. (A-M and N-Z). March 1982 plus updates. Washington, D.C.: The Urban Institute Press.

Webb, Randall L., Bergsman, Anne, Hager, Clara, Murray, Douglas, and Simon, Eric 1986 *TRIM2 Reference Manual: The Framework for Microsimulation.* Working Paper 3069-01. Washington, D.C.: The Urban Institute Press.

Copyright National Academy of Sciences. All rights reserved.

Weinberg, Daniel H. 1987 The distributional implications of tax expenditures and comprehensive income taxation. *National Tax Journal* 40(2):237-253.

Weiss, Carol H. 1977 *Using Social Research in Public Policy Making.* Lexington, Mass.:

Lexington Books, D.C. Heath and Company. 1989 Congressional committees as users of analysis. *Journal of Policy Analysis and Management* 8(3):411-431.

Wertheimer, Richard II Zedlewski, Sheila R., Anderson, Joseph, and Moore, Kristin 1986 DYNASIM in comparison with other microsimulation models. Pp. 187-206 in G. H. Orcutt, J. Merz, and H. Quinke, eds., *Microanalytic Simulation Models to Support Social and Financial Policy.* Amsterdam: Elsevier Science Publishers.

Whalley, John 1988 Lessons from general equilibrium models. Pp. 15-57 in Henry J. Aaron, Harvey Galper, and Joseph A. Pechman, eds., *Uneasy Compromise: Problems of a Hybrid Income-Consumption Tax.* Washington, D.C.: Brookings Institution.

Whiteman, David 1985 The fate of policy analysis in congressional decision making: Three types of use in committee. *Western Political Quarterly* June:294-311.

Wilensky, Gail R. 1970 An income transfer computational model. In *The President's Commission on Income Maintenance Programs: Technical Studies.* Washington, D.C.: U.S. Government Printing Office.

1982 Modeling Changes in the Financing of Health Care Programs. National Center for Health Services Research, Washington, D.C.

Wisconsin Department of Revenue 1979 *The Wisconsin Tax Burden Study.* Madison: State of Wisconsin.

Wolfson, Michael C. 1989a Divorce, homemaker pensions, and life cycle analysis. *Population Research and Policy Review* 8:25-54.

1989b POHEM: A New Approach to the Estimation of Health Status Adjusted Life Expectancy. Paper presented to the First Workshop of the International Network on Health Expectancy, Quebec City.

1991 A system of health statistics: Toward a new conceptual framework for integrating health data. *Review of Income and Wealth* 1(March):81-104.

Wolfson, M., and Rowe, G. 1990 Biased Divorce: Validation of Marital Status Life Tables and Microsimulation Models. Unpublished technical report. Statistics Canada, Ottawa.

Yett, Donald E., Drabek, Leonard, Intriligator, Michael D., and Kimbell, Larry J. 1980 The HRRC health care sector simulation model. Pp. 321-269 in Robert H. Haveman and Kevin Hollenbeck, eds., *Microeconomic Simulation Models for Public Policy Analysis. Vol. 1, Distributional Impacts.* New York: Academic Press.

Zedlewski, Sheila R. 1990 The development of the Dynamic Simulation of Income Model (DYNASIM). Pp. 109-136 in Gordon H. Lewis and Richard C. Michel, eds., *Microsimulation Techniques for Tax and Transfer Analysis.* Washington, D.C.: The Urban Institute Press.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Biographical Sketches of Panel Members and Staff

ERIC A. HANUSHEK *(Chair)* is professor of economics and political science at the University of Rochester. During 1983-1985, he served as deputy director of the Congressional Budget Office. He previously held academic appointments at Yale University and the U.S. Air Force Academy and governmental appointments at the Cost of Living Council and the Council of Economic Advisers. He is a past president of the Association for Public Policy Analysis and Management. His primary academic interests have involved applied public finance and public policy analysis with special reference to schooling and aspects of income determination. He received a B.S. degree from the United States Air Force Academy and a Ph.D. degree in economics from the Massachusetts Institute of Technology.

DAVID M. BETSON is an associate professor in the Department of Economics at the University of Notre Dame. He was previously a research associate at the Institute for Research on Poverty at the University of Wisconsin and a staff economist at the U.S. Department of Health, Education, and Welfare. His research has dealt with the impact of federal tax and transfer programs on the economy and the distribution of income. He received a B.A. degree from Kalamazoo College and a Ph.D. degree from the University of Wisconsin.

LYNNE BILLARD is professor of statistics and associate to the dean of the Franklin College of Arts and Sciences at the University of Georgia. She was formerly head of the statistics and computer science department at the university and has held faculty positions and visiting positions at other U.S. universities and in England and Canada. Her current research interests include time series,

Copyright National Academy of Sciences. All rights reserved.

sequential analysis, stochastic processes, and AIDS. She is a fellow of the American Statistical Association and the Institute of Mathematical Statistics and an elected member of the International Statistical Institute. She has held many professional offices, including president of the Biometric Society, Eastern North American Region, and associate editor and associate book editor for the *Journal of the American Statistical Association.* She is currently a member of the International Council of the Biometric Society and the Council of the International Statistical Institute. She received a B.S. honors degree in mathematics and statistics and a Ph.D. degree in statistics from the University of New South Wales, Australia.

CONSTANCE F. CITRO *(Study Director)* is a member of the staff of the Committee on National Statistics. She is a former vice president and deputy director of Mathematica Policy Research, Inc., and was an American Statistical Association/National Science Foundation (NSF)/Census research fellow at the Bureau of the Census. For the Committee on National Statistics, she has served or is serving as study director for the Panel to Study the NSF Scientific and Technical Personnel Data System, the Panel on Decennial Census Methodology, the Panel on Statistics on Supply and Demand for Precollege Science and Mathematics Teachers, and the Panel to Evaluate the Survey of Income and Program Participation. Her research has focused on the usefulness and accessibility of large, complex microdata files, as well as analysis related to income measurement and demographic change. She is a fellow of the American Statistical Association. She received a B.A. degree from the University of Rochester and M.A. and Ph.D. degrees in political science from Yale University.

MICHAEL L. COHEN is an assistant professor in the School of Public Affairs at the University of Maryland. He was formerly a visiting lecturer at the Department of Statistics, Princeton University, and a research associate at the Committee on National Statistics. His general area of research is the use of statistics in public policy, and his current research concerns whether and how to adjust the census for undercoverage. He received a B.S. degree in mathematics from the University of Michigan and M.S. and Ph.D. degrees in statistics from Stanford University.

SHELDON DANZIGER is professor of social work and public policy and faculty associate at the Institute of Public Policy Studies at the University of Michigan, where he also directs the Research and Training Program on Poverty, the Underclass, and Public Policy. From 1983 to 1988 he was director of the Institute for Research on Poverty, professor of social work, and Romnes Faculty Fellow at the University of Wisconsin. He is the coeditor of several books on poverty and public policy and the author of numerous scholarly articles on poverty, income inequality, and social welfare programs and policies. He is a member of the Committee on Child Development Research and Public Policy

Copyright National Academy of Sciences. All rights reserved.

at the National Research Council and of the Social Science Research Council's Committee for Research on the Urban Underclass. He was a member of the Panel on Employment, Income, and Occupations for the National Research Council's Committee on the Status of Black Americans. He received a B.A. degree from Columbia University and a Ph.D. degree in economics from the Massachusetts Institute of Technology.

EUGENE P. ERICKSEN is a professor and chair of the Department of Sociology at Temple University. He is currently cochair of the Special Advisory Panel on the 1990 Census, which was convened by the U.S. Secretary of Commerce for advice on the possibility of adjusting the census results for the differential undercount. His research interests focus on the estimation of local-area characteristics from sample data, survey sampling, and problems of survey measurement in general. He received a B.S. degree in mathematics from the University of Chicago and an M.S. degree in statistics and a Ph.D. degree in sociology from the University of Michigan.

THOMAS J. ESPENSHADE is professor of sociology and faculty associate of the Office of Population Research at Princeton University. He was formerly director of the Program in Demographic Studies and senior fellow at the Urban Institute in Washington, D.C. His research interests have included the cost of raising children; changes in U.S. family structure; the demographic and economic consequences of slowing population growth in developed countries; and the demographic, economic, and social implications of immigration to the United States. He has written widely on contemporary U.S. immigration and immigrant policy and has testified before Congress on changes in U.S. immigration policy. His current research is related to developing models of undocumented migration to the United States; the role of undocumented migrants in California agriculture; determinants of public attitudes toward undocumented migrants and illegal migration; and proposed reforms of U.S. policy toward legal immigration. He received a B.A. degree in economics from the College of Wooster and a Ph.D. degree in economics and demography from Princeton University.

HARVEY GALPER is a principal in the Policy Economics Group of KPMG Peat Marwick in Washington, D.C. His prior positions include senior fellow at the Brookings Institution, senior public finance resident at the Advisory Commission on Intergovernmental Relations, and director of the Office of Tax Analysis of the U.S. Department of the Treasury. He has also served on the staffs of the Urban Institute and the Board of Governors of the Federal Reserve System and has taught at Dartmouth College, the University of California at Berkeley, Yale University, and the Georgetown University Law Center. He is a past member of the board of directors of the National Tax Association-Tax Institute of America and of the Advisory Group of the Commissioner of the

Copyright National Academy of Sciences. All rights reserved.

Internal Revenue Service, and currently serves on the editorial board of the *National Tax Journal.* He received a B.A. degree from Dartmouth College and M.A. and Ph.D. degrees, both in economics, from Yale University.

LOUIS GORDON is professor of mathematics at the University of Southern California. Previously, he was statistician at the ALZA Corporation and at the Energy Information Administration of the U.S. Department of Energy. His interests are in the analysis of computation-intensive statistical algorithms. He is a member of the Committee on National Statistics at the National Research Council and a fellow of the Institute of Mathematical Statistics. He received a B.S. degree in mathematics from Michigan State University and a Ph.D. degree in statistics from Stanford University.

KEVIN M. HOLLENBECK is senior economist at the W.E. Upjohn Institute for Employment Research. He was formerly with Mathematica Policy Research, Inc., where he developed a number of components of a microsimulation model. He has also been a consultant to a number of federal agencies to conduct studies with and to evaluate specific microsimulation models. His current research interests are in analysis and evaluation of education and training programs. He received a B.S. degree in mathematics from Michigan State University and M.S. and Ph.D. degrees in economics from the University of Wisconsin.

GORDON H. LEWIS is associate professor of sociology at Carnegie Mellon University, with appointments in the School of Urban and Public Affairs and in the Department of Engineering and Public Policy. His major research interest involves income transfers, especially modeling of the interactions among transfer programs; his recent work has focused substantively on areas of child support, tax rebates, and marriage policy. Other research interests include organizational design, organizational management, and individual decision making. He received both undergraduate and graduate degrees in sociology from Stanford.

ROBERT MOFFITT is a professor of economics at Brown University. He is a fellow of the Institute for Research on Poverty at the University of Wisconsin and a research associate of the National Bureau of Economic Research. His research interests cover the economics of welfare and other government benefit programs, statistical methods for the evaluation of public policy programs, and the study of labor force and demographic trends of American women. He has served on a committee for the Social Science Research Council on the Survey of Income and Program Participation and on the U.S. Department of Labor's National Advisory Panel for the National Longitudinal Study, and has been a consultant to the National Research Council's Panel on the Evaluation of AIDS Programs. He is chief editor of the *Journal of Human Resources.* He received a B.A. degree from Rice University and M.A. and Ph.D. degrees from Brown University.

Copyright National Academy of Sciences. All rights reserved.

CHRISTINE M. ROSS, who served as research associate during the first year of the study, is currently an economist with Mathematica Policy Research, Inc., in Princeton, N.J. Before working with the panel, she was associate analyst with the Human Resources and Community Development Division of the Congressional Budget Office, where she worked with microsimulation models on food stamp and Aid to Families with Dependent Children program policy. She received a B.A. degree in economics from Williams College and M.A. and Ph.D. degrees in economics from the University of Wisconsin.

GAIL R. WILENSKY, who served as a member of the panel during 1989, is now administrator of the Health Care Financing Administration of the U.S. Department of Health and Human Services (HHS). Previously, she was vice president for the division of health affairs of Project Hope and a senior research manager at the HHS National Center for Health Services Research where she designed and directed the analysis of the National Medical Care Expenditure Survey. She has served on the faculties of the University of Michigan and George Washington University and held a senior research appointment at the Urban Institute. She has published extensively on health economics and health policy. She is a member of the Institute of Medicine, and was a member of the Physician Payment Review Commission and the Health Advisory Committee of the General Accounting Office. She received a Ph.D degree in economics from the University of Michigan.

MICHAEL C. WOLFSON is director general of the Analytical Studies Branch, Statistics Canada. At Statistics Canada he led the group that developed a personal computer-based public-use microsimulation model of Canada's income tax/transfer system, the Social Policy Simulation Database and Model (SPSD/M). More recently, he has been engaged in a fundamental review of the health statistics program. From 1974 to 1985 he held a variety of positions with the Canadian federal public service, including the Treasury Board Secretariat, Department of Finance, and Privy Council Office, with responsibilities in the areas of program evaluation and tax and pension policy. He is also an appointed fellow of the Canadian Institute for Advanced Research in the Program in Population Health. He received a B.Sc. degree in computer science and economics from the University of Toronto and a Ph.D. degree in economics from Cambridge University.

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Index

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.

Improving Information for Social Policy Decisions -- The Uses of Microsimulation Modeling: Volume I, Review ...

Copyright National Academy of Sciences. All rights reserved.

T

Copyright National Academy of Sciences. All rights reserved.

Copyright National Academy of Sciences. All rights reserved.