Policy Analysis Through Microsimulation: The STATS Model*

by Bernard Wixon, Benjamin Bridges, Jr., and David Pattison

This article gives a nontechnical introduction to a microsimulation model developed by the Social Security Administration to consider the effects of tax and benefit proposals on the economic status of population groups. This model, the Simulated Tax and Transfer System (STATS) model, uses information from a nationally representative survey of households to calculate taxes or benefits for thousands of persons or families, case by case. Then the model aggregates across individual cases to produce estimates for economic and demographic groups of interest to policymakers. Groups are defined by such characteristics as family income, age, sex, and marital status. The article explains the model's microsimulation approach and its emphasis on the economic status of population groups. It describes the procedures for simulating taxes and benefits and then illustrates the use of the model by considering a proposal to change the way Social Security benefits are taxed. The article also discusses the primary data source and the extent to which it limits the types of estimates that the model can produce.

How would elderly widows be affected if lawmakers were to enact a specified change in the income tax treatment of Social Security benefits? How many retired couples would fall into poverty if scheduled cost-of-living adjustments to Social Security benefits were delayed? To what extent could increases in the Social Security tax be offset for low-income workers by liberalizing the earned income tax credit? Policy simulation allows analysts to answer such questions—that is, to consider the effects not only of enacted policies but of envisioned alternatives.

The Social Security Administration's (SSA's) Simulated Tax and Transfer System (STATS) model focuses on how policy alternatives affect population groups, rather than on overall costs or savings.[1] A proposal's effects on population groups—its distributional effects—are often critical to policymakers in assessing proposals to change tax or benefit programs. The STATS model is used to estimate the changes in taxes or benefits that would occur following enactment of a proposal. These changes are then examined in light of the demographic traits and economic status of the families affected.

The model uses a microsimulation approach. Information on individual households is used to recalculate taxes or cash benefits, case by case. In a sense, each person or family undergoes a simplified version of the tax filing or benefit calculation process, except in two respects. First, the information used by the model to determine taxes or benefits is collected in a nationally representative survey of households, rather than on administrative forms. Second, taxes or benefits are often calculated according to a policy proposal, instead of current law.

The STATS model has been developed by analysts within the SSA.[2] It can be used to evaluate several

*Division of Economic Research, Office of Research and Statistics, Office of Policy, Social Security Administration. The authors wish to acknowledge the thoughtful comments of a number of present and former colleagues and to thank Delores Johnson and Carolyn Fulmore for their assistance.

[1] While the model can be used to produce cost or revenue estimates, their reliability is limited by problems associated with the underlying survey data.

[2] The earliest version of the model was developed in the 1970's by analysts within SSA's Division of Economic Research, under the direction of Dorothy S. Projector. More recently, the development of the model has been directed by Benjamin Bridges, Jr. For a brief discussion of the model's development, see Roen (1982a). A number of microsimulation models of the same general type are in use elsewhere, both inside and outside the Federal Government.

tax or cash benefit programs separately or in combination, including Social Security payroll taxes, some proposals relating to Social Security benefits, and Federal individual income taxes. Current programs or proposed alternatives can be evaluated for a single year or over a period of several years. The data used represent the characteristics of the U.S. population as observed in the recent past or as projected several years into the future.

This article provides a nontechnical introduction to the model. Section I describes the model's microsimulation approach, its focus, and the survey data used. The steps of a typical simulation are summarized in sections II and III. A sample simulation, involving the income tax treatment of Social Security benefits, is presented in section IV. Section V discusses limitations and prospects for improvement and, finally, section VI summarizes the model's key features.

I. The Approach, Focus, and Data Source

Approach: Microsimulation

Those who plan tax or benefit policy must evaluate legislative alternatives that have not been enacted and, therefore, whose effects cannot be observed. Administrative or survey data, taken alone, are insufficient. Observations from these sources, no matter how detailed, reflect policies in effect when the data were collected. Policy simulation, by comparison, offers a means of evaluating alternatives to current policies by estimating effects of *changes* in the way taxes or benefits are determined.[3]

The STATS model and other microsimulation models of the same general type estimate changes in taxes or benefits for persons or families, one by one, based on their observed economic and demographic traits.[4] By operating at the individual level, the model is able to utilize data on the combinations of individual traits that determine various taxes and benefits, thereby producing more reliable distributional estimates of changes in these taxes and benefits than those based on aggregated data.[5] Then, by summing over many thousands of persons or families, the model can derive summary measures for population groups of interest to policymakers. Operating at the individual level gives the model the ability to easily tailor its summary output to the needs of policymakers and others.

One of the proposals that has been simulated by the model—a particular proposal to change the income tax treatment of Social Security benefits—illustrates the microsimulation approach. Under current law, up to 50 percent of Social Security benefits received by taxpayers whose incomes exceed certain thresholds are included in the income tax base. The thresholds ($25,000 for single taxpayers and $32,000 for married taxpayers filing jointly) are not indexed by the inflation rate. The specific proposal would decrease (or increase) these thresholds by specified amounts. The STATS model's income tax calculation procedure (a complex and lengthy computer program) would estimate tax liability for each tax unit within each family, following a step-by-step procedure dictated by current statutes as modified by the proposal. The procedure uses a large number of individual economic and demographic traits (for example, income amounts by source, marital status, number of children, and age). Then the model produces various summary measures. Results could be shown for those in various after-tax income classes, for those with tax increases, and for those whose benefits become taxable.

The advantages of processing thousands of records individually could not be realized were it not for obvious advances in high-speed data processing. These advances allow analysts to use more evidence in evaluating policy options within short time frames.

Focus: The Current Economic Status of Population Groups

Policymakers, in evaluating tax or benefit options, often consider the demographic groups affected and their current economic status. The STATS model's focus reflects this distributional perspective.[6] For example, policymakers targeted taxation of benefits for a particular group—beneficiaries with high incomes—while others remained exempt. In other instances, concern has been shown for vulnerable groups such as aged widows. The model permits a focus on such groups and, in fact, can narrow the focus by also considering economic status (for example, the model can show how a proposal affects aged widows with low family income).

Economic status or economic well-being depends on the extent to which financial resources are available to

[3] Because the data used by the model do not include tax liability, the STATS model has also been used to enrich the original data by adding estimates of tax liability. This permits analysis of current tax policies. Simulations can be used in this way to obtain information that is, in principle, observable but is not available in a given data source.

[4] Although the STATS model estimates the changes in taxes or benefits implied by proposals, behavioral adjustments to such proposals (for example, a decision to retire or apply for benefits) are usually not simulated. Ad hoc assumptions about such adjustments have been used on occasion.

[5] The 1970 Projector study is a distributional study of children's allowances and income-tested supplement proposals that used grouped data rather than individual data. The study used survey data for 15 family types distributed by family money income classes. This study made the shortcomings of the aggregate data approach very apparent, prompting the development of the STATS model.

[6] The STATS model typically examines the effects of a proposal from the perspective of the current income of those affected. Other models take a lifetime perspective; they might, for example, compare lifetime benefits with lifetime taxes.

meet needs. A change in taxes or cash benefits increases or decreases the resources available, thus affecting economic status.

Although it is difficult to measure economic status in a straightforward way, there are some useful principles. Income from all sources is a better measure of resources than labor income alone. Asset information would also be helpful, though it is not available in the data used by the STATS model. In addition, a family perspective is preferable to an individual perspective because an individual's economic status depends both on the needs of other family members and on their contributions to family income. Hence, in grouping families that are roughly equivalent in terms of economic status, family income and family composition are basic. If two families are otherwise similar, the one with more income or fewer dependents is better off financially. As a result, for each group considered, a proposal's effects are usually exhibited by family income and, in many cases, by family composition.

Data Source: The March Current Population Survey

The primary source of information for the STATS model is the March supplement to the Current Population Survey (CPS). The CPS is a nationally representative file based on interviews of more than 50,000 households containing more than 100,000 persons. The sample is large enough for reliable estimates on most groups of interest for Social Security policy analysis. It represents the civilian noninstitutionalized population of the United States living in housing units and members of the Armed Forces living in civilian housing. The Bureau of the Census conducts the survey every March and the resulting file is available within a few months; hence, the CPS is a dependable source of up-to-date information. It provides data on the relationships among persons within each household, prior-year employment, prior-year income from various sources for those aged 15 or older, and demographic traits such as age, sex, and race. The CPS is the main source of government statistics on income, poverty, and labor-force characteristics.[7]

Because the CPS represents the general population and includes rich detail on income and demographic traits, it offers special advantages for simulating taxes and cash benefits:

Multiple programs. The survey data can be used to analyze several major tax or cash benefit programs separately or in combination. This feature is particularly useful when a proposal affects more than one program and policymakers wish to estimate net effects (for example, suppose an increase in the Social Security payroll tax were to be offset for low-wage workers by a liberalization of the earned income tax credit). Also, for such proposals, generating estimates for each program using the same population improves the consistency of the estimates.

New entrants. When considering a proposal that would increase the number of those liable for taxes or eligible for benefits, a file that represents the general population can often provide estimates for those not previously liable for taxes or eligible for benefits.

Comparison with general population. In considering how a proposal affects economic status, it is sometimes helpful to compare the economic status of affected groups with the rest of the general population (such as comparing poverty rates or income distributions). Such comparisons are usually more convenient and accurate if based on a file that includes the total population.

II. Simulation Step One: Estimating Taxes and Cash Benefits

A simulation has two major steps. In step one, taxes or benefits are estimated for persons within each family according to proposed or enacted statutes. Step two involves summarizing the effects of a program or proposed alternative on population groups.

Conceptually, step one is similar to having persons or subgroups within each family file taxes or undergo benefit calculations, but the calculations are based on information collected in CPS interviews. This step involves two tasks. First, household members are grouped into units appropriate for the program under consideration. Second, taxes or benefits are estimated for each unit in turn. In some instances, assumptions are employed to simplify tax or benefit provisions or to fill in data not available in the CPS.

Forming Units

The information in the CPS is collected in interviews of households. Interviewing those who live together is a cost-effective way of gathering information about persons and families. However, tax and benefit statutes mandate that family members be grouped in other ways to determine taxes or benefits. How they are grouped varies from program to program, reflecting differences in program objectives. Under the Social Security payroll tax, individual covered workers are considered liable for tax; under the income tax, liability is assessed to family members grouped, in large part, according to financial dependency. When simulating Social Security benefits, the model considers covered workers and their dependents.

Family members considered liable for taxes or entitled to benefits as a group under a given program are

[7] U.S. Department of Commerce (1987) includes documentation of the CPS.

referred to as a "program unit." CPS information on the relationships of members within a family, along with information on age, marital status, income, education, and labor-force behavior, is used to group members of CPS families into program units appropriate for the programs simulated by the STATS model.[8]

Estimating Taxes and Benefits for Each Unit

Once program units have been formed, taxes or cash benefits are estimated unit by unit. The information needed to determine taxes or benefits is not always reported in the CPS. In such instances, the necessary information is inferred from CPS data or, in a few cases, taken from other data sources. For some proposals, using administrative data would be more straightforward, at least for those already paying taxes or receiving benefits.[9] The CPS, however, allows analysts to consider policy options against the background of the income and composition of affected families, as well as their demographic traits. Because the CPS represents the general population, a number of programs can be simulated alone or in combination.

Tax Estimation

Neither payroll taxes nor income taxes are reported in the CPS. Nonetheless, the CPS is suited for tax simulation because the period covered by the survey—the calendar year—coincides with the accounting period generally used to calculate liability under both taxes. The CPS includes information on annual earnings and annual income from other sources. These income data, combined with CPS information on the relationships of household members and demographic and labor-force characteristics, provide the information necessary to simulate both taxes.

The simplest application involves analyzing existing programs rather than proposed alternatives. In this case, the payroll tax (or income tax) is calculated for each worker (or tax unit) according to currently enacted legislation. This permits an examination of the distribution of liability—under either tax or both simultaneously—among selected groups for a single year or for several years.

By comparison, proposals to change current law are simulated by estimating taxes twice for each worker or tax unit. The first estimate is based on current statutes and the second on the proposed alternative. The estimates reflect tax liability under different statutes, but for the same time period.[10] The difference between the two estimates provides a measure of the effect of the proposal.

Payroll tax. The STATS model simulates Federal Insurance Contributions Act (FICA) and Self Employment Contributions Act (SECA) contributions from wage and self-employment income.[11] For each adult, the CPS includes the necessary information on wage, salary, and self-employment income.

Some workers, however, do not work in covered employment and therefore have no payroll tax liability; the CPS does not indicate whether or not a worker's earnings were from covered employment. To deal with this problem, Social Security administrative data are used to estimate the coverage status of each worker represented in the CPS, based on the worker's industry and wage level.

For each covered worker, the employee share of the payroll tax is estimated based on that worker's covered earnings and the current (or proposed) statutory tax rate and taxable maximum. Estimating the employer share of the tax is more complex. Analysts often assume that employers shift the burden of the employer share to employees by paying lower wages. The STATS model can be used to estimate how the burden of the employer tax is distributed under this assumption.

Federal personal income tax. For each family member, gross income is estimated from CPS income information on earnings, interest, dividends, rent, and other selected income types. These data on gross income, along with data on the relationships of family members, are then used to group family members into tax filing units. There may be more than one such unit in a family—for example, the family may include working students. For each unit, determinations are made as to which units are required to file, which units file joint returns, and which persons can be considered as tax dependents of filers. For each tax filing unit, adjusted gross income is calculated and the itemized or standard deduction is estimated with the help of Internal Revenue Service (IRS) data from tax returns.[12] Income tax liability is then determined from the appropriate tax schedule. The earned income tax credit is also calculated for eligible units, including those not required to file.

Applications. The tax estimation procedures have been used to evaluate general revenue financing proposals involving simultaneous increases in income taxes

[8] Roen (1982b) documents unit formation for an earlier version of the model.

[9] Administrative data may yield more reliable estimates of changes in taxes or benefits or aggregate revenue effects if: the proposal does not involve new taxpayers or new beneficiaries; no new data (that is, data not already on the administrative file or that can be inferred from data on the file) are needed; and the proposal involves a single program.

[10] By holding other things constant, this procedure serves to isolate the effects of tax law changes.

[11] For a description of an earlier version of the estimation procedure, see Bridges and Johnston (1976).

[12] For documentation of the tax return data and summary statistics, see Internal Revenue Service (1986).

and decreases in payroll taxes. In addition, the payroll tax liability of families at various income levels has been compared with their income tax liability.[13] The income tax estimation procedure has also been used for several analyses of the taxation of Social Security benefits and to estimate the effects of the earned income tax credit.[14]

Benefit Estimation and Other Procedures

Social Security benefits. Social Security benefits, unlike Social Security taxes, are reported in the CPS. However, the CPS is not well-suited for dealing with many types of benefit proposals. A complete benefit simulation requires historical information on earnings, health, marital behavior, family composition, and program participation, whereas the CPS focuses on a single calendar year. Benefits under some proposals can be estimated if they are related to benefits under current law reported in the survey. The model has been used to simulate simple proportional and more complex nonproportional changes in the cost-of-living adjustments to benefits as well as changes in the earnings test.[15]

Other programs. The STATS model has a limited capability for dealing with financing proposals involving consumption-based taxes. The model can also be used to evaluate some types of proposals relating to the Supplemental Security Income (SSI) program.

Short-term projection. Policymakers are frequently interested in the effects of a proposal several years into the future. As the first step in evaluating such proposals, the most recent CPS data are projected so that the resulting population reflects expert projections of what the U.S. population will look like in the desired year with respect to variables such as aggregate income by source and the age/sex mix of the population. To achieve this result, the composition of the initial CPS population is altered. In effect, adjustments are made to the mix and number of households in the original population, as well as to the income from various sources of adults in the households, so that the resulting population exhibits the projected characteristics.[16]

III. Simulation Step Two: Summarizing Distributional Effects

In step two, summary statistics are formulated to show how a program or proposed alternative affects groups of interest to policymakers. A range of measures is possible and the measures can be varied from one simulation to the next. Program units can be summed, yielding family or household estimates, or decomposed to show effects on persons. Families—or whatever unit is selected—can be grouped by income or by demographic traits. Statistics are selected to summarize a proposal's effect on taxes, benefits, family income, or poverty status. Examples of such statistics are percentage change in benefits and change in taxes as a percentage of income.

The Role of Families

For program administration or for comparison with administrative data, policymakers may need estimates by program units, such as covered workers, tax units, or beneficiaries. To evaluate programs in other contexts, however, family estimates are often preferred. CPS detail on the relationships of household members offers the flexibility to estimate taxes or benefits using program units and then to exhibit results in terms of persons or families.

In most cases, family estimates are employed because of the emphasis on economic status. How well-off an individual is depends not only on his or her income, but on how income is shared among individuals. A family—defined in the CPS as household members related by blood, marriage, or adoption—typically pools the resources of its members to meet combined needs. Hence, measures based on the family, rather than the program unit or person, are less likely to misstate available resources or those supported by the resources.[17]

Population Groups

Most analysts of taxes and benefits recognize that measures of average effects for large groups may give too little detail. Such averages can be misleading if, among those represented by the average, there is substantial variation in income or in the treatment received under a given program.

For example, in evaluating how delays in cost-of-living adjustments (COLA's) to Social Security benefits would affect low-income beneficiaries, SSI recipients should be distinguished from those who do not receive SSI payments. The total incomes of SSI recipients are unaffected by Social Security COLA delays (assuming the SSI COLA is not deferred), while nonrecipients suffer at least temporary income losses.

[13] See Bridges (1981). Also, Johnston and Wixon (1978) consider changes in the payroll tax liability of families over time and Projector et al. (1974b) consider payroll tax relief plans.

[14] See, for example, Dye (1981).

[15] Some earlier work on benefits is reported in Projector et al. (1982).

[16] Projector et al. (1974a) and Millea (1982) document earlier versions of the projection procedure.

[17] In most cases, households include only one family. (Unrelated individuals are considered to be one-person families.) For multifamily households, the model assumes that in most cases the family offers a better approximation of the income-sharing group than does the household.

By use of information in the CPS, families or other units can be classified by a measure of the available resources (such as family income), and/or by demographic traits (such as family size, or the age, sex, race, or marital status of the head). Demographic traits are used to distinguish groups of special concern to policymakers (such as aged widows). Demographic and income information in the CPS are also employed to identify program-related categories (such as aged survivors, covered workers, or nonaged beneficiaries).

IV. Sample Simulation

In 1985, at the request of policymakers, the STATS model was used to simulate the distributional effects of seven proposals to change the income tax treatment of Social Security benefits and Railroad Retirement benefits. Tables 1 and 2 are condensed versions of tables produced from the simulation of one of the proposals.

Under the income tax law, up to 50 percent (the inclusion percentage) of Social Security and Railroad Retirement benefits received by taxpayers whose incomes exceeded certain thresholds was included in the tax base. The thresholds were $25,000 for single taxpayers and $32,000 for married taxpayers filing jointly. The proposal was to increase the inclusion percentage to 85 percent, but to leave the thresholds unchanged.[18]

The proposal was simulated using a 1983 CPS data file that had been adjusted for income underreporting and projected to 1986.[19] The changes in tax liability resulting from the proposal were estimated for each tax unit. Because the tables were to refer to families, the results for tax units were aggregated within families to get family-level variables; then family variables were aggregated across families.

Table 1 classifies families by income. Family income, which measures economic resources, includes nontaxable as well as taxable income receipts. For each income class, table 1 gives the numbers of affected families (for example, those with tax increases) and statistical measures that indicate the magnitude of the proposal's effects (for example, the average tax increase). This table is designed to provide answers to questions such as the following:

(1) What is the increase in the number of families with taxable benefits and how are those families distributed by income class? (The table shows that the number of families with taxable benefits increased by about 570,000. Of this number, 520,000 (more than 90 percent) had incomes in the $20,000-$39,999 range.)
(2) How many families experience tax increases and how are they distributed by income class? (Approximately 5 million experience tax increases; almost two-thirds of them have incomes of $40,000 or more.)
(3) How does the ratio of average tax increase to average benefit vary with income? (For families with tax increases, the ratio of average tax increase to

[18] Under the proposal, 85 percent of the benefit is added to the modified adjusted gross income, and 50 percent of the amount by which this sum exceeds the threshold is added to the tax base, up to a maximum of 85 percent of the benefit.

[19] The projection incorporates the February 1985 baseline aggregate economic assumptions of the Congressional Budget Office and 1984 population projections of SSA's Office of the Actuary.

Table 1.—Effects on Social Security beneficiary families of a proposal[1] to change the taxation of benefits, by family income class, 1986

Family income class	Total beneficiary families (in thousands)	Number of families (in thousands)			Families with tax increases			
		With taxed benefits under current law	With benefits becoming taxable under proposal	With tax increases	Average income tax under current law	Average increase	Increase as a percent of before-tax income	Increase as a percent of benefit
Total	24,691	4,487	571	5,057	$10,398	$662	1.2	8.7
Less than $20,000	14,770	0	0	0	...	...	...	...
$20,000-$24,999	2,047	13	63	76	2,202	53	.2	1.1
$25,000-$29,999	1,577	256	98	353	2,828	218	.8	3.7
$30,000-$34,999	1,302	475	312	787	3,209	240	.7	3.5
$35,000-$39,999	971	634	47	681	4,216	411	1.1	5.2
$40,000-$44,999	683	481	10	491	5,283	495	1.2	6.4
$45,000-$49,999	649	471	13	484	6,701	629	1.3	8.2
$50,000 or more	2,687	2,157	28	2,186	18,388	1,031	1.2	12.7

[1] The proposal is to increase the maximum percent of Social Security benefits included in the tax base from 50 percent to 85 percent; the threshold amounts remain unchanged.

Note: Dollar amounts in 1984 dollars.
Source: STATS model estimates based on the March 1984 CPS.

Table 2.—Effects on Social Security beneficiary families of a proposal[1] to change the taxation of benefits, by age of head, family size, and family income class, 1986

Family income class	Total beneficiary families (in thousands)	Families with tax increase		
		Number of families (in thousands)	Average tax increase	Tax increase as a percent of benefit
	Nonaged families[2]			
Total	4,294	671	$464	7.8
Less than $20,000	2,084	0	...	...
$20,000-$24,999	357	9	48	.8
$25,000-$29,999	351	34	187	2.8
$30,000-$34,999	277	119	158	3.2
$35,000-$39,999	229	81	288	5.2
$40,000-$44,999	195	80	391	6.3
$45,000-$49,999	176	68	404	8.4
$50,000 or more	625	282	724	10.8
	Aged 1-person families[2]			
Total	9,221	1,442	$579	10.1
Less than $20,000	7,318	0	...	...
$20,000-$24,999	516	66	53	1.2
$25,000-$29,999	316	308	227	3.9
$30,000-$34,999	247	245	411	6.9
$35,000-$39,999	170	170	590	10.5
$40,000-$44,999	122	122	639	11.3
$45,000-$49,999	99	99	833	13.5
$50,000 or more	433	433	924	16.0
	Aged 2-or-more person families[2]			
Total	11,176	2,944	$749	8.4
Less than $20,000	5,373	0	...	...
$20,000-$24,999	1,175	1	49	2.3
$25,000-$29,999	910	12	76	1.2
$30,000-$34,999	779	424	164	2.0
$35,000-$39,999	572	430	364	4.0
$40,000-$44,999	366	289	464	5.1
$45,000-$49,999	373	317	613	7.0
$50,000 or more	1,629	1,471	1,121	12.4

[1]The proposal is to increase the maximum percent of Social Security benefits included in the tax base from 50 percent to 85 percent; the threshold amounts remain unchanged.

[2]Aged is defined as age 62 or older and refers to the age of the family head.

Note: Dollar amounts in 1984 dollars.

Source: STATS model estimates based on the March 1984 CPS.

average benefit is much higher for those with incomes of $35,000 or more than for those with lower incomes.)

Table 2 classifies families by income size and type of family. Type of family is defined by age of head and family size. This table is designed to provide answers to questions such as the following:

(1) What proportion of families with tax increases have aged heads? (More than 80 percent have aged heads.)

(2) For the three family types, how did the ratio of tax increases to benefits vary by income? (The pattern is similar for each type of family. In each case, for families with tax increases the ratio is much higher for those with incomes of $35,000 or more than for those with lower incomes.)

V. Limitations

The STATS model has two types of limitations. First, not all of the possible effects are simulated. In many cases, a policy change might change the behavior of those affected by the policy. For example, a policy that increases Social Security benefits might cause some 64-year-old persons to retire earlier than they would have with unchanged benefits. The model, however, typically assumes that such behavioral changes will not occur, and deals only with the direct effects of policy changes on income sources. For many policy proposals,

the effects induced through behavioral changes are small enough to ignore. For those policy proposals with substantial induced effects, the distributional estimates of the direct effects provide a useful starting point for analyzing the total effects. The time and effort required to build a behavioral effect, such as the retirement decision, into the simulation model are usually quite large. Often there is no consensus on the magnitude of such effects. Incorporation of such effects typically increases substantially the complexity of both the simulation procedure and the interpretation of the simulation results.

The second type of limitation results from deficiencies in the CPS data upon which the simulation is made. Income in the CPS is often underreported by interview respondents, and high-income responses are truncated at prespecified amounts. Capital gains and noncash income are not reported. The lack of asset information hampers the simulation of asset-tested programs. The CPS has no Social Security benefit information other than the annual benefit amount and lacks other information such as earnings histories and dates of retirement. The absence of such information makes it impossible to simulate some proposals to change benefits.

Techniques have been developed for adding some of this missing information. It can often be supplied by the statistical imputation of the missing data, using information from other sources. Incomes are adjusted for truncation by using Bureau of the Census information on the aggregate income lost due to truncation. Similarly, income is adjusted for underreporting by using information on aggregate national income by source. The coverage status of wage earners is simulated using information from SSA administrative data. Personal income tax itemized deductions are simulated using data tabulated from tax returns.

In principle, almost any CPS data deficiency could be remedied by such imputations, given enough information on the correlation of the missing data with other simulation variables in the CPS file. In practice, there is usually little such information. It is known, for example, approximately how much is missing in aggregate Social Security benefits and, therefore, how much has to be imputed to the file. However, much less is known about the extent to which missing benefits tend to be concentrated among those who reported some benefits or among those who reported no benefits, or the extent to which the tendency to underreport benefits is correlated with other variables, such as the presence of other sources of income. Such information can sometimes be developed from statistical analysis of other data files, but a judgment has to be made each time as to whether the improvement that such an analysis will offer is worth the time and effort.

A promising development is the recent release, by the Bureau of the Census, of data files from the Survey of Income and Program Participation (SIPP). This survey seeks to overcome many of the CPS's deficiencies. The SIPP provides improved measures of income, assets, and health, and should provide a rich source of information for the improvement of simulation imputations.[20] Although the SIPP file is also potentially a simulation base file itself, the STATS model continues to rely on the CPS, which has a much larger sample size and, because of its longer history and more established status, a more predictable future.

Another technique for supplying missing information is that of file matching. In the past, administrative records on Social Security benefits, Social Security earnings histories, and income tax returns have been matched to corresponding CPS survey records. One such matched file, containing earnings histories, has been used to evaluate benefit proposals related to earnings sharing.[21] Because the creation of matched files requires a large commitment of time and resources, and because confidentiality provisions often seriously limit access to such files, it is difficult to rely on these files as regular and up-to-date simulation files. They do, however, provide another source of data for the assessment and adjustment of CPS data.

VI. Summary of Key Features

Using Simulation Techniques

Simulations allow policy analysts to evaluate legislative alternatives that have not been enacted and, hence, whose effects cannot be observed. In a typical application, several tax or benefit options may be proposed—perhaps to target effects of a policy change toward some groups, but not others. The STATS model's procedures for determining taxes or cash benefits can be adapted to a range of legislative proposals, so that changes in taxes (or benefits) and economic status can be determined according to each.

Using Data on Individual Households

The STATS model estimates changes in taxes or benefits for individual persons or families, one by one, based on their observed traits. By operating at the level of the individual, the model has access to data on the

[20] See Nelson et al. (1985).

[21] See U.S. Congress (1985). The earnings-sharing analysis employed another microsimulation model called MICROSIM. MICROSIM is descended from a model developed at the Urban Institute. It has been adapted at SSA for evaluation of the effects of benefit proposals up to 50 years into the future.

combinations of traits that are critical for determining taxes and benefits. As a result, microsimulation often yields more reliable distributional estimates than those based on grouped data, since grouped data—statistical tables, for example—too frequently lack information on key combinations of traits.

Beginning with information on individual persons or families also offers an advantage in deriving summary measures. Such measures can easily be tailored to the needs of policymakers.

Economic Status of Population Groups

Policymakers, in evaluating tax or benefit options, often consider the groups affected and their current economic status. There are obvious reasons for this interest in distributional effects. First, there are observable differences among groups in factors such as resources and ability to work. Second, in response to these differences, tax and benefit policies are often targeted toward particular groups. In other instances, policymakers may want distributional estimates to ensure that the burden of a tax increase or benefit reduction is spread among many groups. The model's focus reflects such concerns for distributional effects.

In most cases, a proposal's effects are considered by examining the economic status of the families affected as well as their demographic characteristics. A key premise is that since family members share income to meet common needs, an individual's economic status depends on the income and composition of the family as a whole. Hence, using the STATS model, changes in taxes or benefits are evaluated in terms of the income and composition of the families affected.

List of References

Bridges, Benjamin, Jr. "Family Social Security Taxes Compared With Federal Income Taxes, 1979," **Social Security Bulletin,** Vol. 44, No. 12, 1981, Office of Research and Statistics, Social Security Administration.

_____ and Mary P. Johnston. **Estimation of Social Security Taxes on the March Current Population Survey** (Studies in Income Distribution), Office of Research and Statistics, Social Security Administration, March 1976.

Dye, Richard F. **Analysis of the Advisory Council's Proposal to Tax One-Half of Social Security Benefits** (ORS Working Paper Series), Office of Research and Statistics, Social Security Administration, October 1981.

Johnston, Mary P. and Bernard Wixon. **Payroll Tax Liability and Its Relation to Family Unit Income: 1971, 1973, and 1974** (Studies in Income Distribution), Office of Research and Statistics, Social Security Administration, September 1978.

Millea, Mary T. **Short-Term Projection Methodology** (Documentation of Simulated Tax and Transfer System Microsimulation Model Report), Office of Research and Statistics, Social Security Administration, June 1982.

Nelson, Dawn; David McMillen; and Daniel Kasprzyk. **An Overview of the Survey of Income and Program Participation** (SIPP Working Paper Series No. 8401, update 1), Population Division, Bureau of the Census, 1985.

Projector, Dorothy S. "Children's Allowances and Income-Tested Supplements: Costs and Redistributive Effects," **Social Security Bulletin,** Vol. 33, No. 2, 1970, Office of Research and Statistics, Social Security Administration.

_____; Mary T. Millea; and Kenneth M. Dymond. **Projection of March Current Population Survey: Population, Earnings, and Property Income, March 1972 to March 1976** (Studies in Income Distribution), Office of Research and Statistics, Social Security Administration, November 1974(a).

_____; David Podoff; and Mary P. Johnston. **Payroll Relief and its Distribution in the Population—Method of Estimation and Results for 1975** (Studies in Income Distribution), Office of Research and Statistics, Social Security Administration, November 1974(b).

_____ and Mary P. Roen. **Family Demography and Transfer Payments During the 1970's** (Studies in Income Distribution), Office of Research and Statistics, Social Security Administration, September 1982.

Roen, Mary P. **Introduction and Overview** (Documentation of Simulated Tax and Transfer System Microsimulation Model Report), Office of Research and Statistics, Social Security Administration, May 1982(a).

_____. **Unit Formation** (Documentation of Simulated Tax and Transfer System Microsimulation Model Report), Office of Research and Statistics, Social Security Administration, June 1982(b).

U.S. Congress, House Committee on Ways and Means, Subcommittee on Social Security. **Report on Earnings Sharing Implementation Study,** Washington, D.C.: U.S. Government Printing Office, 1985.

U.S. Department of Commerce, Bureau of the Census. **Money Income of Households, Families, and Persons in the United States: 1985** (Consumer Income Series P-60, No. 156), August 1987.

U.S. Department of Treasury, Internal Revenue Service. **Individual Income Tax Returns: 1984,** Publication 1304, Washington, D.C.: U.S. Government Printing Office, 1986.