ESTIMATING IN-KIND INCOME USING THE CPS: SURVEY VS PURE MICROSIMULATION APPROACHES

Timothy M. Smeeding, ASA/Census Research Fellow

Over the past fifteen years there has been a tremendous explosion in the value of income in-kind in the form of government in-kind transfers. Major government in-kind benefits which are predominantly substitutes for private market goods, i.e., food, housing and medical transfers, grew about $2.1 billion in 1965, to more than $76.1 billion in 1980. (Budget of the U.S. Government, FY 1982, 1981.) Sometimes early in the 1970's the market value of in-kind benefits aimed at the poor began to exceed the more commonly known cash public assistance benefits or "welfare" expenditures of i.e., SSI; and general assistance, or GA) which are routinely recorded in the CPS and other income surveys. By 1980 more than two of every three dollars of means-tested government aid to the poor ($23.6 billion in 1980) was in the form of in-kind food, housing, or medical transfer benefits. In particular, medical care transfer benefits increased most dramatically over this period. Medicaid is clearly our largest means-tested income transfer program, far outweighing both other means-tested transfers ($18.9 billion in 1980) and all other major food and housing in-kind transfers ($77 billion in 1980).

Currently several budget directives, and public interest in general, are leading the Census Bureau to begin to collect data on these benefits and to estimate their impact on poverty. An initial report on this topic based on the March 1980 Current Population Survey (CPS) will be published later this fall. However the March 1980 CPS-based estimates of the value and size distribution of these in-kind benefits will not be the first such estimates. For the last seven years, several microsimulation models have been applied to the CPS in order to estimate the value of benefits for these same in-kind transfer programs (Smeeding, 1975; Doyle, et.al, 1980) What the March 1980 CPS adds to these models is a wholly new set of survey based data on in-kind transfers which was never before available, and thus a chance to compare current simulation results to those based on actual survey data.

The purpose of this paper is to trace out the microsimulation methodologies which have been developed to estimate the size distribution of in-kind benefits and to begin to assess the impact of actual CPS data on these methodologies. In particular, the paper will briefly compare two different approaches to this estimation procedure: first, pure microsimulation (PM) efforts whereby both eligibility status and benefit amounts are assigned to individual income records on a large microdata file using imputation methodologies based almost entirely on exogenous administrative data. Second, this PM approach is compared to a simulation process which begins with survey (S) data on recipiency for these programs, and then imputes benefit amounts to those who have reported eligibility for benefits. Microsimulation of benefit amounts is still required by the S based approach. Thus microdata survey cannot avoid the necessity of applying a microsimulation strategy to impute in-kind transfer benefits.

The second section of this paper briefly discusses how, in general, microsimulation models of both types operate. The third section of the paper compares the results of the two approaches based on three important criteria: target efficiency (percentage of program beneficiaries who are poor, and/or the percent of total benefits received by the poor); multiple benefit recipiency (receipt of zero, one, two, or more of these benefits) and finally, microdata adjustment models for survey income reporting problems. The final section of the paper suggests additional comparisons and discussion for contributions which surveys can make to improve our estimates of recipiency and for valuing noncash income, regardless of whether the S or the PM approach, or (as is recommended) an integrated S and PM approach is taken.

II. MICROSIMULATION MODELS

In the early 1970's several researchers began to adjust official Census income estimates as reported in the Current Population Survey (CPS), for its shortcomings. Census money income is underreported (relative to other estimates of the value of various income types), omits personal taxes, and by definition, excludes in-kind income. In particular the basic microsimulation model (Beebout and Bonina, 1971) emerged to adjust for net cash income underreporting. When adjusting for underreporting of cash income, recipients who failed to report income of a given type were assigned amounts based on the economic and demographic characteristics of similar persons or families who actually did report. The weighted number of recipients (reported and assigned) equaled an administrative or control benchmark number of recipients which was presumably more accurate than the CPS. Differences in CPS vs. administrative data recipient-universe definitions, and recipient-unit definitions, were adjusted for before making these assignments. If the total value of a particular income type obtained in the CPS still fell short of the administrative aggregate dollar value for that income type, all reported and imputed incomes were increased to make up the difference. One could consider also using a variant of this model to adjust for in-kind benefit underreporting. Thus, it does have an important indirect bearing on the outcome of the in-kind transfer simulation modeling process which will become apparent later in this paper.

In the mid 1970's first Smeeding (1975) and then Mathematica (in 1976) began to microsimulate both eligibility and benefit amounts for major in-kind transfer programs using the March CPS. In general, the pure microsimulation (PM) process for in-kind transfers works as follows: First, obtain an administrative estimate of the number of families (individuals) who benefitted from a given program during the year in question and the amount of benefits paid out. In addition, obtain any administrative information which is available concerning the eligibility rules from the programs and characteristics of the recipient population. Second, use these eligibility rules and other administrative data to assign benefits to individu-
als on CPS for each program such that aggregate benefits, recipients, and other characteristics match up with administrative records as closely as possible. These rules have been and are now being applied with more or less success (as outlined below) to the Food Stamp, School Lunch, Medicaid and Medicare programs, and to several types of public housing programs.

In recent years Mathematica's "MATH" PM model has continued to refine and update these microsimulation estimates (e.g. see Doyle, et.al., 1980). In fact, most major federal agencies responsible for evaluating in-kind transfer programs and for forecasting their cost (e.g. the Congressional Budget Office) use the MATH model, or some other form of the CPS, augmented to include the in-kind programs which the agency is interested in, and using the procedure outlined above to simulate these benefits.

The second or survey (S) based approach is something new and different. Beginning with the March 1980 CPS, a battery of questions on recipiency of in-kind benefits were asked. Respondents were asked whether they (or members of their families) benefitted from Food Stamps or School Lunch (both "free or reduced" price, and paid "full price"); whether they lived in public or subsidized housing, and whether or not they were covered by Medicaid or Medicare any time during 1979. In general the CPS comes fairly close to full reporting for most programs when CPS respondent totals are compared to administrative data based estimates. The Food Stamp program is a bit lower than average (75 percent of administrative total), while no control estimate was available for the paid School Lunch question. Other programs are at or above the 88 percent of administrative estimate reporting level.

Benefit amounts from in-kind transfer programs are less easily collected in sample surveys like the CPS. The level of benefit is often unknown to the recipients. Only for the Food Stamp program, where benefits are measured in dollar amounts, can such a survey question be reasonably answered by the beneficiary. Thus even using the S approach to determine recipiency patterns, a microsimulation model is usually necessary to assign benefit levels to those recipients.

This concludes the outline of the two basic microsimulation strategies for assigning in-kind transfer benefits to individuals (and also the basic microsimulation adjustment strategy for income underreporting). The next step is to compare some aspects of the results of both types of models.

III. PURE MICROSIMULATION VS SURVEY BASED APPROACHES

There are at least three important criteria for comparing PM and S based approaches to imputation of in-kind benefit amounts: target efficiency, multiple benefit recipiency, and ease or difficulty of underreporting adjustments. We shall treat each in turn.

Target Efficiency. Two important distributive characteristics of in-kind transfer programs are the percent of program beneficiaries which are poor and the fraction of program benefits which accrue to the poor. The greater the number of beneficiaries who are poor, and/or the greater the percent of total benefits received by the poor, the greater the number of persons who will be

<table>
<thead>
<tr>
<th>Table 1: TARGET EFFICIENCY: PERCENT OF RECIPIENT HOUSEHOLDS, PERCENT OF TOTAL MARKET VALUE OF BENEFITS RECEIVED BY THE POOR. 1/</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Percent of Benefits Received by Official Census Poor</strong></td>
</tr>
<tr>
<td>Food Stamps</td>
</tr>
<tr>
<td>School Lunch</td>
</tr>
<tr>
<td>Medicare</td>
</tr>
<tr>
<td>Medicaid</td>
</tr>
<tr>
<td>Public Housing</td>
</tr>
<tr>
<td><strong>B. Percent of Beneficiary Households Who Are Official Census Poor</strong></td>
</tr>
<tr>
<td><strong>Program:</strong></td>
</tr>
<tr>
<td>Food Stamps</td>
</tr>
<tr>
<td>School Lunch</td>
</tr>
<tr>
<td>Medicare</td>
</tr>
<tr>
<td>Medicaid</td>
</tr>
<tr>
<td>Public Housing</td>
</tr>
</tbody>
</table>

Notes: NA = estimate not available or not comparable.
1. Household poverty status is based on "Official" CPS money income, and is determined by the poverty status of the primary family or individual in the household.
2. Based on March 1980 CPS data tapes.
3. Based on CBO-MATH model, i.e. March 1979 CPS aged to FY 1980 (Doyle et.al, 1980).
5. Based on March 1975 CPS.
moved out of poverty--i.e. the greater the efficiency of the program in targeting its benefits to the poor. Table #1 compares the target efficiency of several PM models to that of the March 1980 CPS. The models compared here are: the March 1980 CPS; the March 1978 CPS based MATH model (which is projected forward or "aged" to fiscal year 1980); the spring 1975 Survey of Income and Education (SIE) aged forward to 1979 by the Department of Health and Human Services (HCIS); and Smeeding's March 1975 CPS-based simulations for 1974.

In the top panel of Table 1 the percentage of the aggregate market value of benefits received by poor households is presented. Overall the results are remarkably similar, while the MATH estimates are based on aged data. Both the SIE and Smeeding results indicate a higher degree of target efficiency than the CPS (and MATH), but the differences are not large and may be explained by the time period differences (Smeeding's estimates are for 1974) and by data ageing (SIE). The bottom panel presents similar estimates for households and again the results are similar, particularly when we compare the CPS and MATH models. Considering the difficulty of simulating intra-year income eligibility, estimating asset and income eligibility, selecting participants from the eligible pool, and even in estimating benchmark numbers of persons who were ever for a given program, these similarities are indeed surprising.

Multiple Benefit Recipiency. The second area of comparison between the PM and S models concerns patterns of multiple benefit recipiency. For some time there has been a major public policy interest in overlap between various cash and in-kind transfer programs. However, no major study of multiple benefit recipiency has been undertaken prior to March 1980 CPS, or prior to the 1979 and 1980 March CPS. The Survey of Income and Program Participation (SIPP) 2/ while PM models have been in existence and widespread use for at least five years now, none have been employed for this purpose. The major reason why PM models have not been able to fill this gap is because in PM models, each in-kind program is, in most cases, separately and independently simulated. Because there exist no administrative data which contain, for instance, an estimate of the number of Food Stamp households which are benefitting from Medicaid and/or Public Housing as well, PM models have not been able to "control" for multiple in-kind benefit in their simulations. In effect, multiple benefit recipiency patterns for in-kind transfers emerge from PM models largely as a statistical artifact. With PM S based model, subject to underreporting error, one can put some confidence in the multiple benefit patterns which are reported. Yet, because receipt of multiple benefits can often make the difference between being poor or nonpoor, it is important to compare the PM and S models on this basis.

Table 2 makes these comparisons for both all households and poor households only, for all five major in-kind transfer programs: Food Stamps, Medicare, Medicaid, School Lunch, and Public Housing. Leaving the most striking comparison for last, we first compare the CPS and JEC studies in the top right quadrant. Both sets of data refer to all poor and nonpoor households. But while the JEC (U.S. Joint Economic Committee, 1972) study was conducted for all households, it covered only six low income areas and only about 2100 total households. Clearly any study which was based on the JEC paper would tremendously overstate multiple recipiency among the entire population, e.g. Paglin (1979).

The top left hand quadrant indicates even more striking differences between the CPS and the MATH model. Adding public and subsidized housing and School Lunch to the three major in-kind programs, MATH indicates that only 7 percent of the poor did not benefit from any program while 47 percent benefitted from three or more. The CPS indicates 28 percent of poor households received none of these transfers, while only 23 percent benefitted from three or more. Clearly the MATH-PM model outcomes are quite different from the CPS-S model results. Table 3 is designed to shed some additional light on these differences. While the MATH model is based on the CPS, and while it is for nearly the same period (i.e. October 1979-September 1980 vs. the CPS January-December 1979 annual period), one major difference is the different number of persons who were ever for a given program is the MATH-CPS difference in the CPS-S model results. It is clear that they reduced the total number of official CPS poor households by 14.7 million or by 20 percent (Table 3). More research into the data ageing process is needed. However, because the multiple recipiency data in Table 2 are relative, i.e. the percent of poor with a given number of benefits, it may be argued that the MATH-CPS difference in the absolute number of poor is of little consequence for purposes of comparing multiple benefit recipiency.

Table 3 also presents comparisons between the MATH and the CPS in terms of the total number of poor and the percent of the poor who receive each type of benefit. While the percentage of beneficiaries from each program who were poor (Table 1) was similar, the percentage of the poor who receive benefits from each program clearly is not. In particular the MATH model finds almost twice as large a percent of poor receiving food stamps, and nearly half again as large a percent receiving Medicaid as does the CPS. Differences between the other programs are fairly small. 2/ Can the reason for these differences be CPS underreporting? Suppose we take the percent underreporting estimates for all Food Stamps Medicaid recipients and the number of percent of CPS poor receiving each type of benefit by these fractions to reach control. If so, we would find 48.9 percent of CPS poor with Food

362
Table 2: MULTIPLE IN-KIND TRANSFER BENEFIT RECIPIENCY: PERCENT OF HOUSEHOLDS (POOR HOUSEHOLDS) RECEIVING A GIVEN NUMBER OF BENEFITS

<table>
<thead>
<tr>
<th>Number of Programs Participates In:</th>
<th>Poor Households Only</th>
<th>All Households</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MATH3/ (1979-80)</td>
<td>CPS4/ (1979)</td>
</tr>
<tr>
<td>zero</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>one</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>two</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>three or more</td>
<td>47</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Notes: 1. Household poverty status is based on "Official" CPS money income and is determined by the poverty status of the primary family or individual in the household.
2. Includes: Food Stamps; School Lunch; Medicare; Medicaid; Public Housing.
4. Based on March 1980 CPS reported data.

Stamps, and 45.7 percent with Medicaid--still far short of the MATH estimate of 72.3 and 63.6 percent for each program. The much larger number of poor units receiving each of these types of benefit, at the top of Table 3, probably explains much of this discrepancy in Table 2. Either the estimated CPS underreporting estimate is far off the mark, or tremendous changes in each program took place between 1979 and 1979-80, or the MATH-PM model has overestimated the number of poor Medicaid and Food Stamp beneficiaries by a large amount. Additional research should be undertaken to explore these differences.

One key explanation of these differences may be their treatment of cash welfare. In several current PM models, e.g. the MATH model, CPS reported recipiency and benefit amount for cash public assistance (CPA)--AFDC, GA and SSI--are assigned households using a PM approach. In other words, the reported CPS data which is collected on the CPS is ignored. Instead, just as if these data were not collected, the CPA population is estimated using a PM model. One problem with this approach is the fact that welfare agencies do not have a count of the number of persons who ever benefitted from the various CPA programs during a given year (e.g. for AFDC). Thus one must first estimate the size of the control total number of beneficiaries. After obtaining aggregate benefit amounts (for which there are detailed records of annual expendi
tures) the CPA population and their individual benefits are estimated. A judgement that the PM approach, with all of its potential errors, is still superior to reported S results (which are, as is well known, underreported by about 25 percent) has, of course, implicitly been made in selecting this strategy.

However, one cannot stop at this point in evaluating the PM vs. S models for estimating CPA. One of the largely available and most widely used characteristics of administrative data for Medi-
caid, Food Stamps, and public housing which is used in PM models that estimate these in-kind ben-
efits, is the percentage of beneficiaries who also receive CPA. Thus PM estimated CPA benefits are used to control PM estimated in-kind transfer ben-
efits. All AFDC recipients, and virtually all SSI beneficiaries are eligible for Medicaid. Between 40 and 60 percent of all Food Stamps recipients and about half of all public housing beneficiaries receive CPA. In effect then a PM model which as-
signs CPA benefits to a household, is simultane-
ously assigning medical benefits and greatly in-
creasing the probability that the unit also re-
ceived food and/or public housing transfers as well. Viewed from this perspective the accuracy of the PM assignment of CPA is crucial to the ac-
curate assignment of recipiency status and benefit amounts for the three largest means tested in-kind benefit programs as well. Again, in contrast, be-
cause the S based approach produces program bene-
ficiaries directly, corecipiency of CPS and other in-kind transfers emerges directly. The combined effect of simulating CPA and in-kind benefits may have significant effects on multiple benefit re-
cipiency. For instance, an October 1979 adminis-
trative Food Stamp Survey found that 42 percent of Food Stamp households also received AFDC or GA. More recent administrative data for July 1980 puts this figure at 38 percent. The March 1980 CPS estimate was just over 49 percent for calendar 1979. In contrast, in the MATH model for 1979-80, on an average monthly basis, 57.2 percent of Food Stamp recipients also received cash public assistance (Doyle, et.al, 1980:191). This same MATH data indicates that 83.7 percent of all AFDC units also received stamps during an average month. In the CPS, almost exactly 70 percent of all such units received stamps over the entire year. These figures suggest the possibility that the MATH model may have overestimated the number of units with CPA and Food Stamps. If these results are combined with the fact that virtually all AFDC units also receive Medicaid, multiple benefit re-
cipiency totals for these two programs may also be overestimated. But further research need be undertaken to investigate these suspicions.
Table 3: THE MARCH CPS 1980 AND THE 1979-80 MATH MODEL: FURTHER COMPARISONS

Comparison Element

A. Number of Poor Households (Millions) Receiving Each Type of Benefit:

<table>
<thead>
<tr>
<th>Benefit</th>
<th>MATH (1979-80)</th>
<th>CPS (1979)</th>
<th>MATH-CPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food Stamps</td>
<td>5.551</td>
<td>3.575</td>
<td>1.976</td>
</tr>
<tr>
<td>School Lunch</td>
<td>2.296</td>
<td>2.602</td>
<td>-.306</td>
</tr>
<tr>
<td>Public Housing</td>
<td>1.377</td>
<td>1.170</td>
<td>.207</td>
</tr>
<tr>
<td>Medicare</td>
<td>3.028</td>
<td>3.330</td>
<td>-.302</td>
</tr>
<tr>
<td>Medicaid</td>
<td>4.876</td>
<td>3.799</td>
<td>1.077</td>
</tr>
<tr>
<td>Total Poor Units</td>
<td>7.877</td>
<td>9.549</td>
<td>-1.678</td>
</tr>
</tbody>
</table>

B. Percent of Poor Households Receiving Each Type of Benefit:

<table>
<thead>
<tr>
<th>Benefit</th>
<th>MATH (%)</th>
<th>CPS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food Stamps</td>
<td>72.3 %</td>
<td>37.4 %</td>
</tr>
<tr>
<td>School Lunch</td>
<td>29.9</td>
<td>27.2</td>
</tr>
<tr>
<td>Public Housing</td>
<td>18.0</td>
<td>12.3</td>
</tr>
<tr>
<td>Medicare</td>
<td>39.4</td>
<td>34.8</td>
</tr>
<tr>
<td>Medicaid</td>
<td>63.6</td>
<td>39.8</td>
</tr>
</tbody>
</table>

Source: March 1980 CPS data tapes and U.S. Congressional Budget Office tabulations.

Notes:
1. "Public housing" includes public and subsidized housing for low income families under various public programs including: Low Rent Public Housing, and Sections 8, 235, 236, 101, and 202b of the 1937 Housing Act.
2. Large discrepancy between CPS and MATH units is due to aging process by which March 1978 CPS data for 1977 was projected to fiscal year 1980. See text for explanation.

Underreporting Adjustments. The final area to be discussed is adjustment for underreporting error. In general the PM approach produces recipient counts and benefit amounts which match administrative data. On the other hand, S based estimates typically fall short of administrative estimates. The PM and S approaches are not, however, mutually exclusive. For instance, one could begin with CPS survey reported recipiency characteristics for in-kind benefits and then attempt to adjust for underreporting using the PM underreporting strategy outlined earlier. The problem is that, as far as I can determine, one cannot maintain the proportional multiple benefit recipiency characteristics observed in the CPS and simultaneously raise all in-kind benefit recipients and types to their control totals. Either the control totals are incorrect, or the multiple recipiency characteristics of nonreporters differ from those who have already reported receipt of an in-kind benefit on the CPS.

At this time it is not possible to say which is the major problem. Control totals for Medicaid, Food Stamps, and School Lunch are only estimates. Moreover, it is impossible to identify which public housing program those who report recipiency on the CPS are actually benefitting from. If better administrative control estimates were available, and if, maintaining multiple recipiency characteristics while making underreporting adjustments, imputation of recipiency (benefit amounts) still did not produce control totals of recipients for all programs, one could be sure that the multiple beneficiary characteristics of nonreporters and reporters differed. But we need develop better alternative program control estimates before one can reach this conclusion.

IV WHERE TO GO FROM HERE

The major point of this paper is to suggest that the PM and S approaches to estimating the size distribution of noncash benefits can be expected to yield fairly different results, particularly in the case of multiple benefit recipiency. More research need be devoted to further comparisons of these models and their results. Since beauty is in the eye of the beholder, and since the ultimate policy purposes to which one will put in-kind income estimates may independently influence one's choice of in-kind benefit simulation technique, it is not possible to argue that either the S or the PM technique is, in general, better than the other. However, it is fair to conclude that the ageing process used to create the 1979-80 MATH file from the March 1978 CPS produces too few poor people. Whether it is the overly optimistic macroeconomic assumptions used to age the data, or the ageing model itself that causes these discrepancies is not known. However estimates of the number and percent of persons in poverty from this MATH model are definitely called into question. The 4.1 percent of persons who are in poverty after in-kind transfers are counted at market value, as estimated by Hoagland (1980) using the MATH-PM model is likely to be a serious underestimate of the true extent of poverty due to the ageing process itself (not to mention the multiple benefit recipiency differences shown in Tables 2 and 3).
At some point in the future, researchers may be able to combine the best features of both models to serve their purposes. At the same time that this research progresses, we must continue to remind ourselves of the small fraction of total income in-kind which we are engrossed with. Because in-kind transfers are less than one third of all types of food, housing and medical income in-kind, we must continue to expand both types of models to include estimates of a wider range of the various types of income in-kind as we simultaneously improve the estimates of in-kind benefits which policy makers now rely on.

Footnotes
- Readers interested in a more complete description of these and also other microsimulation models should consult Haveman and Hollenbeck (1980).
- The March 1980 CPS and the ISDP both provide this capability for the 1979 income year. While there are few multiple benefit studies in earlier years to begin with, e.g. U.S. Joint Economic Committee (1973), National Urban League (1980), and Lyon, et al (1976), none of those covered the entire U.S. population. The Joint Economic Committee study covered only six "low income areas" in 1971, while the Urban League study covered only blacks in 1978, and the Lyon et al. study covered only New York City. A multiple benefit reciprocity study for Food Stamp participants using the ISDP data is currently in preparation (MacDonald, 1981).
- Preliminary reports from the March 1981 CPS indicate even larger differences between MATH and the 1980 count of poor households. The reader should be careful to note that the MATH results have not been adjusted for income underreporting. Such an adjustment could account for part of the differences in the number of poor households in Table 3.
- In PM model such as MATH it is difficult to simulate asset ineligibility due to lack of CPS data on assets. Using 1979 ISDP data in a recent paper, MacDonald (1981) found that 30 percent of all poor households were asset ineligible for Food Stamps. If MacDonald's estimates are accurate, the MATH estimate that 72.3 percent of all poor households received Food Stamps is an impossibility, even if 100 percent of Food Stamp eligible poor households participated in the program.
- A major problem with adjustment for underreporting error in the CPS is the lack of administrative estimates of the annual "ever-received" population which are comparable to the CPS. These estimates must be compiled by the researcher, often leading to different "control" or "benchmark" estimates for different researchers.

References