

Measuring and Modeling Taxes in the Survey of Income and Program Participation

**Jeff Sisson
Kathleen Short**

**Prepared for the 2001 Joint Statistical Meetings of the American Statistical Association
August 2001**

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau Publications. This report is released to inform interested parties of research and to encourage discussion.

Measuring and Modeling Taxes in the Survey of Income and Program Participation

Introduction

The ability to accurately estimate taxes based on survey responses has gained importance since a recommendation was made by the National Academy of Sciences (NAS) to change the method used to measure poverty.¹ The current official poverty measure was established in the early 1960's and has gone through very few revisions since it was adopted in 1965. The poverty measure is based on the cost of a minimum diet multiplied by three to account for other expenditures. The income side of the poverty measure is based on before-tax money income and does not include a valuation of non-cash benefits such as food stamps or subsidized housing. Since that time, taxes have become a much more important item in household budgets. The current tax code now includes the earned income credit (EIC). The EIC is an anti-poverty element of the tax code that provided \$21.6 billion to low income families in 1997. Social security taxes have also had significant changes since the mid-60's. Social security taxes are now more than double their levels at the time the current poverty measures were established.

The NAS panel identified a number of weaknesses of the current poverty measures and recommended changes to the poverty measures that would address some of these deficiencies. Among the recommendations was to use the Survey of Income and Program Participation (SIPP) as the survey for generating the poverty measures. The panel also recommended that taxes should be subtracted from family resources and that some measure of non-cash subsidies be included since these free resources to be used on other necessities.

The need to generate accurate tax estimates led to an examination of the current methods of generating tax information and an exploration for more accurate methods of calculating tax liability. The current poverty measures are based on the March Supplement of the Current Population Survey (CPS). The CPS has a current tax model that was developed in the early 1980's, but the model has a number of deficiencies as described below. The SIPP also collects tax information through an annual topical module administered shortly after the end of tax season. The topical module asks for detailed information on tax returns, but the response rates to many of the questions are too low to make the information useful.

Given the NAS recommendation and the necessity for a more accurate tax model, a new model was developed that would provide more detailed tax calculations. The model was developed to work with CPS and will work with SIPP after some data manipulation. The remainder of this paper discusses the need for the tax model and the initial results from the model.

Current Methods for Measuring Taxes

As mentioned in the previous section, it is necessary to estimate tax liability for the experimental poverty measures. Tax liability is presently calculated for the Current Population Survey (CPS) using a tax model developed in the early 1980's. The tax model calculates adjusted gross income, tax liability, a number of tax credits, and state taxes. The model also imputes capital

¹ Citro, Constance F. and Robert T. Michael (eds.), *Measuring Poverty: A New Approach*, Washington, DC: National Academy Press, 1995.

gains and IRA deductions based on total income. The current CPS tax model has been updated continuously to account for changes in tax rates and tax law.

While updated to reflect current tax laws, the current tax model either does not estimate or provides only gross estimates for some income and expense categories. State level taxes, capital gains and losses, the earned income credit, child care credits and statutory adjustments are areas that are trying to be improved with the new tax model. State level taxes are the most significant area for improvement to the tax model. The current model does not consider federal itemized deductions when calculating state taxes. Instead it assumes that everyone takes the standard state or federal deductions and does not give credit for potential deductions focused on age or disability. The current model also does not include details on the state tax returns involving credits or exclusions. The state tax calculator in the current model also does not simulate either state earned income credits or childcare credits. These credits have been implemented by many states in the past 5 years and need to be included in a detailed state tax calculator. These areas are addressed in the new tax model being presented in this paper.

The second area for improvement to the current tax model is capital gains. The current model uses a Monte Carlo technique to assign the presence of capital gains or losses and then assigns the amount based on mean values within a matrix defined by adjusted gross income, filing status and age. The result of this method is that the current tax model tends to assign a larger mean capital gain to a smaller number of houses than the new model (see Table 8). A change in this method of imputing capital gains could provide a more accurate picture of capital gains and losses than the current model provides.

The earned income credit and childcare credits are another area where the current tax model can be improved. While the current model does calculate the earned income credit, it does not include the simulation of the detailed worksheets that test for investment income and the use of modified adjusted gross income (AGI). The worksheets do not effect a large number of filers claiming the earned income credit, but including them in the calculation does provide a more accurate calculation of the tax credit. The current tax model does not include any calculation for the childcare tax credit.

Statutory adjustments to income are the last area where the current tax model could provide a more accurate calculation of taxes. The 1997 tax code identifies 8 adjustments to income. The adjustments include:

- IRA deductions
- Medical savings account deductions
- Moving expenses
- One-half self-employment tax
- Self-employed health insurance deduction
- Keogh and self-employed SEP and SIMPLE plans
- Penalty for early withdrawal of savings
- Alimony paid

Of the eight adjustments mentioned above, the current tax model includes only the IRA deduction in its calculation of taxes. A new model should attempt to include as many of these deductions as possible.

The current method for collecting tax information for the SIPP is through a topical module administered each year. The module asks respondents if they filed taxes for the previous year and, if so, collects information on exemptions, adjusted gross income, filing status, tax liability, itemized deductions, tax credits, and capital gains. The module is administered every year near the end of the tax-filing season.

The SIPP tax module is very thorough in the items it attempts to collect, however, non-response rates for some of the qualitative items and most of the quantitative items are fairly high. The low response rates combined with some inconsistency of answers between the tax module income questions and core responses for the calendar year make the tax module information less useful in determining tax liability.

Table 1 and Table 2 show the non-response rate problems associated with the SIPP Tax Module. Table 1 shows that the main qualitative questions (did you file taxes?, what form was filed?, filing status?) receive fairly high response rates. However the questions asking about more specific or detailed filing items receive a much lower cooperation rate.

Table 1. Non-Response Rates for Tax Module Respondents Who Responded That They Filed Taxes - Qualitative Questions

Question	Percent Non-Response*	Percent Responded
Total Exemptions	8%	92%
What Tax Form Filed	13%	87%
Filing Status	2%	98%
Filed Schedule A (Itemized Deductions)	55%	45%
Filed Schedule D (Capital Gains)	56%	44%
Claimed Child Care Expense Credit	84%	16%
Claimed Earned Income Credit	41%	59%

*Includes responses of don't know

Table 2 shows the much higher non-response rates associated with the quantitative questions on the tax topical module. As you can see in the table, only 53% of respondents who said they filed taxes for the previous year provided their adjusted gross income. The response rates are similarly low for the other quantitative measures in the topical module.

Table 2. Response Rates for Follow-up Quantitative Amounts Based on Positive Responses To Screening Question For Those Amounts.

Screening Question	Quantitative Question	Non-Response	Response of \$0	Response Greater Than \$0
Filed Taxes	Adjusted Gross Income	47%	0%	53%
Filed Taxes	Total Tax Liability	21%	47%	32%
Filed a Schedule A	Total Itemized Deductions	5%	65%	30%
Claimed Capital Gains	Total Capital Gains/Losses	5%	63%	32%
Claimed Earned Income Credit	Earned Income Credit	58%	0%	42%

Development of New Tax Model

The low response rates in the tax topical module for SIPP and the less accurate results associated with the current tax model for CPS created a need for a more complex and accurate tax model. A new tax model has been completed that addresses many of the needs identified in the above discussion. The new tax model has a more complete and accurate calculation of state and local taxes, imputes the presence and amount of capital gains simultaneously using IRS data, and calculates more exclusions and deductions than the current model.

The new tax model was built specifically to function with the CPS March Supplement data, but can be adapted for use with the SIPP data by making some minor changes to the model and by transforming the SIPP data into an annualized file. The remainder of this paper discusses the model’s methodology for both CPS and SIPP and presents initial results from the model. The tax model is still relatively new and enhancements will continue to be made in the future.

Description of the New Tax Model Using the CPS March Supplement²

The new tax model initially creates both a person and household level file from the CPS March Supplement. The data that are extracted focus mainly on income and family and household structure. The first step in the process of calculating taxes is to complete a statistical match of the CPS with the American Housing Survey data. The statistical match allows a variable flagging the presence of a mortgage to be added to the CPS data. The additional variable will allow the deductions for mortgage interest to be imputed later in the process. The statistical match uses age of householder, household income and household size as the main matching criteria for determining the presence of a mortgage.

The second step in calculating taxes is to split the CPS file into potential tax units and complete a statistical match with the Statistics of Income (SOI) file from the IRS. The statistical match is done to provide the following information and append it to the CPS file:

- Presence and amount of capital gains
- Presence and amount of itemized deductions
- Presence and amount of IRA deductions

² Coder, John. 2001. “Summary Comparisons Between IRS Published Statistics and Current and New Tax Simulation Models for Income Year 1997.” Sentier Research, LLC. Unpublished Manuscript.

- Presence and amount of child care expenses
- Presence and amount of self-employed health insurance cost deductions
- Presence and amount of Keogh-SEP/SIMPLE deduction

These items are necessary to generate a detailed and complete tax return for each tax unit. The main variables used in the statistical match vary based on the universe you are matching. Gross income and individual state of residence are used for all the matching routines. Other variables that are used based on the universe being matched include:

- Presence of wage and salary income
- Itemized or standard deduction
- Presence of social security
- Number of age exemptions
- Presence of non-farm self employment income
- Partner of household head
- Number of child exemptions
- Presence of mortgage interest
- Type of return

A more detailed description of the statistical matching procedure is provided by John Coder.³

The final process in the tax model is the calculation of the actual tax liability. The model has recreated the actual tax filing process as closely as possible, including rules for dependents and claiming exemptions. The various potential tax units are pushed through the process of calculating tax liability and tax units are recreated as individuals pass or fail various tests for dependency or exemptions. The final outcome of the tax model is a CPS person level file that contains the major tax variables and identifiers for all tax units. This file can be used to create person level, tax unit level, family level, or household level analyses.

Description of the New Tax Model Using SIPP

The tax model that has been developed can be used for both CPS and SIPP. The actual model works in the same manner for both surveys, but SIPP requires a significant amount of data manipulation prior to being run through the model. The CPS March Supplement collects data on a calendar year basis. All respondents report income amounts for the previous calendar year, and all of their income for the year is included in the file. SIPP collects and reports income on a monthly basis, and this creates a number of issues when using the tax model.

The first step in applying the tax model to SIPP is to create a SIPP calendar year file that resembles the CPS March Supplement file. The creation of the calendar year file requires adding all monthly income sources into a calendar year total, and creating new family and household variables that represent a person's situation at the end of the calendar year. The calendar year

³ Coder, John. 2001. "Using A Statistical Match Of Survey Observations and the IRS Statistics of Income Public Use File to Enhance a Tax Simulation Model Based on the March CPS Survey Data." Sentier Research, LLC. Unpublished Manuscript.

file that is created closely resembles the CPS file that the tax model was based on. While not all variables in the CPS file can be exactly recreated in the SIPP file, the variables can be generated in sufficient detail to allow the tax model to run correctly.

The other major issue in using SIPP with the new tax model is dealing with attrition over the course of the calendar year. CPS respondents complete interviews based on the entire calendar year, and because they are interviewed only once, there are no problems with attrition of the course of the year. The SIPP has issues with respondents who leave the sample at some point during the calendar year and with respondents who join the sample at some point during the year, but are not present for the entire calendar year. The respondents that leave during the year and are not present in December present problems with assignment of their income and associating the respondent with the correct family and household unit. Because family and household identifiers can change on a monthly basis in SIPP, it is difficult to assign a respondent who drops from the survey in mid-year to the appropriate family in December.

The second problem is with respondents who join the survey in the middle of the year and are present in December. These respondents can be assigned to the appropriate end-of-year family and household, but they will not have a full year of income. These respondents will end up under-representing their income for tax purposes. Table 3 shows the change in person level means that occurs by including the part-year respondents.

Table 3. Comparison of Income Means for SIPP Full Calendar Year Versus Part Calendar Year Respondents

	Part-Year Respondents		Full-Year Respondents	
	N	Mean	N	Mean
Wage & Salary Income	88,041	\$10,760	81,049	\$11,303
Self-Employment Income	88,041	\$1,653	81,049	\$1,744
Interest Income	88,041	\$619	81,049	\$656
Social Security Income	88,041	\$1,050	81,049	\$1,115
Total Household Income	88,041	\$46,009	81,049	\$48,169

The issue of part-year respondents and how to handle these people will be addressed in the future. We have limited the SIPP sample to those respondents present for the entire calendar year. Limiting the respondents in this manner will allow the most consistent comparison of the SIPP and CPS model outputs and will provide the best initial evaluation of the model.

Results of the New Tax Model

The new tax model has been run on both the CPS March Supplement and the SIPP file created for calendar year 1996. The new tax model is based on the 1997 tax year files, so the SIPP income levels and comparison to the SIPP tax module responses will be slightly off because of changes in tax laws and general inflation. The tax model was applied to the 1998 CPS March Supplement which collected information for calendar year 1997.

The test of the new model is to examine the AGI distribution for the current model, new model, and IRS published data. Table 4 shows the distribution for these groups. The table shows similar distributions from both models and the distribution from the models are consistent with the distribution from IRS published figures.

Table 4. Adjusted Gross Income (AGI) Distribution for IRS Published Data, Current Tax Model and New Tax Model Using Both CPS and SIPP.

AGI Range	IRS Published Data	Current Model Using CSP	New Model Using CPS	New Model Using SIPP
No Adjusted Gross Income	1%	0%	0%	8%
\$1 to \$999	2%	1%	7%	3%
\$1,000 to \$2,999	5%	6%	6%	4%
\$3,000 to \$4,999	5%	5%	5%	4%
\$5,000 to \$6,999	4%	4%	4%	4%
\$7,000 to \$8,999	4%	4%	4%	4%
\$9,000 to \$10,999	4%	4%	4%	4%
\$11,000 to \$12,999	4%	4%	4%	4%
\$13,000 to \$14,999	4%	3%	3%	4%
\$15,000 to \$16,999	4%	4%	4%	3%
\$17,000 to \$18,999	4%	3%	3%	3%
\$19,000 to \$21,999	5%	5%	5%	5%
\$22,000 to \$24,999	5%	4%	4%	4%
\$25,000 to \$29,999	7%	6%	6%	6%
\$30,000 to \$39,999	11%	11%	10%	10%
\$40,000 to \$49,999	8%	8%	7%	8%
\$50,000 to \$74,999	12%	14%	13%	12%
\$75,000 to \$99,999	5%	1%	6%	5%
\$100,000 to \$199,999	4%	7%	5%	4%
\$200,000 and over	3%	0%	1%	1%

Table 5 shows the results of the new tax model compared to the current tax model and IRS published information. The means are based on tax units created by the new tax model. The new tax model produces tax liability for the CPS data files that is 8% higher than the current model, but still 7% below the published IRS data. The new tax model using the SIPP data produces a much lower tax liability than both the current model and the IRS published data.

Table 5. Tax Unit Level Comparison of AGI, Taxable Income and Federal Income Tax for IRS Published Data, the Current Tax Model, and the New Tax Model Using Data from CPS and From SIPP.

	IRS Published Data	Current Model Using CPS	New Model Using CPS	New Model Using SIPP
Mean AGI	\$50,043	\$42,470	\$48,089	\$45,434
Mean Taxable Income	\$34,528	\$29,674	\$34,672	\$31,715
Mean Federal Income Tax	\$7,824	\$6,737	\$7,289	\$6,349

The differences in the mean taxable income and lower mean federal income tax may be attributable to both a different distribution of filing status for returns using the SIPP file and a difference in the mean amounts for different filing statuses. Table 6 shows the percent distribution of returns by filing status. As you can see in the table, the SIPP file shows a lower percentage of single returns and a higher percentage of married, filing jointly returns.

Table 6. Distribution of Types of Returns for Current Tax Model and New Tax Model Using Both CPS and SIPP Compared to IRS Totals Tax Returns With Taxable Income

	IRS Published Returns	Current Model Using CPS	New Model Using CPS	New Model Using SIPP
Single	N/A	47%	46%	41%
Married, Filing Jointly	N/A	46%	47%	52%
Head of Household	N/A	7%	7%	7%

Table 7 shows the mean AGI, taxable income, and tax amount by filing status. As you can see in the table, the AGI, Taxable Income and Federal Income Tax are all higher for the new model using CPS than for the model using SIPP data. The table shows that the lower mean values generated by the new model using SIPP data is not due to differences in the distribution of filing status. The higher means for CPS data can be partially attributed to higher wage and salary figures found in CPS. This difference is discussed in more detail below.

Table 7. Tax Unit Level Comparison of AGI, Taxable Income and Federal Income Tax for the New Tax Model Using Data from CPS and SIPP

	New Model Using CPS	New Model Using SIPP
Single		
Mean AGI	\$29,068	\$27,791
Mean Taxable Income	\$21,624	\$19,928
Mean Federal Income Tax	\$4,441	\$3,885
Married, Filing Jointly		
Mean AGI	\$69,228	\$62,071
Mean Taxable Income	\$49,978	\$43,547
Mean Federal Income Tax	\$10,682	\$8,866
Head of Household		
Mean AGI	\$32,357	\$30,548
Mean Taxable Income	\$18,254	\$16,264
Mean Federal Income Tax	\$3,304	\$2,787

We have also examined the results of the new model with regards to some of the data generated through the match with the IRS SOI file. Table 8 shows the results of the capital gains from the current model, the new model, and IRS published figures. The table shows that the current model creates too few returns with capital gains, but assigns returns with capital gains to much higher mean amounts than the new model for both SIPP and CPS. The new model generates capital gains that are much more consistent with the IRS figures, although the model does underestimate the mean value. The underestimation of the mean is most likely caused by the

manipulation done by the IRS to the SOI file. The IRS topcodes and does mean replacement on some records of high-end taxpayers to protect identities. This manipulation causes capital gains values of the high-end taxpayers to be lower than would otherwise be expected.

Table 8. Tax Unit Level Capital Gains for the Current Tax Model and the New Tax Model for Both CPS and SIPP Compared to IRS Figures

	IRS Published Data	Current Model Using CPS	New Model Using CPS	New Model Using SIPP
Percent With Capital Gains/Losses	20%	14%	18%	23%
Mean Capital Gain/Loss	\$14,690	\$10,129	\$4,968	\$4,016
Median Capital Gain/Loss	\$325	\$4,963	\$482	\$372

The results for the Earned Income Tax Credit (EITC) are much more consistent between the new model and the current model, but both models are lower than the IRS published figures. Table 9 summarizes the results of the EITC generated by the two models as compared to IRS published figures. The calculation of both the eligibility and the actual calculation of the credit is one area that is being examined for potential improvement or enhancement.

Table 9. Tax Unit Level Earned Income Tax Credit (EITC) for the Current Tax Model and the New Tax Model for Both CPS and SIPP Compared to IRS Figures

	IRS Published Data	Current Model Using CPS	New Model Using CPS	New Model Using SIPP
Percent With EITC	16%	13%	13%	12%
Mean EITC	\$1,567	\$1,358	\$1,244	\$1,232
Median EITC	\$1,341	\$1,190	\$920	\$936

In addition to a comparison of the new model to the old model and IRS published data, we have examined how the new model runs on the SIPP data compared to results from the SIPP tax topical module. Because of the inconsistencies with the tax topical module and amount of missing data, we used a subset of the topical module households. The subset was created by choosing respondents that used a copy of their tax return when answering the topical module and had AGI from the topical module that was within +/- 10% of the total household income from the core files. We also completed the analysis on a household basis because of potential differences in how the tax model might create tax units and how the tax units are reported in the topical module.

Table 10 shows the results of the comparison between the tax model and the topical module. The comparisons of AGI, after credits tax liability, EITC, and capital gains are shown by AGI income groups. The table demonstrates that the tax model is working well. The AGI amounts

and the tax liability generated by the model are very consistent with the reported figures from the tax module. The amounts for EITC and capital gains are less consistent than we would like and we will continue to refine the method for calculating these figures.

Table 10. Comparison of Mean Values for Tax Module Responses and New Tax Model Calculations

Base of Respondents Who Used Tax Return When Answering Topical Module and Had Adjusted Gross Income From Tax Module Within 10% of Total Household Income

Adjusted Gross Income Group	N	Tax Module Means				New Tax Model Means			
		Adjusted Gross Income	Taxes After Credits	Earned Income Credit	Capital Gains	Adjusted Gross Income	Taxes After Credits	Earned Income Credit	Capital Gains
Less Than \$15,000	100	\$11,095	\$604	\$102	\$40	\$11,505	\$659	\$18	\$12
\$15,000 to \$24,999	183	\$20,604	\$1,864	\$5	\$91	\$22,547	\$2,355	\$4	\$771
\$25,000 to \$39,999	291	\$32,021	\$3,780	\$0	\$169	\$34,485	\$3,999	\$8	\$1,214
\$40,000 to \$54,999	207	\$47,008	\$6,467	\$0	\$189	\$50,701	\$6,706	\$4	\$1,606
\$55,000 to \$74,999	144	\$62,890	\$9,065	\$3	\$735	\$65,176	\$9,101	\$0	\$1,598
\$75,000 to \$99,999	61	\$85,022	\$13,521	\$0	\$1,111	\$88,925	\$13,155	\$9	\$2,743
\$100,000 and Over	49	\$145,307	\$28,867	\$0	\$13,364	\$140,811	\$28,534	\$5	\$1,435

Income Differences in CPS and SIPP

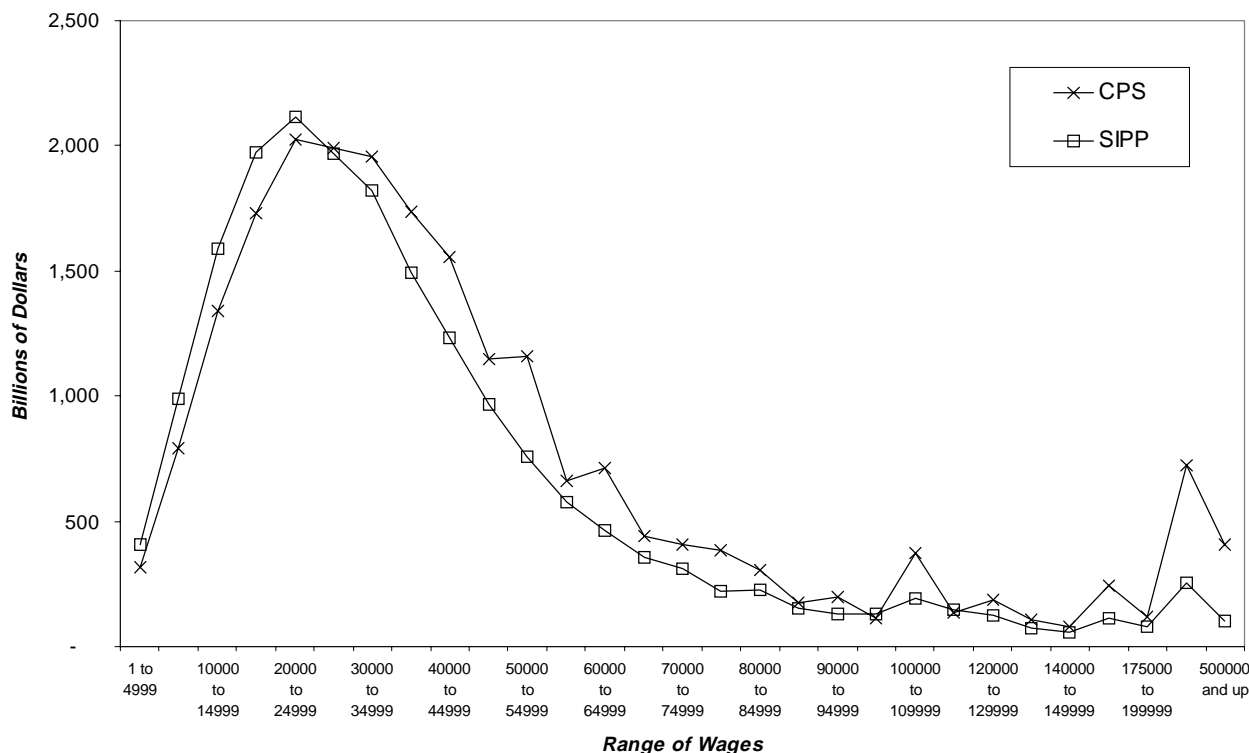
As mentioned above, some of the differences in the tax model between SIPP and CPS occur because of differences in incomes between the two surveys. Previous work has shown a general under reporting of wages in the SIPP survey compared to CPS.⁴ Chart 1 shows the aggregate wage differences between the two surveys. While SIPP collects higher wage amounts for lower income workers, it falls below CPS for wages over \$25,000 a year. The higher incomes will have a larger effect on taxable income and tax liability.

The wage differences between the two surveys are thought to be due to response error and the method of survey administration. Because SIPP respondents are surveyed more frequently and are asked to report monthly amounts, it is believed that SIPP collects more accurate information for low wage earners. Low wage earners are more likely to have multiple jobs and change jobs more frequently. SIPP is more likely to collect all wage information since the respondents are interviewed closer in time to the actual work.

The difference in higher end wage information may be caused partially by reporting error of individuals. SIPP asks for wage amounts on a monthly basis and it is believed that some respondents report take-home pay rather than gross pay on a monthly basis. SIPP may also miss periodic payments such as bonuses or awards that may be reported on a yearly basis.

⁴ Roemer, Marc I., "Assessing the Quality of the March Current Population Survey and the Survey of Income and Program Participation Income Estimates, 1990-1996", U.S. Census Bureau, Housing and Household Economic Statistics Division, Staff Papers, 200.

Chart 1
1996 Comparison of Aggregate Wages for SIPP and CPS



These differences in wage reporting explain some of the variation found in the tax model between SIPP and CPS.

Future Tax Model Enhancements

The initial results of the new tax model demonstrate that it can generate more accurate and detailed tax information than the current model. It also provides better tax information for SIPP respondents than is available from the tax topical module. However we want to continue to examine methods of improving the new model.

One method we are examining is the use of some of the information from the tax topical module rather than completing a statistical match with the SOI for all cases. The respondents from the topical module that use their tax returns while responding generally have more complete and, we believe, more accurate responses to the quantitative items. It may be more accurate to use the respondent answers in this situation rather than a statistical match. It could also be beneficial to use the flags indicating whether the person claimed various items or deductions, even if we don't have accurate information on the amounts. A flag for a respondent could be used to force a match to SOI in which the respondent would be assigned an amount, rather than the possibility that the respondent could be assigned a zero dollar value for the item. The use of tax module information will need to be explored more thoroughly since there may be changes in the questionnaire that could effect the availability of data. We would need to weigh any changes in

the questionnaire against the use of this information for the tax model before implementing any changes.

Appendix I

Comparison of Current and Revised Tax Simulation Models for the March Current Population Survey (CPS)

For the past 15 to 20 years the Census Bureau has maintained a set of FORTRAN-based programs that were designed to simulate the payment of taxes based on information collected in the March Current Population Survey (CPS). A revised set of SAS-based programs has recently been written to provide an alternative tax simulation model that can be used for both the March CPS and the Survey of Income and Program Participation (SIPP).

Both the “old” (model currently used for the March CPS) and new (SAS-based model) tax simulation models attempt to compute taxes paid by households based on the income and related data collected in the surveys. Both include models for computing payroll taxes, federal individual income taxes, and state individual income taxes. While there are great similarities between the two models there are important differences in the details on how the simulations are carried out. The similarities and differences are outlined in this document.

Payroll and Self-Employment Taxes

Simulations of FICA, self-employment taxes, and federal civil service retirement contributions are relatively straightforward given that they are applied on individuals without regard to their living arrangements and because the tax rates and the amounts to which they apply are not complicated. Some complications arise in the simulations because not all workers are in “covered employment” and are thus not subject to the tax. Workers not covered include federal civilian employees hired before January 1, 1984, certain employees of government and state and local governments, railroad workers, and household and farm workers whose earnings do not meet certain minimums.

Old Model

The old model contains provisions to simulate coverage for certain groups of workers noted above. These simulations of coverage include household and farm workers, state and local workers, and federal workers hired after January 1, 1984. Monte Carlo simulations are used to assign coverage to household and farm workers and state and local workers. Earnings levels are used to define the probabilities of coverage within groups. Unfortunately, the probabilities of coverage are based on very outdated information and for which more recent information is not available. For federal workers, those under age 32 are assigned coverage under FICA rather than the Civil Service Retirement System (CSRS). For those federal workers assigned coverage the FICA amount was computed in a routine manner. For those not assigned coverage, a separate amount was computed that reflects the rules and rates governing the CSRS system. Aside from the use of outdated coverage probabilities, there are two other “holes” in the old payroll tax simulation. The first is that no payroll tax is simulated for persons working in the railroad industry. This was an oversight that has never been corrected. The second is that no alternative “mandatory payroll tax” (FICA equivalent) is computed for state and local workers who are not assigned coverage in the Monte Carlo simulations. Since all state and local workers must either

be covered by FICA or by some other state or local alternative, all must make mandatory contributions of some type to a pension system.

The old model includes a simple simulation of self-employment taxes. It does not simulate a self-employment tax amount but rather integrates the computation with that of the FICA tax on wage and salary income. As such it does not include the provision that reduces the self-employment tax by multiplying by .9235 (contained in Schedule SE).

New Model

In terms of coverage under FICA the new model assumes all household and farm workers to be covered. While some non-covered employment still exists, it must be very small. Coupled with the lack of data needed to simulate undercoverage any attempt to simulate the low level of undercoverage was not attempted. Coverage of state and local workers was simulated using Monte Carlo techniques. The proportions of workers covered under FICA were used on a state by state basis (1992 data available from the Green Book) to assign coverage. The simulation did not employ different coverage rates by earnings category as is the case with the old model since no such data were found. State and local workers who were assigned coverage under FICA had their simulated tax amounts computed in a straightforward manner. For those who were not assigned coverage, a “mandatory payroll tax” amount was computed using the same rate as the FICA tax. As all must make mandatory contributions to a pension plan, this simulation was used to assure that some payroll taxes were accounted for. The amount of this simulated mandatory payroll tax was created separately from FICA so that a “clean” FICA amount could be derived from the simulation. For federal civil service employees coverage under FICA was simulated using Monte Carlo techniques. The age of the worker was not used in the simulation as it was in the old model. Office of Personnel Management data showed 44.5 percent of federal workers were covered under the CSRS system in 1997 so the simulation assigned coverage to that pension system at that rate and at the 55.5 percent rate to FICA. Railroad workers were treated as if they were covered under FICA even though their pension system may have slightly different tax rates. Throughout this simulation, separate simulated amounts were maintained for the separate elements of the FICA tax. Unlike the old model, separate amount were computed for each of the three pieces of the FICA tax; old age and survivors insurance, disability insurance, and hospital insurance.

The computation of self-employment tax was integrated into the simulation of the federal individual income tax. Most aspects of the Schedule SE were implemented in the simulation including the deduction of one half the self-employment tax from adjusted gross income.

Federal Individual Income Taxes

The old and new approaches to simulating federal individual income taxes diverge in a number of important areas. In general, the new simulation model was designed to follow the IRS forms and worksheets much more closely than the old model. In addition, the simulation of key items not available from the survey, such as capital gains or losses, itemized deductions, IRA and Keogh contributions, child care expenses, etc. were handled in a very different manner. The old method used Monte Carlo simulations based on data provided by the IRS to simulate the presence of these items and mean values supplied by the IRS when appropriate. The new

method used statistical matches of the survey-based tax units to the IRS Statistics of Income (SOI) public use file to obtain this information. Also of importance is the provision in the old model that declares some potential tax units as non-filers because the earnings or income levels are below certain thresholds. The new model makes no attempt to define filers and non-filers but simply models the tax liability for all potential units even though it may be zero for a large number of those with low earnings. This may be important since the old model does not provide any probability for these low-earner simulated tax units to receive the earned income tax credit.

Neither model attempts to simulate amounts of deferred earnings (earnings contributed to employer sponsored IRA's thrift plans, etc.). This is potentially a very important fault in these models as the number and amount of deferred earnings has grown dramatically in recent years. Since the survey collects gross annual earnings amounts, the simulations overestimate the amount of earnings used to compute adjusted gross income.

Both models simulate three types of filing units: 1) single, 2) married, joint, and 3) head of household and assume that a Form 1040 is used to file.

Old Model

This simulation method, which has been in place with only minor modifications and updates for many years, handles the basic elements of the federal individual income tax filing process in developing simulated federal tax liabilities for survey households. Potential tax filing units (and corresponding dependents) are formed by examining the income, family relationship, and age of each household member. Total income is computed by summing the amounts of taxable income by source coming directly from the survey and any amount of capital gain or loss derived from the Monte Carlo simulation. Next, adjusted gross income is computed by subtracting only one statutory deduction, that being IRA contributions. These contributions were also derived through a Monte Carlo simulation. Amounts of capital gains (losses) and IRA contributions were derived from a table of current year mean values provided by the IRS and based on returns processed through mid-summer. The computation of taxable income is then made by subtracting 1) either the standard deduction or itemized deduction whichever is larger and 2) the personal exemptions amount. The presence of an itemized deduction amount is also simulated using a Monte Carlo method and amounts are assigned from a table of mean ratios of itemized deductions to adjusted gross income provided by the IRS. Cells of this table of mean deduction amounts are defined by type of return and adjusted gross income category, presence of a mortgage, and mortgage amount category. Taxes before credits are computed using the applicable tax rates. Both the Earned Income Credit (EIC) and the child care credit are simulated. The EIC simulation covers the basic elements of the of the computation but does not include a number of tests including the asset income test which limits eligibility to those receiving less than \$2,250. There is also no attempt to compute a "modified" adjusted gross income amount. This amount removes the effects of losses and other deductions involved in the computation of adjusted gross income. Presence of childcare expenses is simulated using Monte Carlo methods and the amount of the childcare credit is assigned from a table of mean values. This table is defined by number of children and adjusted gross income level.

The simulation of itemized deductions does not fully integrate the simulated state individual income taxes or the property tax amounts. The itemized deduction is simulated independently of

either of these items but is replaced by the sum of these two items if the initially assigned value lower than this sum.

New Model

The SAS-based simulation of federal individual income taxes is more closely tied to the details of the forms, schedules, and work sheets than the older simulation methods. For many of these the simulation is structured on a line-by-line basis. The method for establishing potential tax units differs somewhat from that of the older model. The regulations for determining dependency adhere more closely to the IRS rules and no attempt is made to define a potential unit as a “non-filer”. Options have also been included that permit some of the rules of dependency to be changed by simply changing a parameter line. The methods for obtaining exogenous items such as itemized deductions, capital gains, and statutory deductions differs significantly as did the way in which this information was integrated into the simulation. All of this information was obtained based on a statistical match between the IRS’s 1995 SOI public use file and the tax units derived from the survey (this is the latest SOI file available). In addition, a statistical match is made with the American Housing Survey to generate a variable indicating whether or not owner occupied households had a mortgage. This is then used to construct the matching keys linking the SOI and the CPS tax units. Unlike the old method, computation of itemized deductions is fully integrated with the simulation of state income taxes. The itemized deductions generated from the statistical match exclude amounts of state income taxes as these are taken directly from the state tax simulation. Simulations of the EIC and child care credits include more details of the actual work sheets than the old model. Simulation of the dependent care credit and the credit for the elderly and disabled are included, but absent from the old model. Statutory deductions were expanded to include Keogh and SEP contributions, the deductible portion of health insurance premiums paid by the self-employed, and the deduction of 50 percent of the amount of self-employment taxes paid, all not simulated in the old model.

State Individual Income Taxes

The new SAS-based simulations of state individual income taxes simulations are very much different than the old model. It is more detailed and up-to-date in many ways. The details incorporated into the new model include many exemptions, deductions, and credits related to retirement, disability, and age that are not simulated at all in the old model. They also include state EIC and state childcare credits where applicable, neither of which were included in old model used for the 1997 tax year.

Old Model

The latest simulation of state taxes was based on rules, regulations, and laws applicable to tax year 1995 (income received in 1995) even though the income year was 1997. These simulations are updated to reflect changes in tax law only once every five years. The basic approach starts with the adjusted gross income derived from the federal tax simulation. Exemptions and deductions are made in a very simple way to arrive at taxable income and then tax rates are applied as needed to compute the simulated tax amount. There is no use of itemized deductions in the computation of taxable income. Only standard deductions are applied. There is also no adjustment made to federal adjusted gross income to remove the taxable portion of social security benefits. Virtually all states exclude social security income from taxation.

New Model

The new model was constructed based on a detailed review of each states tax return forms. This review lead to the simulation of many of the details that were not incorporated in the old model. Some of these include 1) exclusion of military pay from taxation in some states, 2) exclusion of social security income from taxation, 3) exclusion of all or part of pension and disability income from taxation in many states, 4) simulation of special New York City tax, 5) simulation of state EIC and child care credits, 6) simulation of special aged exemption amounts in many states, and 7) inclusion of itemized deductions in computation of taxable income. The new model also uses the amount of state income tax as a direct component in the computation of the amount of itemized deductions on the federal income tax return. As noted earlier, the old model appears to make no use of the simulated state income tax amount in the computation of itemized deductions.