New Methods for Simulating CPS Taxes

Amy O'Hara
Modeling and Outreach Branch

This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on methodological issues are those of the author and not necessarily those of the U.S. Census Bureau.

INTRODUCTION

The Census Bureau produces federal, state and payroll tax estimates for the Current Population Survey Annual Social and Economic Supplement (CPS ASEC). These estimates are used when calculating the alternative income definitions used in the Income in the United States report series and the National Academy of Sciences (NAS) income definitions used in the Experimental Poverty Measures report series. Many of the variables are also released on the CPS public use file.

In the fall of 2004, the Census Bureau has implemented a new model to produce tax estimates. The new model closely follows the Internal Revenue Service (IRS) 1040 tax form and rules on filing unit formation and dependent assignment. More statutory adjustments and credits are simulated in the new model compared to the old tax model. The Census Bureau will continue to release the thirteen tax-related variables on the public use person file that have historically been included.

While the old tax model had been updated annually to account for changes in marginal tax rates, the underlying methodology had not been revised since the model's inception in the early 1980s. The new model will be continuously updated to reflect annual changes in the tax code and will investigate the simulation and inclusion of more items on the individual tax return. The new model will also be used to develop tax estimates for the Survey of Income and Program Participation (SIPP).[1]

The remainder of this paper provides an extensive explanation of the new tax model. The methods used to sort the data into filing units, identify exemptions from relationship codes, and calculate taxes are described. Note that filing units are constructed from person-level survey data; they are not equivalent to households as defined by the surveys. IRS rules are used to determine filing requirements and dependent status, which determine the number of exemptions a return may claim. After the new model is described, comparisons of the old, new and IRS data for tax year 2002 are presented.

NEW MODEL OVERVIEW

The new tax model simulates three taxes using a revised methodology: 1) payroll, 2) federal individual income, and 3) state individual income. At the current time, the methodology for estimating property taxes has not changed from the old model.

The new model first calculates payroll taxes for every person with earned income. Next, potential filing units are formed based on marital status and household

[1] SIPP calendar year files will be created, and filing units classified using the new CPS method. Including SIPP topical module data will necessitate fewer imputations, and tax calculations will follow the same iterative process described in this paper for the CPS estimates.

relationships, and return types are assigned. A statistical match to the Statistics of Income (SOI) file from the IRS allows tax fields such as capital gains, statutory adjustments, and itemized deductions not collected in the survey to be imputed. The individual income tax calculations follow an iterative process. Preliminary federal taxes and credits are computed, and then used in the state tax calculations (which often require fields such as federal adjusted gross income (AGI), itemized deductions, earned income tax credit (EITC)). After state taxes are computed, the final federal estimates are generated, substituting the new state tax numbers in the itemized deduction calculator for itemizing filers.

Payroll Taxes

Payroll taxes are computed for each person over age fourteen reporting income in the CPS ASEC. The Medicare portion of payroll taxes is calculated for all earners. To calculate values for the retirement (OASI) and disability (DI)[2] portions of Federal Insurance Contributions Act (FICA) taxes, covered workers are identified.[3] The proportions of federal, state and local employees who are covered under FICA are estimated. The model assumes that 29.5% of federal workers are still covered under the Civil Service Retirement System (CSRS) in 2003 (32.1% in 2002, according to the Office of Personnel Management). The proportions of covered state and local government workers, for each state, were obtained from the Green Book.[4] The FICA variable still represents payroll taxes for all covered employees, but the FED_RET variable now includes payroll tax deductions for state and local as well as federal government employees. The old model did not model payroll taxes for non-covered state and local employees.

Tax Unit Formation

A person-level set of potential filers is created, consisting of persons over age 14 who are neither the reference person nor spouse. Subfamilies (related and unrelated) are flagged, as are secondary individuals. All are assigned single filing status with one potential tax exemption.

A subfamily-level set of potential filers is created, consisting of related and unrelated subfamilies, secondary individuals, and primary families living in group quarters. Each subfamily tax head[5] is identified. Married filing joint status is assigned to those who are married or married with spouse absent in Armed Forces. All others are assigned single filing status. Exemptions are tabulated for these subfamily filers.

---

[2] OASI refers to the Old Age and Survivors Insurance component of Social Security. DI refers to the Disability Insurance component.

[3] The model assumes that "news vendors" under 18 and work-study or student assistants under 25 are not covered by FICA.

[4] Committee on Ways and Means, U.S. House of Representatives, 2000 Green Book 106-14.

[5] The tax head is the person assumed to file the tax return. Tax head status is determined by applying filing IRS rules to CPS relationship codes.

A primary family set of potential filers is created, consisting of primary families and non-family householders (and non-family reference persons in group quarters). Potential exemptions and children are counted for each head.

Exemptions for self and spouse are assigned. Exemptions for dependents are identified as: 1) Persons in unit under age 19,
2) Persons in unit under age 25 who are enrolled in school, or
3) Persons in unit over 18 earning less than the IRS threshold

Filing status (single, married or head of household) is set as follows: Married is assigned to those who are married or married with spouse absent in Armed Forces. All other reference person tax heads with more than one exemption are assigned head of household filing status. Single status is given to non-reference person tax heads, all with one exemption.

Imputation of Missing Tax Fields

The most recent Statistics of Income (SOI) public use file (2000) is used in a non-constrained statistical match[6] to impute variables necessary to complete the tax return simulation. These variables are capital gains and losses, itemized deductions, IRA contributions, self-employed health insurance premium deductions, self-employed SEP and SIMPLE[7] deductions, and childcare expense amounts.

The SOI file is limited to its year 2000 responses, and only observations with assigned states are used. This effectively removes the IRS oversample of extremely wealthy filers. The IRS removes the state assignment for those with very high income or losses. The distribution of returns with states included parallels the CPS distribution and includes enough high-income returns for matching purposes. SOI cases with losses exceeding negative $10,000 in business income (Schedule C), supplemental income (Schedule E) or farm income (Schedule F) are omitted from the donor pool to better align with the CPS distribution. This amount was chosen because the CPS variables for self-employed income, self-employed farm income, and rent income are bottom-coded at negative $9,999. Negative values are not permitted in the "other income" variable, OI_VAL, from which Schedule E components are drawn (estate/trust income and royalties).

To prepare for the statistical match, the SOI sample is restricted to single, married joint and head of household filers. Partitions within these return types are made. In the CPS, the set of potential primary filing units includes those records that

---

[6] See "Summary Comparisons Between IRS Published Statistics and Current and New Tax Simulation Models for Income Year 1997, " by John Coder, Sentier Research, LLC for details.
[7] SEP refers to Simplified Employee Pensions; SIMPLE refers to Savings Incentive Match Plans for Employees of Small Employers. Keogh plans are also included in this summary variable of self-employed savings plans.

filed single, married or head of household. Subfamily units could file either single or married. Person-level units, all single filers, are divided into two groups: 1) tax heads under age 23 who are assumed to be dependents on another return, and 2) tax heads age 23 and over.

Common variables between the two data files are condensed into categorical variables for the statistical match. These variables include: income, filing status, presence of earned income, presence of self employment income, presence of unearned income, presence of social security income, presence of mortgage, number of child exemptions, state of residence, presence of pension income, and whether person is a dependent on another return.

For all three match groups (household, subfamily, and person), the main income variable contains wages and salaries, farm and non-farm self-employment income, total interest income, Social Security income, dividends, alimony, total IRA and pension distributions, rental income, royalties, estate/trust income, and unemployment compensation. This amount is roughly equal to IRS total income minus capital gains, other gains, refunds, and other income.[8]

The subset of singles under 23 is composed of dependents (children, grandchildren, siblings or other relatives) on other tax returns. They are not permitted to have a mortgage and are assumed to take the standard deduction. Persons over 18 who have total income exceeding the $7700 filing threshold are not viewed as dependents and are considered in the 23 and over statistical match. The mortgage flag is not currently used in this match.[9]

A separate statistical match is developed for potential subfamily filers. Dependent filers are omitted from the SOI pool. SOI head of household filers are recoded as singles; the Census tax model assumes that the primary household takes priority in filing married or head of household. The primary filer is assumed to provide at least fifty percent of household needs. The statistical match for potential primary filers restricts the donor pool in the same manner as the subfamily match: no dependent filers are permitted and certain large losses are excluded. However, SOI filers retain their single, married or head of household status for the primary family match.

Note that all statistical matching techniques used in the model are under review for improvement. Details on the current technique are in Appendix One.

---

[8] Differences in income reporting in CPS and SOI complicate the construction of income match keys. Income responses to the IRS and Census differ substantially; see Appendix One for a comparison of income components between the two agencies.

[9] Exception: If the tax head is a partner/roommate who owns a home, has a mortgage, and has income past the filing threshold, they are assigned half of reported property taxes. They keep their mortgage flag, and are not coded as dependents.

The statistical match determines a SOI donor observation for most CPS cases. Values for capital gains, itemized deductions, IRA contributions, self-employed health insurance premiums, self-employed SEP and SIMPLE deductions, and childcare expense amounts are assigned from the designated SOI donor. Prior to the statistical match, capital gain and loss amounts from the SOI 2000 were aged forward to tax year 2002 using IRS data grouped by AGI class. The number of people receiving capital gains and losses changed dramatically between 2000 and 2002. IRS aggregate statistics indicate that 52% fewer returns had capital gains in 2002 than in 2000.[10] Thus, the incidence of capital gain assignment had to be adjusted as well. IRS data on the number of returns with capital gains was used to randomly deselect capital gains from CPS observations for tax year 2002. Adjustments to the number of returns with capital losses were not attempted. While the number of returns with losses nearly doubled between 2000 and 2002, the size of such losses could range from negative $3000 to zero. In the future, the model may be further adjusted to randomly assign losses in that range to filers by income class.

Values for itemized deductions are also adjusted before the statistical match to age the 2000 data forward. Again, IRS tabs from the SOI by AGI class were used. After the match, these donor values were assigned to the CPS observations. The incidence of itemizing was not adjusted, as the increase between 2000 and 2002 was not very large. The components of itemized deductions (medical expenses, interest paid, taxed paid, gifts to charity, casualty and theft losses, job expenses, and miscellaneous expenses) used in the model were also aged.

IRA contributions assigned from the statistical match to working filers are prorated between married filers if both are employed. In accordance with IRS rules, a $2000 limit per person is applied. In 2002, IRA contributions were nearly 29% higher than in 2000. Before the statistical match, the SOI values were adjusted by this factor. After the match, if the donated amount exceeded the reported CPS wage amount, it was limited to 25% of wages.[11]

Statutory adjustments on self-employed health and retirement savings are aged by their aggregate growth between 2000 and 2002. When both the SOI and CPS matched cases report self-employment income, the amount of self-employed health insurance premiums paid and SEP/SIMPLE deductions are assigned from the donor record. Donor values are adjusted by the ratio of CPS to SOI self-employment income. If the values exceed the reported CPS total income amount, they are limited to 25% of total income.

Childcare expenses in the SOI are amplified by 25% to age forward to 2002. This follows the IRS allowable limit on deductible expenses, which increased from

---

[10] "Individual Income Tax Returns, Preliminary Data, 2002." *SOI Bulletin*, Winter 2003-2004.
[11] In the SOI 2000 public use file, 90% of those reported IRA contributions of one-quarter their wages or less.

$2,400 per child to $3,000 per child. After these imputations, all fields currently modeled on the IRS 1040 form are available.

Initial Federal Tax Calculation

The model first counts tax exemptions for each tax unit. Exemptions for self and spouse (if present) are counted. Next, other persons in the tax unit meeting the following criteria are counted as exemptions: persons under 18, students ages 19-24, and those over age 18 with income less than $3000 in 2002, $3050 in 2003. Internal flags are established for children eligible for the childcare expense credit (if under age 13), for the EITC (if under age 19 or students under 24), and for the child tax credit (if under 17). Persons under 65 who are disabled and have no income are marked as eligible for the elderly or disabled care expense credit.

Federal taxes are simulated for each potential tax unit, closely following IRS Form 1040. Separate programs for each schedule and worksheet calculate income, credit and tax amounts. As mentioned before, the tax model makes an initial estimate of federal taxes. In the first pass of the federal model, AGI, taxable income, taxes and credits are determined. Taxes are calculated in one of three ways: standard using marginal tax rates, from Schedule D using maximum capital gains rates, or from Form 6251 using the Alternative Minimum Tax rules. These values are then used as input in the state tax model.

State Tax Calculation

Once the model has calculated the federal tax amounts, state taxes are simulated according to each state's individual income tax return. Most states start with the federal AGI and credit amounts. Some have different allowable itemized expenses or calculate credits differently than the federal model. States also vary in their offering of refundable credit programs. Appendix Two offers an overview of what was modeled for each state, and a comparison of aggregated state tax estimates from the old and new CPS tax models for 2002.

Final Federal Tax Calculation

Once the model has calculated the state tax amounts, the estimates are merged back to the filing records. The same programming reruns the federal estimates, this time incorporating the state tax estimates for those who itemize.

Final filing units were established for subfamilies, primary families, and persons. When subfamilies were processed, any persons not marked as exemptions were removed and offered to the primary family processing. When IRS rules allowed, these persons were counted on the primary returns. Those not permitted were released to the person level processing. In this manner, the exemption status of each CPS person record was evaluated and assigned in accordance with IRS rules during final processing. Tax estimates were produced on the final tax units

having incorporated as much information as possible from the statistical match and iterative tax calculating process.

To determine who has a filing requirement, five conditions were evaluated: 1) Units with income exceeding the filing threshold for their filing status must file; 2) Units who do not meet the filing threshold but would receive an earned income credit must file; 3) Units with $400 or more of self-employment income must file; 4) Units with negative gross income must file;[12] and 5) Units with negative reported (net) self-employed or farm self-employed income must file.

Filing status (FILESTAT) was assigned to every unit with a filing requirement. Knowing that some people file though they are not required to, an exception was made to allow single and head of household returns with gross income over $3,000 to file.

KEY CHANGES

The new tax model differs from the old tax model in several important areas. An overview of the old tax model can be found in Appendix Three. The key changes between the two models involve: tax unit formation, imputing missing tax variables, simulating statutory adjustments, and estimating state taxes.

The new tax model allows non-relatives as dependents in the filing units created from the person level survey data. The new model also allows single returns to have more than one exemption. These changes impact adjustments to AGI when computing taxable income, and may alter eligibility for credits.

The method for imputing missing tax fields such as capital gains and itemized deductions has changed the most between the old and new models. The new model employs a statistical match to the SOI file; the old model imputed means. The values obtained from the statistical match add variability to the variables that was missing in the past. The new model also adjusts the 2000 SOI data forward based on additional IRS data to improve the value ranges; the old model used unadjusted data on a one-year lag.

For the first time, the new model simulates several statutory adjustments. In the old model, AGI was computed by adding imputed capital gains to reported income amounts. The new model includes IRA contributions, self-employed health insurance deductions, and self-employment SEP/SIMPLE deductions.

State tax estimates now include more information specific to each state's individual income tax return. More refundable and non-refundable credit

[12] Gross income must be estimated for the CPS because it collects only net amounts for business and farm business income. Ratios of gross to net income in the SOI were computed and applied to CPS for farm and non-farm self-employed. The ratios were computed separately for those reporting positive and negative amounts.

programs are also simulated. The old model estimated an amount for state EITC, but truncated the amount at zero. The new model allows the variable (STATEEITC) to be refundable, thus some state tax values are negative. The changes in the new model produce very different estimates in some states.

RESULTS

Table 1 illustrates differences in AGI classes between the old and new CPS models and preliminary IRS statistics for 2002. The CPS AGI includes simulated items such as capital gains and statutory adjustments. The large group of filers in the lowest AGI class in the IRS data includes those who file a tax return despite not having a filing requirement. These could include persons filing to recapture withholding, or those who need to file a federal return to become eligible for a state credit program.

Table 1. Adjusted Gross Income (AGI) Distribution, Tax Year 2002

AGI Range	CPS Old Model	CPS New Model	IRS Published Data[13]
Less than $15,000	26.5%	23.1%	29.3%
$15,000 under $30,000	22.1%	23.1%	23.0%
$30,000 under $50,000	19.3%	19.9%	18.9%
$50,000 under $100,000	22.0%	23.4%	20.5%
$100,000 under $200,000	8.2%	8.7%	6.5%
$200,000 and over	1.9%	1.9%	1.9%

As seen in Table 2, aggregate AGI in the new model is now closer to the IRS value. AGI is impacted by the imputation of capital gains and losses, IRA contribution deductions, and several self-employed deductions that were modeled for each reported self-employed person in the CPS this year.

The new tax model completes a tax form for each filing unit including credits not simulated in the old model such as the dependent care expense credit and the credit for the aged and disabled. The new model also updated dependency rules in unit formation, allowing the inclusion of more non-relatives. These changes have lowered the mean taxable income and federal tax estimates compared to the old model. When comparing them to IRS estimates, it is important to recall that the Census tax estimates assume that the tax unit will take advantage of every available credit to its legal limit. The IRS data reveal what taxpayers actually filed, not what they were eligible to file.

[13] Statistics in this and other Results section tables are from "Individual Income Tax Returns, Preliminary Data, 2002." *SOI Bulletin*, Winter 2003-2004. Per capita amounts are shown from this source, as weighted means are unavailable from the aggregates reported.

Table 2. Tax Unit Level Comparison of AGI, Taxable Income and Federal Taxes After Credits, Tax Year 2002

	CPS Old Model	CPS New Model	IRS Published Data
Mean AGI	$48,867	$48,678	$46,385
Median AGI	$30,309	$32,397	--
Aggregate AGI	$6,265,476,178	$ 6,004,516,003	$6,039,405,382
Mean Taxable Income	$41,020	$37,528	$40,005
Median Taxable Income	$23,825	$23,550	--
Aggregate Taxable Income	$4,166,707,118	$3,893,238,117	$4,099,015,901
Mean Federal Tax	$7036	$6,452	$8,759
Median Federal Tax	$2,445	$2,472	--
Aggregate Federal Tax	$767,943,297	$724,473,825	$797,791,644

Comparing return types across the old model, new model, and IRS published data, Table 3 indicates that the old and new models are very similar. Slightly fewer units are given single filing status, and slightly more receive head of household status. This is due to the revised method of assigning dependents to returns; the new model does not require them to be related, just to meet the IRS criteria. Both Census models differ from the published IRS statistics. The Census tax model builds from survey responses collected at the household level. Filing units are created using data reported on who lives at the address, not whom is dependent on those living at the address. Therefore, the Census filing units will inaccurately build filing units for persons who support children or parents not living in the household.

Table 3. Types of Returns, Tax Year 2002

	CPS Old Model	CPS New Model	IRS Published Data[14]
Single	42.9%	42.5%	39.6%
Married, Filing Jointly	9.9%	10.0%	16.5%
Head of Household	47.2%	47.5%	43.9%

Table 4 compares amounts generated by the old and new tax models for each filing class. For single returns, the new model produces slightly higher AGI due to more extensive modeling of capital gains and statutory adjustments. Taxable

[14] To compare with the CPS filing categories, IRS return types were collapsed by adding qualifying widows to head of household filers, and married-separate filers to single filers.

income and federal taxes are lower for the new model due to the inclusion of more credit programs and better adherence to the IRS rules.

For married-joint returns, the new model has lower mean and median taxable income compared with the old model. This could be due to the unit construction (more dependents are eligible for inclusion in the new model which would impact the amount of allowable exemptions), the additional credits simulated, and broader assignment of itemized deductions from the statistical match.

For head of household filers, the new model estimates of AGI, taxable income and federal taxes are larger than in the old model. In the new model, 45.6% of head of household units have a negative federal tax amount; the median after-credit tax amount for non-zero responses was negative $116. In the old model, 57.4% of such returns had a negative federal tax amount; the median after-credit tax amount was negative $840. Another credit impacting the range of tax estimates is the childcare expense credit. The new model assigns childcare expenses from the SOI statistical match; the old model assigned a mean value randomly. In the old model, more than twice as many head of household returns were given a childcare expense credit to align the aggregate with the IRS aggregate. This lowered FED_TAX before deducting EIT_CRED, and explains the lower values in the old model.

Table 4. Tax Unit Level Comparison of AGI, Taxable Income and Federal Taxes After Credits by Return Type, Tax Year 2002

	CPS Old Model		CPS New Model	
Single	Mean	Median	Mean	Median
AGI	$27,870	$19,030	$27,999	$20,000
Taxable Income	$23,607	$15,568	$22,090	$15,300
Federal Tax	$4,196	$1,930	$3,776	$1,809
Married, Filing Jointly				
AGI	$76,712	$58,085	$76,512	$61,250
Taxable Income	$61,507	$42,635	$55,802	$41,850
Federal Tax	$11,290	$4,871	$10,270	$4,806
Head of Household				
AGI	$28,136	$21,700	$28,797	$22,575
Taxable Income	$20,841	$13,021	$20,936	$14,100
Federal Tax	$556	-$840	$1,030	-$116

Imputing capital gains and losses is a challenging part of the Census tax simulation. The old tax model imputed means from IRS tables on a one-year lag. The new model relies on a statistical match to the most recent SOI public use file. Tax year 2002 used the 2000 SOI, which was released in 2004. The SOI data had to be aged forward to account for both lower filing rates and lower amounts of capital gains.

As seen in Table 5, the new model has far lower mean capital gains than the value computed from the IRS data (aggregate capital gains divided by the number of filers). The old tax model was, by design, generating gains in alignment with the IRS published results. This method failed to assign many gains at the lower end of the income distribution, and assigned excessive values to select cases at the upper end of the income distribution. The old model randomly assigned a mean value to units depending on their AGI class; the new model uses the adjusted value from the SOI donor chosen in the statistical match. The match allows substantial variation in the distribution. In the new model, the statistical match excluded IRS-designated high-income returns, lowering aggregate value for the new model. The IRS mean is quite high due to the presence of these high-income returns that claim gains that elevate the mean and total amounts.

Table 5. Comparison of Capital Gains and Losses, Tax Year 2002

	CPS Old Model	CPS New Model	IRS Published Data
Percent with Capital Gain	12.3%	8.8%	8.3%
Percent with Capital Loss	10.5%	8.1%	10.2%
Mean Capital Gain	$22,832	$7,611	$22,821
Median Capital Gain	$6,184	$2,258	--
Aggregate Capital Gain	$320,977,654	$82,695,595	$246,831,535
Mean Capital Loss	$2,122	$2,399	$2,248
Median Capital Loss	$2,112	$3,000	--
Aggregate Capital Loss	$26,173,735	$23,978,666	$29,898,639

The statistical match was also used for the assignment of itemized deductions in the new model. As indicated in Table 6, estimates from the new model are closer to those published by the IRS. The new model has higher mean and median values compared with the old model, and the aggregate amount is closer to the IRS published data.

Table 6. Comparison of Itemized Deductions, Tax Year 2002

	CPS Old Model	CPS New Model	IRS Published Data
Percent who Itemized	75.5%	37.1%	35.0%
Mean Itemized Deduction	$14,548	$16,990	$19,293
Median Itemized Deduction	$11,780	$13,180	--
Agg. Itemized Deductions	$1,254,713,253	$777,544,393	$879,237,828

In the old tax model, AGI was equivalent to IRS total income, because no statutory adjustments were modeled. The new tax model simulates several statutory adjustments that may reduce total income. Values for IRA contributions, and the deductions for self-employed health insurance and pension contributions are derived from the statistical match. Table 7 shows that the simulations from the new model differ from reported IRS amounts, but incorporate information missing in the previous model.

Table 7. Comparison of Simulated Statutory Adjustments, Tax Year 2002

	CPS Old Model	CPS New Model	IRS Published Data
Percent with IRA Contributions	--	2.8%	2.6%
Mean IRA Contribution	--	$1,664	$2,900
Aggregate IRA Contributions	--	$5,662,625	$9,639,868
Percent with SE HI	--	2.2%	11.2%
Mean SEHI Deduction	--	$2,431	$1,235
Aggregate SEHI Deductions	--	$6,643,912	$10,019,154
Percent with SEP/SIMPLE	--	1.3%	0.9%
Mean SEP/SIMPLE Deduction	--	$7,398	$13,200
Aggregate SEP/SIMPLE Deduct.	--	$11,657,334	$15,590,116

More robust results were obtained through the statistical match on childcare expenses. A CPS variable asking whether the household paid for childcare allowed a targeted match to the SOI. Resulting numbers and amounts, as seen in Table 8, indicate the new model's improvement over the old estimates.[15] In the new model, childcare expenses were imputed during the SOI statistical match. The credit was calculated from these expenses based on IRS Form 2441.

Table 8. Comparison of Childcare Expense Credit, Tax Year 2002

	CPS Old Model	CPS New Model	IRS Published Data
Percent with Childcare Expense Deduction	12.0%	5.3%	4.8%
Mean Childcare Expense	$402	$407	$438

[15] The old model assigned the childcare expense credit amount based on aggregate tabulations from the IRS. The assignment of means failed to allow any returns to claim the maximum allowable amount of the credit. Also, the step amounts (a maximum $480 credit for one child, $960 for two or more children for anyone with wages over $28,000) were lacking in the distribution.

Credit			
Median Childcare Expense Credit	$396	$480	--
Aggregate Childcare Credit	$ 5,672,657	$2,689,499	$2,710,432

Evaluating simulated EITC estimates is difficult because the IRS published data are unaudited, and many claims for the EITC are denied or adjusted downward. The new tax model carefully applies IRS rules for eligibility for single person and those with dependents. As noted earlier, the new model imputes capital gains from the SOI statistical match. This process assigns capital gains to more filers at lower income levels than in the past. This affects EITC modeling because capital gains are considered in the investment income test, which disqualifies cases with imputed capital gains of $2550 (in 2002) from the credit, regardless of earned income.

Table 9. Comparison of Earned Income Tax Credit

	CPS Old Model	CPS New Model	IRS Published Data
Percent with EITC	13.2%	13.3%	16.8%
Mean EITC	$1,523	$1,517	$1,767
Median EITC	$1,301	$1,231	--
Aggregate EITC	$25,758,259	$24,816,475	$38,687,554

FUTURE REFINEMENTS

Beyond annual maintenance to keep the simulation current with IRS tax law, some refinements to the model are planned. All imputations used in the tax model are currently under review. The SOI statistical match is greatly limited by the income responses in the survey. CPS aggregates on unearned income, business income, interest and dividends are far below their SOI counterparts. Other imputed values under review are those obtained from the AHS statistical match. These variables, presence of mortgage and property tax, are essential in the SOI statistical match to identify filers likely to itemize their deductions.

Once match donors are identified, the time lag issue remains. The SOI public use file will always lag the CPS data by two or three years. The methods currently used to age the data values forward will be investigated.

The possibility of modeling additional statutory adjustments will be analyzed. Other fields such as moving expenses and student loan interest deduction may be simulated. CPS variables on mover status and age/education profiles could be used to increase match probabilities.

The modular nature of the new federal and state tax models makes annual updates possible. Currently, tax year 2003 state models were used for tax year 2002. In the future, annual updates are expected for both components.

APPENDIX ONE: STATISTICAL MATCH DETAILS

Imputing values from the SOI involved a hierarchical match. For each match attempted, five levels of increasing leniency sought a donor value. Partitions used in the match are listed below. The match keys used for each partition are also listed. The number of classes for the income match keys varies by match level. The highest level match uses thirty-three income classes; the lowest level match uses eight income classes. The wage amount key used nine classes. The number of child exemptions ranged from five classes to two classes. Some CPS cases in the subfamily and dependent person partitions were permitted to remain unmatched.

Subfamily partitions:

1) Self-employed, high income
   Match on state, high income classes, number of child exemptions, marital status

2) Self-employed, not high income
   Match on state, income classes, number of child exemptions, marital status, presence of Social Security

3) Not self-employed, high income
   Match on state, high income classes, number of child exemptions, marital status

4) Not self-employed, not high income
   Match on state, wage classes, number of child exemptions, marital status, presence of Social Security

Single partitions:

1) High income
   Match on high income class, income class, presence of unearned income, presence of wages

2) Not high income
   Match on income classes, state, presence of wages, presence of unearned income

Dependent partitions:

1) High income
   Match on state, high income classes, presence of Social Security, self-employed flag, presence of unearned income, presence of mortgage, presence of wages

2) Not high income
Match on state, income classes, presence of Social Security, presence of wages, presence of mortgage, presence of unearned income

Primary partitions:

1) Social security, high income
Match on state, high income classes, number of child exemptions, marital status, presence of mortgage, presence of unearned income

2) Social security, not high income
Match on state, income classes, number of child exemptions, marital status, presence of wages, presence of mortgage, presence of unearned income

3) No social security, self-employed, high income
Match on state, high income classes, having paid for childcare, number of child exemptions, marital status, presence of mortgage, presence of unearned income

4) No social security, self-employed, not high income
Match on state, income classes, having paid for childcare, number of child exemptions, marital status, presence of mortgage, presence of wages

5) No social security, not self-employed, with children, high income
Match on state, income classes, having paid for childcare, number of child exemptions, marital status, presence of mortgage, presence of wages, presence of unearned income

6) No social security, not self-employed, with children, not high income
Match on state, income classes, having paid for childcare, marital status, presence of mortgage, presence of wages

7) No social security, not self-employed, no children, high income
Match on high income classes, marital status, presence of wages, presence of mortgage, presence of unearned income

8) No social security, not self-employed, no children, not high income
Match on state, income classes, marital status, presence of wages, presence of mortgage

Table A1. SOI Total Income Components, Tax Year 2002

	IRS[16]	CPS Person
Wages	4,594,558,226	5,078,462,210
Interest	193,177,625	145,427,277
Dividends	98,758,800	58,542,066
Business Income	248,994,633	307,858,945
Business Loss	34,527,253	5,276,256
Pensions	363,178,764	229,222,057
Unemployment	43,411,772	37,912,012
Rent income	49,055,657	63,914,139
Rent loss	35,643,737	8,578,340
Royalty income	6,501,063	575,631[17]
Royalty loss	127,260	0
Estate/trust income	10,556,522	2,641
Estate/trust loss	633,567	0
Farm income	6,537,448	22,322,767
Farm loss	20,219,702	1,872,380

Table A1 compares preliminary IRS national aggregates to the total CPS person file as reported for Tax Year 2002. The CPS numbers are for all persons, not for tax units. While wages and business income are higher in the CPS, unearned income variables are grossly underreported. Farm income seems to differ diametrically between the surveys. This table does not include other important sources of income, such as partnership/S-corporation income because a comparison between the data sets was not available; the CPS does not collect information on passive income.

[16] "Individual Income Tax Returns, Preliminary Data, 2002." *SOI Bulletin*, Winter 2003-2004.
[17] OI_OFF=6 includes royalties and rents not included in RNT_VAL above.

APPENDIX TWO: STATE TAX MODELING[18]

State	Simulated items for Tax Year 2002 (Using 2003 programs)
Alabama	Federal and military pensions excluded No 50% self-employment tax deduction No self-employed health insurance deduction Different itemized deductions
Arizona	Pension exclusion Non-refundable family income tax credit
Arkansas	Spouse pension exemption separated Armed forces earnings exempt Unemployment compensation exempt Taxable Social Security exempt Non-refundable personal credit Non-refundable childcare expense credit
California	Taxable Social Security exempt Different itemized deductions Exemption credit Refundable childcare expense credit
Colorado	Non-taxable pensions excluded Refundable childcare expense credit
Connecticut	Adjusts taxable Social Security Non-refundable personal credit
Delaware	Pension exclusion based on age Aged-disabled exclusion Social Security exempt Split assets between spouses to apply pension exclusion Non-refundable personal credit Non-refundable childcare expense credit
District of Columbia	Social Security exempt Non-refundable low income credit Non-refundable childcare expense credit Refundable EITC[19] $3000 pension exclusion Disability exclusion
Georgia	Social Security exempt No self-employed health insurance deduction Retirement income exclusion Refundable low income credit
Hawaii	Social Security exempt Pensions exempt

[18] Note that seven states (Alaska, Florida, Nevada, South Dakota, Texas, Washington, and Wyoming) are not included in the table because they do not levy individual state income taxes.
[19] Cannot claim both earned income credit and low income credit.

	IRA exempt if over age 70 Refundable childcare expense credit Refundable low income credit Refundable low income renter credit
Idaho	Pension exclusion Social Security exempt Refundable grocery credit Non-refundable childcare expense credit
Illinois	Adds Federal non-taxable interest Social Security exempt Excludes taxable IRA distributions Excludes Armed Forces income Non-refundable EITC
Indiana	Exemption for aged and blind Exemption for aged with AGI below $40,000 Exemption for aged spouse Social Security exempt Disability exclusion Refundable unified credit for elderly Can deduct part of federal pension, military pay Part of unemployment is taxable Non-refundable EITC County credit
Iowa	Pension exclusion Adjusted Social Security exemption Armed forces income exempt Disability exempt Alternative tax for low income Non-refundable EITC Refundable childcare expense credit Non-refundable exemption credit
Kansas	Pension deduction Food sales tax refund Non-refundable childcare expense credit Refundable EITC
Kentucky	Separate income for spouses Non-refundable childcare expense credit Non-refundable low income credit Pension deduction
Louisiana	Social Security exempt Public pensions exempt Portion of dependent care expense credit allowed Childcare expense credit (refundable if low income)
Maine	Low income credit Add back ¼ of self-employed health insurance expense Non-refundable EITC

	Non-refundable elderly credit Social Security exempt Pension exclusion Partially refundable childcare expense credit
Maryland	Military pension excluded Disabled pension excluded Social Security exempt Dual income exclusion Partially refundable EITC Non-refundable poverty level credit Non-refundable childcare expense credit
Massachusetts	Part of FICA is deductible Assume 75% of income is not state generated Separate tax on capital gains and dividends Non-refundable limited income credit Refundable EITC
Michigan	Part of private pensions deductible Part of unearned income for aged deductible Military pay exempt Part of unemployment compensation excluded No self-employment tax deduction Social Security exempt Pension exclusion
Minnesota	Aged and disabled exclusion Married joint earners credit Refundable childcare expense credit Refundable EITC
Mississippi	Social Security exempt No self-employment tax deduction All pensions excluded
Missouri	Social Security exempt Pension exclusion
Montana	Unemployment compensation exempt No self-employment tax deduction No self-employment health insurance deduction No self-employment pension deduction Interest exclusion for aged Pension exclusion Military pay for active duty excluded Childcare expense credit
Nebraska	Standard deduction added back to high income returns Self-employment health insurance deduction Non-refundable personal exemption credit Non-refundable disabled/elderly credit Refundable childcare expense credit
New Hampshire	Only federal interest and dividends are taxed

New Jersey	Social Security exempt Unemployment compensation excluded Pension exclusion Retirement exclusion Refundable EITC
New Mexico	Low income rebate
New York	Social Security exempt Pension exclusion Disability pay exclusion limit Alternate tax for high income returns Non-refundable state household credit Non-refundable New York City household credit Refundable EITC Refundable New York City school tax credit
North Carolina	Portion of elderly/disabled credit allowed Pension exclusion Refundable childcare expense credit Social Security exempt Refundable child tax credit Refundable elderly/disabled credit
Ohio	Non-refundable credit for elderly Social Security exempt Disability exclusion Non-refundable childcare credit Refundable joint filing credit
Oklahoma	Social Security exempt Pension exclusion Active military pay exclusion Non-refundable childcare expense credit Non-refundable EITC
Oregon	Social Security exempt Non-refundable retirement credit 60% Federal pension exclusion Military pay exclusion Non-refundable working family credit Non-refundable elderly/disabled credit Non-refundable EITC Non-refundable childcare expense credit
Pennsylvania	IRA distributions exclusion Pension exclusion Unemployment compensation exclusion Social Security exempt Refundable Taxback credit
Rhode Island	Non-refundable childcare expense credit Non-refundable dependent care expense credit Partially refundable EITC

South Carolina	Social Security exempt No disabled pension deduction Pension exclusion Additional self-employed health insurance deduction Limit on long-term capital gains Extra exemption for children under 6 Non-refundable dual earner credit Non-refundable childcare expense credit
Tennessee	No tax if aged and meet AGI test
Utah	Federal tax deduction Retirement income exclusion Pension exclusion
Vermont	Non-refundable child and disability credits Refundable EITC
Virginia	Social Security exempt Age 62+ tax exclusion Age 65+ tax exclusion Low income exemption credit Childcare expense deduction Refundable low income credit
West Virginia	Military pension deduction (after 20 years service) Government pension exclusion Senior citizen or disabled deduction Low income earner exclusion
Wisconsin	Unemployment compensation partially deductible Married couple dual earner credit Working family credit Pensions deductible Disabled exclusion Itemized credit Refundable EITC

Aggregate State Taxes After Credits, Tax Year 2002

State	CPS New Model	CPS Old Model
Alabama	2,328,647	2,989,728
Arizona	1,768,796	4,044,731
Arkansas	1,621,882	2,645,179
California	26,397,473	45,262,551
Colorado	3,295,689	3,716,649
Connecticut	3,003,978	1,670,606
Delaware	645,173	1,172,870
DC	1,077,110	1,537,612
Georgia	6,252,275	8,096,388
Hawaii	1,255,730	3,333,512
Idaho	943,207	1,242,793
Illinois	6,909,328	7,609,010
Indiana	4,801,429	4,063,269
Iowa	2,369,366	3,477,697
Kansas	2,021,416	2,276,963
Kentucky	3,304,108	3,573,628
Louisiana	1,872,798	1,891,530
Maine	1,050,495	1,275,131
Maryland	5,083,168	8,173,764
Massachusetts	7,358,233	9,307,659
Michigan	6,809,171	9,078,280
Minnesota	5,690,882	8,415,059
Mississippi	1,127,911	1,448,720
Missouri	4,031,047	4,937,908
Montana	665,908	817,115
Nebraska	1,490,612	1,503,521
New Hampshire	35,798	46,503
New Jersey	6,601,764	10,977,837
New Mexico	977,069	1,293,370
New York	21,951,480	26,034,025
North Carolina	7,408,893	9,009,874
North Dakota	198,978	175,472
Ohio	7,445,496	10,209,681
Oklahoma	2,700,289	3,211,935
Oregon	4,131,654	5,117,747
Pennsylvania	6,846,105	7,907,634
Rhode Island	885,676	869,602
South Carolina	2,501,158	3,525,396
Tennessee	152,455	208,308
Utah	1,467,088	1,937,693
Vermont	397,113	481,745
Virginia	6,394,038	8,249,269

West Virginia	975,759	1,048,843
Wisconsin	5,277,915	7,071,255

APPENDIX THREE

OLD MODEL OVERVIEW

The old tax model simulated four taxes separately: 1) federal individual income, 2) state individual income, 3) property taxes on owner-occupied housing, and 4) payroll taxes.

Federal Individual Income Tax Simulation

The old model formed tax units based on household relationships, marital status and dependency rules. A filing unit was assigned to any individual or married couple with at least $1000 taxable income from a) wage or salary income, b) combined interest, rental, royalty, estate or trust income, c) pension income, or d) a combination of these categories.

All primary family householders and spouses were listed as dependents on their own tax return. Children under fifteen years of age living in the filing unit, those over fifteen with income less than $1000, and children who were students were designated as dependents on the householder's tax return. Other family members with taxable income less than $1000 were also given dependent status on the primary family householder's tax return. Related subfamilies eligible to form their own tax unit were assigned dependents as described for the primary families. Members of related subfamilies without a filing unit were assigned as dependents on the primary tax return. Primary and secondary unrelated persons aged 15 and over were treated as dependents on their own tax returns.

Married joint filing status was assigned to married persons and those whose spouses were absent in the Armed Forces. Head of household status was assigned to unmarried persons with dependents and to separated married persons with dependents. All other persons meeting the filing criteria were assigned as single filers.

In the old model, AGI was computed by summing reported income amounts with imputed capital gains. Capital gains were imputed using statistics from the IRS on the number of filers and the aggregate amount of gains or losses by AGI class. The proportion of filers with gains or losses could then be determined, and a Monte Carlo technique utilized to randomly assign mean values for the appropriate AGI class.

Taxable income was computed by subtracting deductions and exemptions from AGI. Itemized deductions were imputed using IRS tables in the manner described for capital gains. For non-itemizing returns, the standard deduction was assigned based on the unit's filing status. The exemption amount was based on the number of dependents established during tax unit formation.

Total taxes were computed in each year, using the appropriate tax schedule for each return type. The old model simulated four credits to deduct from the total tax liability: 1) the credit for child and dependent care expenses, 2) the child tax credit, 3) the additional child tax credit, and 4) the earned income credit. The dependent childcare expense credit was simulated and deducted from the total tax liability. IRS data on the number and aggregate amount of reported childcare expenses by AGI allowed mean imputation of credit amounts. The child tax credit and additional child tax credit were computed based on IRS rules. The CPS variable FED_TAX reports federal taxes after the tax liability has been adjusted by these credits (1-3). The refundable nature of the additional child tax credit allows FED_TAX to range negative. The final credit simulation is the earned income tax credit. This variable, EIT_CRED, is subtracted from FED_TAX to determine federal taxes after all credits. Computed federal taxes after credits (1-4) could be negative from the refundable additional child tax credit and/or the refundable EITC.

State Individual Income Tax Simulation

State income taxes were simulated, where applicable,[20] using the filing units and AGI determined in the federal simulation. State tax rates and brackets were updated annually. State earned income credits were modeled for some states but not as refundable (thus, state taxes after credits were bounded at zero). Along with Federal taxes, state tax estimates were subtracted from money income to construct alternative income measures for Census Bureau reports.

Property Taxes on Owner-Occupied Housing

Property taxes on owner-occupied housing were simulated using a statistical match to the 1995 American Housing Survey (AHS). Property taxes are used in determining imputed home equity in one of the alternative income definitions.

Payroll Taxes

Payroll taxes on workers covered under the Federal Insurance Contributions Act (FICA) were computed up to the specified income thresholds ($84,900 in 2002 and $87,000 in 2003). Payroll taxes on self-employed workers were calculated according to Self-Employment Contributions Act (SECA) rules. These estimates were included in the FICA field. Mandatory retirement payroll deductions for federal employees covered under the Civil Service Retirement System (CSRS) were estimated separately and reported in the FED_RET field.

[20] Alaska, Florida, Nevada, South Dakota, Texas, Washington, and Wyoming do not levy individual state income taxes.

APPENDIX FOUR: Updated Tables

Table A4.1 National Aggregate Estimates, Tax Years 2002 and 2003[1]

	Mean		Median		Aggregate	
	2002	2003	2002	2003	2002	2003
FICA payroll taxes	2,583	2,651	2,066	2,111	372,216,523	382,337,945
Public employee payroll taxes[2]	2,843	3,018	2,673	2,754	17,538,543	18,264,061
State taxes after credits	2,206	2,316	1,195	1,226	180,524,558	190,031,620
Federal taxes after credits	6,452	6,488	2,472	2,499	724,473,825	729,767,186

[1] These estimates were produced using the new tax model described in this working paper. Tax year 2002 results differ from public use microdata produced using the old tax model. CPS ASEC 2003 was used to produce tax year 2002 estimates. CPS ASEC 2004 was used to produce tax year 2003 estimates.

[2] The variable FED_RET estimates payroll taxes for public sector employees not covered by FICA. State, local and federal government employees are included in this category. See page three of this paper for a discussion of payroll tax estimation in the new tax model.