Evaluation of CPS Tax Simulation Using Administrative IRS Data

Amy O'Hara

U.S. Census Bureau
amy.b.ohara@census.gov

Introduction

A new methodology has been developed to simulate federal and state taxes for the Current Population Survey (CPS). These tax estimates are used in Census Bureau reports, such as the poverty and income reports, and are released as a public use file. The Census Bureau has developed a new tax model that simulates an individual income tax return more closely and estimates more variables and credits than the previous tax model. The CPS data used in this paper have been generated using the March 2000 person file with the updated methodology. Table 1 shows weighted aggregate values and how they differ between the new and old models. These aggregates are based on the March 2000 public file, comparing the new and old methodologies.

Table 1: Tax Year 1999 New and Old CPS ASEC Tax Model Comparisons, weighted aggregates in thousands of dollars

	New Model	Old Model	New/Old Ratio
Adjusted Gross Income	$5,662,718,414	$5,783,239,614	0.98
State Income Tax	161,567,838	223,498,157	0.72
Earned Income Credit	20,100,173	22,925,583	0.88
Capital Gains	421,380,858	463,462,657	0.91
Itemized Deductions	795,514,591	1,059,102,737	0.75
Federal Tax after Credits	564,799,015	827,996,639	0.68

This paper evaluates aspects of the new tax model estimates by matching Census microdata and administrative records from the Internal Revenue Service (IRS) for tax year 1999. The results of linking the CPS records with individual income tax records are used to gauge how well return classifications and the number of exemptions were assigned in the tax model. These results will allow better modeling of filing incidence in the future.

A match to an IRS individual tax return filed for tax year 1999 was attempted for each adult (age fifteen and older) in the CPS Annual Social and Economic Supplement (CPS ASEC) person file for 2000. To attempt the match, a CPS ASEC observation must have a validated Social Security Number (SSN). The person file for income year 1999 was validated through an agreement between the Census Bureau and Social Security Administration. Of the 103,226 adults in the CPS ASEC person file, 71,137 were validated (69 percent). Since 1999, the validation rate has increased due to the implementation of an automated system.

This paper only discusses cases that were matched to an IRS return, and whether or not they had a filing requirement as modeled by the new Census Bureau tax model. The matched group with a filing requirement informs CPS ASEC users on the validity of the publicly released tax variables. The discussion of cases designated as non-filers by the tax model also informs users as to possible limitations of the CPS ASEC data, and certifies the need for more research. The paper does not discuss CPS ASEC records that did not match to an IRS return. This includes the subset of persons whose SSN was validated, whose information implied a filing requirement, but who did not file or appear in the linked data set. The purpose of this analysis is to evaluate the tax model, not to address potential tax code compliance issues.

Data

This paper uses the 2000 CPS ASEC for income year 1999. The U.S. Census Bureau conducts the survey, collecting demographic and economic information on the U.S. non-institutional population. A probability sample was used to select approximately 65,000 households, including information on 103,226 persons aged 15 and over.

The IRS Individual Master File (IMF) for tax year 1999 was used to attempt the record linkage to the validated CPS ASEC file. The IRS file includes a limited number of 1040 form variables governed by formal agreements between the Census Bureau and the IRS. Return type (single, married, etc.) and exemptions (quantity and type) are included. Several included variables are actual dollar values (AGI, wages), but most other variables are indicators (whether certain schedules were filed).

The IRS IMF data is not to be confused with the IRS Statistics of Income (SOI) data. The IRS releases a public use SOI file on an annual basis. The new Census Bureau tax model uses this data source to impute variables in the tax model. These include capital gains and losses, several statutory adjustments, itemized deductions, and child care expenses. The Census Bureau uses the SOI public use file and reported IRS aggregates to test model estimates on an annual basis, a very different evaluation than what is described in this paper: here, administrative tax data from the actual CPS ASEC respondent are compared with the modeled information.

Tax 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 relationships. This information is the basis for forming tax units and assigning filing status. The CPS ASEC model includes single, married-joint, and head of household returns. Because CPS ASEC does not collect capital gains, statutory adjustments, and other tax variables, a statistical match to the SOI public use file is used to impute this information. Because the new tax model follows the IRS 1040 Form, the individual income tax calculations follow an iterative process to arrive at a final estimate of taxes. Preliminary federal taxes and credits are computed, and then used in the state tax calculations. State taxes often require federal inputs such as adjusted gross income (AGI), itemized deductions, and the earned income tax credit (EIC). 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.

The model creates tax filing units according to IRS rules. That is, all persons aged 15 and over with reported income are potential filers in the three filing classes (single, married-joint, and head of household). Because the model uses tax exemptions to determine some of the categories for filing status, exemptions are defined first. Exemptions are tabulated for these potential filers according to the following rules:

1. Exemptions for self and spouse are assigned.
2. Exemptions for dependents are identified as:
    a. Persons in unit under age 19,
    b. Persons in unit under age 24 who are enrolled in school, or
    c. Persons in unit over 18 earning less than the IRS threshold

Persons that are in primary families are treated differently than those in subfamilies or secondary individuals. For primary families, units may consist of primary families and non-family householders (and non-family reference persons in group quarters). Filing status 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[1] with more than one exemption are assigned head of household filing status. Single status is given to non-reference person tax heads. All single filers are initially assigned one exemption, but the model also checks to see if they are dependents on other returns. If so, they will be assigned single filing status with zero exemptions.

In all subfamilies (related, unrelated, secondary individuals, and primary families living in group quarters) a subfamily tax head is identified. As in primary families, married filing joint status is assigned to those who are married or married with spouse absent in Armed Forces. Unlike primary families, all others are assigned single filing status. No head of household returns are permitted for subfamilies, as it is assumed that they fail to provide more than 50 percent of the cost of maintaining the household.

Persons not in primary families or subfamilies are compiled into a single filer group, each with one exemption. Later in the processing, these persons will be evaluated to determine whether they are dependents on other returns. Those found to be dependent on other returns are assigned a zero exemption single return.

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

To determine who must file, the model evaluates five conditions. A tax unit has a filing requirement if:

1) Their income exceeds the filing threshold for their filing status;
2) The income does not meet the filing threshold but would receive an earned income credit;
3) Self-employed income exceeds $400;
4) Gross income is negative;[2] and
5) Reported (net) self-employed or farm self-employed income is negative.

Filing status (FILESTAT on the public use CPS ASEC file) was assigned to every unit with a filing requirement. Knowing that some people file though they are not required to, the new CPS ASEC tax model makes an exception to allow additional single and head of household returns. This adjustment allows the tax model aggregates to align better with published IRS results. For tax year 2003, those with gross income over $3,000 are assigned a filing requirement. However, for this exact match analysis, these are not considered to have a filing requirement, and the cases discussed below follow the IRS rules and not this artificial $3000 constraint.

The CPS ASEC tax model generated 55,706 returns with filing requirements for tax year 1999. The IRS file contains 39,989 returns that link to persons in the CPS ASEC file. When weighted using the CPS ASEC March Supplement person weight, this corresponds to 82.5 million tax returns. Because of the 69 percent validation rate for adults on the CPS ASEC file, this number is lower than 127.1 million tax returns reported by the IRS for 1999. The IRS number also includes non-filers, while the Census Bureau tax model reports estimates for required filers.

Of the returns in the common data extract, where a CPS ASEC person (or persons if married) matched to an IRS filer by SSN, the CPS ASEC tax model expected 34,999 to have a filing requirement according to the IRS tax code (72.1 million weighted returns). The validated cases with filing requirements were evaluated to determine how well the Census tax model set the filing status and number of exemptions. Cases that matched on both measures were further described, and cases that did not match were investigated to see where the discrepancies might have arisen. A second set of results briefly describes cases in the exact match data that did not have a filing requirement, but still filed a return. Figure 1 shows how the unweighted counts in the linked data matched on filing status and number of exemptions. The rest of the paper discusses the returns in these nodes.

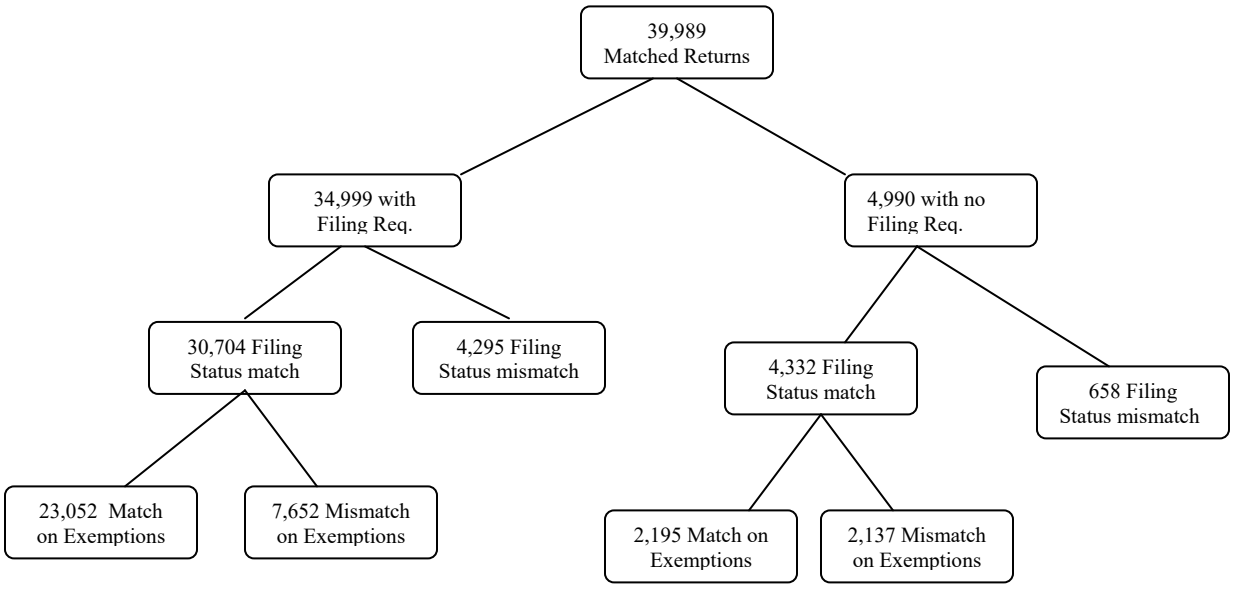

Figure 1: Breakdown of Matched Cases, unweighted.

[2] Gross income must be estimated for the CPS ASEC 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 ASEC amounts of farm and non-farm self-employed income to modify the reported net amounts to gross amounts. The ratios were computed separately for those reporting positive and negative amounts.

Results on Cases with Filing Requirements

Of the 34,999 returns that had a filing requirement according to the tax model, the correct filing status was assigned to 30,704 (88 percent) of the returns. Correctly designating filing status is very important: filing thresholds, exemption amounts, and credit eligibility limits are based on filing status. Knowing that this initial step was completed correctly for the majority of constructed returns is a positive finding for the evaluation of the new tax model.

The majority of these returns were married-joint filers, as seen in Table 2, which also presents information on the fraction of returns that did not match on filing status. IRS statistics indicate that single returns outnumber married-joint returns, 45 percent to 41 percent, as did the CPS ASEC tax model.[3] In this table, the married-joint returns outnumber the single returns for those whose filing status was correctly assigned. The two columns describing the incorrect filing status group of 4,295 filers show how the Census Bureau model expected the returns to be filed, compared to how taxpayers actually filed. The model expected nearly the opposite of what occurred: The Census Bureau model assigned many single returns (2,664 of 4,295) that actually filed as married-joint or head of household (1,413 and 2,104 of 4,295 respectively). This could be due to the changes in marital status occurring between the survey date and end of the tax year, to the presence of exemptions that were unobserved in the survey data, or to inaccurate filing with the IRS.

Table 2: Matched Returns with Filing Requirement, unweighted

	Returns assigned correct filing status	Incorrect filing status, CPS ASEC estimates	Incorrect filing status, IRS actual
Single	11,684	2,664	778
Married Joint	16,691	871	1,413
Head of Household	2,329	760	2,104
Total	30,704	4,295	4,295

Table 3 indicates the number of cases where exemptions matched according to filing status, and shows the proportion of exemption matches to the correctly assigned filing status cases from Table 2. Exemptions on single filer returns were correctly identified most often, followed by married-joint returns. About two-thirds of head of household returns matched on the number of exemptions; this low percentage may be due to changes in household composition or support that occurred during the calendar year after the CPS ASEC survey data was collected.

Table 3: Matched Returns with Filing Requirement and Correct Filing Status, unweighted

	Number of exemptions match	Percent of correct filing status
Single filers	9,435	9,435/11,684 = 80.8%
Married Joint filers	12,048	12,048/16,691 = 72.2%
Head of Household filers	1,569	1,569/2,329 = 67.4%
Total	23,052	23,052/30,704 = 75.1%

For those who matched on status and exemptions, Table 4 illustrates how well the Census Bureau model performed on other measures. The tax model must obtain itemized deductions and capital gains from a statistical match because those values are not collected in the survey. Looking at the cases where both filing status and exemptions agree with the IRS administrative records, Table 4 presents weighted percentages of returns filing schedules A or D[4] and weighted adjusted gross income aggregates. Schedule A is for itemized deductions and Schedule D is for capital gains and losses. The table shows considerable agreement on the incidence of filing Schedules A or D between the CPS ASEC and IRS. The Census Bureau is still revising the statistical match used to impute these fields. The AGI summaries also indicate a solid correspondence

[3] Table 1.3 in the IRS Publication 1304, Individual Income Tax Returns 1999, indicates that 41 percent of all returns filed married joint/separate/surviving spouses, 14 percent filed head of household, and 45 percent filed single returns. The full, weighted CPS ASEC model for 1999 (not limited to those cases in the exact match), estimated that 42 percent of all returns filed married, 8 percent filed head of household, and 49 percent filed single returns.

[4] The IRS data does not include amounts for these fields, only whether or not the return included the schedule. The Census model does compare its imputed aggregates to IRS tables annually to benchmark amounts.

between the modeled and reported amounts. Overall, the Census Bureau modeled 101 percent of IRS AGI. The AGI aggregate for single CPS ASEC filers was 109 percent of the reported amount; for married filers, the CPS ASEC amount was 98 percent of the IRS amount, and for Head of Household filers, the CPS ASEC AGI amount was 109 percent of the IRS amount. Table 4 also validates the new tax model estimates, though additional research is needed to determine which components of AGI contributed to the CPS ASEC overestimates. It is possible that imputed variables for capital gains and statutory adjustments were contributing factors.

Table 4: Additional Tax Return Information on CPS ASEC Returns in agreement with IRS Returns on Filing Status and Exemption Count, weighted

	CPS ASEC	IRS	CPS ASEC/IRS Difference
Schedule A	46.2%	41.5%	1.11
Schedule D	24.3%	22.9%	1.06
Aggregate AGI[5]	$2,650,551,011	$2,628,881,939	1.01
Single AGI	650,712,538	596,494,666	1.09
Married-Joint AGI	1,908,017,613	1,947,906,012	0.98
Head of Household AGI	91,820,860	84,481,261	1.09

The returns assigned the correct filing status, but with a different number of exemptions, are analyzed in Table 5. The number of exemptions affects tax estimates in many ways. In tax year 1999, $2,750 could be deducted from AGI for each personal and dependent exemption (subject to the AGI limit). Correct assignment of dependent exemptions can also affect earned income credit eligible children, persons eligible for the child and dependent care expense credit, and the child tax credit.

Table 5 indicates that the Census Bureau model is generally undercounting exemptions. These persons are probably not included in the survey information, thus are not counted on the returns that the model forms. The table presents weighted percentages of exemptions by filing status for both CPS ASEC and IRS. The model undercount could include persons not present at the time of the interview, persons who do not typically reside with the survey respondent (as in the case of non-custodial parents who are authorized to claim exemptions), or any other related or unrelated persons supported by the filer who did not appear in the CPS ASEC data.

Focusing on cases where the predicted filing status in the CPS ASEC records matched the filing status in the IRS administrative records, 75 percent (23,052 unweighted observations) matched on the number of exemptions. As the first row of Table 5 shows, the IRS has more dependent returns than the CPS ASEC model predicted. The persons who file a return with zero exemptions can also be claimed on another return. The second panel of the table shows that the CPS ASEC tax model requires that married-joint filers each carry an exemption (one for primary filer and one for spouse); the IRS cases show that some married filers only claimed one exemption.[6] Head of household returns in the third panel show another model-based difference: the CPS ASEC tax model requires that these filers have personal exemptions plus at least one dependent exemption. The model disallows single exemption head of household returns, which comprise 16 percent of exemptions on the IRS returns for this filing status. While the CPS ASEC and IRS percentages for the rows indicating two and three exemptions appear to be similar, it is worth noting again that these are the cases where the number of exemptions is mismatched. There may have been a similar proportion of 2- and 3-exemption head of household returns, but they did not refer to the same exact match filers.

Table 5: Matched Returns with Filing Req., Correct Filing Status and Incorrect Exemption Count, weighted percentages

Exemptions for Single Filers	CPS ASEC	IRS
0	13.4%	18.3%
1	84.6%	14.2%
2	1.5%	61.7%

[5] Aggregate AGI amounts in thousands of dollars.

[6] One case in the IRS file was a married return that claimed zero exemptions; the IRS data received are unaudited and lack the depth of information to determine the cause for such an unusual filing.

3+	0.5%	5.8%
Total	100.0%	100.0%
Exemptions for Married Filers	CPS ASEC	IRS
1	0.0%	7.4%
2	60.2%	7.9%
3	16.7%	29.8%
4	12.5%	41.6%
5	6.4%	8.4%
6	2.7%	3.2%
7+	1.5%	1.7%
Total	100.0%	100.0%
Exemptions for Head of Household Filers	CPS ASEC	IRS
1	0.0%	16.1%
2	33.3%	33.9%
3	37.5%	35.2%
4	19.5%	10.5%
5	6.5%	2.7%
6+	3.2%	1.6%
Total	100.0%	100.0%

Results on Cases without Filing Requirements

According to the CPS ASEC tax model, some people did not have a filing requirement but appear in the linked data because they filed anyway. This would generally include cases where income was less than the filing threshold for 1999,[7] those without an earned income credit, or those with self-employment income less than $400. There were 4,990 returns in the matched data set that did not have a filing requirement according to the CPS ASEC tax model. The CPS ASEC tax model expected 70 percent of these returns to have zero AGI.[8] The IRS showed a positive or negative AGI amount for nearly all of these returns (99.8 percent), indicating a possible discrepancy in income reported to the IRS and CPS ASEC or difference between imputed values from the SOI and reported amounts.

Table 6 shows that the majority (69 percent) of the 4,990 filed a single return. Almost one-quarter filed married-joint, and the remaining 7 percent filed head of household. The correct filing status was modeled for 4,332 or 87 percent of these returns. To reiterate what "correct filing status" means for these cases that have no filing requirement: All CPS ASEC persons are sorted into potential filing units and evaluated to see if their income or circumstances merit a filing requirement. If the CPS ASEC unit does not meet the IRS criteria for having a filing requirement, no filing status[9] is assigned to the person record and no tax estimates will be released. This is why Table 6 calls the CPS ASEC filing status "unused" – it was generated in the tax model but non-filer status was assigned because the returns' income levels fell below the IRS threshold.

The percentage of dependent returns was similar between the CPS ASEC prediction and IRS result; more analysis of dependent returns is required to determine what prompts a non-required dependent to file. The number of modeled exemptions differed from reported exemptions for half of these returns. As shown in Figure 1, 2,195 matched on exemptions and 2,137 did not match on exemptions. The discrepancies on filing status and number of exemptions may again be due to outmoded or misinterpreted survey responses.

Table 6: Comparison of Actual to Predicted Filing Status for Returns with No Filing Requirement, weighted		
	IRS Filing Status	Unused CPS ASEC Filing Status

[7] In 1999, the filing thresholds for persons under age 65 were: $7,050 for single filers, $12,700 for married-joint filers, and $9,100 for head of household filers. Amounts were higher for persons over 65, and separate thresholds exist for children and dependents.

[8] The remaining 30 percent of CPS ASEC modeled returns had AGI values below the filing threshold for their filing status, as described in Footnote 7 above.

[9] FILESTAT will equal 6 for non-filers.

Single	69%	75%
Married Joint	24%	18%
Head of Household	7%	7%
Dependents	35%	32%

For the returns that were assigned the correct filing status, reported IRS AGI is much larger than the computed amount from the CPS ASEC model. Table 7 illustrates the differences by filing status. The low amounts in the CPS ASEC AGI column reveal why the Census model did not assign a filing requirement to these returns. In this subset of non-required returns, the married filers in the CPS ASEC model had no AGI. As seen in the IRS AGI column, this was the largest missing amount of AGI. Additional investigation of the AGI components for married filers is needed to determine how the CPS ASEC model could better calculate income for these filers.

Table 7: Aggregate AGI Comparison of Non-Required Filers with Correct Filing Status, weighted aggregates in thousands of dollars

	IRS AGI	CPS ASEC AGI
Single	$68,110,132	$11,986,380
Married-Joint	130,450,243	0
Head of Household	5,306,714	1,064,665
Total	203,867,088	13,051,045

Conclusion:

As seen from the exact match data, the CPS ASEC tax model is generating results consistent with filer behavior. The results on filing status are especially promising, as the model is assembling filing units in line with reported results. Overall, the model correctly assigned filing status and the number of exemptions for 23,052 of the 34,999 required filers, a 66% success rate. Aggregate AGI amounts also compare favorably by filing status. Further research into the components of AGI should better inform the statistical match used to impute tax variables. Comparing income levels reported to the CPS ASEC and IRS will also inform the modeling process: the CPS ASEC tax model must use survey responses as its inputs, and if they differ from amounts reported to the IRS, the model cannot address the discrepancy.

More analysis is required on the group of dependent filers as a whole. These were highly concentrated in the group that lacked a filing requirement, but still filed. The CPS ASEC and IRS percentages of these returns were similar, showing good logic in the model, but these returns never received a filing requirement due to their income falling below the IRS threshold. Attempting to determine what could cause these people to file would improve the CPS ASEC tax model, particularly the FILESTAT variable. Attention to the non-dependent returns who did not have a filing requirement is also in order. However, this will not lead to a large difference in aggregate totals because this is a relatively small group.