Evaluating Tax Data Generated Using the Survey of Consumer Finances and TAXSIM

Robert Argento[1], Kevin Moore[2]

[1]Freddie Mac, 1551 Park Run Drive, McLean, VA 22102

[2]Federal Reserve Board, 20th and C St NW, Washington DC 20551

Abstract

This paper details the methods used to generate federal and state tax rates and liabilities using the Survey of Consumer Finances (SCF) and NBER TAXSIM. The detailed information in the SCF on income, assets, and liabilities allows us to create the variables necessary for the TAXSIM calculator. One critical issue in computing tax rates and liabilities is that the unit of observation in the SCF is the household, not the tax unit. We take two approaches to dealing with this issue; the first is to assume the household and tax unit are identical, while the second attempts to split households into tax units. The paper compares the tax data generated by the two methodologies on the number of filers by filing status, the number of itemizers, tax rates and tax liabilities to estimates published by the Internal Revenue Service (IRS). We find that assuming that the household and tax unit are identical generates too few single and head of household filers relative to the methodology that splits households into tax units. The results for tax liabilities and rates show that although both methods generate higher tax liabilities compared to IRS data, splitting households into tax units better matches the distribution of marginal taxes rates across filing status.

Key Words: Survey of Consumer Finances, TAXSIM, tax rates

1. Introduction

Federal and state tax liabilities often represent a significant part of the household balance sheet. As a source of detailed information on income, assets, and liabilities, the Survey of Consumer Finances (SCF) is in an ideal position to provide accurate estimates of a household's tax liabilities and tax rates, estimates that can be used to supplement all types of tax policy research.

Since a household's expected tax liability is not asked about directly, we construct an estimate with the help of the TAXSIM calculator created by the National Bureau of Economic Research. TAXSIM takes variables derived from responses in the SCF as input and returns estimates of federal and state tax liabilities based on existing law for that year.[1] Although the SCF contains much of the information necessary for computing taxable income, the unit of observation is the household, not the tax filer. A household in the SCF consists of the household head and any individuals financially dependent on the household head; this is referred to as the primary economic unit (PEU). For most

---

[1] This paper uses the internal version of the SCF data, which is only available to SCF staff. A similar analysis can be done using the public version of the SCF data, which is available at http://www.federalreserve.gov/econresdata/scf/scfindex.htm.

households, the PEU comprises the entire household and translates readily into a filing status. Unfortunately, households undergoing a change in marital status and households with financially independent members, referred to as the non-primary economic unit (NPEU), tend to be problematic. Additionally, there is no obvious process for determining which households did not file a return. Complications like these confound the process for determining the filing status of a household.

One method for dealing with this problem is to make a few simplifying assumptions in the SCF in order to proceed as if the traditional SCF concept of a households can be treated as a tax unit. The most important of these assumptions is that all married or otherwise living with partner couples file as married filing jointly. The rational is that it is difficult to split up income and assets in households with a financially interdependent head and spouse. The NPEU is ignored under this methodology, owing to the limited amount of information provided. We also assume that every household files a tax return in 2009. These assumptions allow the standard SCF to be used to generate tax data.[2]

A second method is to develop an algorithm for breaking the SCF households into tax filer units by making use of the information available in the SCF. Recent marital history is used to infer the filing status the household would have qualified for in 2009. If a household did not file as married filing jointly, then the household's income, deductions, and exemptions are split into two separate tax filing units. Persons residing in the NPEU are appended separately as single filers. Households that fall below the income thresholds required for filing are assumed to not have filed a return. The process alters the values of many variables relevant to tax analysis and increases the number of tax units in the SCF.

Both methods are explored in this paper. In the accompanying tables, the first method is labeled as simply the "SCF". The method that makes material modifications to the structure of the SCF is labeled as the "Tax-SCF". We then compare and contrast the results from the two methods to data published by IRS Statistics of Income (SOI)

2. Measuring Tax Concepts

For each year of the SCF, there is a SAS program that generates the variables for the TAXSIM calculator using the extensive asset, liability, income and demographic data available in the SCF. In this analysis, we focus on the 2010 SCF. Taxable income is defined as total income less nontaxable income less exemptions less deductions. Calculating total income is fairly straightforward since many of the SCF income questions refer to the IRS Form 1040 (Figure 1), where the tax year for the SCF data is one year prior to survey year. Although most of the SCF income variables match the IRS definition, there some amounts that are reported as total rather than taxable. Also, there are no questions on adjustments to total income in the SCF.

Exemptions include both personal exemptions and any dependents under the age of 18. Deductions include either the standard deduction or itemized deductions. Only certain itemized deductions can be calculated, specifically the deductions for mortgage interest, property tax, investment interest expense, charitable contributions, and state income

---

[2] SAS programs for this method that use the public SCF data are available at http://users.nber.org/~taxsim/to-taxsim/scf/.

taxes. According the IRS data, these five deductions account for about 85 to 90 percent of total itemized deductions.[3]

The mortgage interest deduction varies based on the estimated interest paid in the year prior to the survey on qualifying first liens, second liens or home equity loans on the primary residence. Home equity lines of credit on the primary residence that qualify under the IRS rules are also included. For households with other properties, we assume that the household chooses the qualifying property with the largest amount of interest paid in the prior year. The total amount of mortgage interest paid is subject to the IRS limits. The property tax deduction is simply the amount of property tax paid on the primary residence. Information on property taxes paid on other properties owned by the household is not collected.

The investment interest expense deduction consists of the amount of interest paid on loans used for investments. Only loans intended to invest in real estate or invest in other asset are counted. For loans against properties other than the primary residence, the interest is counted as investment interest expense if the household receives rental income. Per IRS regulations, the amount of investment interest expense deduction is limited to the reported amount of interest and dividend income.

The SCF asks households making more than $500 in charitable contributions about the amount made in the prior year. We assume that all contributions are cash contributions to organizations that meet the IRS requirements for charitable organizations.[4] The state tax deduction is calculated by the state tax module of TAXSIM and added back into the federal tax calculation. All results focus on federal tax liabilities and rates.[5]

Some adjustments to the SCF variables are necessary to more accurately map those variables into the TAXSIM categories. A flag is created for receiving the age exemption. Total wage and salary income is reported for the household using the TAXSIM variable for the head, ignoring the variable for the spouse. TAXSIM also has categories for long-term and short-term capital gains, but the SCF only asks about overall capital gains or losses. We apportion overall capital gains or losses by using the IRS Individual Return data to determine the share of gains and losses in each category for three broad AGI classes for a given year. The AGI classes are less than $50,000, $50,000 to less than $100,000, and $100,000 or more. Two adjustments are made to the SCF Social Security, pensions, and annuities income variable to fit the TAXSIM categories. First, a separate variable for gross Social Security is created and that amount is subtracted from the overall Social Security, pensions, and annuities income variable. Second, withdrawals in the prior year from tax deferred retirement accounts (defined contribution pension and IRAs / Keoghs) reported in questions on the value of these assets are added to the pensions and annuity income variable.

3. Creating Tax Units

For the first method, "SCF", no further adjustments are needed to the data beyond those detailed above. Here we focus on the "Tax SCF" method which adds an algorithm to convert households into tax units. In creating tax units, we first address three issues that

---

[3] Authors calculations from IRS Individual Returns table 2.2 for 2009

[4] TAXSIM applies the limit on charitable contributions of 50% of adjusted gross income.

[5] The state of residence variable is not available in the public SCF data.

affect filing status: recent changes in marital status, a partnership that is not legally recognized, and the inclusion of NPEU filers.

Creating tax units involves a complex algorithm that splits the tax information of households headed by a couple that should file separate returns. The couples that likely file separate returns are couples that legally married or remarried less than a year from the interview date, legally married couples that live apart, and unmarried couples. This represents 2,550 (unweighted) observations out of 32,460 observations in 2010. The algorithm also estimates a return for each person in the NPEU, assuming each person represents a single filer. In 2010, 3,495 of those same 32,460 households have at least one person in the NPEU. By including the NPEU, 4,610 single filer households are added to the dataset. This process adds additional observations to the standard SCF dataset and revises the auxiliary variables needed for TAXSIM, creating an SCF based on tax units rather than households.

The most complex series of adjustments revolve around allocating income. For the split households, wages and salary are reallocated by multiplying the total by the fraction of wage and salary income associated with the respondent or the spouse over the combined value of the respondent and the spouse together. A similar process is used for income from social security. Business income is assigned to the respondent or spouse who works in the business or split evenly between them if both or neither work in the business. Similarly, unemployment benefits are assigned to the respondent or spouse who reported having a spell of unemployment, while child support and alimony is allocated using marital history data. Income from short and long term capital gains, other business, royalties, trusts, and other forms of passive income is simply split evenly.

All withdrawals and benefits from IRAs or employer-provided pension plans are associated with either the respondent or spouse, so total withdraws and payouts are aggregated as separate totals. The separate totals are used to recalculate total pension income for TAXSIM as the separate contribution to pension income less social security plus withdrawals. All other forms of income are split evenly.

For the NPEU the amount of income is split evenly between the reported income categories, with the amount per category distributed across each person in the NPEU.

For most deductions, we assume a tax-minimization approach by assigning the mortgage interest, property tax, charitable contributions and interest expense deductions to either the respondent or the spouse based on who has the highest income. For this purpose, income is the total share contributed by the respondent or spouse to labor and pension income. The NPEU is assumed to be claiming the standard deduction.

Having adjusted total income and deductions, the only adjustment remaining is the number of exemptions. Obviously, the exemption for the other spouse is removed. We assume that both the respondent and the spouse claim all children under 18 in the household as dependents on both returns. Doing so appears to fit the data better, even though the practice is technically illegal. NPEU households are assumed to be taking only one personal exemption. The exemption for being over 65 is recalculated based on the new household membership for both the split households and the NPEU households.

Using the results of these adjustments in addition to the marital history information, a filing status can be imputed. Households where the respondent and the spouse are legally

married for longer than a year and living together are taken to be filing jointly. Households filing married filing separately are those that have remarried, divorced, or separated within the last year along with married couples living in different places. The remaining households are classified as single filers if the number of personal exemptions equals one, and head of household if greater than one.

After assigning a filing status to each tax unit, we calculate an estimate of adjusted gross income (AGI), total deductions, and the total exceptions. AGI is recalculated for split households and NPEU households based on the adjusted values for total income, including pension withdrawals but excluding nontaxable income. AGI is used together with the filing status when calculating the total dollar amount of exemptions. The exemption phase-outs are recalculated using standard IRS rules, as are the standard deductions. For split and NPEU tax units, any tax variable that depends on income or filing status will need to be recalculated.

As a final step, we flag some cases that likely did not file taxes in 2009. Tax units that had an AGI lower than the filing requirements for their filing status and age are reclassified as "nonfilers." This is a purely income-based approach to estimating non-filers and ignores households who opted to file a return when one was not required. A 2003 report by the Tax Policy Center (Orzag and Hall, 2003) estimated that 18 million tax units out of 139 million did not file a tax return. We estimate that roughly 25 million tax units out of 146 million did not file a tax return (Table 1). As the number of nonfilers is not measured by SOI data by design, the inferred number of nonfilers in the SCF cannot be verified by comparing the two datasets.

4. Results

Table 1 presents the results for the number of filers by filing status for the two SCF methods and the SOI data. Compared to the SCF method, the Tax-SCF matches closer to the SOI data for the two married filing statuses and for single filers. Both methods only generate about one-half of the head of household filers compared to the SOI data. This shortfall in head of household filers may be due to misclassification of some actual head of household filers as nonfilers. About 20 percent of nonfilers had an original filing status of head of household prior to applying the income limits used to flag nonfilers. Reclassifying these nonfilers as head of household would eliminate about one-half of the gap between the SCF estimates and the SOI data. Although both methods generate a similar total number of filers, the Tax-SCF method is better match across filing status categories.

Figure 2 shows the number of filers by size of AGI, where the two AGI groups are less than $50,000 and $50,000 or more. The graph shows the ratio of the estimates from SCF method and Tax-SCF method to the SOI data. This format will be used in throughout the analysis. As evident from the figure, both methods generate too few filers in the lower AGI class and too many filers in the higher AGI class. This result is partially driven by how AGI is computed in the SCF data. The SCF has no information on adjustments to total income, which could lead to an upward bias in AGI.

Figure 3 compares the components of total income generated under the two methods to the SOI data. There is little difference between the two methods across the components of income, as evidenced by the estimates of total income under both methods being less than 10 percent higher than the SOI data. However, there is sizeable variation in how well the

components match up with the SOI data. The SCF estimates of interest, dividends, pension, annuities and Social Security income are only about 60 percent and 80 percent of SOI values, but this deficit is more than offset by the SCF estimate of broad business income, which is 160 percent of the SOI value.[6] For details on why the SCF income components may not match the SOI data, see Johnson and Moore (2008). Examining total income by AGI class reveals that the Tax-SCF method estimate is about 85 percent of the SOI value for the lower AGI group, and about 115 percent for the high AGI group. The SCF method has a similar ratio for the high AGI group, but only captures about 70 percent of total income in the lower AGI group. This difference is driven by the creation of separate tax units for cohabiting couples and for any individuals in the NPEU.

Figure 4 shows the ratio of the number of exemptions under the two methods to the SOI data. Overall, the total number of exemptions under both methods is about 90 percent of the SOI value, and the Tax-SCF method appears to match better across filing status. Both methods have far too few exemptions for head of household filers, which is not surprising given that we understate the number of filers in this group by about 50 percent.

Figure 5 provides results on the number of filers that itemize deductions. Determining itemization status is important part of calculating taxable income, and both SCF methods underestimate the number of itemizers by about 20 percent. Although neither method dominates in determining itemization status, both have the highest ratio to the SOI value for the married filing jointly filers. For the other filing status categories, the number of itemizers may be low relative to the SOI data due to deductions not captured in the SCF data. For example, the SCF does not have data on medical and dental expenses, casualty and theft losses, and moving expenses. These deductions are a larger share of itemized deductions for filers in the lower AGI group.

Figure 6 shows the ratio of the value of itemized deductions by filing status for each method to the SOI data. The pattern across filing status categories is similar to the number of itemizers, but both SCF methods are a closer match to the SOI data. For total deductions, the ratio of both methods to the SOI total is about 90 percent, with the ratio being nearly 100 percent for married filing jointly filers. Head of household filers have the lowest ratio, similar to results seen in earlier comparisons. However, there is a stark contrast when we examine the value of deductions by AGI class. For the higher AGI group, the ratio is almost 100 percent for both methods, while the ratio is only about 45 percent for the lower AGI group.

Figure 7 compares taxable income under the SCF methods to the SOI data across filing status. Overall, total taxable income for all filers is 20 percent higher under the SCF method and almost 40 percent higher under the Tax-SCF method. This result is not surprising given that although the SCF income data matches the SOI data quite closely, the number of exemptions and amount of itemized deductions are lower under both SCF methods.

Figure 8 shows the ratio of tax liability by filing status for each SCF method to the SOI data. Under both of the SCF methods, total tax liability is about 35 percent higher than the SOI value, and the ratio is over 100 percent for all filing status categories except head of household filers. The higher estimate of tax liability is a direct result of SCF income

---

[6] Broad business income includes business income from sole proprietorships, farms, capital gains and losses, income from rent, royalties, and other businesses reported on Schedule E.

matching the SOI data fairly closely, while exemptions and deductions are lower under both SCF methods than the SOI data.

Figure 9 presents the distribution of federal bracket rates for the two SCF methods and the SOI data. Unlike the previous figures, this chart shows the fraction of filers in each federal bracket rate category.[7] Although estimated tax liability is higher under both SCF methods, the distribution of filers across federal brackets is similar to the SOI data. Both SCF methods have too few filers in the 10 and 15 percent rates, and too many filers in the rates about 25 percent. Of the two SCF methods, the Tax-SCF method more closely matches the SOI data. The differences between the two methods are even more apparent if we examine the distribution of bracket rates by filing status. Figure 10 shows the distribution of rates for married filing jointly filers and the differences between the two SCF methods is very small. However, the differences in figure 11, which shows the distribution of rates for single filers, are much larger. Figure 11 is evidence of the advantages to the Tax-SCF method.

5. Conclusions

The results of comparing the two SCF methods for defining tax units shows that the Tax-SCF method is a better match with the SOI data for the number of filers and the distribution of federal bracket rates. Both methods generate estimates of tax liability that are about 35 percent higher than the SOI data. The higher tax liability is driven by the lower number of exemptions and lower amount of itemized deductions generated by the SCF data. An area for future work is to add more types of itemized deductions, possibly by imputing values from the SOI data by AGI group. Another issue is that both SCF methods generate only about 50 percent of number of head of household filers reported in the SOI data, and for the Tax-SCF changes to the how we classify nonfilers may help close that gap. Even with these issues, estimates of federal bracket rates from the Tax-SCF method match the SOI distribution fairly closely. Reasonable estimate of tax rates combined with the extensive household balance sheet data in the SCF provide a rich resource for the analysis of changes in and the distributional effects of tax policies.

Acknowledgements

References

Feenberg, D., & Coutts, E. (1993). An introduction to the TAXSIM model. Journal of Policy Analysis and Management, 12(1), 189-194.

Johnson, Barry, and Kevin B. Moore (2008). Differences in Income Estimates Derived from Survey and Tax Data, *Proceedings of the Survey Research Methods Section*, American Statistical Association. American Statistical Association, pp. 1495-1503.

---

[7] The federal bracket rate is the highest rate paid on taxable income.

Orszag, Peter R. and Matthew G. Hall. (2003). Nonfilers and Filers With Modest Tax Liabilities." *Tax Facts*. Tax Policy Center.

Individual Income Tax Returns Publication 1304 (2009). Tables 1.2, 1.3, 2.1, 2.2, 2.4, and 3.4, Internal Revenue Service.

Figure 1: Mapping SCF variables to IRS Form 1040

Table 1: Number of Filers by Filing Status in 2009, SCF vs. SOI

Filing Status	SCF	Tax-SCF	SOI
Married filing Jointly	65,595,407	53,812,417	53,639,038
Married filing Separately	0	2,342,006	2,539,588
Head of Household	10,290,017	11,639,532	21,496,275
Single	41,723,994	52,814,896	62,819,226
Nonfilers	0	25,571,086	0
Total filers	117,609,418	120,608,851	140,494,127
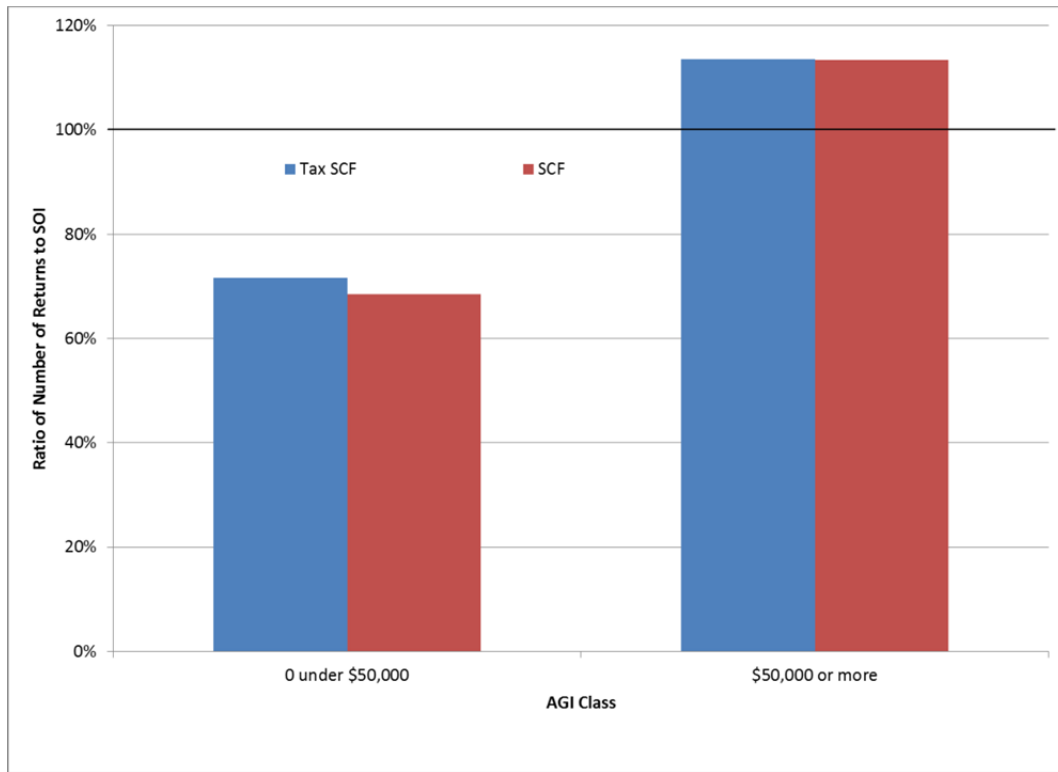
Figure 2: Ratio of Number of Returns to SOI Data, by AGI Class (2009)
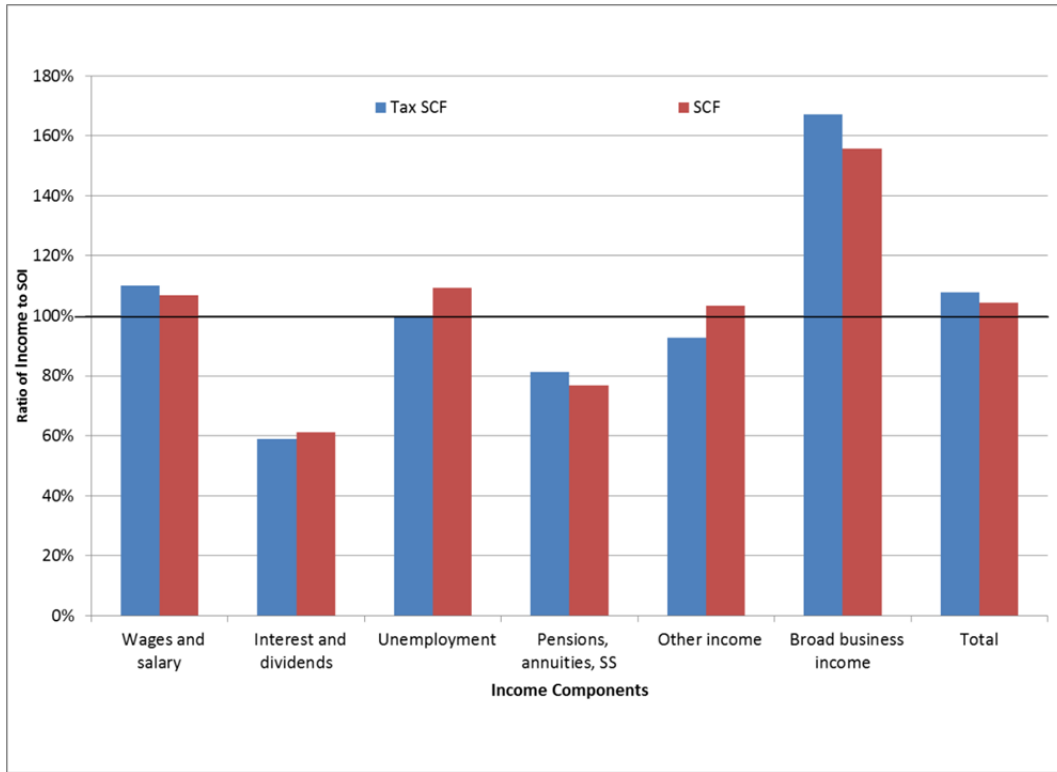
Figure 3: Ratio of Income Components to SOI Data, All Filers (2009)
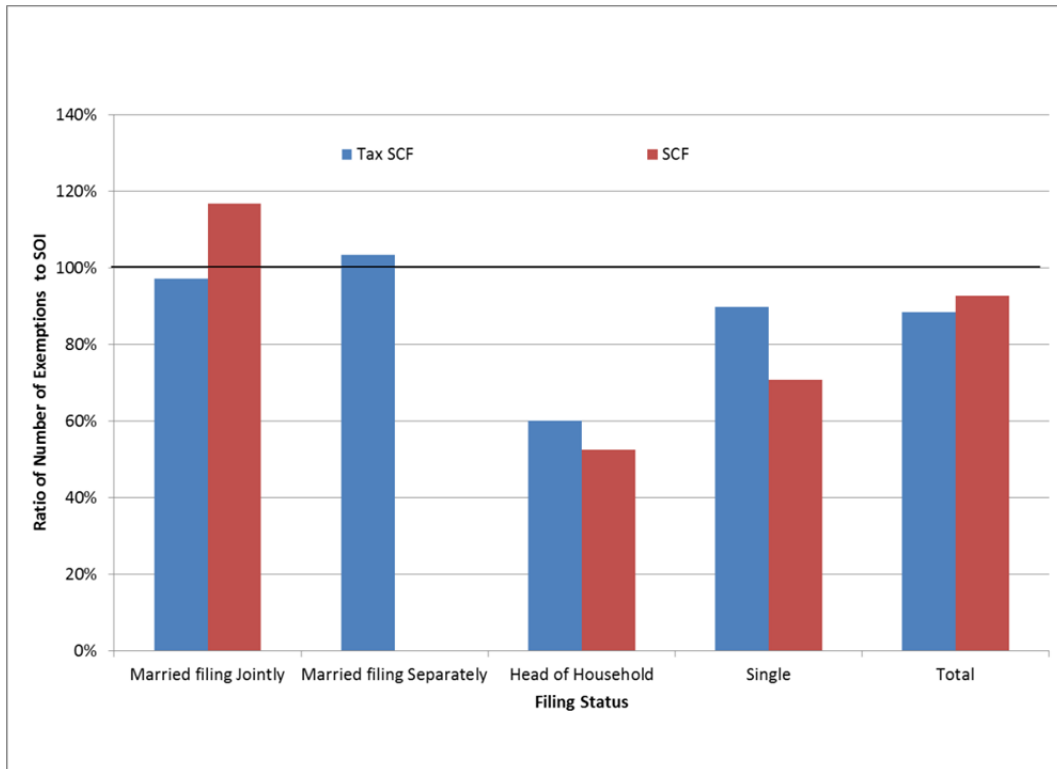
Figure 4: Ratio of Number of Exemptions to SOI Data, by Filing Status (2009)
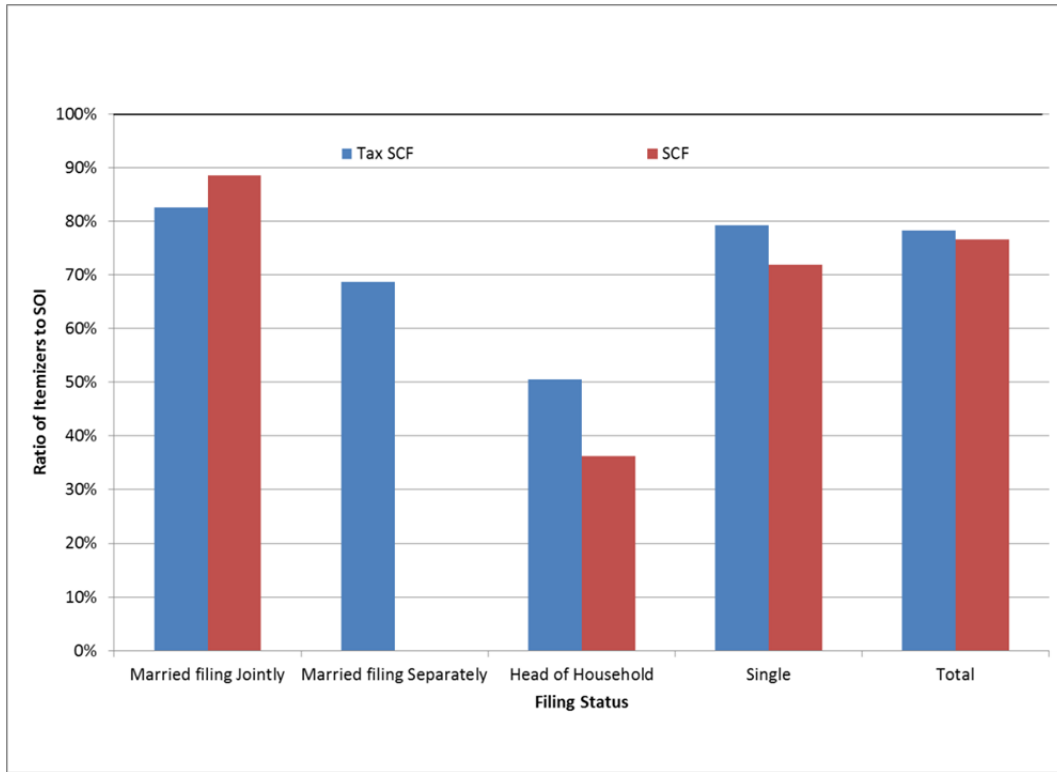
Figure 5: Ratio of Itemizers to SOI Data, by Filing Status (2009)
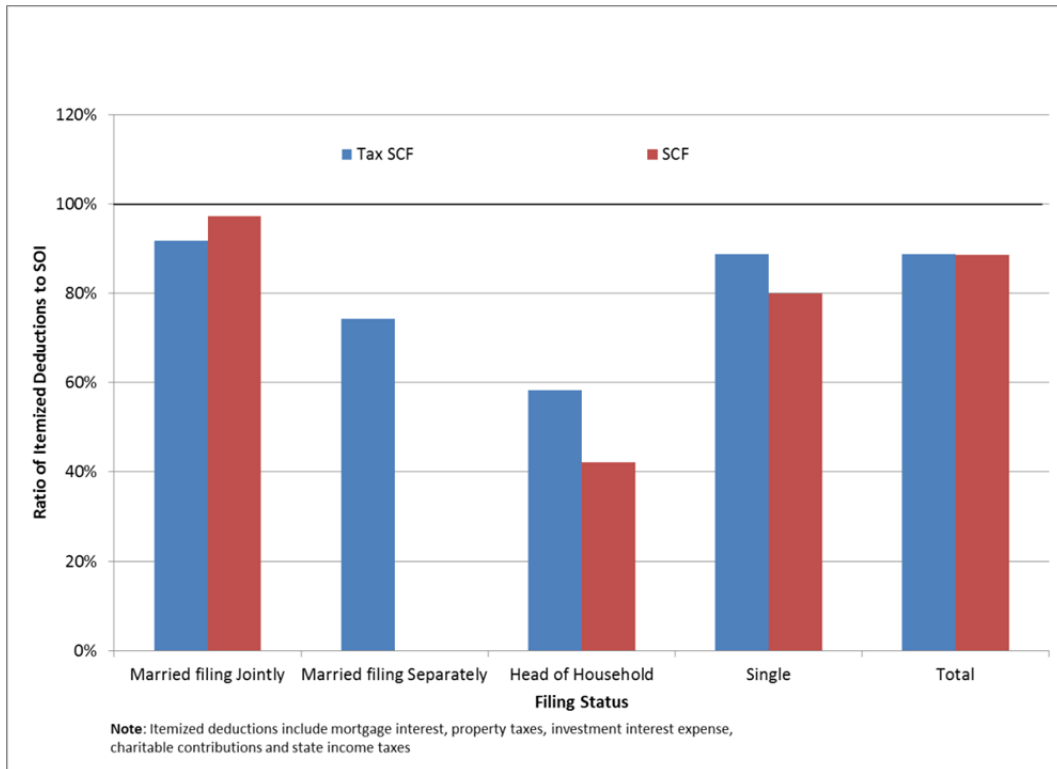

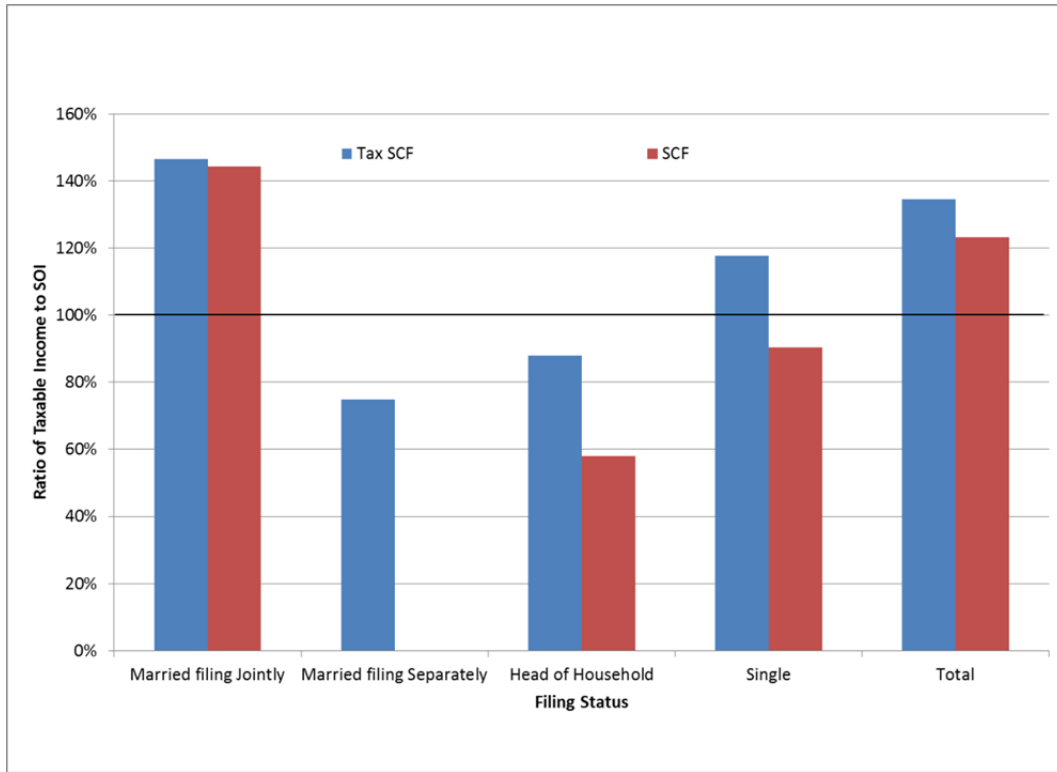**Figure 6:** Ratio of Value of Itemized Deductions to SOI Data, by Filing Status (2009)

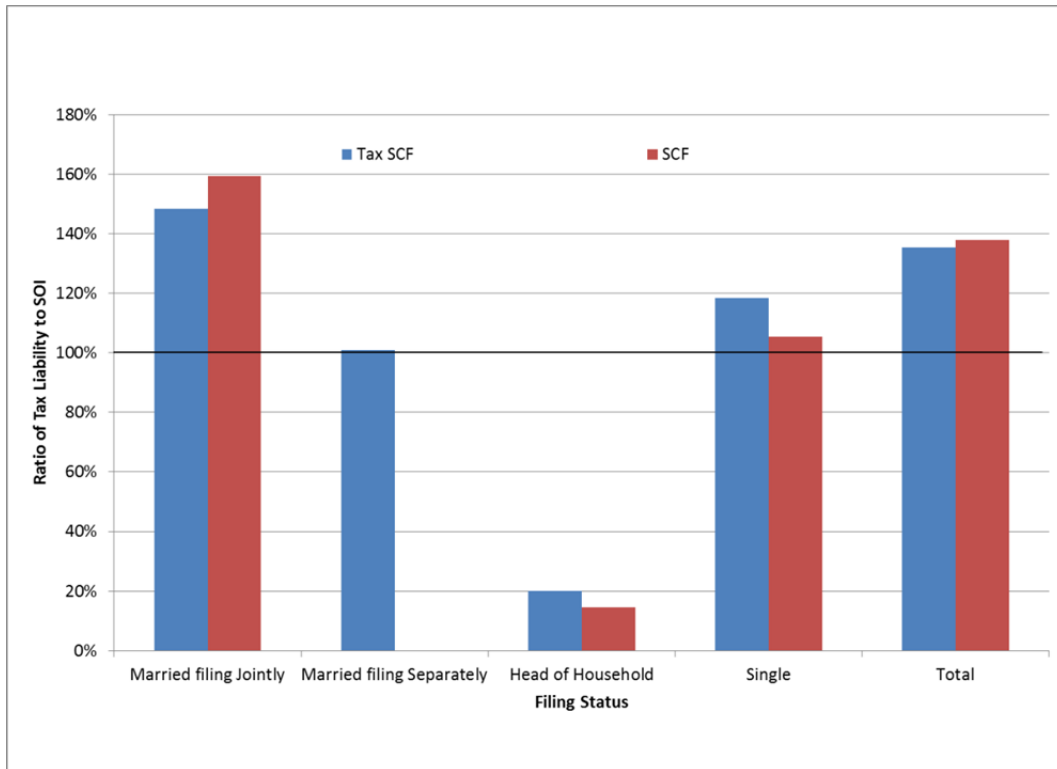**Figure 7:** Ratio of Taxable Income to SOI Data, by Filing Status (2009)

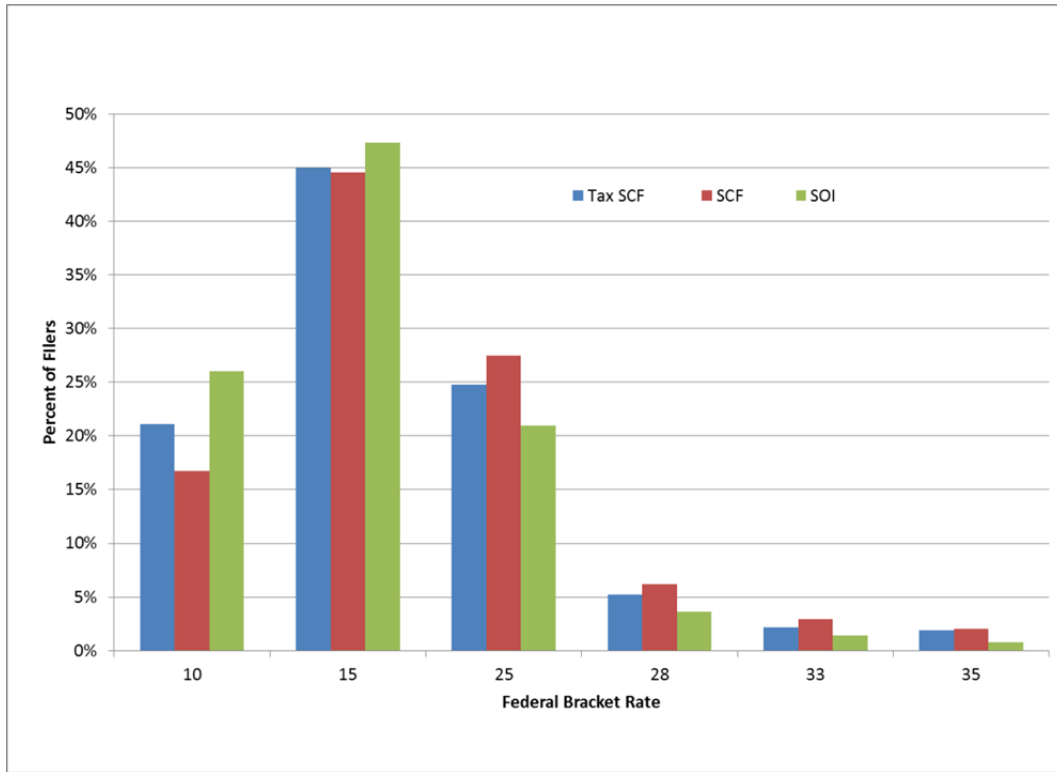**Figure 8:** Ratio of Tax Liability to SOI Data, by Filing Status (2009)

Figure 9: Distribution of Filers in Federal Bracket Rates (2009)
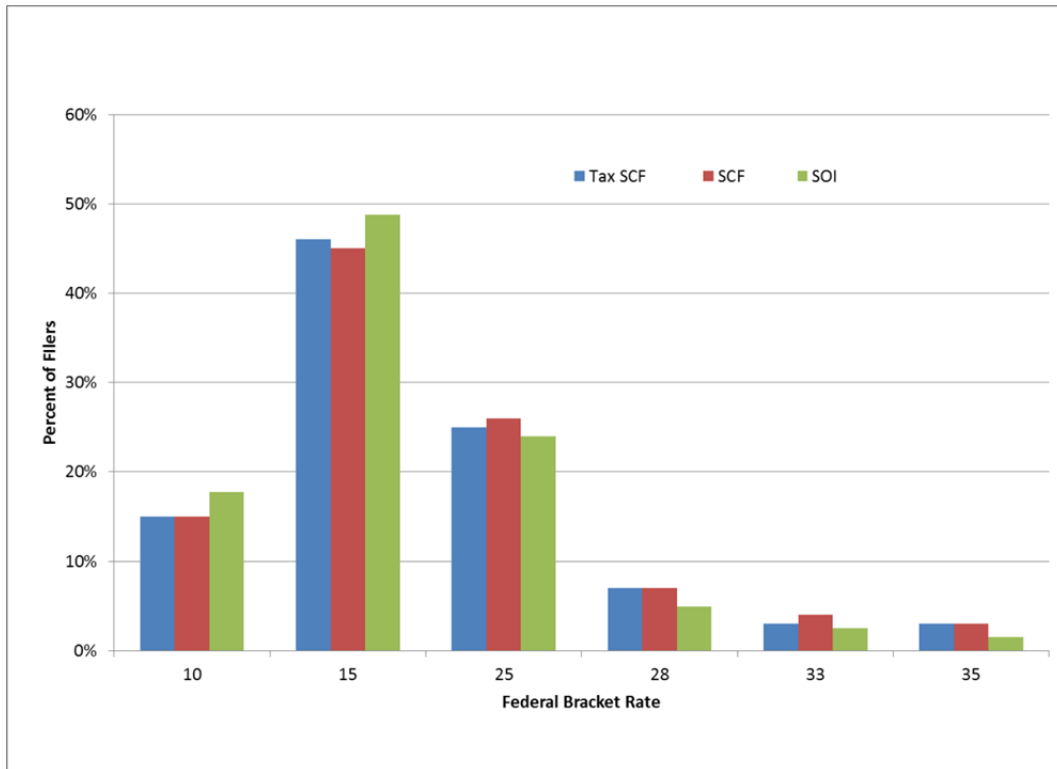

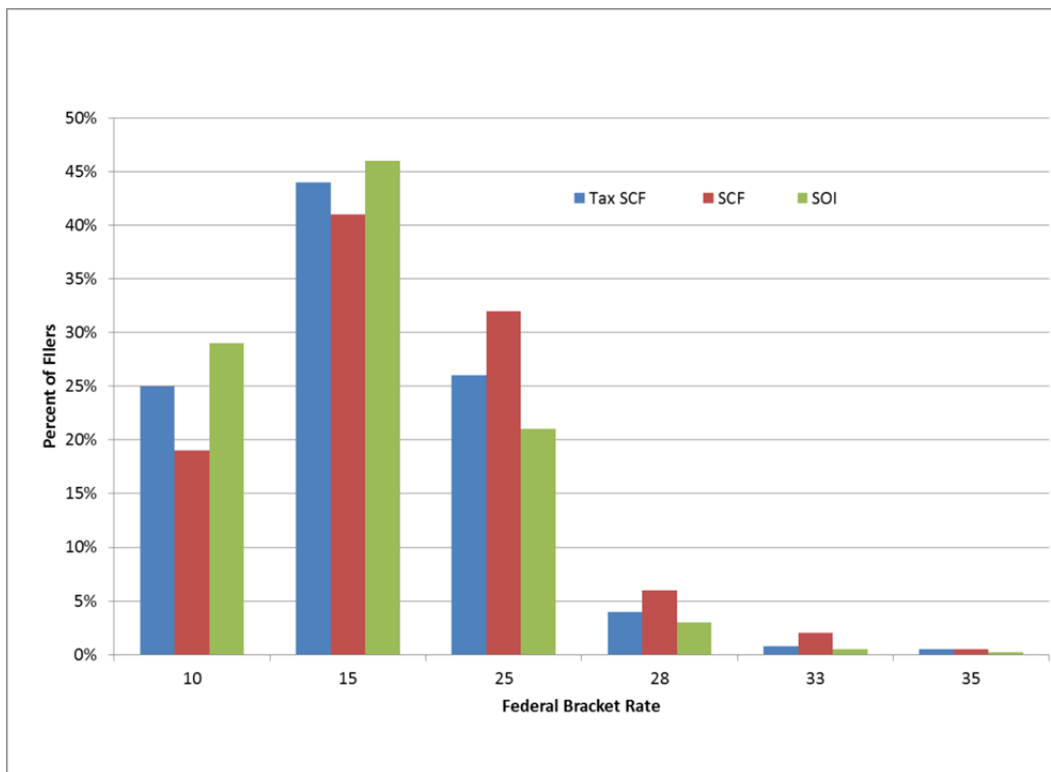**Figure 10:** Distribution of Married Filing Jointly Filers in Federal Bracket Rates (2009)

Figure 11: Distribution of Single Filers in Federal Bracket Rates (2009)