# An Updated Method for Calculating Income and Payroll Taxes from PSID data using the NBER's TAXSIM, for PSID Survey Years 1999 through 2011

Sara Kimberlin, Jiyoon Kim, & H. Luke Shaefer

June 2015

**Overview**

This paper describes a method to calculate income and payroll taxes from Panel Study of Income Dynamics data using the NBER's Internet TAXSIM version 9 (http://users.nber.org/~taxsim/taxsim9/), for PSID survey years 1999, 2001, 2003, 2005. 2007, 2009, and 2011 (tax years n-1).

These methods are implemented in two Stata programs, designed to be used with the PSID public-use zipped Main Interview data files: PSID_TAXSIM_1of2.do and PSID_TAXSIM_2of2.do. The main program (2of2) was written by Sara Kimberlin (skimberlin@berkeley.edu) and generates all TAXSIM input variables, runs TAXSIM, adjusts tax estimates using additional information available in PSID data, and calculates total PSID family unit taxes. A separate program (1of2) was written by Jiyoon (June) Kim (junekim@umich.edu) in collaboration with Luke Shaefer (lshaefer@umich.edu) to calculate mortgage interest for itemized deductions; this program needs to be run first, before the main program. Jonathan Latner contributed code to use the programs with the PSID zipped data.

The overall methods build on the strategy for using TAXSIM with PSID data outlined by Butrica & Burkhauser (1997)[1], with some expansions and modifications.

Note that the methods described below are designed to prioritize accuracy of income taxes calculated for low-income households, particularly refundable tax credits such as the Earned Income Tax Credit (EITC) and the Additional Child Tax Credit. Income tax liability is generally low for low-income households, and the amount of refundable tax credits is often substantially larger than tax liabilities for this population. Payroll tax can also be substantial for low-income households. Thus the methods below focus on maximizing accuracy of income tax and payroll tax calculations for low-income families, with less attention to tax items that largely impact higher-income households (e.g. the treatment of capital gains).

Key features of the tax calculation approach include:

**Tax units**

The method described here follows Butrica & Burkhauser (1997) and differs from more simplified PSID TAXSIM interface approaches in that multiple tax units are identified within each PSID family unit. Thus cohabiting couples are treated as two separate tax units, with children assigned to the appropriate tax unit (head or cohabitor) using relationship codes.

Sub-households comprised of "other family unit members" (ofums) who are not dependent children of the head or cohabitor are also treated as their own tax units. However, all ofum adults and all children not assigned to head or wife are combined into a single ofum tax unit (i.e. no creation of multiple ofum tax units based on relationship codes). In PSID data available to public users, all ofum income is combined, not reported by individual, so there is no way to separate ofum income into multiple ofum tax units.[2]

---

[1] Butrica, B. & Burkhauser, R. (1997). Estimating federal income tax burdens for Panel Study of Income Dynamics (PSID) families using the National Bureau of Economic Research TAXSIM model. Aging Studies Program Paper No. 12. Syracuse, NY: Center for Policy Research, Maxwell School of Citizenship and Public Affairs, Syracuse University.

[2] Note that the methods described in Butrica & Burkhauser (1997) appear to make use of individual-level income data for "other family unit members," but this level of detail for ofum income is not available in PSID public use or restricted use data

The basic strategy for building the tax units is as follows. For each PSID family unit, four sets of TAXSIM variables are calculated representing four different potential tax units: a head-only tax unit (with head dependents), a "wife"/cohabitor-only tax unit (with cohabitor dependents), a joint/married tax unit (with head and legal wife dependents), and an ofum tax unit (with all adults besides the head, wife, and "wife"/cohabitor and all children not assigned to the head, wife, or "wife"/cohabitor). Taxes are calculated for each tax unit using TAXSIM, and then each tax unit's tax amounts are adjusted to account for additional information about self-employment and immigrant legal status available in the PSID. Finally, total PSID family unit taxes are then calculated by summing the taxes for the tax units actually present within the family unit, namely:

- For a head not residing with a wife or "wife"/cohabitor, with or without other family unit members – head-only taxes + ofum taxes

- For a head legally married to a wife, with or without ofums – joint taxes + ofum taxes

- For a head residing with a "wife"/cohabitor, with or without ofums – head-only taxes + "wife"/cohabitor-only taxes + ofum taxes

- Note that if there are no ofums in the PSID family unit, then the TAXSIM output for ofum taxes is equal to zero.

For comparison purposes, an additional set of TAXSIM output variables is created for a "simple" tax unit that combines all PSID family unit members into a single tax unit and uses only very basic income and family composition variables available in the Family File without any intermediary calculations. These "simple" variables can be used to see the effect on estimated family unit taxes of using more complex tax units and income and expense inputs.

Also note that the approach here uses move-in/out dates to determine the number of PSID family members present during the tax year for purposes of calculating number of dependents and seniors and number of tax units within the PSID family unit.

A variety of strategies are used to generate specific TAXSIM inputs:

**Dependents**

- Only child dependents are considered, not adult dependents. This means no adults are counted as dependents, even if their income is below the filing threshold. This is done for two reasons:

  o TAXSIM uses a single input variable –depx– to calculate both the dependent exemption *and* the value of the EITC. The tax calculation method described here prioritizes accurate calculation of the EITC, thus only dependents who could be EITC qualifying children are counted.[3,4]

---

files accessible to public PSID users as of 2014, as confirmed by PSID staff (C. Brown, personal communication, October 2014).
[3] Adult dependents could be added using variables generated in the accompanying Stata program for more accurate dependent exemption amounts, with the caveat that EITC amounts would then be skewed.

- o Another reason adult dependents are not included is that income of all adults besides head and wife, and all children, is lumped together as "other family unit member" income in PSID data available to public users, so it can be hard to tell if a particular adult could qualify as a dependent.

- This program counts all children age 0 to 18 who were present at some time during the tax year as dependents, without considering whether they lived in the family at least half of the year or whether they had income too high to qualify as dependent. Again, individual child income is not reported separately from "other family unit member" combined income in PSID data, so it is difficult to know if a particular child had income too high to qualify as dependent.

- All children age 0 to 16 who were present at some time during the tax year are counted as qualifying children for the child tax credit.

- In cases where the ofum tax unit does not include any adults, or where the ofum tax unit has no earned income, the ofum tax unit will not be eligible for the EITC and will have limited eligibility for the Child Tax Credit, including the refundable portion of the CTC. In these cases, any children from the ofum tax unit who are related to the head or cohabitor are reassigned to the head or cohabitor tax unit so that they can serve as qualifying child dependents for the EITC, Child Tax Credit, and Additional Child Tax Credit (the refundable portion of the CTC). These children include siblings, grandchildren, cousins, etc. of the head or cohabitor; biological and adopted children, step children, and foster children are already assigned to the head or cohabitor tax units.

- Head of household status is assigned based on presence of any related children, without considering whether the tax filer paid more than half of the child's expenses (not clear if this can be determined for particular children in PSID data, and support test not required for EITC or CTC qualifying children).

**Income**

- The most detailed income and expense information available in the PSID is used – meaning highly detailed information for head, somewhat less detailed for wife/cohabitor, much less detailed for "other family unit members" (ofums).

- It is worth noting that ofum income for PSID survey years 1999 through 2011 is reported in annual amounts (total income received by the ofum during the calendar year). This is in contrast to the reporting of ofum income for survey years up until the mid-1990s, in which reported ofum income was prorated to account for the amount of time the ofum lived with the PSID family.[5]

---

[4] Adult children who meet the qualifying child relationship test and are disabled, as well as those who are up to age 24 and full-time students, are also considered qualifying children for the EITC. However, individual-level data on "other family unit member" disability and school enrollment status are not available in the PSID, so these types of EITC qualifying children are ignored here.

[5] PSID staff confirmed that ofum income for the years 1999 through 2011 is not prorated in either the individual ofum income variables or when summed with head and wife income to generate total family income (C. Brown & F. Campbell, personal communication, November 2014).

*Labor income*

- Income earned through employment, unincorporated businesses, and farms is included.

- Both business/farm "earnings" and "asset" income are treated as wages/salary (business/farm "asset" income is not treated as property income for TAXSIM purposes). This is done because when the head or wife puts in any actual work hours in the business/farm, the PSID business/farm earnings and asset variables are calculated by arbitrarily dividing the total business/farm income by two, counting half the income as "earnings" and half as "assets." The IRS does not follow that process for taxing an individual's business/farm income.

*Asset/property income*

- A variety of separate property/asset income items are summed to generate the TAXSIM property income and dividends input variables.

- Short-term and long-term capital gains are set to zero. It might be possible to estimate these from asset and other data in PSID, but as they are generally an insignificant source of income for low-income households, they are ignored here.

*Retirement income*

- Multiple pension and retirement income items are used to generate the TAXSIM taxable pension input variables. Social Security benefits are drawn from direct reports in the PSID.

*Transfer income*

- A variety of separate transfer income items are summed to generate the TAXSIM transfer income and unemployment input variables.

**Expenses**

- Directly reported data from the PSID are used to generate TAXSIM input variables for rent paid, child care expenses, and alimony paid (used to adjust property income).

*Itemization*

- Itemized deductions are calculated for married and head-only tax units where the family reported itemizing, based on the " wtr itemize for taxes" variable in the PSID Family File. Mortgage interest is calculated (see appendix below for details on calculation method) and reported values for property tax, itemized charitable contributions, and itemized medical expenses are used.

- Other itemized deductions that are a preference for the Alternative Minimum Tax (e.g. local income tax) are ignored, as they are not readily available in PSID data and AMT generally does not affect low-income households.

**Item-missing values**

While some variables needed to calculate taxes include PSID-generated imputed values for item-missing data, many do not. Thus item-missing values for income and expense items are generally imputed here (particularly for years prior to 2005) using the simple approach of substituting the median non-zero value by family unit. This follows the convention for PSID-provided imputed values for most income items from 2005 on (except labor income, for which PSID uses a more complex imputation strategy).[6]

Exceptions to this missing data strategy include missing values for property tax paid, charitable gift deduction, and medical expense deduction, which are not imputed. Instead they are set to zero in the TAXSIM input variables (as these items are generally small amounts for low-income households that do not substantially impact their tax liabilities/credits). Missing values for the mortgage interest deduction are also effectively set to zero (as no value is calculated) if any components required to calculate total mortgage interest are missing.

**Adjustments and corrections to TAXSIM output**

After using the input variables to run TAXSIM, the output tax estimates are adjusted in a few ways to account for information available in the PSID that cannot be incorporated in TAXSIM.

*Self-employment*
With Internet TAXSIM version 9, self-employment income is not entered as a separate variable but is rather combined with wage/salary income from external employment. As a result, TAXSIM does not correctly calculate payroll taxes for tax filers who have both wage/salary income through outside employment as well as a loss from self-employment in an unincorporated business or farm. Thus payroll taxes are manually re-calculated for these filers before calculating total family unit taxes.

TAXSIM also does not calculate the employer portion of payroll tax for individuals with profit from self-employment in an unincorporated business or farm (paid as half of self-employment tax). Thus this additional payroll tax is calculated for these tax filers, discounted by the marginal federal-plus-state income tax rate in order to approximate the allowed deduction of the employer portion of self-employment tax.

*Immigrant ineligibility for the EITC*
Unauthorized immigrants and those whose immigration status does not legally allow them to work are not eligible for the EITC. The PSID contains information about immigrant legal status for the head, wife, and "wife"/cohabitors of families added to the PSID in 1997 and 1999 as the "immigrant refresher sample." These variables are used to mark tax filers who are ineligible for the EITC, and their EITC amounts are set to zero before calculating total family unit taxes.

---

[6] See PSID documentation: Duffy, D. (2011). 2007 PSID Income and Wage Imputation Methodology. PSID Technical Series Paper #11-03. Survey Research Center, Institute for Social Research, University of Michigan. Retrieved from http://psidonline.isr.umich.edu/Publications/Papers/tsp/2011-03_Duffy.pdf#page=1.

Immigrants not legally permitted to work are eligible to claim the Child Tax Credit and Additional Child Tax Credit (the refundable portion of the CTC) if their qualifying children are legal U.S. residents, which is frequently the case (e.g. if their children were born in the U.S.). Thus this program does not adjust CTC or refundable CTC amounts calculated for immigrants; they are assumed eligible to claim these credits.

**Differences compared to simplified tax calculations**

This more complex approach to accounting for multiple tax units within PSID family units, more accurate counts of family members present during the tax year, and more detailed income and expense information produces some notably different tax estimates for low-income households than a more simplified approach (using the "simple" tax variables calculated here) that assumes a single tax unit per PSID family unit and uses only summary taxable income variables and interview-time family composition. The weighted number of PSID family units estimated to receive any refundable tax credit benefit (EITC and/or refundable portion of CTC) with this more complex approach is an average of 1.6 times the number estimated to receive any refundable credit with the simpler approach across the years examined. In addition, the weighted total amount of refundable tax credits estimated to be claimed in each year is an average of 1.2 times the estimated amount claimed using the simpler approach.[7]

In addition, adjusting for immigrant ineligibility for the EITC results in changed federal tax estimates for an average of 11.6% (weighted) of PSID families who are part of the immigrant refresher sample or their descendant families (though this is a small absolute number of families). These families' annual tax liabilities increased by an average of $1,918.

Calculating the additional self-employment tax resulted in an average $1,111 (weighted) increase in annual payroll taxes for self-employed family unit heads (though again this affects a small absolute number of tax filers).

---

[7] In terms of the EITC alone, the more complex method results in 1.6 times as many PSID family units receiving the credit and 1.3 times the total amount of credit claimed. In terms of the refundable portion of the CTC alone, the more complex method results in 1.1 times the number of PSID family units receiving the credit and 1.1 times the total amount of credit claimed.

## TAXSIM input variables for each type of tax unit

| TAXSIM variable | | Head-only tax unit | "Wife" (cohabitor)-only tax unit | Joint (married) tax unit | Ofum tax unit | Simple tax unit |
|---|---|---|---|---|---|---|
| 1) id | unique identifier | generated variable personid | generated variable personid | generated variable personid | generated variable personid | generated variable personid |
| 2) year | year | = survey year - 1 | = survey year - 1 | = survey year - 1 | = survey year - 1 | = survey year - 1 |
| 3) state | state | state recoded to SOI code | state recoded to SOI code | state recoded to SOI code | state recoded to SOI code | state recoded to SOI code |
| 4) mstat | filing status | default = single. recoded as head of household if generated variable sumheadchild18 >=1 | default = single. recoded as head of household if generated variable sumwifechild18 >=1 | =joint | default = single. recoded as head of household if generated variable sumofumtaxchild >=1. recoded as joint if generated variable sumofumtaxadult >1. | default = single. recoded as head of household if number of children (variable from Family File) >1. then recoded as joint if couple status of head = married |
| 5) depx | number of dependents | generated variable sumheadchild18 | generated variable sumwifechild18 | sum of sumheadchild18, sumwifechild18 | generated variable sumofumtaxchild | number of children (variable from Family File) |
| 6) agex | number of seniors | =1 if age of head >=65 | =1 if age of wife >=65 | sum of agex for head, wife | generated variable sumofumtaxsenior | from age of head, age of wife |

| TAXSIM variable | | Head-only tax unit | "Wife" (cohabitor)-only tax unit | Joint (married) tax unit | Ofum tax unit | Simple tax unit |
|---|---|---|---|---|---|---|
| 7) pwages | wages of primary taxpayer | sum of head labor income, head labor income from business, head asset income from business, family farm income | sum of wife labor income, wife labor income from business, wife asset income from business | value of pwages for head | ofum taxable income (pre-2005), ofum labor income (2005 on) | sum of head-wife taxable income, ofum taxable income (variables from Family File) |
| 8) swages | wages of secondary taxpayer | =0 | =0 | value of pwages for wife | =0 | =0 |
| 9) dividends | dividend income | head dividends (for 1999, 2001 – from 2003 on this TAXSIM variable is qualified dividends only, so set to zero here and dividends reported under other property income) (pre-2005 summed monthly income) | wife dividends (for 1999, 2001 – from 2003 on this TAXSIM variable is qualified dividends only, so set to zero here and dividends reported under other property income) (pre-2005 summed monthly income) | sum of head dividends, wife dividends | =0 | =0 |

| TAXSIM variable | | Head-only tax unit | "Wife" (cohabitor)-only tax unit | Joint (married) tax unit | Ofum tax unit | Simple tax unit |
|---|---|---|---|---|---|---|
| 10) otherprop | interest and other property income | sum of head interest income, head rent income, head trust fund income, head alimony received, head annuities, head dividends from 2003 on, minus head alimony paid<br><br>(pre-2005 summed monthly income for each income item) | sum of wife interest income, wife rent income from 2003 on (not available pre-2003), wife trust fund income, wife dividends from 2003 on<br><br>(pre-2005 summed monthly income for each income item) | sum of value of otherprop for head, otherprop for wife | =0 pre-2005 (not available), ofum asset income from 2005 on | =0 |
| 11) pensions | taxable pension income | sum of head retirement/ pensions, head other retirement<br><br>(pre-2005 summed monthly income for each item) | wife pensions<br><br>(pre-2005 summed monthly income) | sum of value of pensions for head, pensions for wife | =0 pre-2005 (not available), ofum pensions from 2005 on | =0 |

| TAXSIM variable | | Head-only tax unit | "Wife" (cohabitor)-only tax unit | Joint (married) tax unit | Ofum tax unit | Simple tax unit |
|---|---|---|---|---|---|---|
| 12) gssi | gross Social Security income | family Soc Sec income pre-2005 (individual amounts not available pre-2005), head Soc Sec income from 2005 on | =0 pre-2005 (individual amounts not available pre-2005), wife Soc Sec income from 2005 on | family Soc Sec income pre-2005 (individual amounts not available pre-2005), sum of head Soc Sec income plus wife Soc Sec income 2005 on | =0 pre 2005 (individual amounts not available pre-2005), ofum Soc Sec income from 2005 on | =0 |
| 13) transfers | non-taxable transfer income | sum of head TANF, head SSI, head other welfare, head VA pension, head workers comp, head child support received (pre-2005 summed monthly income for each item) | sum of wife TANF, wife SSI, wife other welfare, wife workers comp, wife child support received (pre-2005 summed monthly income for each item) | sum of value of head transfers, wife transfers | =0 | =0 |
| 14) rentpaid | rent paid | rent paid x rent time period | =0 | rent paid x rent time period | =0 | =0 |
| 15) proptax | property tax paid | real estate taxes paid | =0 | real estate taxes paid | =0 | =0 |

| TAXSIM variable | | Head-only tax unit | "Wife" (cohabitor)-only tax unit | Joint (married) tax unit | Ofum tax unit | Simple tax unit |
|---|---|---|---|---|---|---|
| 16) otheritem | other deductions that are preferences for the AMT | =0 | =0 | =0 | =0 | =0 |
| 17) childcare | child care expenses paid | childcare paid prorated by proportion of child dependents assigned to head | childcare paid prorated by proportion of child dependents assigned to cohabitor | childcare paid prorated by proportion of child dependents assigned to head + legal wife | childcare paid prorated by proportion of child dependents assigned to ofum tax unit | =0 |
| 18) ui | unemployment benefit income | head unemployment (pre-2005 summed monthly income) | wife unemployment (pre-2005 summed monthly income) | sum of head unemployment, wife unemployment | =0 pre-2005 (not available), ofum unemployment from 2005 on | =0 |
| 19) depchild | number of dependent children under age 17 | generated variable sumheadchild16 | generated variable sumwifechild16 | sum of generated variables sumheadchild16, sumwifechild16 | generated variable sumofumtaxchild16 | number of children at time of interview (variable from Family File) |

| TAXSIM variable | | Head-only tax unit | "Wife" (cohabitor)-only tax unit | Joint (married) tax unit | Ofum tax unit | Simple tax unit |
|---|---|---|---|---|---|---|
| 20) mortgage | mortgage interest paid, plus charitable contributions and other deductions not preferences for AMT | sum of mortgage interest (see separate memo on how calculated), itemized charitable contributions, itemized medical expenses | =0 | sum of mortgage interest (see separate memo on how calculated), itemized charitable contributions, itemized medical expenses | =0 | =0 |
| 21) stcg | short-term capital gain/loss | =0 | =0 | =0 | =0 | =0 |
| 22) ltcg | long-term capital gain/loss | =0 | =0 | =0 | =0 | =0 |

# APPENDIX:
# Constructing mortgage interest payment for the past calendar year from PSID data for PSID survey years 1999 to 2011

Mortgage interest paid in the past calendar year can be included as an itemized deduction. The PSID includes sufficient data on mortgages for detailed estimation of mortgage interest paid. This appendix describes the method used to generate the mortgage interest variable, as one component of the TAXSIM itemized deduction input variable –mortgage-.

**Mortgage variables in PSID**

| *Available in PSID* | *Used in Mortgage interest calculation* |
|---|---|
| Remaining principal amount | |
| Monthly mortgage payment | V |
| Current interest rate on the loan | V |
| Year obtained | V |
| Years to pay | V |

**Steps to construct intermediary variables**

1) *Monthly interest rate*
   Converted two-variable APR into single-variable monthly rate

   r= [(whole percent) + (decimal fraction x 0.001)] / 100 /12

2) *Adjusted monthly payment*
   Exclude annual value of property tax and/or insurance if those were included in monthly payment.

   Monthly payment = [(monthly payment x 12) - (insurance if included in mortgage) - (property taxes if included in mortgage)] / 12

3) *Calculate mortgage principal, $P_m$, for each month of past calendar year*

$$(1) \quad P_m = Adjusted\ Monthly\ payment \cdot \frac{(1+r)^n - 1}{r(1+r)^n}$$

   *n - the loan's remaining terms (in months) at month m past calendar year*

We assume that loan was initiated in JULY of the year when it was obtained.

Calculating remaining months n at each month m:

First, compute: diff=survey year – year obtained

- if diff = 0, then interest rate past year = 0
- if diff>1, n at January = 12 + (interview month-1) + years to pay * 12

  n at February = 11 + (interview month-1) + years to pay * 12

  n at March = 10 + (interview month-1) + years to pay * 12

  …..

  n at December = 1 + (interview month-1) + years to pay * 12

- if diff=1

  - Mortgage interest for Jan – June = 0
  - Since the mortgage is assumed to be initiated in July past year
  - Therefore,

    n at July = 6 + (interview month-1) + years to pay * 12

    n at August = 5 + (interview month-1) + years to pay * 12

    …..

    n at December = 1 + (interview month-1) + years to pay * 12

## 4) *Calculate the interest for each month of past calendar year*

$$(2) \quad I_m = interest\ payment\ in\ month\ m\ past\ year = P_m \cdot (1 + r)$$

## 5) *Mortgage interest sum past calendar year*

Sum of (2) from January to December

$$(3) \quad I_{past\ year} = \sum_{m=1}^{12} I_m$$