

Calculating Inputs for the NBER TAXSIM model, using the Consumer Expenditure Survey (CE) Public Use Microdata (PUMD)

Scott Curtin
Bureau of Labor Statistics
May 23, 2017

Background:

TAXSIM is a FORTRAN program created and supported by NBER that calculates liabilities under the US Federal and State income tax laws from individual data. The program is updated annually for both current year tax code changes and improvements to previous years' code.

Defining a Tax Unit:

The CE Tax Estimation process uses CE demographic and income from the CE Interview survey, to group each individual Consumer Unit (CU) into a Tax Unit (or multiple Tax units) by grouping CU members based upon being the tax head, spouse, or dependents. The primary tax unit will be defined as the tax unit which contains the reference person. All other tax units will be defined as secondary tax units. Tax Units income information is loaded into TAXSIM to produce the tax liability, by Federal and State, for both the current and previous tax year.

Each member of a CU will be identified as a dependent or a non-dependent based upon their income, age, relationship, and whether or not the member supports her/himself. Then, each member of the CU is assigned to a tax unit either with other CU members or by themselves (by assigning a tax unit identification number TAX_UNIT). The TAX_UNIT will serve as the basis by which unique records are created in the NTAXI files.

There will be one or more tax units in a consumer unit. One tax unit will be defined as the primary tax unit that will include the reference person, and other members. Any other tax units will be defined as secondary tax units.

Each tax unit will be a data row that is an input into the TAXSIM program. Data collected at the CU level in CE will be used to calculate the TAXSIM input variables only for the primary tax unit except where specified. Data collected at the member level in CE will be used to calculate the TAXSIM input variables for the tax unit containing that member. This member-specific information will be used for both the primary and secondary units.

TAXSIM Inputs:

There are 22 variables that will need to be created for each tax unit. The TAXSIM program requires all 22 input variables to have a value in order to run. “Blanks” (valid or invalid) and tax information that the CE survey does not capture (i.e. Unemployment Compensation, Short term capital gains or losses, Long term capital gains or losses) are set to zero.

For most cases, there are specific CE variables that match the TAXSIM variables. In other cases, adjustments must be made to the CE variables (i.e. the creation of temporary variables for the purpose of inputting into TAXSIM, but not to be stored by CE). Moreover, two groups of TAXSIM variables information will be gathered for both the previous tax year (denoted by variables with “_PY” at the end) and the current tax year (denoted by variables with “_CY” at the end) surrounding the reference period.

In addition to creating these variables for the TAXSIM program, the tax information is stored in the NTAXI PUMD files, along with the outputs.

Variable 1: Unique Tax ID

“TAXID” – this will identify the unique tax unit with a CU identifier (NEWID) and a tax unit identifier within the CU. The tax unit identifier within the CU (“TAX_UNIT”) will be a number starting with 1, for the primary tax unit and incrementing with each tax unit. TAXID will be a concatenation of the NEWID and TAX_UNIT

EX:

	NEWID	TAX_UNIT	TAXID
Primary Tax Unit	0111234	1	01112341
Secondary tax unit 1	0111234	2	01112342
Secondary tax unit 2	0111234	3	01112343

Variable 2: Tax Year “TAX_YR”

This is the year that taxes are calculated for. Tax code and levels for that year will be used in the program. TAXSIM will run for each tax year in the reference period (i.e. current and prior years), defined as the previous 12 months from the interview month and year.

Variable 3: SOI State “SOI_ST”

The state parameter that is used in TAXSIM is the state code that is used in the IRS Statistics of Income (SOI) data.

Variable 4: Filing Status (or “Marital Status” in TAXSIM documentation) “FILESTAT”

Filing status is determined for the tax unit by the members in that tax unit. The Filing Status code is the code that will be used in the TAXSIM program.

- Single (Filing Status = 1): Tax units that contain one member

- Married filing jointly (Filing Status = 2): Tax units with a head and spouse, may or may not have dependents
- Head of household (Filing Status = 4): Tax units with a head and dependents, but no spouse
- Dependent Tax Payer (Filing Status = 8): Tax unit that has a member who is a dependent of another tax unit. (Note: This tax unit is run through TAXSIM claiming no exemptions in Variable 5)

Variable 5: Dependent Exemptions “DEPCNT”

This is the count of dependents assigned to the tax unit. It summed across members in the same tax unit. Dependents are assigned during the tax unit creation and stay the same regardless of the year. Dependent tax payers will count toward their initially assigned tax unit as a dependent and will have a value of zero for DEPCNT.

Variable 6: Taxpayers 65 and over “T65CT_PY” & “T65CT_CY”

Count of taxpayers from (head of tax unit or spouse) that are 65 and older during the tax reference period. It is summed across members in the same tax unit, but the actual value can only equal 0, 1, or 2 since it only pertains to the head and spouse.

Variable 7: Wage and salary income of taxpayer “WAGE_HD”

Sum of the imputed salary (MEMI.SALARYXM) of the tax head, the positive imputed self-employment income (MEMI.SEMPFRMM > 0). Otherwise, value is zero.

Variable 8: Wage and salary income of spouse “WAGE_SP”

Sum of the imputed salary (MEMI.SALARYXM) of the tax head spouse (if present), the positive imputed self-employment income (MEMI.SEMPFRMM > 0). Otherwise, value is zero.

Variable 9: Dividend income “DIVINC”

Dividend income cannot be isolated in CE data, therefore all tax units will have zero for this variable.

Variable 10: Other income “OTHTXINC”

This is a catch all for other income in the tax unit. For the primary tax unit, OTHTXINC will include the family and specific member level variables listed below; for the secondary unit, it will only contain the member level variables that apply to those members listed in the particular secondary unit.

The value here will be the additions and subtractions of the below inputs.

Sum the following:

- Imputed interest and dividends paid (FMLI.INTRDVXM) – primary tax unit only, zero for others
- Other regular contributions, including alimony, child support, unemployment compensation and worker’s compensation (FMLI.OTHREGXM) – primary tax unit only, zero for others

- Negative imputed self-employment income (MEMI.SEMPFRMM <0) – for tax unit head and spouse (Note: The amount is stored as a negative, so technically it's adding the negative amount)
- Income from roomers and boarders and other rental income (FMLI.NETRENTM) – primary tax unit only, zero for others
- Lump sum financial income (FMLI.LUMPSUMX) – primary tax unit only, zero for others
- Other income (FMLI.OTHRINCM) – primary tax unit only, zero for others

Subtracting the following:

- Alimony paid (CNT.CONTEXPX where CNT.CONTCODE = “120”) – primary tax unit only, zero for others
- Payments made to private pensions (MEMI.ANPRVPM) subject to specific thresholds and income limit that are recalculated each year.
- Payments made to individual retirement account (MEMI.INDRETX) subject to specific thresholds and income limits recalculated each year and dependent on age and whether or not the individual made contributions to a retirement plan.

Variable 11: Taxable pensions “TAXPENS”

If there is more than one tax unit in the CU with retired members, then the imputed pension income at the CU level (FMLI.RETSURVM) is divided proportionally among the tax units that have retired members in the CU (Members where MEMI.INCNONWK=1). If there are no retired persons in the CU, then the amount should be divided proportionally amongst the tax units that have disabled member in the CU (Members where MEMI.INCNONWK=4). If there are no retired or disabled members, then the entire amount will be included in the Primary Tax Unit.

- Example: Consumer Unit has a Total Yearly Amount of \$3,000.00 and 3 total retirees. $\$3000.00 / 3 \text{ retirees in CU} = 1,000.00$ for each retiree. The amounts will be dispersed among the Tax Units; TU1 has 1 retiree, TU 2 has 2 retirees so TU1 taxable pension value is \$1,000, TU2 is \$2,000

Variable 12: Social Security benefits “SOSSECB”

Social security benefits at the member level (MEMI.SOCRRXM) - for tax unit head and spouse. It summed across those members per tax unit.

Variable 13: Other non-taxable transfer income such as welfare municipal bond interest, and child support (affects state estimated taxes only) “NONTXINC”

Sum the following:

- Imputed welfare (FMLI.WELFAREM) - primary tax unit only, zero for others

Variable 14: Rent paid (used for calculating state property tax rebates) “RNTPAID”

Annualized rent paid (RNT.QRT3MCMX*4) - primary tax unit only, zero for others

Variable 15: Property taxes paid “PROPTXPD”

Real estate tax for all owned properties (OPB.QADPTAX) - primary tax unit only, zero for others

Variable 16: Itemized deductions – Alternative Minimum Tax (AMT) preference “AMTDEDCT”

The deductions that can be used in AMT will go here. These will all be assigned only to the primary tax unit. All other tax units will have a zero.

Sum of the following:

- Totalled medical expenses (MDB.MEDPMTX) that are between 7.5% and 10% of income (income = NTAXI.WAGE_HD + NTAXI.WAGE_SP + NTAXI.DIVINC + NTAXI.OTHTXINC [everything except ‘refunds from state taxes paid’ {FMLI.SLRFUNDX}] + NTAXI.TAXPENS + NTAXI.SOSSECB)
- Personal property taxes for vehicles (VLR.VOPREGY = 405)
- Miscellaneous taxes (FMLI.MISCTAXX)
- Annualized, business adjusted home equity loan mortgage interest paid on first and second properties. Sum (HEL.QADINT1G, HEL.QADINT2G, HEL.QADINT3G)*4.
Annualized, business adjusted home equity line of credit interest on first or second home (OPH.JINTPDX*4)

Variable 17: Child/Dependent Care Expenses “CHLDCARE”

Annualized child or dependent care expense of the household (MIS.MISCEXPX*4 where MIS.MISCCODE= 225). Divide proportionally amongst tax units with dependent children less than thirteen years old.

Example: Consumer Unit has a Total Quarterly Expense of \$300.00 and 3 total kids. $\$300 * 4 = \$1,200.00$ / 3 kids in CU = 400.00 for each kid. This amount have to be dispersed among the Tax Units; TU1 has 1 kid, TU 2 has 2 kids so TU1 child expense value is \$400, TU2 is \$800

Variable 18: Unemployment Compensation (“UNEMPCOMP”)

- Enter zero for all tax units (not available in CE)

Variable 19: Number of dependents under age 17 (for child credit) “DEPUND17”

- Count of dependents in the tax unit whose ages are less than 17

Variable 20: Non-AMT deductions “OTHDEDCT”

These will be remaining deductions NOT included in Variable 16 (Itemized AMT deductions), but are available in CE data. These will all be assigned only to the primary tax unit. All other tax units will have a zero.

Sum the following:

- Annualized charitable giving to educational institutions, religious organizations, and charities (CNT.CONTEXPX*4 where CNT.CONTCODE IN ('130' '150' '160')). If amount is a continuous expense (CNT.CONTMO=13), assume same amount given each month (multiply it by 12), otherwise assume same amount given each quarter (multiply it by 4)
- Annualized, business adjusted mortgage interest paid on first and second properties, excluding interest on home equity loans and lines of credit. Sum (MOR.QADINT1X, MOR.QADINT2X, MOR.QADINT3X)*4
- Totaled medical expenses (MDB.MEDPMTX) that are greater than 10% of income (income = NTAXI.WAGE_HD + NTAXI.WAGE_SP + NTAXI.DIVINC + NTAXI.OTHTXINC [everything except 'refunds from state taxes paid' {FMLI.SLRFUNDX}] + NTAXI.TAXPENS + NTAXI.SOSSECB)

Variable 21: Short term capital gains or losses

- Enter zero for all tax units (not available in CE)

Variable 22: Long term capital gains or losses

- Enter zero for all tax units (not available in CE)