STaTa

Statistics/Data Analysis

---

help for **taxpuf**

NBER TAXSIM model for federal and state income taxes – Full

Description

> **taxsim**[,**full** **o**utput **s**econdary|**interest**|**long** **t**emp]

calculates federal and state income tax liability from a transformed version of the SOI public use file. Where the Stata procedure **taxsim** uses fewer than a score of input variables likely to be available in a survey, **taxpuf** uses all the data available in the public use files, about 200 values per taxpayer.

The TAXSIM version of the public use file is documented elsewhere, but includes variables named **data1** through **data210** for various income, deduction, and demographic characteristics. data100 is the taxpayer id variable, data11 is wages, etc.

One of the TAXSIM PUF files must be in the workspace before calling {hi:taxpuf}. The program returns your Stata workspace after creating a file taxsim_out.dta with values for the various liabilities and marginal tax rates. The two files can be merged for further analysis.

Here is an example of a complete job to calculate year 2000 tax liabilities with taxsim and compare them to taxpayer reported liabilities:

. **use /home/data/soi/taxsim/dta/s2000**
. **taxpuf**
. **merge data100 using taxsim_out**
. **replace fiitax = max(fiitax,0)**
. **reg fiitax data16**

This loads the year 2000 2% subset, calculates taxes, merges taxsim output with the original data, truncates tax liability at zero (to match SOI conventions), and regresses the TAXSIM calculated value of federal liability on the taxpayer reported value. You should see an r-squared value of better than .99.

The tax calculator itself is the same FORTRAN program that the NBER has been updating annually since 1974. This interface converts the data to ASCII and executes the tax calcultor against it, then reads the output of the tax calculator and converts it to a Stata dataset. The full tax calculator is available only while logged onto the NBER Unix cluster, not via the Internet.

Data

Input files are available for all years from 1960 through 2000, except for 1961, 1963, and 1965. Each year is available as a 2% subset (about 2,000 taxpayers) or a full version. File x1999.dta would be the full dataset, where s1999.dta is the subset. All files are kept in /home/data/soi/taxsim/dta.

Returned values

fiitax: Federal tax liability, after regular, minimum, and maximum tax, and after refundable credits.

siitax: State tax liability after refundable credits.

fica: FICA tax liability

frate: Marginal federal tax rate wrt a weighted average of the rates on the primary and secondary earners, or equal weights if both are non-workers. You can change this to the secondary earner, property income, or long term gains with the options, "Secondary" or "Interest" or "Long".

srate: Marginal state tax rate. Same options apply here.

ficar: Marginal FICA rate. The estimate of FICA tax includes both employer and employee portions, and ignores the favorable treatment of self-employment income.

Options

full: Return many intermediate calculations (taxable income, regular tax, individual credits, etc. All the returned variables have labels defined in taxpuf.ado – use "describe" to get the full list.

output: Specify the name of the output dataset. The default is taxsim_out.dta in the current directory.

secondary: Calculate marginal tax rates with respect to the secondary wage earner. The default is a weighted average of the primary and secondary wage earners.

interest: Calculate marginal tax rates with respect to interest income.

long: Calculate marginal tax rates with respect to long term gains.

temp: Save temporary files to disk.

Notes:

Dollar amounts are rounded to the nearest dollar before transmission to the calculator, and calculated amounts are similarly treated.

A general description of Taxsim is given in http://www.nber.org/taxsim/feenberg-coutts.pdf.

Daniel Feenberg (topical)   Jean Roth (technical)
feenberg@nber.org   jroth@nber.org
617-588-0343

Online: help for taxpuf