README
"Income Inequality in the United States: Using Tax Data to Measure Long-Term Trends"
by Gerald Auten and David Splinter

Overview
The SAS code for this paper is a single file. Output is entered into the spreadsheet to produce all figures and tables. The code uses individual tax microdata to estimate shares of fiscal and national income going to top income groups. Other input data is hardcoded. These data are from the spreadsheet using data from IRS SOI, NIPA, Piketty and Saez (2003), the Federal Reserve, the Survey of Consumer Finance, Census, and tax population data, as described in the online appendix.

Tax Data
The tax data file used are the "public use" versions of this data from 1960 to 1978 (excluding 1961, 1963, and 1965). These datasets are available from the IRS, see https://www.irs.gov/statistics/soi-tax-stats-individual-public-use-microdata-files. Confidential tax data files (called "INSOLE" files) are used for years since 1979. To access these confidential data, a researcher would generally need to apply to the IRS Statistics of Income Joint Statistical Research Program (www.irs.gov/uac/soi-tax-stats-joint-statistical-research-program). The public use versions of tax data since 1979 should give similar results as the confidential data, although they do not have unmasked TINs or state of residence and so cannot be matched to year of birth or remove non-residents, set state=1 and yob_dm1=&yr-40.

Running the Program
1) Name the folders at the top based on where your tax microdata files are located and your desired output folder.

2) Update the tax file names to your file names. They are currently puindiv&yr (1960-1978 public use files) and soi&yr (1979-2019 confidential files).

3) Uncomment the macro statements at the bottom of the code and run to estimate income shares with various adjustments.

4) Main output is TOTAL_INCxxx, where xxx is the income type. For all income definitions including main measures and intermediate steps, these are entered in the spreadsheet tab 'Output'. Breakouts of income sources are in TOTAL_ALLxxx. For the main income definitions, these are entered in the spreadsheet tab 'Output2'. Also, the output ZTOTAL_INC stacks results for multiple income types.