Supplemental information of Stata programs for *Estimating Confidence Intervals in a Tax Microsimulation Model*

This document contains supplemental information prerequisite to running the Stata programs. The information is divided into three sections: (1) Stata commands required to run the program, (2) codebook for the data sets used in the Stata programs, and (3) the sequence of Stata programs.

(1) Stata commands required to run the program

   The Stata commands bsweights and bs4rw are required.
   For bsweights: (a) In the Stata command box, type search bsweights, (b) Click on bsweights from http://staskolenikov.net/stata, (c) Click on click here to install
   For bs4rw: (a) In the Stata command box, type search bs4rw, (b) Click on bs4rw from http://www.stata.com/users/jpitblad, (c) Click on click here to install

(2) Codebook for the data sets used in the Stata programs
   There are three datasets required to run the set of Stata programs.
   - **bsw0.dta** contains four variables:
        o newseq (integer) is the unique observation ID.
        o ECI_lvlgroup (integer) is the income group.
        o WSAMP (integer) is the observation's stratification ID.
        o tabwt (real) is the observation's analytic weights.
   - **baseline_2019cl.dta** contains six variables:
        o newseq (integer) is the unique observation ID.
        o tabwt (real) is the observation's analytic weights.
        o depind (binary) identifies dependent tax filers.
        o agi_cl (real) is the observation's Adjusted Gross Income under the baseline.
        o eci_cl (real) is the observation's broader Income under the baseline.
        o cl_burden (real) is the observation's tax burden under the baseline.
   - **Txexint_2019pr.dta** contains two variables:
        o newseq (integer) is the unique observation ID.
        o alt_burden (real) is the observation's tax burden under the proposal (in this case Txexint_2019pr).

(3) The sequence of Stata programs
   The Stata programs must be run in a correct sequence to obtain necessary datasets.
        (a) **bsw_gen.do** will generate the set of replicated weights. Users need to specify **num_reps** -- the number of replicated weights they want to produce. **num_reps** is set to 201 currently. As a result, two datasets bsw21.dta (contains 21 replicated weights) and bsw201.dta (contains 201 replicated weights) are produced.
        (b) **main_reps##.do** where ## is the relevant **num_reps** that users want to use (either 21 or 201 in the provided programs). This program will call **bs_dist_cl_prog.do** which produces bootstrapped statistics for the baseline and **bs_dist_alt_prog.do** which produces bootstrapped statistics for the relevant proposals defined by the global PROP_CUR.