This file contains instructions on how to generate figures and tables of the article “Optimal Tax Progressivity: An Analytical Framework”, by Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante. For more details, please contact one of the authors.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Note for Figure 1.**

The starting dataset is *PSID\_sampleB\_DedSelf.dta*. This is the PSID dataset that contains the calculations of deductions explained in Appendix A.

The STATA file *Progressivity\_US\_PSID\_final.do* reads the dataset *PSID\_sampleB\_DedSelf.dta* and constructs the variables needed for the social-security correction of gross income and disposable income. Next, it estimates the progressivity parameter tau (the slope of the regression is 1-tau) and the constant lambda. Finally, it creates the bins used for Figure 1.

This code needs two inputs, *hdatamatlab.dta* and *wdatamatlab.dta*, computed by the MATLAB file which in turn, needs as input the Life Tables In the excel file *lifetables.xlsx.*

The code *plottau\_final.m* generates Figure 1a and Figure 1b. It reads the percentiles for gross and disposable income generated by the file *Progressivity\_US\_PSID\_final.do* and copied manually to the Excel file *tau\_US\_plotdata.xlsx*.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Note for the regressions in Section 8.**

The file *xcountryreg\_final.do* reads the various cross-country datasets described in Appendix C (*WID.dta, WTI.dta, PWT.dta, and SWIID.dta*), merges all data together and runs the regressions of Table 5 and Table C1. The code also needs the file *DATA\_SIGMA.dta*. This is the file that contains the variances of the LogNormal component of the Pareto-LogNormal distribution for the subset of countries included in the WID data. For each country/year, this variance is computed with a nonlinear equation solver from the closed-for expression of the Gini coefficient derived by Hajargasht-Griffiths (2013), given a value for the Pareto scale-parameter which we obtained from the WID data.