README file for Malamud and Kaestner (2014)

This file provides information about the data and code used for Malamud and Kaestner (2014) "Self-Selection and International Migration: New Evidence from Mexico" REStat Volume 96, Issue 1. Note that all of the dofiles and data should be placed in the same directory and all of the analysis was conducted with Stata version 13.1.

Data

The data for this study come primarly from rounds 1 and 2 of the Mexican Family Life Survey (MxFLS). However, we also use data from the 2000 U.S. and Mexican Censuses to calculate the relative returns to education and other characteristics.

Mexican Family Life Survey (MxFLS) 2002 (Round 1): http://www.ennvih-mxfls.org/english/index.html
Mexican Family Life Survey (MxFLS) 2005 (Round 2): http://www.ennvih-mxfls.org/english/index.html
U.S. Cenus 2000: https://usa.ipums.org/usa/
Mexico Cenus 2000: https://international.ipums.org/international/

All of these data sets are available to users after a simple registration process.

Setup Files

These dofiles are used to create some intermediate data files and the final data used to generate the tables in the paper:

Dofiles creating intermediate data files:
setup_proxy.do – creates mxfls02_proxy.dta which contains all the proxied responses ready for merging
setup_migration2002.do – creates mg2002.dta and mt2002.dta which reshapes the 2002 migration data
setup_migration2005.do – creates mg2005.dta and mt2005.dta which adds individuals who migrated and returned between round 1 and 2 of the MxFLS surveys (in addition to migrants in the migus.dta file)

Dofiles creating final data files:
setup_selection_fullsample.do – creates mxfls02_fullsample.dta used to generate Tables 1 and 2
setup_selection_imputed – creates mxfls02_imputed.dta used to generate Tables 3 and 4

The files above require the source files downloaded from the MxFLS website. Note, it is important to execute the dofiles which create the intermediate data files before executing the dofiles which create the final data files. Also, note that setup_migration2005 uses renamed versions of the original files "iiia_mg1.dta" and "iiia_mt1.dta" files as "iiia_mt1_2005.dta" and "iiia_mg1_2005.dta" respectively (in order to avoid confusion with similarly named files from the 2002 round).

Census Setup Files

These dofiles are used to input data from extracts created using IPUMS data and then calculate the differential returns between the US and Mexico:

Dofiles to input Census data
usa_input.do – uses an extract from the 2000 U.S. Census to generate us2000.dta

mexico_input.do – uses an extract from the 2000 Meixco Census to generate mexico2000.dta

Dofiles used to calculate differential returns
usa_regs.do – creates us_pwages.dta which contains calculated returns in the US
mexico_regs.do – creates mex_pwages_urban_state.dta which contains calculated returns in Mexico
setup_returns – creates returns_mexico_us.dta which contains differential returns between the US and Mexico

Output Files

These dofiles are used to generate the main tables in the paper using the final data files created by the setup dofiles.

generate_Tables1_2.do – creates Tables 1 and 2 as xml files which can be viewed in Excel
generate_Tables3.do – creates Table 3 as an xml file which can be viewed in Excel
generate_Tables4_panelA.do – creates Panel A of Table 4 as an xml file which can be viewed in Excel
generate_Tables4_panelB.do – creates Panel B of Table 4 as an xml file which can be viewed in Excel
generate_AppTables_weighted.do – creates the Appendix Table as an xml file which is viewable in Excel