Eduardo Souza-Rodrigues, *"Deforestation in the Amazon: A Unified Framework for Estimation and Policy Analysis"*

The primary data set is "main_data.dta". The sources of this data are described in detail in the supplemental document of the paper.

The do-file "Prepare_Dataset.do" organizes the "main_data" set, generates the summary statistics presented in tables 1 and 2, and creates all data sets that are used in the MATLAB routine.

The MATLAB files are:

- dem_def_main.m: executes all estimation and counterfactual calculations
- dem_def_reg.m : runs all land-use regressions: OLS, IV, QR, and IVQR estimators
- dem_def_compute_demand.m: computes demand for deforestation and save results
- dem_def_plot_predict_vs_tc.m: generates figure 3, predicted share of deforestation vs transportation costs
- dem_def_carbon_emissions.m: computes avoided emissions of CO2 and save results
- dem_def_first_stage_reg.m: estimates first-stage regression and generates table 3
- dem_def_reg_tables.m: generates tables with the estimated land-use regressions and marginal effects by farm sizes (tables 4 and 5)
- dem_def_lost_surplus.m: computes and shows farmers' lost surpluses from alternative policy interventions
- dem_def_plot_demand.m: plots the demand for deforestation curves – the total demand and demand by farm sizes (figure 4)
- dem_def_plot_demand_monitoring.m: plots demand for deforestation for different levels of monitoring: actual monitoring vs fines set at pre-2004 levels (figure 6)
- dem_def_plot_emissions.m: plots supply of avoided carbon emissions (figure 7)
- dem_def_median_reg_tables_app.m: generates land-use median IVQR regression tables (appendix)
- dem_def_emissions_monitoring.m: computes avoided carbon emissions when fines are set at pre-2004 levels (appendix)

- cluster_bootstrap.m: implements the (geographic) pair cluster bootstrap to calculate standard errors (table 7 in the appendix)

- dem_def_spqiv.m: estimates the semiparametric quantile IV model based on the penalized sieves minimum distance estimator proposed by Chen and Pouzo (2012)

- dem_def_median_spqiv_tables_app.m: generates table with the median SPQIV estimates (table 10 in appendix)

- neural_net.m: calculates the artificial neural network basis function used in the SPQIV estimation procedure

- npqiv_ann_log.m: computes the criterion function to estimate the SPQIV model using the artificial neural network basis function

- dem_def_reg_yields.m: estimates yields regressions and shows the results (tables 11 and 12 in the appendix)

- dem_def_sat_reg.m: compares census vs satellite deforestation measures (table 13 in the appendix)

- dem_def_plot_fixedprop_vs_vertical_demand.m: plots the demand for deforestation curves, comparing the 'fixed-proportions' and the 'vertical' models (figure 2 in the appendix)

- rq.m: quantile regression estimator, written by Daniel Morillo, Paul Eilers and Roger Koenker

- vcqr.m: computes standard errors for QR, written by Chernozhukov and Hansen

- inv_qr.m: IVQR estimation routine, written by Chernozhukov and Hansen

- vciqr_oid.m: standard errors for the IVQR estimator, written by Chernozhukov and Hansen

- ksrmv.m: multivariate kernel smoothing, written by Yi Cao

- splibas.m: computes polynomial spline basis, written by Paul L. Fackler and Mario J. Miranda

- splidop.m: creates differential operator matrices for polynomial splines, used in “splibas.m”, written by Paul L. Fackler and Mario J. Miranda

The software versions on which the computer codes are run are STATA15 and MATLAB2018a.

It takes approximately 5 minutes to generate all results presented in the paper using a computer with the following configuration:

System Type: x64-based PC

Processor: Intel(R) Xeon(R) CPU E5-1607 v2 @ 3.00GHz, 3000 Mhz, 4 Core(s), 4 Logical Processor(s)

Installed Physical Memory (RAM): 16.0 GB

It takes approximately 4.5 hours to compute the bootstrap standard errors presented in the supplemental document of the paper.