

# Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings

Raj Chetty, John Friedman, and Emmanuel Saez

This zip file contains the Stata programs and datasets used to generate the results reported in the paper.

[1] *analysis\_final.do* produces all the tables, figures, and numbers reported in the paper. The datasets it utilizes are described at the top of the file. These data are accessed under contract with the U.S. Internal Revenue Service. These datasets as well as the code run internally on IRS workstations cannot be posted publicly due to confidentiality restrictions.

[2] *sharpbunching.dta* contains our measure of sharp bunching. The list of variables is as follows:

tax_yr	Tax year
zip3	Three-digit ZIP code
count	Number of EITC claimants with children observed in year-zip3 cell
b_zip3	Degree of sharp bunching. Defined as percentage of EITC claimants with children who report total earnings within \$500 of the first EITC kink and have non-zero self-employment income.
count_selfemp_only	Number of EITC claimants with children and non-zero self-employment income observed in year-zip3 cell
b_zip3_selfemp_only	An alternative measure of the degree of sharp bunching, described in footnote 14, which only includes self-employed individuals in the denominator. Defined as percentage of EITC claimants with children and non-zero self-employment income who report total earnings within \$500 of the first EITC kink.

[3] *zip3\_state\_crosswalk.dta* is a crosswalk which matches three-digit ZIP codes to the state that contains them.

[4] *state\_eitc.dta* contains data describing state EITC top-up regimes, which supplement the federal EITC. Details are provided in footnote 17. The list of variables is as follows:

state	State
tax_yr	Tax year
state_topup	State top-up as a percentage of the federal EITC
refund	Indicates whether the state top-up is a refundable tax credit
eligible	Indicates whether workers without children are eligible for the top-up

[5] *app\_figure\_schedules.dta* contains the data used to create Appendix Figure 1, which plots the EITC Refund Schedule against Total Tax Liabilities for Single Filers with One Child. See the notes to that figure for further details. The list of variables is as follows:

incbin	Income bin, in \$100 intervals
eic_refund	EITC refund amount
ctc	Child Tax Credit (including Additional Child Tax Credit)
incometax	Federal income tax amount (excluding payroll taxes)
total_refund	EITC refund + Child Tax Credit – Federal income tax

[6] *maps/zip3\_coords.dta* and *maps/zip3\_database.dta* are the datasets we use with the Stata command *spmap* to create ZIP-3 maps of the United States (e.g. Figure 2). These datasets were generated from “TIGER” shapefiles distributed by the U.S. Census Bureau. See the code for Figure 2 in *analysis\_final.do* for details on how these datasets are used to generate a map figure.

[7] *binscatter.ado* and *discscatter.ado* are locally-written Stata commands used to draw binned scatter plots (e.g. Appendix Figure 2). *discscatter* is an older program, and deprecated in favor of *binscatter*. *disclog.ado* is a companion program, which *discscatter* requires in order to run.