Replication Package for "Police Discretion and Public Safety"

Felipe Goncalves, fgoncalves@econ.ucla.edu
Steven Mello, steve.mello@dartmouth.edu

Data availability statement

All data required to replicate our paper is publicly available and is provided in this replication package.

Instructions for replication

1. Make the following edits to the STATA do-file _replicate.do:
   a. Edit line 6 so that the global macro **`proj`** corresponds to the location of the replication package on your system.
   b. If necessary, edit line 10 so that the global macro **`rlib`** corresponds to the directory from which R packages should be loaded from on your system. If using a PC or Mac, you should be able to leave this blank. However, on some Unix/Linux server systems, you will need to provide a directory here.
   c. If necessary, uncomment lines 18-19 to run the Stata and R setup files which install the necessary packages (see below for additional guidance on packages).
2. Execute the STATA do-file _replicate.do

A detailed walk-through of each subprogram that is run and what it does is provided below.

Computing

This replication kit uses both Stata and R and was most recently executed in batch mode on a Red Hat Enterprise Linux 8 server (128-CPU, 64-bit, 1.5TM memory) using Stata 18 MP-Parallel edition and R version 4.4.2 (2024-10-31) – "Pile of Leaves" (Platform: x86_64-redhat-linux-gnu).

The provided Stata code relies on the following external packages, which can be installed by uncommenting line 18 of _replicate.do:

- egenmore (https://ideas.repec.org/c/boc/bocode/s386401.html)
- reclink2 (https://journals.sagepub.com/doi/pdf/10.1177/1536867X1501500304)
- strdist (https://ideas.repec.org/c/boc/bocode/s457547.html)
- reghdfe (https://github.com/sergiocorreia/reghdfe)
- ranktest (https://ideas.repec.org/c/boc/bocode/s456865.html)
- ivreg2 (https://ideas.repec.org/c/boc/bocode/s425401.html)
- ivreghdfe (https://github.com/sergiocorreia/ivreghdfe)
- binsreg (https://nppackages.github.io/binsreg/)
- rscript (https://github.com/reifjulian/rscript)
- estout (http://repec.org/bocode/e/estout/estout.html)

The `rscript` package enables one to call for the execution of an R program (.R source file) from Stata. In this replication kit, all R programs are run from Stata using the `rscript` command. Here are a few things to note about the usage of this package:

- By default, this package will find and use the Rscript executable on your system. If this is not working on your computer, you can set the path directly using a global macro **`RSCIPT_PATH.`**
- The `rscript` command does not work in batch mode on Windows. Hence, if you are trying to execute the full replication procedure on Windows, you will need to run Stata in interactive mode.

See https://julianreif.com/rscript/ for additional information about this package if necessary.

The R code relies on the following external packages, which can be installed by uncommenting line 19 of _replicate.do:

- rio (https://cran.r-project.org/web/packages/rio/index.html)
- dplyr and tidyverse (https://dplyr.tidyverse.org/; https://www.tidyverse.org/)
- fixest (https://cran.r-project.org/web/packages/fixest/index.html)
- broom (https://cran.r-project.org/web/packages/broom/index.html)

Note that the provided R programs load additional packages during the preamble if the global macro **`rlib`** is defined. These additional packages are "dependencies" for the above packages and are installed and loaded by default when installing the about four packages. On a typical Mac or PC configuration, these dependency packages will automatically be loaded when loading the main packages. However, the Linux server where we ran the code has a small bug which requires the loading of certain dependencies prior to loading the main packages This is the reason for the manual loading of the dependency packages if an **`rlib`** value is passed to R.

You should be conscious of both disk space and computing time/requirements when using this replication package. The raw data directory is 41GB. The data construction routine (see below for details) stores several very large datasets in a temporary directory, which briefly exceeds 50GB during the replication process. Hence, prior to downloading and executing this replication kit, you should ensure that you have ~100GB of available disk space on your computer.

Total runtime for most recent execution: ~23 hours

Raw data

The following raw datasets are provided in the folder data/raw:

subdirectory	files	description
census	Names_2010Census.csv	File from U.S. Census Bureau with race/ethnicity distribution associated with surnames in 2010 census: https://www2.census.gov/topics/genealogy/2010surnames/names.zip
crash	See description →	This directory stores raw auto accident data provided by the Florida Department of Transportation. We have preserved the directory structure and file names as these data were originally provided to us. There is a subdirectory for each year (2006-2018). For 2015 and 2016, there are additional subdirectories for quarters. Each "terminal" directory (years for 2006-2014, 2017-2018; years and quarters for 2015-2016), includes the following data files: • event-level file (with "event" appearing somewhere in the file name). • driver-level file (with "driver" appearing somewhere in the file name). • passenger-level file (with "passenger" appearing somewhere in the file name). • pedestrian-level file (with "pedestrian" appearing somewhere in the file name for 2006-2010 and "non_motorist" appearing somewhere in the file name for 2011-2018.
FDLE	agencies.xlsx	List of police agencies and their identifier codes, provided by the Florida Department of Law Enforcement (FDLE)
	employment.xlsx	Dataset of employment histories for Florida police officers provided by FDLE
	person_data.xlsx	Officer personnel information provided by FDLE
	person_data2.xlsx	Updated officer personnel information provided by FDLE
FDOC	INMATE_RELEASE_ROOT_Jan_2021.txt	Dataset of prison releases as of January, 2021 from Florida Department of Corrections (FDOC) OBIS database: https://www.fdc.myflorida.com/statistics-and-publications/public-records-requests-for-the-obis-database
geo	CENSUS_COUNTY_2010.csv	List of county names and 2010 FIPS codes : https://www.census.gov/library/reference/code-lists/ansi.html
	USPS_ZIP_COUNTY_2010.xlsx	Crosswalk from zip codes to 2010 county FIPS codes: https://www.huduser.gov/apps/public/uspscrosswalk/login
	UTC_CountyNum.xlsx	List of county names and their codes in the UTC data, constructed from UTC appendix B
	CountyTroopDistrict.dta	Constructed file which crosswalks UTC county codes to FHP troops and districts. Based on: https://www.flhsmv.gov/florida-highway-patrol/troop-boundaries-information/
geocode	utc_togeo_05_18_geo_MT.csv*	Constructed file with geocoded road segments for the subset of citations with GPS coordinates
IRS	zipcode05.csv zipcode06.csv	Zip-level tax filing statistics from the IRS statistics of income (SOI) files:

	zipcode07.csv 08zpall.csv 09zpallagi.csv 10zpallagi.csv 11zpallagi.csv 12zpallagi.csv 13zpallagi.csv 14zpallagi.csv	https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi
reference	basemap_route_road	Shapefiles of Florida roads used for geocoding: https://www.fdot.gov/statistics/gis/default.shtm
	MT_geocode.mp4	Video demonstrating geocoding method
	utc_togeo_05_18.do	Stata do-file to produce the data file that was geocoded (i.e., the input to the geocoding depicted in video above)
	utc_appendix_c.pdf	PDF version of UTC Appendix C: https://www.flhsmv.gov/pdf/courts/utc/appendix_c.pdf
TCATS	goncalves_{year}.txt	Yearly extracts from the TCATS (Traffic Citation Accounting Transmission System) database (traffic court dispositions) provided by the Florida Clerks and Comptrollers. One file for each {year}, 2005-2017
UTC	citation_{year}.txt	Yearly extracts from the UTC (Uniform Traffic Citation) database provided by the Florida Clerks and Comptrollers. One file for each {year}, 2005-2018
vehicle	make-list.xls	Crosswalk from vehicle make abbreviations in the UTC and full vehicle make names, provided by the Florida Clerks and Comptrollers
	true_car_listings.csv	Used car listings as of 09/24/2017 posted on TrueCar.com: https://www.kaggle.com/jpayne/852k-used-car-listings/data. These data are licensed under a CC0 Public Doman license.
violation	utc_appendix_c_CLEAN_FLAG.xlsx	Constructed file with descriptions for violation codes, detailed information on sanctions, and offense classifications. Constructed from UTC appendix C.
	viol_codes.xlsx	Constructed file displaying violation code shares and characterizations in the UTC data

*In “data/raw/reference”, we provide the Florida roads shapefile used for geocoding (basemap_route_road), the do-file that produces the file that we geocoded (utc_togeo_05_18.do), and a video illustrating how the geocoding was done (MT_geocode.mp4). Note that the geocoded data is only used to produce figure B-3 in the paper.

Detailed walk-through of replication process

Directory structure

Lines 23-34 of _replicate.do set up the following directory structure (nested within the directory of the replication kit, set via the global macro **`proj`**):

directory	subdirectory	description
code	data_construct	Stores subprograms for data construction routine
	iv_checks	Stores subprograms for instrument validity analyses
	iv_analysis	Stores subprograms for main 2SLS/IV analyses
	extrap_main	Stores subprograms for main extrapolation-based analyses
	extrap_hetero	Stores subprograms for extrapolation-based heterogeneity analyses
data	raw	Stores raw data files
	out (created by _replicate.do)	Stores constructed datasets
estimates (created by _replicate.do)		Stores output from estimation routines
logs (created by _replicate.do)		Stores log files
output (created by _replicate.do)	main (created by _replicate.do)	Stores output appearing in main text
	apx_iv (created by _replicate.do)	Stores output appearing in appendix A
	apx_deter (created by _replicate.do)	Stores output appearing in appendix B
	apx_extrap (created by _replicate.do)	Stores output appearing in appendix C
_temp (created by _replicate.do)		Stores temporary files

At the end of the process, lines 46-49 of _replicate.do delete remaining files in the _temp directory and then line 50 deletes the _temp directory itself using the `rmdir` command. On some systems, this line may not work if you have spaces in the path to the replication directory (set via the global macro **`proj`**).

It is convenient to think of the replication process proceeding in 5 steps, with each step corresponding to a step-level master program called by _replicate.do (e.g., code/01-data_construct.do), an associated directory of subprograms which are run by this master program (e.g., code/data_construct), and the production of an associated log file in the logs directory.

Step 1: Data construction

The first step is to construct analysis datasets from the raw data. This task is performed by **code/01-data_construct.do**, which calls the following subprograms in code/data_construct:

program	description	output
01-process_citation.do	Read and append raw citations data	

02-process_court.do	Read and append raw courts data	
03-process_crash.do	Read and append raw crash data	
04-process_officer.do	Match FHP citations to issuing FHP officers	
05-process_driver.do	Construct motorist information	
06-process_vehicle.do	Construct vehicle information	
07-process_gps.do	Merge geocoding information for citations with GPS	
08-process_violation.do	Construct violation information	
09-build_offending.do	Build master "offending" dataset with all citations and crashes for all motorists	data/out/1-offending.dta
10-build_stopsfhp.do	Build master dataset of all FHP stops for FHP officers included in analysis sample	data/out/2-stopsfhp.dta
11-build_officer.do	Build	data/out/3-officer.dta
12-build_main.do	Build main analysis dataset	data/out/4-main.dta

This process produces the four primary datasets used in our analyses (listed as outputs in the table above), stored in the directory data/out.

Total runtime for data construction: ~3 hours

Step 2: Instrument validity analyses

The next step is to check the validity of the officer instrument. This task is performed by **code/02-iv_checks.do**, which calls the following subprograms stored in code/iv_checks:

program	**description**	**output**
01-getfe.R	Estimate "bunching" fixed effects for officers, calculate R2 decomposition	estimates/getfe.dta output/apx_iv/rsquared.txt
02-propensity.do	Plot distribution of officer propensities, within-officer propensity correlations, check relationship between propensity and officer characteristics, show common support of instrument (figure A-1, figure A-2, table A-1, figure C-1)	output/apx_iv/propensity_{}.pdf output/apx_iv/corrs_{}.pdf output/apx_iv/table_officers.tex ouput/apx_extrap/support.pdf
03-balance.do	Test instrument balance (figure 2a, figure 2b, table A-2)	output/main/validity_{}.pdf output/apx_iv/table_balance.tex
04-first_stage.do	Estimate first stage (figure 2c, figure A-3)	output/apx_iv/fs_{}.pdf output/main/validity_fs.pdf
05-fll_test.R	Implement exclusion and monotonicity test from Frandsen et al. (2023)	estimates/fll_full.dta estimates/fll_sub.dta estimates/fll_judges.dta
06-monotonicity.do	Construct figure showing output of FLL test, estimate first stage by driver subgroups (table A-3, figure B-5)	output/apx_iv/fll_test.pdf output/apx_iv/table_mono.tex
07-speeds.do	Additional randomization checks for distribution of stopped speeds (figure A-5)	output/apx_iv/speeds_predicted.pdf output/apx_iv/speeds_slow.pdf
08-bunching.do	Create bunching plots (figure 1, figure A-4)	output/main/bunch.pdf output/apx_iv/bunch_radar.pdf

09-summary.do	Create summary statistics table (table 1)	output/main/table_summary.tex

This process produces figure 1, table 1, figure 2, figure C-1, and all of appendix A.

Total runtime for instrument validity analysis: ~20 minutes

Step 3: Instrumental variables analyses

The next step is to conduct the main IV analyses. This task is performed by **code/03-iv_analysis.do**, which calls the following subprograms stored in code/iv_analysis:

program	**description**	**output**
01-naive.do	Estimate naïve OLS regressions (table B-1)	output/apx_deter/table_naive.pdf
02-reduced_form.do	Plot reduced form estimates (figure 3, figure B-1, figure B-2)	output/main/reduced_form.pdf output/apx_deter/rf_{}.pdf
03-iv_tables.do	Produce tables with first stage, reduced form, and IV estimates (table 2, table B-2)	output/main/table_iv.tex output/apx_deter/table_fsrf.tex
04-selection.do	Sample selection robustness analysis (figure B-3)	output/apx_deter/selection_{}.pdf
05-alt_instruments.do	Estimates using alternative versions of stringency instrument (table B-3)	output/apx_deter/table_instruments.tex
06a-mono_data.do	Create dataset for estimation of IV estimates using group instrument	_temp/mono_data.dta
06b-mono_output.do	Estimates using group instrument (table B-4)	output/apx_deter/table_grouped.tex
07-heterogeneity.do	IV estimate heterogeneity (figure B-4)	output/apx_deter/heterogeneity.pdf
08-learning.do	Tests for driver learning (figure B-6)	output/apx_deter/learn_exposure.pdf output/apx_deter/learn_location.pdf
09a-mte_bootstrap.do	Estimate MTE's with bootstrapped standard errors	estimates/mte_bootstrap.dta
09b-mte_output.do	Produce MTE output (figure C-7, table C-6)	output/apx_extrap/table_mte.tex output/apx_extrap/mte_figure.pdf

This process produces table 2, figure 3, table C-6, figure C-7, and all of appendix B.

Total runtime for instrumental variables analyses: ~2 hours, 15 minutes

Step 4: Main extrapolation analyses

The next step is to conduct the main extrapolation-based analyses. Except for estimation using the Cattaneo et al. (2024) optimal bin width (table C-5), all extrapolation estimation is conducted in R. Parameter estimates and bootstrapped standard errors are stored as .csv or .dta files in the estimates directory. Output is always constructed using Stata programs which read in these estimate files. This step is performed by the program **code/04-extrap_base.do**, which calls the following subprograms in code/extrap_base:

program	**description**	**output**
01-data_extrap.do	Construct datasets for extrapolation estimates	_temp/data_extrap.dta _temp/speeds_data.dta

02-outcomes_q2.R	Estimate quadratic extrapolations for all outcomes	estimates/outcomes_q2.csv
03-outcomes_b5.R	Estimate tail mean extrapolations for all outcomes	estimates/outcomes_b5.csv
04-recid_poly.R	Estimate polynomial extrapolations for varying polynomial order	estimates/recid_poly.csv
05-recid_bins.R	Estimate tail mean extrapolations for varying bin widths	estimates/recid_bins.csv
06-recid_pscore.R	Estimate extrapolations using propensity score instead of officer stringency instrument	estimates/recid_pscore.csv
07-recid_troop.R	Estimate extrapolations within troops	estimates/recid_troop.csv
08-recid_county.R	Estimate extrapolations within counties	estimates/recid_county.csv
09-recid_spc.R	Estimate within-locations extrapolation based on Feigenberg & Miller (2022) approach	estimates/recid_spe.csv
11-recid_cov.R	Estimate within-covariate extrapolations	estimates/recid_cov.csv
11-recid_addcov.R	Estimate extrapolations with covariate controls	estimates/recid_addcov.csv
12-speeds_q2.R	Estimate quadratic extrapolations with reoffending at different speeds as outcome	estimates/speeds_q2.csv
13-speeds_b5.R	Estimate tail mean extrapolations with reoffending at different speeds as outcome	estimates/speeds_b5.csv
14-adjust_q2.R	Estimate quadratic extrapolations which extrapolate away from Z=0	estimates/adjust_q2.csv
15-adjust_q2_drop.R	Estimate quadratic extrapolations which extrapolate away from Z=0 (subsample)	estimates/adjust_q2_drop.csv
16-optimal_bins.do	Estimate tail mean extrapolation using optimal bin width from Cattaneo et al. (2024) – varying bin width across bootstraps	estimates/optimal_bins.dta
17-optimal_fixed.do	Estimate tail mean extrapolation using optimal bin width from Cattaneo et al. (2024) – fixed bin width across bootstraps	estimates/optimal_fixed.dta
18-output_outcomes.do	Produce output: extrapolation results for different outcomes (table 4, table C-1, and figure C-4)	output/main/table_outcomes.tex output/apx_extrap/table_outcomes_bins.tex output/apx_extrap/outcomes_{}.pdf
19-output_main.do	Produce output: main extrapolation results (figure 4 and table 3)	output/main/baseline_extrap.pdf output/main/baseline_est.pdf output/main/table_robustness.tex
20-output_fform.do	Produce output: sensitivity to functional form assumptions (figure C-2)	output/apx_extrap/fform.pdf
21-output_groups.do	Produce output: within-group extrapolation estimates (figure C-3)	output/apx_extrap/groups_cov.pdf output/apx_extrap/groups_loc.pdf output/apx_extrap/groups_est.pdf
22-output_optimal.do	Produce output: optimal bin choice (table C-5)	output/apx_extrap/table_bootstrap.tex
23-output_speeds.do	Produce output: reoffending speeds as outcome of interest (figure C-5)	output/apx_extrap/speed30_extrap.pdf output/apx_extrap/speed30_est.pdf output/apx_extrap/speeds_gains.pdf output/apx_extrap/speed_levels.pdf
24-output_adjust.do	Produce output: extrapolations away from Z=0 (table C-4)	output/apx_extrap/table_adjust.pdf

This process conducts the underlying estimation and then produces figure 4, table 3, table 4, figure C-2, figure C-3, figure C-4, table C-1, figure C-5, table C-4, and table C-5. Note that his step is very time-

consuming because each estimation program (numbers 02-17) performs a Bayesian bootstrap to compute standard errors.

Total runtime for main extrapolation analyses: ~12 hours

Step 5: Extrapolation-based heterogeneity analyses

The final step is to conduct extrapolation-based heterogeneity analyses. This process functions similar to the one above, where estimation is conducted using R, estimation output is stored in the estimates directory, and then output is constructed using Stata. This step is performed by the program **code/05-extrap_hetero.do**, which calls the following subprograms in code/extrap_hetero:

program	**description**	**output**
01-history_q2.R	Estimate quadratic extrapolations by motorist offending history	estimates/history_q2.csv
02-history_b5.R	Estimate tail mean extrapolations by motorist offending history	estimates/history_b5.csv
03-exper_data.dta	Create dataset with officer experience measure	_temp/data_exper.dta
04-exper_q2.R	Estimate quadratic extrapolations by officer experience	estimates/exper_q2.csv
05-exper_b5.R	Estimate tail mean extrapolations by officer experience	estimates/exper_b5.csv
06-exper_binscatter.do	Estimate binscatters by officer experience	_temp/exper_binscatter.dta
07-covar_q2.R	Estimate quadratic extrapolation by motorist subgroups	estimates/recid_troop.csv
08-output_history.do	Produce output: heterogeneity by motorist offending history	estimates/recid_county.csv
09-output_exper.do	Produce output: heterogeneity by officer experience	estimates/recid_spe.csv
10-output_subgroups.do	Produce output: tables by subgroup	estimates/recid_cov.csv
11-output_covar.do	Produce output: relationship between harshness, levels, and gains by motorist subgroup	estimates/recid_addcov.csv

This process conducts the underlying estimation and then produces figure 5, figure 6, figure C-6, table C-2, and table C-3.

Total runtime for extrapolation heterogeneity analyses: ~5 hours

Utility programs

There are four additional programs provided in this replication kit beyond those covered above, which are stored in the "code/utility" directory:

program	**description**
setup_stata.do	Installs required STATA packages. This is run on line 18 of _replicate.do (if you uncomment this line, as explained in the instructions)

setup_r.do	Installs required R packages. This is run on line 19 of _replicate.do (if you uncomment this line, as explained in the instructions)
ebayes.do	Utility program to compute the Morris (1983) empirical Bayes shrinkage estimator, based on the implementation in Chandra et al. (2016). Downloaded from: https://sacarny.com/programs/ This program is used in /code/iv_checks/02-propensity.do
create_levels_gains.do	Utility program that sets up data for the "levels and gains" plots (e.g., panel b of figure 4). This code is called by all programs which produce this type of plot: code/extrap_base/output_main.do, code/extrap_base/output_groups.do, code/extrap_base/output_speeds.do, code/extrap_hetero/output_history.do, code/extrap_hetero/output_exper.do, code/extrap_hetero/output_covar.do

Additional notes for replicators

Although a randomization seed is set in all programs, we have noticed that a small subset of our programs produce slightly different results across different systems (e.g., different operating systems, different versions of R, different versions of the binsreg Stata package). Here is where these differences may occur:

- Standard errors in panels (b) and (c) of table C-2.
- Standard errors in panels (b) and (c) of table C-3.
- Bin locations chosen by `binsreg` in figure C-3, panels (a) and (b).
- Bin locations chosen by `binsreg` in figure C-4, all panels.

References

Cattaneo, M., Crump, R., Farrell, M. and Feng, Y. (2024). On Binscatter. *American Economic Review* 114(5): 1488-1514.

Chandra, A., Finkelstein, A., Sacarny, A., and Syverson, C. (2016). Healthcare Exceptionalism? Performance and Allocation in the U.S. Healthcare Sector. *American Economic Review* 106(8): 2110-44

Feigenberg, B. and Miller, C. (2022). Would Eliminating Racial Disparities in Motor Vehicle Searches have Efficiency Costs? *Quarterly Journal of Economics* 137(1): 49-113.

Florida Court Clerks and Comptrollers (2019). Uniform Traffic Citation (UTC) database [dataset]. See https://www.flclerks.com/ for contact information.

Florida Court Clerks and Comptrollers (2019). Traffic Citation Accounting Transmission System (TCATS) database [dataset]. See https://www.flclerks.com/ for contact information.

Florida Department of Highway Safety and Motor Vehicles (2019). Uniform Traffic Citation Appendix C [dataset]. Available at: https://www.flhsmv.gov/pdf/courts/utc/appendix_c.pdf.

Florida Department of Highway Safety and Motor Vehicles (2019). Uniform Traffic Citation Appendix B – DHSMV county codes [dataset]. See https://www.flhsmv.gov/ for contact information.

Florida Department of Law Enforcement (2018). Employment database for Florida law enforcement officers [dataset]. See https://www.fdle.state.fl.us/ for contact information.

Florida Department of Law Enforcement (2018). Personnel database for Florida law enforcement officers [dataset]. See https://www.fdle.state.fl.us/ for contact information.

Florida Department of Law Enforcement (2018). Directory of Florida law enforcement agencies [dataset]. See https://www.fdle.state.fl.us/ for contact information.

Florida Department of Transportation (2019). Florida Traffic Crash Records Database [dataset]. See https://www.fdot.gov/legal/publicrecords/prrhomepage.shtm for contact information.

Florida Department of Transportation (2019). Basemap Routes with Measures [dataset]. Available at: https://www.fdot.gov/statistics/gis/default.shtm#Designated

Frandsen, B., Lefgren, L., and Leslie, E. (2023). Judging Judge Fixed Effects. *American Economic Review* 113(1): 257-277.

Morris. C. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association* 78(381): 47-55.

Payne, E. (2017). 852k Used Car Listings. Available at: https://www.kaggle.com/datasets/jpayne/852k-used-car-listings/data?select=true_car_listings.csv

U.S. Census Bureau (2012). Frequently occurring surnames from the 2010 Census [dataset]. Available at: https://www2.census.gov/topics/genealogy/2010surnames/names.zip

U.S. Census Bureau (2012). ANSI and FIPS codes. Available at: https://www.census.gov/library/reference/code-lists/ansi.html.

U.S. Department of Housing and Urban Development (2012). HUD USPS Zip Code Crosswalk Files [dataset]. Available at: https://www.huduser.gov/apps/public/uspscrosswalk/login

U.S. Internal Revenue Service (2018). Individual Income Tax ZIP Code Data, yearly extracts 2005-2014 [dataset]. Available at: https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi