Replication Instructions for "Would Eliminating Racial Disparities in Motor Vehicle Searches Have Efficiency Costs?"

Benjamin Feigenberg (bfeigenb@uic.edu)
Conrad Miller (ccmiller@berkeley.edu)

Overview

This README file is divided into two parts. In the first part of the README file, we describe the structure and contents of the replication folder and we provide instructions for replicating the paper's analyses. We also identify the subset of datasets that include personally identifiable information (PII) and cannot be publicly posted. The second part of the README file provides instructions for accessing these datasets with specific references to the programs that call in each confidential data file.

Part 1: Replication Folder Contents

This folder contains the Stata do-files (in the **code** sub-folder) that generate the tables and figures from the paper. In the **data** sub-folder, we include one folder (**data/raw**) that contains data files that can be publicly posted. We also include a folder (**data/raw_notincluded**) that serves as a placeholder for those data files that cannot be publicly posted. The **data/intermediate** folder contains intermediate data files produced by the included programs. Tables and figures are stored in the **output** sub-folder. The **logs** sub-folder stores the log files produced by the do-files.

Hardware Requirements

This code was last run on a 6-Core Intel-based laptop with Mac OS 10.15.7.

Software Requirements

All included analyses were conducted in Stata 16.

Data

List of datasets included (**data/raw**):

File	Description
`Names_2010Census.dta`	File from U.S. Census Bureau with racial and ethnic distributions associated with surnames appearing in the 2010 Census
`income.dta`	Constructed file that maps each stop to the median household income for the Census Block Group corresponding to the motorist's address
`zip_demographics_2011.dta`	Median household income by zip code from the 2011 American Community Survey

stops_all_loc.dta	Constructed file that maps each geocoded stop to a sergeant patrol area (mapping constructed in ArcGIS)
intersections.dta	Constructed file of intersections between sergeant area borders and highways (constructed in ArcGIS). Used in RD exercise.
driving_distances.dta	Constructed file of driving distances between stop locations and nearby intersections (constructed in ArcGIS). Used in RD exercise.
trooper_demographics.dta	File with trooper demographic characteristics
countypres_2000-2016.csv	File with county-level data on vote counts by presidential candidate in the 2016 presidential election
maps/THP_SGTArea_outlines_20200422	Sergeant area boundaries shapefile
maps/TxDOT_Roadways	Texas roadways shapefile
shp2longlat.dta	Constructed manually in ArcGIS to harmonize coordinate systems across shapefiles

List of datasets not included (**data/raw_notincluded**):

File	**Description**
/TX_original/csv/	Folder with raw stops data obtained from Stanford Open Policing Project
final_ids.dta	This file maps stops to constructed ID for each motorist. Constructed using proprietary Infutor data on address histories and restricted access criminal history data.
arrest_hist.dta	Criminal history for each motorist constructed using restricted access criminal history data.
arrests.dta	Maps stops to arrests (where applicable) using restricted access criminal history data.

Programs

The file master.do is the master file that sets the directory structure, describes the function and output of each component do-file, and calls in each do-file sequentially. Note that the do-files called in by master.do are located in the **code** sub-folder. These do-files cannot be run without first obtaining access to the subset of data files with personally identifiable information that are listed above and described in more detail in Part 2 of this README file.

A set of globals are declared at the start of master.do. You will need to change the global **basedir** to correspond to the location of your replication folder. In addition, after acquiring the subset of data files that are not posted in the repository, the corresponding global **sensdir** will need to be changed. Below, we list the program that generates each table/figure as well as the names of the relevant output files produced. We also list those additional programs referenced in master.do but not included in "List of Tables and Figures."

List of Tables and Figures

Table/Figure #	**Program**	**Output File**
Table 1	3-desc_analysis.do	desc_stats.tex
Table 2	6-spc_pool.do	selection_models.tex
Table 3	10-wm_design.do	Produced inline in log file
Table 4	12-bt_counterfactuals.do	Produced inline in log file
Table 5	13-wt_design.do	Produced inline in log file
Table 6	14-wt_counterfactuals.do	Produced inline in log file
Figure 3	9-spc_figures.do	spc_pool.pdf
Figure 4	9-spc_figures.do	spc_black.pdf spc_hisp.pdf spc_white.pdf
Figure 5	10-wm_design.do	wd_psearch_pool.pdf wd_search_pool.pdf wd_rf_pool.pdf
Figure 6	11-rd_design.do	rd_trp_asgn.pdf
Figure 7	11-rd_design.do	rd_search_trp.pdf rd_psearch.pdf rd_search.pdf rd_contra.pdf
App. Table A1	2-create_working.do	Produced inline in log file
App. Table B1	3-desc_analysis.do	search_results.tex
App. Table B2	3-desc_analysis.do	search_contra_det_log.tex
App. Table B3	6-spc_pool.do	desc_stats_spc_pool.tex
App. Table B4	8-spc_race.do	desc_stats_spc_race.tex
App. Table B5	6-spc_pool.do	selection_models_contra.tex
App. Table B6	9-spc_figures.do	Produced inline in log file
App. Table B7	10-wm_design.do	desc_stats_wm_pool.tex
App. Table B8	10-wm_design.do	desc_stats_wm_race.tex
App. Table B9	11-rd_design.do	desc_stats_rd.tex
App. Table B10	12-bt_counterfactuals.do	desc_stats_btcounterfactual.tex
App. Table B11	17-hypo_spc.do	search_det_byq.tex
App. Table B12	15-spc_trooper_cat.do	Produced inline in log file
App. Table B13	15-spc_trooper_cat.do	trooper_chars.tex
App. Table B14	13-wt_design.do	Produced inline in log file
App. Table B15	13-wt_design.do	Produced inline in log file
App. Table B16	14-wt_counterfactuals.do	Produced inline in log file
App. Table B17	14-wt_counterfactuals.do	Produced inline in log file
App. Table B18	16-gaps_by_loc.do	Produced inline in log file
App. Table B19	16-gaps_by_loc.do	by_trp_race.tex
App. Table B20	16-gaps_by_loc.do	Produced inline in log file
App. Figure B1	18-maps.do	spc_sample_pooled.png spc_sample_byrace.png
App. Figure B2	9-spc_figures.do	search_dist.pdf
App. Figure B3	9-spc_figures.do	search_comps_pooled.pdf

App. Figure B4	6-spc_pool.do	psearch_dist.pdf
App. Figure B5	6-spc_pool.do	psearch_offloc_dist.pdf
App. Figure B6	6-spc_pool.do	psearch_select.pdf psearch_select_stop.pdf psearch_select_speed.pdf
App. Figure B7	9-spc_figures.do	spc_slope_robust.pdf
App. Figure B8	9-spc_figures.do	spc_slope_robust_race.pdf
App. Figure B9	9-spc_figures.do	spc_pool_fe.pdf
App. Figure B10	9-spc_figures.do	spc_pool_night.pdf
App. Figure B11	9-spc_figures.do	spc_pool_arrest_l.pdf spc_black_arrest_l.pdf spc_hisp_arrest_l.pdf spc_white_arrest_l.pdf
App. Figure B12	9-spc_figures.do	spc_pool_sumlength_l.pdf spc_black_sumlength_l.pdf spc_hisp_sumlength_l.pdf spc_white_sumlength_l.pdf
App. Figure B13	18-maps.do	rd_sample.png
App. Figure B14	12-bt_counterfactuals.do	bt_search_split.pdf
App. Figure B15	13-wt_design.do	wt_search_pool.pdf wt_contra_pool.pdf
App. Figure B16	9-spc_figures.do	hit_rates_split.pdf spc_pool_scatter.pdf
App. Figure B17	15-spc_trooper_cat.do	spc_pool_byexp.pdf spc_pool_byrace.pdf spc_pool_bystopr.pdf
App. Figure B18	19-strong_safety.do	Plot_Searches_Quarterly.pdf Plot_Contraband_Quarterly.pdf
App. Figure B19	17-hypo_spc.do	hypo_spc.pdf

List of Additional Programs (not included above in List of Tables and Figures)

Program	Description
1-stop_construct.do	This program uses the raw stops data from /TX_original/csv/ to construct the stops_all.dta data file
4-calc_stop_rates.do	This program uses stops_all.dta to construct the trp_stop_rates.dta data file that provides stop rates at the officer by location by time level
5-speed_patterns.do	This program uses stops_working.dta to construct the trp_speeds.dta data file that provides cited speeds at the officer by location level and to produce related summary statistics
7-spc_pool_night.do	This program uses stops_working.dta to construct the np_rates_pool_night.dta data file that is used to estimate the between-trooper search productivity curve for night stops

Part 2: Accessing Data Files That Cannot Be Publicly Posted

Administrative data on Texas Highway Patrol stops can be obtained from the Stanford Open Policing Project (openpolicing.stanford.edu). We use the raw data that the Stanford Open Policing Project received from the Texas Department of Public Safety as part of a FOIA request. The raw data include motorist identifying information, including full name and address. These data can be requested by contacting open-policing@lists.stanford.edu. These data correspond to the `/TX_original/csv/` files listed under *List of datasets not included* and used in program `1-stop_construct.do` to construct the `stops_all.dta` data file. Pre-processed data with harmonized variable names are publicly posted by the Stanford Open Policing Project at https://openpolicing.stanford.edu/data/. These data are missing motorist identifying information needed to construct a number of the variables required to run the programs described above.

Administrative data from the Texas Computerized Criminal History System were obtained from the Texas Department of Public Safety under Texas Government Code 411.083(b)(4). Requests for data access can be directed to OGC.Webmaster@dps.texas.gov. These data are used to construct the `arrest_hist.dta` and `arrests.dta` data files listed under *List of datasets not included* and used in program `2-create_working.do` to construct the `stops_working.dta` data file. These data are also used in conjunction with proprietary Infutor data on address histories to construct the `final_ids.dta` data file listed under *List of datasets not included* and used in program `2-create_working.do`.