

Estimate Regression Models

Basic Information

- Reference → Ba, Bocar A., Dean Knox, Jonathan Mummolo, and Roman Rivera. 2021. “The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago.” Science.
- Github (for my code) → <https://github.com/jrisi256/policing>
- Data location → <https://codeocean.com/capsule/8907164/tree/v1>

Steps For Replication

- The data provided by the authors is in its raw form so the end user (in order to successfully replicate) needs to replicate the data cleaning and transformation steps themselves.
- **Step 1:** Merge officer demographic information with officer shift assignments. See `create-officer-assignments/src/merge-officer-assignments-ba.qmd`.
- **Step 2:** Determine which of the outcomes occurred during a shift. I.e., merge stops, arrests, and uses of force to the merged officer demographic/shift assignment data file. See `merge-stops-shifts/merge-stopss-shifts-ba.qmd`, `merge-arrests-shifts/merge-arrests-shifts-ba.qmd`, `merge-force-shifts/merge-force-shifts-ba.qmd`.
- **Step 3:** For each resulting merged file, count the number of outcomes which occurred during each shift and record these data as new columns. These represent our dependent variables. See `create-outcomes/src/create_outcomes_ba_max.R`.
- **Step 4:** The authors for this paper, I believe, are technically employing a four-way fixed effects OLS model. The fixed effects being: **1**) the month-year (e.g., January 2013), **2**) the timing of the shift (i.e., morning, afternoon, or evening), **3**) location of the shift (i.e., each shift is assigned to a specific beat), and **4**) day of the week. First, any groupings which do not have at least 2 racial groups represented (or 2 sex groups when running analyses for men vs. women), are dropped. Second, the independent and dependent variables are demeaned within each grouping.
- **Step 5:** Estimate the OLS regression models using the demeaned variables. The models themselves are quite simple. The outcome variables are the number of stops, arrests, or uses of force. The independent variable is only the race of the officer.

- **Limitations:** The replication of results was much more involved than I thought so I have not been able to replicate their results for officer sex (yet). Similarly because of the four-way fixed effects, they use a Bootstrap method to calculate their standard errors. The Bootstrapping procedure is computationally demanding so I have not been able to replicate the standard errors yet, either. However, I have replicated their point estimates, and I compare them informally to the main figure in their paper. The point estimates (verified more carefully in my code) are equivalent to their point estimates. Assuming they are using a 95% confidence interval in their graph (because that is what I am using in my graph), our standard errors are not actually that different.

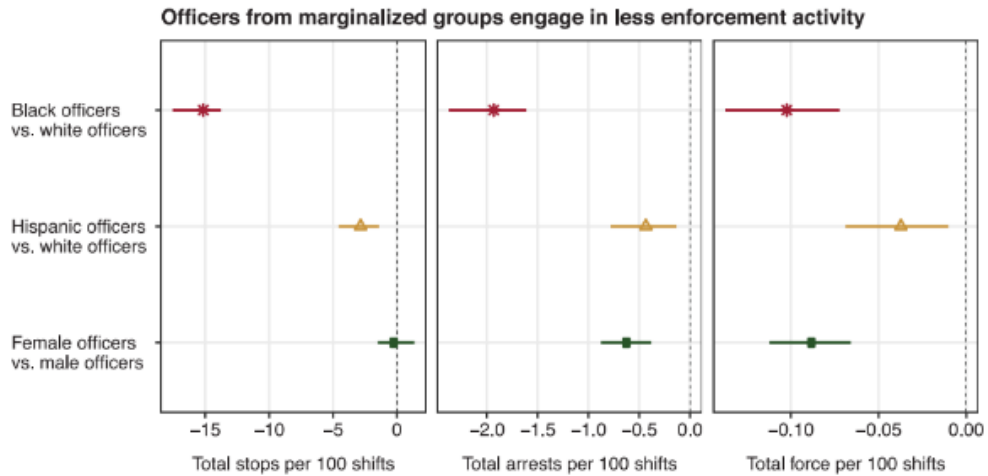
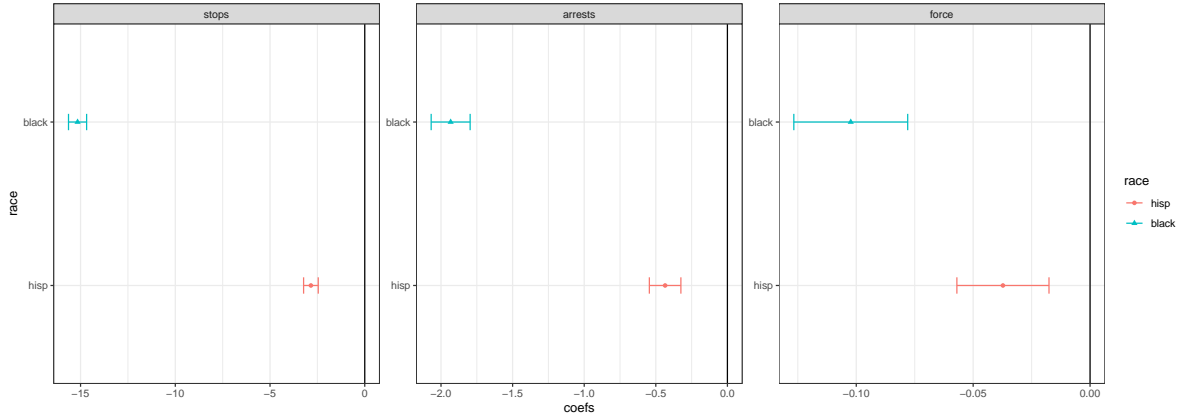


Fig. 3. Effects of deploying officers from marginalized groups on total enforcement activity. Average within-MDSB differences in rates of stops, arrests, and uses of force across officer groups. See tables S4 to S8 for numeric results.

Figure 1: Results from Ba et al. 2021 paper

Nodal attributes

