Project Proposal

Misclassification by Race / Ethnicity and Consequences for Police Stops in Texas
Mariana BV Claudia SR | Feb 19, 2020

Statement & Motivation

There is overwhelming evidence that officers in Texas "misidentify" individuals, categorizing them as white in instances where the individuals do not self-identify as white (see ticketing hispanic as white KXAN investigation 2015). Officers may identify Hispanic individuals as white for many reasons, and with consequences of interest for public policy. Officers may engage in a deliberate deception to alter recorded statistics (as suggested in the KXAN article). The misidentification may be the result of an unintentional misperception of the individual's race or ethnicity, or misunderstanding of the race-as-racial/ethnic-descent procedural definition of race in a police stop (and perhaps confusion with the more common understanding of race as separate from ethnic descent outside of the procedural instructions). Regardless of the specific motivation behind officers' differential misidentification of the Hispanic individual, little research has been done on what effect this might have on outcomes of the stop interaction. The proposed project seeks to understand, relative to individuals who are "correctly classified," whether misidentified individuals are more or less likely to be arrested or searched?

Misidentification of hispanic descent across race categories in police stops uniquely allows for the study of variation in recording statistics, police officer perception, and/or standard of evidence along an additional axis of ethnicity. This is of particular interest to research on discrimination and bias in police stops and other enforcement activity, given the historical precedent for categorization of Hispanic individuals as white in discrimination cases and census records (Lopez 1997). Today, researchers tend to classify all Hispanic individuals as Hispanic only (e.g. Pierson 2020), even though Hispanic individuals often self-identify as white or black (Golash-Boza and Darity Jr 2008). This occurs even in circumstances where there may have been separate racial and ethnic categories available in the data (see Pierson and others forthcoming 2020). While there may be efficiency reasons for researchers to do this type of analysis (e.g. fewer categories to compare), the results suggest that the Hispanic status of stopped individuals is more salient to the officer than their racial status. If this is indeed the case, under what circumstances do officers identify individuals as Hispanic rather than using a racial category? Are there particular areas (e.g. predominantly white neighborhoods, wealthier neighborhoods) where individuals are more likely to be categorized as white?

Our proposed research project seeks to answer two questions, 1) Under what circumstances (neighborhood income levels, racial composition) are hispanic individuals more likely to be identified by racial categories?, and 2) What effect does this labeling have on drivers' outcomes? Specifically, what proportion of searches successfully turn up contraband for each group (Becker outcome test)? For the purposes of this research, we will look into hispanic/white, hispanic/black, hispanic/asian/ other combinations, but this will likely be dependent on the sample size available for each category.

Data

We will use the data made available by the Open Policing Project at Stanford. Because evidence of "mis-identification" is only available for Texas jurisdictions, we will focus our analysis on the Texas stops. We will also restrict the sample to stops occurring after 2015, as this appears to be the period after which officers were retrained in documenting driver ethnic/racial status (Friberg 2015). The data provided by Stanford includes stop latitude and longitude, which can be merged with census tract level data (likely ACS decennial 2010) to generate covariates for the surrounding geographic area.

Methods

Three approaches are considered to answer the research questions:

1. How does hisp/non-hisp mislabel vary by ethnic/racial composition of the surrounding geography and nature of the stop?
   a. Mapping/descriptive exercise (MSA, poverty level)
   b. Logistic regression predicting misclassification with census-tract level predictors
   c. Spatial point process estimation to generate point specific estimates of misclassification for use in 2 (maybe not that useful?)

2. What effect does this have on drivers?
   a. Replicate analysis from Pierson paper adding the inaccurately labeled white non-hispanic group and correctly labeled any-race-Hispanic groups (hereafter "mislabeled hispanic/non-hispanic"). I.e. estimate potential bias in stop decisions by examining whether mislabeled hispanic /non-hispanic vs. white labeled drivers are less likely to be stopped after sunset, when a "veil of darkness" masks one's race
   b. Investigate potential bias in decisions to search stopped mislabeled hispanic/non-hispanic drivers by examining both the rate at which drivers are searched and the likelihood that searches turn up contraband- predict hit rate.
      i. Add control for propensity to be misclassified in a given location (estimated in 1b)
      ii. Show models with and without propensity score control

Timeline

The table below presents a timeline of the anticipated progress on our project over the semester.

Week	Tasks
4 Feb 20	Submit proposal Open datasets; organize shapefiles
5	Finalize proposal

Feb 27	Send first project update Select datasets for usage Begin data cleaning Continue lit review
6 Mar 5	Data cleaning Merge with Census Data Continue lit review
7 Mar 12	Troubleshoot any issues with Census Data or big datafiles Continue lit review
** Mar 12	Spring Break Some prelim models Continue lit review
8 Mar 26	Claudia unavailable Some more prelim models
9 Apr 2	Progress meeting for final project Show preliminary models
10 Apr 9	Mariana unavailable Edits to models Output for charts
11 Apr 16	Writing Draft Presentation Format charts
12 Apr 23	Writing Draft Paper Practice presentation
13 Apr 30	Buffer
14 May 7	Presentation Get/incorporate feedback Paper edits
15 May 11	Paper due 6 pm

References

E. Pierson, C. Simoiu, J. Overgoor, S. Corbett-Davies, D. Jenson, A. Shoemaker, V. Ramachandran, P. Barghouty, C. Phillips, R. Shroff, and S. Goel. (Forthcoming 2020) "A large-scale analysis of racial disparities in police stops across the United States".
http://www.rshroff.com/uploads/6/2/3/5/62359383/100m-stops.pdf

Golash-Boza T, Darity Jr W. Latino racial choices: the effects of skin colour and discrimination on Latinos' and Latinas' racial self-identifications. Ethnic and Racial Studies. 2008 Jul 1;31(5):899-934.
https://www.tandfonline.com/doi/full/10.1080/01419870701568858

Lopez, Ian Haney. White by law: The legal construction of race. Vol. 21. NYU Press, 1997.

Open Policing Data. https://openpolicing.stanford.edu/data/

Surnames. U.S. Census. https://www2.census.gov/topics/genealogy/2000surnames/surnames.pdf
Texas troopers ticketing Hispanic drivers as white (Ben Friberg, David Barer, Rachel Garza, Josh Hinkle, Robert Sims, Calily Bien, Patrick Tolbert and Chad Cross Nov 6, 2015)
https://www.kxan.com/investigations/texas-troopers-ticketing-hispanic-drivers-as-white/

Stanford Police Stop Texas state data [link] (do not share)