

An Analysis of Racial Profiling in the US using Propensity Score Matching

Stephen McLoughlin | Email: c00264570@itcarlow.ie

Supervisor: Dr Oisín Cawley

MSc in Data Science, Institute of Technology Carlow

1. Introduction

Law enforcement in the United States have been strongly criticized for their approach and service to the minority communities which they are intended to protect (Higgins et al., 2012). While no widely recognized meaning of racial profiling is agreed upon, this study defines racial profiling as the act of suspecting, targeting, or discriminating against a citizen based on their race rather than on individual suspicion (Hernandez-Murillo and Knowles, 2004).

According to research, the most common interaction between a police officer and a member of the public arises during a traffic stop (Helfers, 2016). The use of racial profiling by police officers negatively affects those who are stopped, and the mistrust it causes overwhelmingly impacts racial minorities (Epp, Maynard-Moody and Haider-Markel, 2016). This research of analysing traffic stop outcomes will contribute to a better understanding of police officer decision-making processes in relation to racial profiling.

2. Research Objective

This research will analyse police officer decision-making in traffic stop outcomes by using propensity score matching with logistic regression. This will match comparable racial groups and comparable male and female racial groups to identify racial disparities in policing.

3. Literature Review

This dissertation will investigate the outcomes of **search, frisk, citation, and arrest** as they reflect a level of involvement from the officer and could demonstrate differing degrees of racial profiling that may take place during a traffic stop. Previous studies have shown that young Black male drivers are more likely to get a **citation** than White drivers (Tillyer and Engel, 2013). It has also been found that, in comparison to White drivers, Black and Hispanic drivers are more likely to have their vehicles **searched and be arrested** after a traffic stop (Withrow, 2004).

Matching is the process of identifying an individual in a treatment group that shares common measurable characteristics with a person in a control group based on a specific research question (Higgins et al., 2012). Only three studies to date have used **propensity score matching** to analyse racial profiling. None of these have examined the outcomes of frisks and arrests. This will also be the first study known to use propensity score matching to examine traffic stops over a multi-year period and the first to analyse state patrol officer decision-making.

4. The Data

- Obtained from The Stanford Open Policing Project.
- Contains 7,297,538 records of traffic stops by the Florida State Patrol from 2010 to 2018.
- Measurements of interest:

- | | | |
|------------|-----------------|-------------------------------|
| • Search | • Driver Gender | • Officer Race |
| • Frisk | • Driver Race | • Officer Gender |
| • Citation | • Driver Age | • Officer Age |
| • Arrest | | • Officer Years of Experience |

5. Technologies



7. Early Indications and Next Steps

Early indications:

- Using propensity score matching allows for better statistical analysis than standard ordinary least squares or multiple regression analysis because it matches drivers based on many measures so that those in similar situations are compared to see if any racial disparities in traffic stop outcomes truly exist.
- Highlighting the issue of racial profiling by police officers will help lead to policy reform and further training within police departments.

Next steps include:

- Involve an independent researcher in the collection of data to help address the self-report issue that a police officer could under-report information.
- Apply theoretical explanations for racial profiling to help explain police officer decision-making such as focal concerns theory.

6. Methodology

1. The data will be cleaned and manipulated through **Jupyter Notebook** using **Python** and the **pandas** package.
2. The data will be loaded into **RStudio** to initially perform some descriptive statistics to indicate how the data is distributed.
3. **Logistic regression models** will be calculated with race as the treatment variable. This study will focus on White and Black drivers.
4. Individuals will be matched using a **caliper-based nearest neighbour** technique that is available from the **matchit** package in **R**.
5. The quality of the **propensity score matchings** will then be assessed.
6. Finally, the logistic regression models on the **weighted matches** of the propensity scores will be calculated.

References

1. Epp, C.R., Maynard-Moody, S. and Haider-Markel, D. (2016). Beyond Profiling: The Institutional Sources of Racial Disparities in Policing. *Public Administration Review*, [online] 77(2), pp.168–178. Available at: <https://onlinelibrary.wiley.com/doi/full/10.1111/puar.12702>.
2. Helfers, R.C. (2016). Ethnic disparities in the issuance of multiple traffic citations to motorists in a southern suburban police agency. *Journal of Ethnicity in Criminal Justice*, 14(3), pp.213–229.
3. Higgins, G.E., Vito, G.F., Grossi, E.L. and Vito, A.G. (2012). Searches and Traffic Stops: Racial Profiling and Capriciousness. *Journal of Ethnicity in Criminal Justice*, 10(3), pp.163–179.
4. Tillyer, R. and Engel, R.S. (2013). The Impact of Drivers' Race, Gender, and Age During Traffic Stops. *Crime & Delinquency*, 59(3), pp.369–395.
5. Withrow, B.L. (2004). Race-Based Policing: A Descriptive Analysis of the Wichita Stop Study. *Police Practice and Research*, 5(3), pp.223–240.