

POLICE USE OF FORCE AS AN EXTENSION OF ARRESTS:
EXAMINING DISPARITIES ACROSS CIVILIAN AND OFFICER RACE

EMILY K. WEISBURST

ONLINE APPENDIX

This appendix motivates the test of racial bias used in this paper by adapting the economic models in Anwar and Fang (2006) and Fryer (2018). The model applies to officer decisions to use force during the course of an arrest.

Arrestees and officers belong to one of two groups $r_s, r_p \in \{M, W\}$. Conditional on observable characteristics, other characteristics of arrests are similar across officer and arrestee race. However, the unobservable compliance rates of different civilian race groups may vary.

For simplicity, this appendix considers the maximization problem of officers and takes civilian behavior as given.¹ In this framework, arrestees are either compliant or non-compliant. Total levels of compliance can vary by arrestee race group. π^{rs} is the average likelihood of non-compliance for an arrestee race group.

Each arrestee exhibits a signal of non-compliance to police officers, $s \in (0, 1)$. The likelihood of non-compliance is increasing in s . Signals of compliance are viewed at the beginning of an arrest interaction. Actual civilian compliance is revealed after the officer decides whether or not to use force, through the civilian and officer interaction. The distribution of s varies by race and compliance type. This signal captures elements of the interactions between civilians and officers; including verbal exchanges, visible weapons, attitude, posture, clothing, etc.

Let $G_c^{rs}(s)$ be the probability that the signal does not exceed s given that the civilian in group r_s is compliant, and $G_n^{rs}(s)$ represent this probability given civilian non-compliance. $g_c^{rs}(s)$ and $g_n^{rs}(s)$ are the corresponding probability density functions.

Define $\mu^{rs}(s) = \frac{g_n^{rs}(s)}{g_c^{rs}(s)}$ is the likelihood ratio of non-compliance at s . $\mu^{rs}(s)$ is strictly increasing in s , $\mu^{rs}(s) \rightarrow \infty$ as $s \rightarrow 1$ and $G_c^{rs}(s) \leq G_n^{rs}(s) \forall s$. Very high signals provide nearly certain information of non-compliance.

From the officer's perspective, the posterior likelihood of non-compliance is given by Bayes' rule for each suspect group, r_s .

$$\Psi(\pi^{rs}, s) = \frac{\pi^{rs} g_n^{rs}(s)}{\pi^{rs} g_n^{rs}(s) + (1 - \pi^{rs}) g_c^{rs}(s)}.$$

Officers face a common set of costs when interacting with compliant arrestees, as well as a common cost of not exerting force when the arrestee is non-compliant. The benefit of not using force with a compliant arrestee is normalized to equal 0. Officers also receive a benefit when they use force on an arrestee that is non-compliant, and this benefit varies across officer and suspect race.

$$\begin{aligned} \text{force used} &: \begin{cases} t(r_s, r_p) & , \text{ non-compliant} \\ -y & , \text{ compliant} \end{cases} \\ \text{no force used} &: \begin{cases} -z & , \text{ non-compliant} \\ 0 & , \text{ compliant} \end{cases} \end{aligned}$$

¹It is also possible to extend this model to consider civilian returns to compliance and make compliance rates π endogenous.

Each officer maximizes his utility as a choice between exerting force and not exerting force:

$$\max\{\Psi(\pi^{rs}, s) \cdot t(r_s, r_p) - (1 - \Psi(\pi^{rs}, s))y, -\Psi(\pi^{rs}, s)z\}$$

This maximization implies that officers will choose to exert force, when:

$$t(r_s, r_p) \geq \frac{1 - \Psi(\pi^{rs}, s)}{\Psi(\pi^{rs}, s)} \cdot y - z$$

The right-hand side of this inequality is strictly decreasing in s , implying that officers will exert force when $s \geq s^*(r_s, r_p)$, where $s^*(r_s, r_p)$ satisfies equality in the expression above. A higher race-specific benefit of force, $t(r_s, r_p)$, is associated with a lower signal threshold $s^*(r_s, r_p)$.

The total probability of force is decreasing in $s^*(r_s, r_p)$ (and increasing in $t(r_s, r_p)$) and is given by:

$$UOF(r_s, r_p) = 1 - G_n^{rs}(s^*(r_s, r_p)).$$

Following the Anwar and Fang (2006) framework, the following definitions apply:

- 1) Racial Bias: Officers are racially biased if for some officer race, r_p , $t(M, r_p) \neq t(W, r_p)$.
- 2) Monolithic Behavior: Officers are not monolithic in their behavior if officer benefits of using force differ across officer race for a given suspect race, or $t(r_s, M) \neq t(r_s, W)$.
- 3) Statistical Discrimination: Assume all officers are not racially biased, or $t(M, r_p) = t(W, r_p)$. Then r_p officers will exhibit statistical discrimination if $s^*(M, r_p) \neq s^*(W, r_p)$.

If officers are not racially biased and exhibit monolithic behavior, then $t(M, M) = t(M, W) = t(W, W) = t(W, M)$. It follows that use of force rates within suspect race will be constant across officer race, but that total use of force rates may differ across arrestee race groups if $s^*(M, r_p) \neq s^*(W, r_p)$, or there is statistical discrimination. Statistical discrimination is equivalent to officers using race as a signal of compliance. This could be a rational response if different race groups have different distributions of signals of compliance, $G_c^{rs}(s)$ and $G_n^{rs}(s)$, and/or different total rates of compliance, π^{rs} .

If officers do not exhibit monolithic behavior but are also not biased, then the relative ranking of use of force rates across officer race within arrestee race will be independent of arrestee race. For example, allow M officers to have a lower benefit of using force for all arrestee race groups. This will translate to a lower use of force rate for M officers relative to W officers for both arrestee race groups.

Conversely, if use of force rates are higher for W officers interacting with M arrestees, $UOF(W, W) < UOF(M, W)$, and use of force rates are higher for M officers interacting with W arrestees, $UOF(W, M) > UOF(M, M)$, then we can conclude that one or both groups of officers is biased. This is illustrated through the following stylized example:

$$\begin{array}{ll} t(M, M) < t(W, M) & \& t(W, W) < t(M, W) \\ \& t(M, M) = t(W, W) & \& t(M, W) = t(W, M) \\ s^*(M, M) & > & s^*(M, W) \\ s^*(W, M) & < & s^*(W, W) \\ UOF(M, M) & < & UOF(M, W) \\ UOF(W, M) & > & UOF(W, W) \end{array}$$

The opposing rank order in officer arrest rates for different arrestee race groups violates the null hypothesis and implies that $t(M, W) \neq t(W, W)$ and/or $t(M, M) \neq t(W, M)$.

REFERENCES

Anwar, Shamena, and Hanming Fang. 2006. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *The American Economic Review*, 96(1): 127-151.

Fryer, Roland G. Jr. 2018. "An Empirical Analysis of Racial Differences in Police Use of Force." *Journal of Political Economy*, Forthcoming.