Testing for racial bias in searches of motor vehicles

Camelia Simoiu
With Sam Corbett-Davies and Sharad Goel

Stanford University
Traffic stops

- Traffic stops are the primary way in which the public interacts with law enforcement
- Widespread concern of racial bias in police actions
- Seemingly reasonable tests of discrimination can give misleading results
Our contribution

- Novel test for discrimination, “threshold test” to measure racial bias in officers' **decision to search**

- Are minorities subjected to a search on the basis of less evidence than whites?

- Bayesian hierarchical latent variable model
North Carolina Data Set

- 4.5 million stops
- 6 year observation period: 2009-2014
- Largest 100 local police departments
  - account for 90% of local stops
- 4 race groups (White, Black, Hispanic, Asian)
Standard Tests of Discrimination
**Benchmarking Test**

Compare likelihood of being searched across race groups

<table>
<thead>
<tr>
<th>Race</th>
<th>Search Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>4.4%</td>
</tr>
<tr>
<td>Black</td>
<td>8.3%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5.9%</td>
</tr>
<tr>
<td>Asian</td>
<td>2.3%</td>
</tr>
</tbody>
</table>
Outcome Test  [Becker 1957, 1992]

Compare the search success (hit) rate across race groups

<table>
<thead>
<tr>
<th>Race</th>
<th>Hit Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>36%</td>
</tr>
<tr>
<td>Black</td>
<td>32%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>23%</td>
</tr>
<tr>
<td>Asian</td>
<td>29%</td>
</tr>
</tbody>
</table>
Problem of infra-marginality [Ayers, 2002]

It is possible to find lower hit rates and higher search rates for minorities in the presence of no discrimination.

- Two types of white drivers: 5% or 75% chance of carrying contraband
- Two types of black drivers: 5% or 50% chance of carrying contraband

- If officers search drivers who are at least 10% likely to be carrying contraband
  - White hit rate: 75%
  - Black hit rate: 50%
Threshold Model
Modeling a Traffic Stop

- Officer in department $d$ stops a driver of race $r$
- Officer observes a random signal: $x_i \sim \text{Beta}(\Phi_{rd}, \lambda_{rd})$
Modeling a Traffic Stop

- Officer in department $d$ stops a driver of race $r$

- Officer observes a random signal: $x_i \sim \text{Beta}(\Phi_{rd}, \lambda_{rd})$

- Deterministically conduct search $S_i = 1$ iff $x_i > t_{rd}$

- If $S_i = 1$: $H_i \sim \text{Bernoulli}(x_i)$

- Lower $t_{rd}$ indicate discrimination
Problem of infra-marginality [Ayers, 2002]

Discrimination against Blue by construction.

Benchmark and outcome tests fail to identify discrimination against Blue.

<table>
<thead>
<tr>
<th></th>
<th>Red</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search rate</td>
<td>71%</td>
<td>64%</td>
</tr>
<tr>
<td>Hit rate</td>
<td>39%</td>
<td>44%</td>
</tr>
</tbody>
</table>
Parametrizing the Signal Distribution

\[ x \sim \text{Beta}(\Phi_{rd}, \lambda_{rd}) \]

\[ \Phi_{rd} \sim \logit^{-1}(\Phi_r + \Phi_d) \]

Probability that a driver is carrying contraband

\[ \lambda_{rd} \sim \exp(\lambda_r + \lambda_d) \]

Difficulty in distinguishing between guilty and innocent drivers
Simplifying inference

For a given department $d$, race $r$

Observe $N_{rd}$ stops

$x_{rd} \sim \text{Beta}(\Phi_{rd}, \lambda_{rd})$

$\delta_{rd} = P(x_{rd} > t_{rd}; \Phi_{rd}, \lambda_{rd})$

$\gamma_{rd} = E(x_{rd} | x_{rd} > t_{rd}; \Phi_{rd}, \lambda_{rd})$

$S_{rd} = \text{Binomial}(\delta_{rd}, N_{rd})$

$H_{rd} = \text{Binomial}(\gamma_{rd}, S_{rd})$
Race parameters
\( \Phi_r \sim N(0,2) \)
\( \lambda_r \sim N(0,2) \)

Department Parameters
\( \Phi_d \sim N(\mu_d, \sigma_d) \)
\( \mu_d \sim N(0,2) \)
\( \sigma_d \sim N_+(0,2) \)
(same for \( \lambda_d \))

Threshold Parameter
\( t_{rd} \sim \logit^{-1}(N(\mu_{trd}, \sigma_{trd})) \)
\( \mu_{trd} \sim N(0,2) \)
\( \sigma_{trd} \sim N_+(0,2) \)
Performing Inference

- No-U-Turn Sampler (NUTS) in Stan [Hoffman and Gelman, 2014]
- An extension of Hamiltonian Monte Carlo (HMC) that retains efficiency and requires no hand-tuning

Assessing convergence
- Simulate 5 independent Markov chains
- 5,000 iterations (2,500 warmup, 2,500 sampling)
- Inspect potential scale reduction factor $R$, and effective sample size
Results
Results
## Results

<table>
<thead>
<tr>
<th>Race</th>
<th>Search Threshold</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>19%</td>
<td>(18%, 21%)</td>
</tr>
<tr>
<td>Black</td>
<td>5%</td>
<td>(2%, 8%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>8%</td>
<td>(6%, 10%)</td>
</tr>
<tr>
<td>Asian</td>
<td>17%</td>
<td>(14%, 19%)</td>
</tr>
</tbody>
</table>

### Likelihood of carrying contraband

- **Density**
- **White**
- **Black**
- **Hispanic**
- **Asian**

**X-axis:** Likelihood of carrying contraband

**Y-axis:** Density
Posterior Predictive Check

RMS prediction error 0.2%

RMS prediction error 2.7%
Infra-marginality in the wild: Raleigh, NC

Black drivers:

- Higher search rate than whites (5.7% vs. 2.4%)
- Higher hit rate than whites (19% vs. 15%)

<table>
<thead>
<tr>
<th>Race</th>
<th>Hit Rate</th>
<th>Search Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>15%</td>
<td>10%</td>
</tr>
<tr>
<td>Black</td>
<td>19%</td>
<td>5%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>Asian</td>
<td>11%</td>
<td>91%</td>
</tr>
</tbody>
</table>
Conclusions

- Bayesian latent variable model allows for direct estimation of thresholds, overcoming the problems of omitted-variable bias and infra-marginality.
- Find unjustified disparate impact against black and Hispanic drivers in North Carolina.
- Had the white search threshold been applied, 30,000 fewer searches of black drivers and 8,000 fewer searches of Hispanic drivers.
- Cannot prove biased intent, but we can shift the burden of proof.
Questions?
Omitted Variable Test
Testing for heterogeneity in the thresholds