

COMMENTS ON
“MEASURING RACIAL PROFILING”

Victoria Stodden
Columbia University

Conference on Empirical Legal Studies
Northwestern Law School
November 5, 2011

WHAT THE PAPER DOES WELL

- confronts a politically difficult question,
- confronts a statistically difficult problem,
- clarifies the complex nature of the modeling problem itself,
- suggests innovative ways for improving on estimation techniques for problems with “multi-layered” selection bias,
- clarifies modeling approaches with the use of simulation,
- impressive depth of data collection.

SUGGESTIONS

3 categories of improvements

1. Data, including questions on variable choice and definitions,
2. The model and model presentation,
3. General notes and comments.

DATA QUESTIONS

- race variable:
 - Need a theory for inclusion, the one given (that inclusion is justified since race metrics won't be identical by crime) is insufficient - it justifies the inclusion of any variable. Need a reason to suspect it is *significant*.
 - Since race is the driver for decisions to stop/search ideally it would be more nuanced. Race as a spectrum? Could help illuminate what component is race and what are other factors correlated with race e.g. culture.

DATA QUESTIONS

- include plots of variable distributions over time,
- address whether there are other exogenous reasons for police behavior, such as regulatory changes. Discuss omitted variables,
- include continuous variable for police experience rather than a 10-year dummy,
- time dummy,
- surprising the out of state plate makes little difference (Donahue, DeAngelo),
- justify the reduction to one stretch of MD road.

MODELING QUESTIONS

- Lack of clarity in model exposition:
 - need table(s) of variables, data, and their meanings, (including for the aggregate data you have)
 - release of data and code,
- Distributions chosen often for mathematical convenience:
 - can more informative priors be estimated? from the previous work?
 - hierarchical Bayes techniques may be useful

MODELING QUESTIONS

- Summarize approaches of previous work you build upon (ie. Heckman),
- Repeat the simulations for your innovative estimation case (create a true underlying model and test your method for accuracy in recovery),
- For all simulations give how data created, how many observations, how many iterations, etc.; provide code,
- All tables need complete captions, including a comprehensive description of their contents, sample size, associated models..
- Need significance levels of coefficients in the tables,
- Bayesian selection bias inference accurate only insofar as the distributional assumptions are accurate (“exact” estimation of Z),

GENERAL COMMENTS

- Organizational: give relevant legal standards up front in the paper, then the statistical issues,
- All assertions, such as those regarding black youth on page 4, should be referenced,
- General increase in level of precision in exposition (precision in variable definitions, model definitions, estimation steps, footnote precision..),
- Release of data and code vital.