**We want to discuss three issues in this analysis:**

1) Problems of the threshold test;

2) Definition and evaluation of fairness in court;

3) A complementary method to detect racial discrimination.

**Part One: Problems of the threshold test**

The threshold test is indeed a robust statistical model to complement benchmark test and outcome test. The purpose of this section is not to negate the contribution of the threshold test, but rather to offer some new perspectives.

Besides common criticisms like “it is unrealistic that police officers can make correctly calibrated estimates of each person’s probability of carrying contraband”, we identified two problems of the threshold test even after we accept the assumption above. First, there might be a discrepancy between a person’s actual probability of carrying contraband (the probability measured in the paper) and the probability that police offices perceive on the spot. For instance, within each police department’s patrol area, there are always some areas that are more dangerous than average. Any person who wanders in these dangerous areas will have a higher perceived probability of carrying contraband compared with the same person wandering on a safe street, even if nothing intrinsic about this person has changed. By the same token, a shabbily dressed person will have a higher perceived probability of carrying contraband compared with the same person dressing in suits. In general, poor socioeconomic status will make a person look more “dangerous” than he actually is. We don’t know if minorities’ socioeconomic disadvantages increase their perceived probability of carrying contraband, even if police officers intend to set up same threshold for all races. Ideally, the threshold test should also be applied across all income groups, but we don’t have the data.

The second question we want to bring to discussion is the so-called deterrence effect stop-and-frisk advocates proclaim. In the landmark lawsuit *Floyd, et al v. City of New York, et al.,* the defendant argues that hit-rate should not be the sole criterion to evaluate the effect of stop and frisk, since stop and frisk also deters people from engaging in illegal activities in the future. For instance, a police officer searched a person of 60% probability of carrying contraband and this search resulted in a miss, but maybe this search deterred that person from carrying contraband in the future. Let’s assume, in a hypothetical area, race A has a significantly higher crime rate than race B. Is it legitimate for police officers to adopt a relatively lower threshold to deter certain people of race A from committing crimes? Although we don’t agree with this proposition, we think it is worth bringing up.

**Part Two: Definition and evaluation of fairness in court**

After reading court opinions of several stop and frisk cases, we found it nearly impossible to define fairness. Under many cases, the court used a preponderance of evidence to support its ruling as opposed to a single definition. Nevertheless, we found some patterns in court’s deciding making process. To evaluate whether a policy is race neutral, the court requires the defendant/plaintiff to prove/disprove two things: 1) there is no *direct* racial profiling in the designation and implementation of the policy (intention); 2) the policy does not make minorities vulnerable by subjecting them to collective subconscious/indirect racial bias of the policy’s enforcers (result). The first field of contention is usually proved/disproved through direct testimony. For instance, a NYPD insider once revealed that some officer specifically target minorities in order to meet the confiscation quote more quickly. Where most statistical models come to play is usually in the second field contention, namely, whether there is indirect bias involved as a result of a policy. We think all three models, benchmark test, outcome test and threshold test, reveal some truth of whether the result of a policy involves racial bias. Given the pros and cons of each model, we think in a court setting, it is the best to use all models to form a preponderance of evidence (also considering the fact that judges and juries are not as statistically sophisticated as model designers).

**Part Three: A complementary method to detect racial discrimination**

One major criticism of stop and frisk is that, in many cases, police officers stop minorities without a truly reasonable suspicion. We came up with a model to evaluate that claim.

When filling out the post-stop-and-frisk reporting form, police officers can choose from a list of reasons to account for why they stopped a person. Some reasons are generally deemed subjective and result in low hit rates, e.g. furtive movement, evasive response; some reasons are deemed to be more objective and result in high hit rate, e.g. suspicion of carrying weapons.

We can first categorize all reasons of suspicions into three groups: reasons leading to high hit rates, reasons leading to medium hit rates and reasons leading to low hit rates. Then we calculate what percentage of Black/White/Hispanic/Asian people were stopped for each group of reasons. If minorities are more often stopped for trivial reasons than white, it tells things.

Actually, this method has been used in the lawsuit *Floyd, et al v. City of New York, et al.* Furtive movement is arguably one of the most controversial justification to stop a people. The court found that 48% of Blacks, 45% Hispanics and 40% of Whites were stopped under this reason, which add to the evidence of discrimination.