Does Selective Crime Reporting Influence our Ability to Detect Racial Discrimination in the NYPD's Stop-and-Frisk Program?*

Steven F. Lehrer
NYU-Shanghai, Queen's University and NBER

Louis-Pierre Lepage
University of Michigan

July 2018

Abstract

Prior analyses of racial bias in the New York City's Stop-and-Frisk program implicitly assumed that potential bias of police officers did not vary by crime type and that their decision of which type of crime to report as the basis for the stop did not exhibit any bias. In this paper, we first extend the hit rates model to consider crime type heterogeneity in racial bias and police officer decisions of reported crime type. Second, we reevaluate the program while accounting for heterogeneity in bias along crime types and for the sample-selection which may arise from conditioning on crime type. We present evidence that differences in biases across crime types are substantial and specification tests support incorporating corrections for selective crime reporting. However, the main findings on racial bias do not differ sharply once accounting for this choice based selection.

* We are grateful to Decio Coviello, Maxwell Pak and Rosina Rodriguez Olivera for helpful comments and suggestions on this project. We also thank Victor Aguiar and other participants at the Econometrics of Complex Survey Data: Theory and Applications conference for additional comments. NYC Stop-and-Frisk data is publicly available at *https://nycopendata.socrata.com* as well as at the ICPSR website at the University of Michigan. Lehrer thanks SSHRC for research support. We are responsible for all errors.

1 Introduction

In many administrative and survey data sets, researchers must confront the challenge that non-sampling errors due to deliberate bias in providing a response may distort the analyses. Much research has investigated this issue in survey research (see Bound, Brown, and Mathiowetz (2001) for a survey of the literature) and the presence of measurement error has been shown to potentially cause biased and inconsistent parameter estimates thereby leading to erroneous conclusions to various degrees in statistical and economic analyses. Different methods are needed to treat measurement error in survey data since errors can arise from different sources. For example, they might arise from coding errors by surveyors or survey participants may choose to not provide truthful responses.

A specific form of measurement error arises with qualitative data resulting in misclassification. Misclassification occurs when observations are placed erroneously in a different group or category. Within administrative data sources, this erroneous information is provided not from survey responses, but rather in how the records are generated and maintained. Just as certain sampling issues may influence those being surveyed, it may also affect those who create and maintain administrative records. For example, individuals preparing entries in administrative records may rely on rules of thumb in a bid to minimize the burden of completing the underlying forms accurately.[1] These errors in classification not only affect summary statistics on sample proportions but may influence analyses that investigate heterogenous behavioral relationships across these groups or categories.[2] In many settings, economic the-

[1] A large literature has documented evidence that police officers engage in implicit bias and employ rules of thumb (see Fridell and Lim (2016) for a recent survey). In recent work, James (2017) provides evidence that the association between African-Americans and weapons is stronger when officers have less sleep. The NYPD is well aware of this literature. In December 2014, the NYPD announced that they would retrain a significant portion of their police force regarding implicit bias. However, a 2017 Newsweek investigation found that no officer had received such training to that date.

[2] Rauscher, Johnson, Cho, and Walk (2008) present a meta-analysis of studies that measured the race-specific validity of survey questions about self-reported mammography use against documented sources, such as medical and billing records. They found that the specificity of survey questions that measure mammography use is lower among black women than white women.

ory would suggest that we should expect heterogeneity across these groups or categories[3] which may be policy-relevant and could be completely masked when investigating data on the full sample.

We illustrate the importance of considering the consequences of misclassification that can arise from using rules of thumb to determine categories by reevaluating if there is racial discrimination in New York Citys infamous Stop-and-Frisk program. Under this program, officers can stop and frisk anyone they believe has committed, is committing, or might commit a crime. These policing practices often disproportionately target minorities, generating signicant controversy. Details on each of over five millions stops occurring in New York City between 2003-2014 have been collected and are frequently analyzed by researchers. However, these records may also feature misclassification error in the type of crime reported as the basis for the stop, a source of bias that has been implicitly ignored prior analyses of this large administrative dataset.

Advocacy groups have long criticized the New York City Police Department's (NYPD) Stop-and-Frisk program[4] and have even suggested that it has effectively turned some neighborhoods — usually poor and nonwhite ones — into occupied territories rife with unnecessary, tense interactions between neighborhood residents and the police. More generally, across all neighborhoods, surveys increasingly document that the American law enforcement community is coming under increasing scrutiny and criticism and the levels of trust in the police have plummeted.[5] Proponents of the NYPD Stop-and-Frisk program such as former NYPD commissioner Ray Kelly claim it has saved over 7000 lives and played a key role in the city's decrease in crime over the past years. Opponents of the program claim it constitutes a violation of freedom and provides a means for officers to engage in racial profiling.

[3]In work assuming categories are measured accurately, Lehrer, Pohl, and Song (2016) use a simple static labor supply model to motivate why treatment effects from a welfare program that changes work incentives should vary across demographic groups and over the earnings distribution to motivate their empirical tests.

[4]This program has also been targeted by legislative action, including high profile cases and class action lawsuits such as Floyd, et al. v. City of New York, et al.

[5]This finding is not unique to the US, see Bradford, Jackson, and Stanko (2009) for evidence on trends in the UK.

These claims are based in part on unconditional summary statistics such as the fact that the overwhelming majority of those targeted by the program (consistently around 85% of all stops in each year) are minorities. The use of unconditional summary statistics to suggest that there is evidence of racial discrimination has been made by advocacy groups at almost every stage of the criminal justice system.[6]

Testing for discrimination in the NYPD Stop and Frisk program can be challenging since an analysis of disparate impact alone does not constitute evidence of discrimination. Knowles, Persico, and Todd (2001) (henceforth KPT) propose a hit rates test that relies on the assumption that police officers try to maximize successful searches. This test compares the productivity of stops across different racial groups. Stops that are the result of discrimination alone and not the cause of reasonable suspicion should be less likely to lead to arrests or summons, lowering the likelihood of those outcomes for that racial minority.[7] This test has been applied in prior research evaluating the NYPD Stop-and-Frisk program. Coviello and Persico (2015) find no evidence of discrimination against African-Americans in the aggregate sample of all recorded crime types in the whole city over ten years but, along with Goel, Rao, and Shroff (2016), do find evidence of discrimination against African-Americans when restricting the sample to only stops relating to the possession of a concealed weapon.[8]

We add to the existing evidence on the importance of accounting for potential heteroge-

[6]Studies that explore racial discrimination in the economics literature at other stages of the criminal justice system make clear that these may be non-discriminatory and one needs to account for racial differences in crime prevalence. Yet, even after taking this feature into account, there is evidence of racial prejudice at other stages. See Rehavi and Starr (2014), Abrams, Bertrand , and Mullainathan (2012), Anwar, Bayer, and Hjalmarsson (2012), Bushway and Gelbach (2010), Alesina and La Ferrara (2014) and Anwar and Fang (2015) for studies looking at prejudice in prosecution, bail-setting, sentencing, prison releases as well as in judges and juries.

[7]In other words, police officers who stop more members of a certain racial group would not be racially biased if these stops are productive and lead to arrests or summons. In addition, this test can account for the empirical features related to the geographic concentration of crime across neighbourhoods.

[8]For completeness, other research evaluating Stop-and-Frisk include Gelman, Fagan, and Kiss (2007) and Ridgeway (2012) who each use a different subset of the data employed in the Coviello and Persico (2015) study. Lehrer and Lepage (2018) also use the Stop-and-Frisk data to test for discrimination against Arabs and find evidence consistent with racial profiling in periods of high terrorism threat. Regarding the use of non-lethal force, Fryer (2016) finds that blacks are over 50% more likely than whites to have force used against them when stopped.

neity in bias across different types of crimes by showing its potential importance in theory and confirming it empirically. African-Americans constitute the overwhelming majority of suspects arrested for crimes related to drugs or possession of a weapon, two types of crime related to the long-lasting nationwide War on Drugs. Similarly to the NYPD's Stop-and-Frisk program, the War on Drugs has also been suggested to be discriminatory given disparate impacts for African-Americans,[9] motivating our particular focus on these crimes.

Prior research using Stop-and-Frisk data has also treated the reported crime classifications as being exogenous. However, these classifications are selected by individual police officers at the time when they complete the mandated forms indicating why they stopped a given suspect. If individual police officers have perceptions of a specific race being the perpetrators of certain types of crime,[10] these classifications might be subject to unconscious bias. In other words, conditioning on those reported crime categories leads to endogenous stratification, which is well-known to lead to biased estimates.

Prior analyses of the Stop-and-Frisk program not consider the decision we focus on which is taken by the officer at the time of choosing whom to stop and must also define the basis for the stop. While we illustrate the issue of analyzing effects on subgroups that may be misclassified with Stop-and-Frisk data, we should stress that these issues are becoming increasingly prevalent in survey data.[11] Further, misclassification has been shown to have consequences for econometric estimates (Bollinger and David, 1997, 2001), sometimes changing the overall conclusions. Generally, the consequence of misclassification depends on how it occurred. In this setting, it is not simply a measurement error problem but reflects a choice-based sample. If misclassification occurs to the same degree across racial groups and policing outcomes, then it is random and no bias should arise. Conversely, if misclassification of the subgroup

[9] See Nunn (2002) for further details.

[10] See for example Fridell (2008).

[11] For example, research has found high rates of misclassification in categorical variables such as education (Black, Sanders, and Taylor (2003)), labor market status (Poterba and Summers (1995)) and disability status (Benitez-Silva, Buchinsky, Man Chan, Cheidvasser, and Rust (2004); Kreider and Pepper (2008)); among other measures.

varies differently between racial groups, then there is non-random misclassification. Since much research analyzes impacts on subgroups, we argue that there is likely going to be an increase in misclassification of race and ethnicity variables as time progresses given changes in the meaning of these variables to survey respondents. As such, the methods we describe could be applied to various other contexts with both survey and administrative datasets.

In this paper, we first extend the model underlying the hit rates test to account for the potential that police officer bias depends on both the type of crime and the race of the suspect, thereby influencing the type of crime reported as the basis for the stop. This motivates investigating racial bias across different crime types which can conciliate evidence in Coviello and Persico (2015) of no bias in the full sample of stops. Further, it also motivates the need to correct the estimates to account for potential bias in the reporting of crime categories. Second, we reevaluate whether there is evidence of racial discrimination in New York City's Stop-and-Frisk program by modeling the selection process of crime categories as a polychotomous choice. In effect, we implement a sample-selection correction to explicitly account for the relative impact that being African-American may have on the difference in the likelihood of being stopped for certain types of crime when conducting the hit rates tests. To the best of our knowledge, this study presents the first use of a polychotomous selection model to estimate whether there is evidence of racial discrimination in the economics of crime literature.[12]

Our main empirical results indicate that with or without correcting for selective crime categorization, there is robust and conclusive evidence of discrimination. After accounting for both crime-dependent bias and selective classification, African-Americans on average are 2.874% less likely than whites to be arrested when stopped for crimes classified under the US War on Drugs; a difference of approximately 50%. Contrasting estimates from the hit rates test that both account for and ignore selective reporting of crime categories provide evidence that the correction can be important both economically and statistically. Hausman

[12] This is surprising since as Bushway, Johnson, and Slocum (2007) note, issues of selection bias pervade criminological research but econometric corrections have often been misapplied.

tests provide further evidence that this correction is statistically important.

The remainder of the paper is organized as follows. In the next section, we discuss how the Stop-and-Frisk program is implemented in New York and briefly describe the data set. To motivate our empirical tests, section 3 presents a theoretical model that extends KPT to consider crime-dependent payoffs to stopping suspects which can differ by racial group. Section 4 describes the two-step empirical strategy and discusses identification of the selection correction terms. The empirical results are presented and discussed in section 5. A final section summarizes the main findings and concludes.

2 Stop-and-Frisk: Data and Institutional Knowledge

For several decades, police officers throughout the United-States have had the authority to stop someone, ask questions and possibly frisk the suspect if they have reasonable suspicion of a possible crime on the basis of the subject's answers. There has been longstanding backlash against this form of policing which culminated with the case of Terry v. Ohio, 392 U.S. in 1968 that was decided by the United-States Supreme Court. The Supreme court ruled that a police officer must have "a reasonable suspicion" of some wrongdoing to conduct a stop.[13] Most importantly, the Supreme Court's holding required the scope of any resulting police search to be narrowly tailored to match the original reason for the stop. In response to this ruling, New York's Criminal Procedure Law (CPL) was written which authorized police officers to stop private citizens in a public place only if the officer reasonably suspected that the citizen is committing, has committed or is about to commit either a felony or a misdemeanor.[14] Once such a stop has been made, New York law further authorizes a frisk

[13] This decision creates a narrow exception to the Fourth Amendment's probable cause and warrant requirements, permitting a police officer to briefly stop a citizen, question them, and frisk them to ascertain whether they possess a weapon that could endanger the officer. By reasonable suspicion, a police officer "must be able to point to specific and articulable facts" and is not permitted to base their decision on "inchoate and unparticularized suspicion or [a] 'hunch'".

[14] See the 2000 report by the US Commission on Civil Rights (2000) for detailed information on policing in New York.

of the suspect only if the officer "reasonably suspects that they are in danger of physical injury".

While there is no legal requirement that NYPD officers record Stop-and-Frisk encounters with private citizens, the completion of a UF-250 form is common practice. Specifically, NYPD policy requires the completion of a UF-250 form in the following four circumstances: i) a person is stopped by use of force, ii) a person stopped is frisked or frisked and searched, iii) a person is arrested, or iv) a person stopped refuses to identify themselves. This potential sampling issue is noted in Coviello and Persico (2015) who discuss whether one should restrict the analysis to only stops that legally must be recorded. They conclude that imposing this restriction would require the implausible assumption that at the time of choosing whom to stop, the officer could distinguish whether the stop will develop into one that has to be recorded or not. The standard UF-250 form requires officers to document, among other things, the time, date, place, and precinct where the stop occurred; the name, address, age, gender, race, and physical description of the person stopped; factors that caused the officer to reasonably suspect the person stopped; the suspected crime that gave rise to the stop; the duration of the stop; whether the person stopped was frisked, searched or arrested; and the name, shield number, and command of the officer who performed the stop. Each police officer must submit all completed UF-250 forms to the desk officer in the precinct where the stop occurred so the stated factual basis of the "stop" for legal sufficiency can be reviewed. The data is then analyzed for quality assurance and a restricted version that suppresses the shield number and other identifying information is made publicly available.

The primary data used in this paper comprises all 5,028,789 recorded stops in the Stop-and-Frisk program between 2003 and 2014. For each stop, we are provided with all of the characteristics and outcomes listed on the UF-250 form that are permissible in the restricted dataset. In total, over 40% of observed stops in our sample did not have to be reported by law.[15] We first restrict our sample to stops involving only Caucasians or African-Americans

[15]The fraction of mandated stops is higher for African-Americans (60% versus 47%) which likely reflects higher rates of frisking, summons and force used. Our main results are robust to restricting the analysis only

and for which crime categories are recorded on the UF-250 form, yielding 2,649,300 observations. Crime classification was not reported for 2003 and inconsistently reported in both 2004 and 2005, reducing the number of observations for these earlier years.[16] Summary statistics are presented in Table 1. Approximately 84% of this sample is African-American and the vast majority of suspects are Male. Among the 13 categories of crimes reported in Table 1, which account for over 95% of recorded stops, possession of a weapon (27.67%), robbery (17.30%), trespassing (11.84%) and drugs (11.11%) are the most commonly listed as the basis for a stop. As shown in the table, we pool these crime categories into four main classifications.[17] Drugs and weapon crimes are pooled together since they represent felonies linked to the US War on Drugs and account for nearly one-third of the stops. Similarly, we pool other major crimes by whether they are either economic in motivation (i.e. trespassing, burglary, grand larceny and grand larceny auto) or violent (i.e. assault, robbery, murder and rape). The final category consists of less severe offenses which include petit larceny, graffiti and criminal misconduct. Crimes of an economic nature account for over half of the stops, while there are fewer stops associated with either violent or minor crimes.[18]

The first column of Table 2 documents the substantial heterogeneity in the average percentage of stops involving African-American suspects across crime types, which ranges from 46.4%-93.4%. African-Americans have lower stop rates than whites for graffiti and are in-

to those stops which had to be legally reported. Following Coviello and Persico (2015), we do not use this as a sampling restriction since it would condition on ex-post information. The external validity of the results rely on the plausible (yet untestable) assumption that the sample is representative of all stops in the city. If police officers under-report racially-sensitive stops, this would underestimate the number of unproductive stops and our results would constitute a lower-bound. This is consistent with the data since the estimated arrest differential is larger when restricting to this subsample.

[16] We also exclude observations which were related to other crimes than those reported in Table 1 (less than 5% of all crimes) since we group crimes within categories in our analysis. Including these crimes as an additional classification did not change the main results but led to sub substantial computational costs.

[17] Our results do not depend on those categories; the estimates are quantitatively and qualitatively similar whether we only group War on Drugs crimes and leave other crime types either ungrouped or grouped as violent and minor crimes in two categories while leaving others ungrouped. We selected these classifications ex ante and as such present them as the main results.

[18] The low rate of stops for minor crimes may reflect an effort to justify the stops by being "too ambitious" in stating suspected crimes to signal a stronger rationale for the stop. Alternatively, police officers are likely to simply put more weight on serious offenses.

volved in less than 65% of stops for burglary and criminal misconduct. In contrast, over 92% of all stops categorized as weapons, trespassing or murder involve an African-American suspect. The remaining columns provide summary information on the percentage of stops that resulted in an arrest by race and a test of whether there is a significant difference in these proportions between races. In nearly every crime category, African-Americans have on average lower rates of arrest relative to white suspects. Results from tests of equality in arrest rates between groups are presented in the last column and indicate that these differences are statistically significant at the 5% level in every category with the exception of stops for murder, rape and robbery. Most striking is that the rate of arrest for weapon possession stops is 75% higher for white suspects relative to African-Americans, despite the fact that almost 19 of every 20 stops for this category involve a black suspect.

3 Theory

We begin by extending the model that underlies the hit rates test first developed in KPT to directly consider the Supreme Court's holding in Terry v. Ohio that the scope of any resulting police search has to be narrowly tailored to match the original reason for the stop.[19] Since the crime type must be reported and police officers may associate certain racial groups with particular crime types, they may be more likely to be biased for stops related to those classes of infractions.[20] We incorporate this feature within an economic model that describes

[19]The KPT test was adapted to the Stop-and-Frisk setting in Coviello and Persico (2015), whose model we extend. Our extension is also inspired by Anwar and Fang (2006). Note that the KPT test has faced criticism in Dharmapala and Ross (2003) and Gelman et al. (2007), among others. These critiques focus on allowing police officers to consider varying degrees of severity across types of crime, allowing for the fact that officers frequently do not observe potential offenders or accounting for racial and neighborhood heterogeneity in the probability of guilt. We pursue a similar line of inquiry in further investigating heterogeneity along different types of crime, which prior work did not consider within the KPT framework.

[20]While we do not make the distinction, racial bias is likely to be unconscious in nature particularly when police officers have limited time to decide whether or not to stop a pedestrian on the street. Smith, Makarios, and Alpert (2006) provide evidence that police officers can develop unconscious biases along observable characteristics such as gender which affect their propensity to be suspicious of a member of that group under different circumstances.

pedestrian and police officer behavior.

3.0.1 Pedestrians

We categorize pedestrians by their race r and a set of other costlessly-observable characteristics o. There are $Q^{r,o}$ pedestrians in each group (r, o). A pedestrian chooses from $A_{pedestrian} = \{a_0, a_1, ..., a_n\}$ where a_0 is the outside option of not committing a crime (value of 0) and a_i with $k \neq 0$ represents committing a specific crime of type k. Suppose that ϕ_k represents the payoff from crime $k = 1, ..., n$ while c_k represents the cost of being caught for crime k which is constant across pedestrians. Also, let $F_{r,o}\left(\phi_1, ..., \phi_n\right)$ denote the joint conditional distribution of $\phi_1, ..., \phi_n$ given group (r, o) and ω denote the expected number of members of a given group that are searched.

The expected payoff of a pedestrian with type (r, o) and $(\phi_1, .., \phi_n, c_1, ..., c_n)$ who chooses to commit a crime with the expectation that ω other pedestrians of their type will be stopped is:

$$u_{r,o}(\phi_1, ..., \phi_n, c_{1,}, ..., c_n, \omega) = \max_{1 \leq k \leq n} \left\{ \phi_k - c_k \frac{\omega}{Q^{r,o}} \right\}$$

Let crime i be the maximizer of the previous expression. An agent commits a crime if $u_{r,o}\left(\phi_1, ..., \phi_n, c_{1,}, ..., c_n, \omega\right) \geq 0$:

$$P(\phi_i - c_i \frac{\omega}{N^{r,o}} \geq 0) = V_i^{r,o}\left(\phi_1, ..., \phi_n, c_{1,}, ..., c_n, \omega\right)$$

$$V_i^{r,o}\left(\omega\right) \equiv \int V_i^{r,o}\left(\phi_1, .., \phi_n, c_{1,}, ..., c_n, \omega\right) dF_{r,o}(\phi_1, ..., \phi_n |\, \phi_i = \max_{1 \leq k \leq n} \phi_k)$$

where $V_i^{r,o}\left(\omega\right)$ is the crime-dependent crime rate for group (r, o). This is the objective probability that an individual of group (r, o) is guilty of crime type i.

3.0.2 Police Officers

Given a mass M of police officers who, after having exogenously been allocated to a given precinct, receive a type $p \sim U\left[0,1\right]$.[21] Each officer type has a search capacity E_p, receives payoff $\pi_p^{r,i}$ from stopping a suspect of race r for crime i. We denote by $D(p,i)$ the additional benefit that a racially biased officer of type p gains from stopping a suspect of race A for type of crime i.[22]

Assumption I : $\pi_p^{W,i} = \pi_p^W = 1$ (normalized) and $\pi_p^{A,i} = \pi_p^W + D(p,i)$ for $i = 1, ..., n$.

Let $E_p(r,o,i)$ be the number of stops for group (r,o) from an officer of type p and $E(r,o,i) = M\int_0^1 E_p(r,o,i)dp$ be the total number of stops for group (r,o) and crime i. Defining

$$\begin{aligned} W(p,r,o,i) &= P(Guilty\ of\ crime\ i|\, r,o) \\ &= V_i^{r,o}(E(r,o,i)) \end{aligned}$$

the expected payoff of an officer of type p is then given by:

$$\sum_{r,o,i} E_p(r,o,i)\left\{\pi_p^{r,i}W(p,r,o,i) - s_p\right\}$$

where s_p is the cost of performing a stop for the officer. An officer chooses to stop an agent from group (r,o) if $\pi_p^{r,i}W(p,r,o,i) - s_p \geq 0$.

[21] The assignment of police officers across precincts is considered further in Coviello and Persico (2015) and is beyond the scope of this analysis.

[22] This is the same way through which racial bias enters in the original hit rates model from KPT. Our model differs by allowing this additional benefit to vary by crime type and therefore impact the officer's decision to stop pedestrians differently across race and crime dimensions.

3.0.3 Existence of an equilibrium

The existence of an equilibrium for such non-atomic games is established in Schmeidler (1973). For this theorem to hold in our setting, $V_i^{r,o}(\cdot)$ must be continuous and the set of officer types that strictly prefers one action to the other must be measurable for every action. The first condition holds if $F_{r,o}(\phi_{1,}...., \phi_n)$ has no atoms for each (r, o). The second condition holds if, for all w, w', r and r', the set of $p's$ such that $\pi_p^{r,i} w - s_p > \pi_p^{r'} w' - s_{p'}$ and the set of $p's$ such that $\pi_p^{r,i} w - s_p > 0$ are measurable.

3.0.4 Characterization of the equilibrium

Theorem 1 *Suppose that in equilibrium groups (r, o) and (r', o') are stopped for crime $i \in \{1, ...n\}$. If police officers are unbiased, then $V_i^{r',o'}(E(r', o')) = V_i^{r,o}(E(r, o))$. If police officers are biased against race r, then $V_i^{r',o'}(E(r', o')) > V_i^{r,o}(E(r, o))$.*

The proof for Theorem 1 is omitted as it is directly analogous to that from the classical KPT framework, but now simply defined separately for each crime type. Intuitively, as in the original model, it implies that the crime rate for a specific type of crime has to be equal across races if police officers are unbiased. In our case, this implies that the likelihood of arrest conditional on being stopped should be equal across races for a given type of crime. On the other hand, if police officers are biased against race r, then the probability of being arrested when stopped for a specific type of crime must be lower for race r. Critically, this does not rule out that police officers be more or less biased against a specific racial group for different types of crimes.

The model first highlights that, if police officers associate racial groups with certain crime types and in turn derive additional satisfaction from successful stops by racial group and type of crime, then racial bias may not be uncovered using the traditional approach of pooling all crimes together. Rather, racial bias would not only differ across race but also across crime type, requiring the hit rates test to be performed conditioning on the crime type. It also highlights that, since police officers' benefits of successful stops vary across both race and crime classification, the decision to stop a suspect for a specific type of crime may also depend on race and should be accounted for in the hit rates analysis when conditioning on the crime type.

4 Empirical Strategy

The traditional hit rates test examines whether there is a racial difference in the percentage of stops that result in an arrest. When running regressions on subgroups defined by type of crime classification, this requires that the subgroups are exogenous as assumed in both Coviello and Persico (2015) and Goel et al. (2016). This involves estimating an equation for whether a stop s involving a suspect stopped for type of crime c at time t in precinct p, resulted in an arrest. Formally,

$$ARREST_{sctp} = \beta_0 + \beta_1 RACE_s + X_{tp} + \varepsilon_{sctp} \tag{1}$$

where $RACE_s$ is an indicator variable for whether the suspect is African-American, X_{tp} is a set of year and precinct fixed effects and ε_{sctp} is an error term with zero mean. If β_1 is statistically significant, there is evidence of differential success rates of stops by race. Equation (1) corresponds to our baseline specification to investigate bias heterogeneity across types of crime.

In the theoretical model, payoff modifiers to police officers from stopping African-American suspects can differ by crime type, hence we have a choice based sample. The decision to classify a stop under a given crime category may therefore also depend on race if police officers have strong associations between race and potential crime types. Evidence of similar implicit bias has been shown to hold for violent crime (Brigham, 1971; Devine and Elliot, 1995) and more recently in a set of lab experiments. Correll, Park, Judd, and Wittenbrink (2002) show evidence that undergraduate students playing a computer simulation were more likely to misinterpret neutral objects (e.g., wallet, cell phone) as weapons and mistakenly shoot when the suspect was a black person compared to a white person. Since bias may enter in at this classification stage, it is important to account for it and go beyond an analysis which conditions on crime type. An additional rationale for this may be that police officers are "too ambitious" in stating crime types to appear that they have a good reason for the stop. This is consistent with the relatively few number of stops for minor crimes presented in table 1 even though these crime categories are likely to account for a large share of daily criminal activity.[23]

A two-step method allows us to address potential selection bias.[24] In the first step, an unordered multinomial logit model is used to explain the officer's choice of the criminal basis of the stop. Formally, for crime categories $c = 1, ..., C$, we assume logit errors and express the log-likelihood function as

$$lnL(\gamma) = \sum_{s=0}^{S} \sum_{c=0}^{C} 1\{y_s = c\} ln \frac{exp(x_s^{'} \gamma_c)}{\sum_{k=0}^{C} exp(x_s^{'} \gamma_k)} \tag{2}$$

where y_s is a categorical variable representing which type of crime presented in Table 1 is the basis of the stop and covariates include the race of the suspect along with year and precinct fixed effects with a vector Z_{ctp} used to identify the selection correction term.[25] The matrix Z_{ctp} contains measures of the fraction of stops related to each of the four crime

[23]There is also the possibility that police officers can lie about the type of crime for the basis of the stop. This could go in both directions to show effort or to not appear prejudiced. Ultimately, this may lead to bias even when adjusting for selection in crime-classification if officers consciously misreport the causes of their stops across different crime categories but the situation would be worse without the selection correction.

[24]The intuition and mechanics behind the approach we use is proposed in Bourguignon, Fournier, and Gurgand (2007) and parallel the seminal Heckman (1979) two-step estimator.

[25]Note that the inclusion of these fixed effects leads to the well-known incidental parameter problem but since the ratio of the number of observations to number of parameters is very high in our application, this suggests that issues of bias should be limited. On the other hand, by ignoring the fixed effects, the interpretation of the coefficients of the outcome equation would be unclear since a different set of coefficients enter the first stage and outcome equation. Thus, our preferred estimates given the large sample size for each precinct and year include fixed effects. For completeness, we present estimates using both approaches in the results section. The use of a conditional fixed effects estimator is computationally infeasible in our setting.

categories in a given precinct the day prior to the current stop. We argue these variables constitute a valid exclusion restriction given the likely state dependence in policing behavior at the precinct level. These are likely to be correlated to the decision of which crime type to currently stop a suspect for, but should not be related to unobserved factors that lead to individual arrests. The relevance aspect of the exclusion restriction is supported by Table 5 in which the lagged variables are all highly statistically significant. We also conduct the analysis using longer periods between the instruments and the stop date which has limited impacts on the estimates. Specifically we consider the fraction of stops related to each of the four crime categories in a given precinct the week and month prior to the current stop. With increased length, the exogeneity assumption requires that, while the recent history of policing in a precinct informs the decision to stop individuals for certain types of crime, it is not correlated to omitted factors that determine arrests. This alternative assumption is presumably less restrictive for longer lagged periods but the assumption itself remains untestable.

Using estimates of equation (2), Bourguinon et al. (2007) provide formulas to construct $C-1$ selection correction terms that are captured in the vector $\lambda_c(\Gamma)$. Adding this vector of selectivity correction terms to equation (3) generates our estimating equation

$$ARREST_{sctp} = \beta_0 + \beta_1 RACE_s + \beta_2 X_{tp} + \beta_3 \lambda_c(\Gamma) + \eta_{sctp}. \tag{3}$$

Using weighted least squares for each crime category allows us to obtain unbiased and consistent estimation of the coefficients.[26] A nice feature of this estimator is that it has been shown to perform well in correcting selection bias even in settings where the restrictive independence of irrelevant alternatives (IIA) assumption of the multinomial logit model is violated. Last, to conduct inference, we use bootstrapped standard errors to explicitly account for the two-step estimation procedure. We use the strategy proposed in Bourguinon

[26] See Bourguignon et al. (2007) for further details on constructing the selectivity correction terms and weights, which account for potential heteroskedasticity present in the model due to selectivity.

et al. (2007) to estimate the two-step model since it not only relaxes the restriction in Dubin and McFadden (1984) that all correlation coefficients add up to zero, but they present Monte Carlo evidence indicating superior performance of their estimator relative to Dubin and McFadden (1984), Lee (1983) and Dahl (2002). For completeness, Bourguinon et al. (2007) differ from Lee (1983) since the latter estimates a single selectivity effect for all choices as opposed to estimating C-1 selection terms for the C choices we consider. This approach is less restrictive since Lee (1983) requires equal covariances between the unobservables in the arrest rate equation and the unobservables which determine the crime categories and comes with the computational costs of estimating additional parameters. Dahl (2002) differs from Bourguinon et al. (2007) by the functional form used to construct the selectivity correction terms.

Examining estimates of β_3 from equation (3) also provides insight since a positive (negative) coefficient estimate indicates higher (lower) arrest rates for those stopped in this classification relative to a randomly chosen suspect that was stopped. If at least one of the $C-1$ estimates of β_3 enters in a statistically significant manner, then there is suggestive evidence of selection. To formally examine if selectivity correction leads to statistically significantly different estimates, we conduct Hausman tests.[27]

The consequence of misclassification for the analysis depends on how it occurred. The direction of bias depends on the correlation between unobservables in the outcome and selection equations. If police officers are more likely to misclassify subjects that are at high perceived risk of having committed a crime, we would expect that ignoring the selection correction would underestimate the effect of racial differences. After all, when a decision to make a stop occurs rapidly, police officers are more likely to use any implicit bias based on the suspect's characteristics when choosing crime classifications. In other words, where police officers are less certain of the exact crime at the time they make the stop and they are relying on criminal offender profiling to select the crime category, we would underestimate

[27] Bootstrapped Hausman tests would be preferable since they relax the assumption that OLS be fully efficient under the null but are computationally unfeasible in our setting.

the effect of racial discrimination.[28]

5 Results

We first present estimates of the hit rates test across different types of crime in Table 3. The columns of the table differ based on the level of fixed effects that are included. An important point from Coviello and Persico (2015) is that, as shown in the first row of Table 3, once accounting for time and precinct fixed effects in columns 6-7, there is no evidence of discrimination when pooling all crimes together. Further, the inclusion of precinct fixed effects leads to an approximate 50% reduction in the magnitude of race on arrest rates. Rows 2-4 present results by subgroups of different crime classifications as previously defined and shows that the previous result conflates vastly different effects into one which creates the false appearance of no arrest differential. The results indicate that African Americans are significantly less likely to be arrested when stopped for crimes related to the War on Drugs but significantly more likely to be arrested when stopped for other economic crimes with little estimated differential for violent or minor crimes. This is consistent with the conjecture that potential police officer bias differs by crime type which leads to inefficient policing. Further, we note that adding extra pedestrian and stop characteristics to Equation (1) as shown in column 7 has little incidence on the results.

Table A2 and Figure A1 in the appendix also show that the estimated arrest differential for War on Drugs crimes is present in every borough in the city (though there is important heterogeneity) and has increased consistently over the period considered in our sample. Table A1 in the appendix shows that, in the case of summons, the outcome differentials by type of crime are more reflective of the aggregate regression as it is estimated that African Americans

[28] Whenever making statistical corrections for selection bias or endogeneity, there is always the risk that the cure may be worse than the disease (see Bound, Jaeger, and Baker (1995)). That said, the literature on policing (e.g. James (2017) and the references within) suggest that implicit bias may be higher for weapon crimes, which is accounted for by our correction and consistent with our results. As such, it appears unlikely that these differences in categories of crime would be solely due to chance and our selection correction appears to be operating in the desired direction.

are less likely to be issued a summons when stopped for any crime group.

Next, we investigate whether these results may be partly influenced by the endogenous decision of police officers of which type of crime to report. To examine if there is selective classification of crime type in the NYPD Stop-and-Frisk program, Table 4 presents estimates of β_1 from equations (1) and (3) for each crime category defined in Table 1. These specifications include additional pedestrian and stop characteristics which may also define the selective classification of stops. As shown in Table A3 of the appendix, the number of stops differs across racial groups but also various other characteristics which are likely correlated.

Estimates of the selection-correction model in the second column are noticeably different in economic significance from estimates using the standard hit rates presented in column 1. While African-Americans are statistically less likely at the 1% level to be arrested when stopped for War on Drugs related crimes irrespective of whether crime categories are exogenous or a behavioral choice, the estimated coefficient is roughly 15% larger than that which ignores selectivity. For other crime categories, the difference between the two methods are also important both in magnitude and statistical significance. Making corrections for selective crime classifications leads to a 20% reduction in the magnitude of the race coefficient for economic crimes and large changes in magnitude for violent and minor crimes.

While our adjusted estimates do not alter the overall conclusion of racial discrimination for War on Drugs crimes, the estimates obtained from the two-stage procedure do suggest that there is non-negligible sample-selection. The last column in Table 4 reports the p-values from Hausman specification tests of the equality of the estimated coefficient on black between estimates of equations (1) and (3). For War on Drugs, we observe that the p-values from the Hausman tests are less than 0.01, indicating that we can safely reject the assumption that crime categories are exogenous. Similarly, we can clearly reject that crime categories do not reflect a behavioral choice for major economic and violent crimes and for minor crimes at the 6% level. The results provide evidence that the choice of crime that officers report as the basis for individual stops generates endogenous stratification. Table A4 in the appendix applies the two-stage correction in the case of summons and finds that we can reject the hypothesis that sample-selection is negligible for all crime categories.

Marginal effect estimates from the first stage crime classification selection are presented in Table 5 of the appendix. Each of the variables used to identify the selection correction terms in equation (3) are individually and jointly statistically significant with a plausible sign and magnitude. A somewhat striking finding is that African-Americans are statistically significantly more likely to have their stop categorized as a War on Drugs crime. Estimates of equation (2) also find that blacks are significantly less likely to have their stop categorized as other crime types. Since the categories underlying War on Drugs crimes can be viewed as representing police officer speculation that a suspect is either hiding a weapon or drugs as opposed to having committed a robbery or trespassing, it is likely that they are easier to use to justify a stop. Thus, when an officer decides to instantly make a stop based on the suspect's characteristics, the use of this crime classification may also partially reflect implicit bias. Thus, it is not surprising that the estimated effect of racial discrimination on arrest rates for War on Drug crimes increases once the selection correction is used.

Last, we conducted a series of robustness checks shown in Table A5 to investigate how the results of the sample-correction procedure vary depending on various assumptions. We find that using different lagged values of the share of stops related to each crime category as the exclusion restriction, which improves the plausibility of the exogeneity assumption, has little incidence on the conclusion. We also find that excluding first stage fixed effects from the correction does alter the estimates of the correction but does not change the conclusion that there is a large arrest differential between racial groups for War on Drugs crimes.

6 Conclusion

The NYC Stop-and-Frisk program often plays a prominent role in debates surrounding racial profiling. Analyses of this data which condition on reported crime type are necessary to uncover heterogeneity in bias but may lead to biased estimates due to endogenous stratification. This stratification may arise since individual police officers could possess unconscious biases, raising concerns that race not only influences whether a police officer decides to stop a pedestrian but also extends to which crime is reported as the basis of the stop. In this paper, we extend the original model underlying KPT to include police officer bias which can differ across types of crime. We show that the traditional hit rates test can be modified to incorporate a selection-correction term from a polychotomous choice model to account for this consideration. Second, using data from the Stop-and-Frisk program, we show that this correction is empirically important in numerous situations since it reduces the estimated race differentials by 20% - 35%.

However, even after applying the selection correction, we concur with prior research that there is strong and robust evidence of discrimination against African-Americans for weapon and drugs related crimes. Since numerous variables measured in both survey and administrative data sets may contain similar non-sampling errors due to misclassification, the econometric methods utilized in this paper could be used in other contexts to ensure that conclusions are not distorted by this source of bias when conducting analyses on endogenous subgroups. In conclusion, this paper demonstrates the importance of researchers accounting for institutional features that guide police officer behavior in the field by illustrating the challenges that arise when applying tests for racial bias across crime types and locations.

More generally, trends related to social movements calling for fundamental changes in the way the federal statistical system classifies people by race and gender may increase the risk of misclassification in both survey and administrative datasets.[29] Recently, the U.S. federal Office of Management and the Budget passed legislation ensuring that the Census 2000 and subsequent federal statistical documents must allow individuals to identify with as many of the major races as they wish. The U.S. Census Bureau introduced a new race

[29] As Williams (2006) points out, some leading advocates of this change in the United States were white women married to African-American men who found that their children were almost always classified as black by those who collected statistical data or tabulated persons by race.

category in the Census and a Census Bureau interview-reinterview survey in 2000 found that those who identify with multiple races are not consistent in their reporting of race. Further, there was significant heterogeneity in reporting multiple races across both geographic and demographic backgrounds. Thus, the tools we illustrate in our application may be important for future research that considers generating policy-relevant evidence from analyses carried out on subgroups defined by potentially misclassified variables.

References

[1] Abrams, D.S., Bertrand, M., and Mullainathan, S., 2012. Do Judges Vary in Their Treatment of Race? *Journal of Legal Studies*, 41(2), pp. 347–383.

[2] Alesina, A., and La Ferrara, E., 2014. A Test for Racial Bias in Capital Punishment. *The American Economic Review*, 104(11), pp. 3397–3433.

[3] Anwar, S., Bayer, P., and Hjalmarsson, R., 2012. The Impact of Jury Race in Criminal Trials. *Quarterly Journal of Economics*, 127(2), pp. 1017-1055

[4] Anwar, S. and Fang, H., 2015. Testing for Racial Prejudice in the Parole Board Release Process: Theory and Evidence. *Journal of Legal Studies*, 44(1), pp. 1-37.

[5] Anwar, S. and Fang, H., 2006. An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence. *The American Economic Review*, 96(1), pp. 127-151.

[6] Benitez-Silva, H., Buchinsky, M., Man Chan, H., Cheidvasser, S. and Rust, J., 2004. How Large is the Bias in Self-reported Disability?. *Journal of Applied Econometrics*, 19(6), pp.649-670.

[7] Black, D., Sanders, S. and Taylor, L., 2003. Measurement of Higher Education in the Census and Current Population Survey. *Journal of the American Statistical Association*, 98(463), pp.545-554.

[8] Bollinger, C.R. and David, M.H., 1997. Modeling Discrete Choice with Response Error: Food Stamp Participation. *Journal of the American Statistical Association*, 92(439), pp.827-835.

[9] Bollinger, C.R. and David, M.H., 2001. Estimation with Response Error and Nonresponse: Food-Stamp Participation in the SIPP. *Journal of Business & Economic Statistics*, 19(2), pp.129-141.

[10] Bound, J., Brown, C. and Mathiowetz, N., 2001. Measurement Error in Survey Data. In *Handbook of Econometrics*, Vol. 5, pp. 3705-3843. Elsevier.

[11] Bound, J., Jaeger, D.A. and Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American statistical association, 90(430), pp.443-450.

[12] Bourguignon, F., Fournier, M. and Gurgand, M., 2007. Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons. *Journal of Economic Surveys*, 21(1), pp. 174-205.

[13] Bradford, B., Jackson, J. and Stanko, E., 2009. Contact and confidence: Revisiting the Impact of Public Encounters With the Police. *Policing and Society*, 19(1). pp. 20-46.

[14] Brigham, J.C., 1971. Racial Stereotypes, Attitudes, and Evaluations of and Behavioral Intentions Toward Negroes and Whites. *Sociometry*, pp.360-380.

[15] Bushway, S.D., and Gelbach, J. B., 2010. Testing for Racial Discrimination in Bail Setting Using Nonparametric Estimation of a Parametric Model. *mimeo.* Yale Law School.

[16] Bushway, S.D., Johnson, B.D. and Slocum, L.A., 2007. Is the Magic Still There? The Relevance of the Heckman Two-step Correction for Selection Bias in Criminology. *Journal of Quantitative Criminology,* 23(2), pp. 151–178.

[17] Coviello, D. and Persico, N., 2015. An Economic Analysis of Black-White Disparities in NYPD's Stop-and-Frisk Program. *Journal of Legal Studies*, 44(2), pp. 315-360.

[18] Correll, J., Park, B., Judd, C.M. and Wittenbrink, B., 2002. The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals. *Journal of Personality and Social Psychology*, 83(6), p.1314.

[19] Dahl, G.B., 2002. Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica*, 70(6), pp. 2367-2420.

[20] Devine, P.G. and Elliot, A.J., 1995. Are Racial Stereotypes Really Fading? The Princeton Trilogy Revisited. *Personality and Social Psychology Bulletin*, 21(11), pp.1139-1150.

[21] Dharmapala, D. and Ross, S.L., 2004. Racial Bias in Motor Vehicle searches: Additional theory and evidence. *Contributions to Economic Analysis & Policy*, 3(1), Article 12.

[22] Dubin, J.A. and McFadden, D.L., 1984. An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. *Econometrica*, 52(2), pp.345-362.

[23] Fridell, L.A., 2008. Racially Biased Policing: The Law Enforcement Response to the Implicit Black-Crime Association. In Lynch, M.J., Patterson E.B., and Childs K. K. (Eds.). *Racial Divide: Racial and Ethnic Bias in the Criminal Justice System*. Monsey NY: Criminal Justice Press, pp. 39-59.

[24] Fridell, L. and Lim, H., 2016. Assessing the Racial Aspects of Police Force Using the Implicit-and Counter-bias Perspectives. *Journal of Criminal Justice*, 44, pp.36-48.

[25] Fryer Jr, R.G., 2016. An Empirical Analysis of Racial Differences in Police use of Force. *NBER Working Paper No. 22399.*

[26] Gelman, A., Fagan, J. and Kiss, A., 2007. An Analysis of the New York City Police Department's Stop-and-Frisk Policy in the Context of Claims of Racial Bias. *Journal of the American Statistical Association,* 102(479), pp. 813-823.

[27] Goel, S., Rao, J.M. and Shroff, R., 2016. Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy. *The Annals of Applied Statistics,* 10(1), pp. 365-394.

[28] Hausman, J.A., 1978. Specification Tests in Econometrics. *Econometrica*, 46(6) pp. 1251-1271.

[29] Heckman, J.J., 1979. Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), pp. 153-161.

[30] James, L., 2017. The Stability of Implicit Racial Bias in Police Officers. *Police Quarterly*, Vol 21, Issue 1, pp. 30-52.

[31] Knowles, J., Persico, N. and Todd, P., 2001. Racial Bias in Motor Vehicle Searches: Theory and Evidence. *Journal of Political Economy*, 109(1), pp. 203-229.

[32] Kreider, B. and Pepper, J.V., 2011. Identification of Expected Outcomes in a Data Error Mixing Model with Multiplicative Mean Independence. *Journal of Business & Economic Statistics*, 29(1), pp.49-60.

[33] Lee, L.F., 1983. Generalized Econometric Models With Selectivity. *Econometrica*, 51(2), pp.507-512.

[34] Lehrer, S. F. and Lepage, L., 2018. How Do NYPD Officers Respond to General and Specific Terror Threats?, *mimeo,* University of Michigan.

[35] Lehrer, S.F., Pohl, R.V. and Song, K., 2016. Targeting Policies: Multiple Testing and Distributional Treatment Effects. *NBER Working Paper No. 22950.*

[36] Nunn, K.B., 2002. Race, Crime and the Pool of Surplus Criminality: or Why the War on Drugs was a War on Blacks. *Journal of Gender, Race & Justice*, 6, pp. 381-445.

[37] Poterba, J.M. and Summers, L.H., 1995. Unemployment Benefits and Labor Market Transitions: A Multinomial Logit Model with Errors in Classification. *The Review of Economics and Statistics*, pp.207-216.

[38] Rauscher, G.H., Johnson, T.P., Cho, Y.I. and Walk, J.A., 2008. Accuracy of Self-reported Cancer-screening Histories: A Meta-analysis. Cancer Epidemiology and Prevention Biomarkers, 17(4), pp.748-757.

[39] Rehavi, M.M., and Starr, S., 2014. Racial Disparities in Federal Criminal Sentences. *Journal of Political Economy* 122(6), 1320–1354.

[40] Ridgeway, G., 2007. Analysis of Racial Disparities in the New York Police Department's Stop, Question, and Frisk Practices. *RAND Technical Report #534*.

[41] Schmeidler, D., 1973. Equilibrium Points of Nonatomic Games. *Journal of Statistical Physics*, 7(4), pp. 295-300.

[42] Sit, R., 2017. Since Eric Garner's Death, Not One NYPD Officer Has Received Implicit Bias Training, Despite What The Mayor Says. Newsweek. Available at http://www.newsweek.com/eric-garner-erica-nypd-implicit-bias-bill-de-blasio-765165 [Date Published: 12/29/2017] [Date Accessed: 02/21/2018]

[43] Smith, M.R., Makarios, M. and Alpert, G.P., 2006. Differential Suspicion: Theory Specification and Gender Effects in the Traffic Stop Context. *Justice Quarterly*, 23(2), pp. 271-295.

[44] U.S. Commission on Civil Rights, 2000. Chapter 5: Stop, Question, and Frisk. Police Practices and Civil Rights in New York City. Available online at *http://www.usccr.gov/pubs/nypolice/main.htm.*

[45] Williams, K.M., 2008. Mark One or More: Civil Rights in Multiracial America. University of Michigan Press. Ann Arbor, Michigan.

Table 1: Summary Statistics

	Mean	Standard Deviation.
Outcomes		
Arrest Rate*100	5.49	(22.78)
Summons*100	6.14	(24.01)
Demographics		
Suspect is Black	84.55	(36.15)
Suspect is Male	92.84	(25.77)
Age of Suspect	28.34	(12.39)
Suspect is Youth	54.13	(49.83)
Suspect is Tall	23.61	(42.47)
Suspect has Heavy Build	8.52	(27.91)
Stops		
Number made at Night	60.00	(48.99)
Mandated Stops	58.52	(49.27)
Crimes		
War on Drugs	38.88	(48.75)
Drugs	11.21	(31.55)
Weapon	27.66	(44.73)
Other Economic Crimes	35.33	(47.80)
Trespassing	12.11	(32.63)
Burglary	9.25	(28.98)
Grand Larceny	4.49	(20.70)
Grand Larceny Auto	9.48	(29.30)
Violent Crimes	20.76	(40.57)
Assault	3.42	(18.17)
Robbery	17.21	(37.75)
Murder	0.05	(2.20)
Rape	0.10	(3.19)
Minor Offenses	5.02	(21.83)
Petit Larceny	2.62	(15.98)
Graffiti	1.16	(10.70)
Criminal Misconduct	1.24	(11.06)
Observations	2,399,717	

Standard deviations in parentheses. Youth refers to the fraction of suspects aged below 25. Tall refers to the fraction of suspects six feet or taller. Heavy build refers to the fraction of suspects classified as heavy build by the NYPD. Night refers to the fraction of stops performed between 7PM and 6AM. Other economic crimes refers to non-violent crimes including trespassing, burglary, grand larceny and grand larceny auto. Violent refers to violent crimes including rape, murder and assault. Minor refers to minor crimes and includes petit larceny, graffiti and criminal misconduct.

Table 2: Race Differentials by Crime Classification

	% Black	% Arrested Black	% Arrested White	Test of Equality
Overall	0.840	5.864	6.306	0.000
	(0.367)	(23.495)	(24.307)	
Drugs	0.839	10.484	11.927	0.000
	(0.367)	(30.634)	(32.410)	
Weapon	0.935	4.073	7.442	0.000
	(0.247)	(19.767)	(26.246)	
Trespassing	0.936	7.810	10.459	0.000
	(0.245)	(26.833)	(30.604)	
Burglary	0.630	3.794	2.701	0.000
	(0.483)	(19.105)	(16.212)	
Robbery	0.902	3.511	3.543	0.727
	(0.298)	(18.405)	(18.488)	
Assault	0.818	10.864	13.888	0.000
	(0.386)	(31.119)	(34.583)	
Murder	0.921	5.053	7.217	0.358
	(0.270)	(21.914)	(26.011)	
Rape	0.883	6.258	6.980	0.851
	(0.322)	(24.226)	(23.751)	
Grand Larceny	0.800	5.359	4.738	0.000
	(0.400)	(22.520)	(21.246)	
Grand Larceny Auto	0.699	4.216	3.528	0.000
	(0.459)	(20.095)	(18.450)	
Petit Larceny	0.802	8.815	9.431	0.026
	(0.398)	(28.352)	(29.226)	
Graffiti	0.464	4.077	5.445	0.000
	(0.499)	(19.777)	(22.691)	
Criminal Misconduct	0.646	7.542	8.288	0.019
	(0.478)	(26.408)	(27.571)	

Standard deviations in parentheses The last column presents the p-values from tests where the null hypothesis is that of equality in the probability of arrest between African-Americans and whites.

Table 3: Estimates of the Hit Rates Test on Arrests, Overall and by Crime Type

Model	OLS	OLS	OLS	FE	FE	FE	FE
	1	2	3	4	5	6	7
Black							
All Crimes	-0.386***	-0.386***	-0.386	0.164***	0.142***	0.142	0.221
	(0.0416)	(0.0416)	(0.485)	(0.0540)	(0.0540)	(0.211)	(0.204)
War on Drugs	-3.726***	-3.713***	-3.713***	-2.584***	-2.597***	-2.597***	-2.463***
	(0.102)	(0.102)	(0.635)	(0.119)	(0.119)	(0.498)	(0.497)
Other Economic	1.696***	1.704***	1.704***	1.794***	1.771***	1.771**	1.821***
	(0.0521)	(0.0522)	(0.495)	(0.0715)	(0.0715)	(0.219)	(0.222)
Violent	-3.726***	-1.689***	-1.689***	-0.00628	-0.0483	-0.0483	-0.0267
	(0.102)	(0.106)	(0.578)	(0.130)	(0.130)	(0.273)	(0.267)
Minor	0.265	0.264	0.264	0.342*	0.353*	0.353	0.0595
	(0.163)	(0.163)	(0.966)	(0.207)	(0.207)	(0.494)	(0.511)
Clustered SE	no	no	yes	no	no	yes	yes
Time FE	no	yes	yes	no	yes	yes	yes
Precinct FE	no	no	no	yes	yes	yes	yes
Extra Controls	no	no	no	no	no	no	yes

Standard errors are presented in parentheses. *0.1 **0.05 ***0.01. The dependent variable is the probability of being arrested conditional on being stopped and is multiplied by 100. Extra controls refer to the inclusion of indicators for gender, youth, suspect height and build as well as time of day as defined in Table 1.

Table 4: Estimates of Racial Discrimination for Arrests by Crime Category that Both Ignore or Account for Selective Crime Categorization

Model	Hit Rates Test Exogenous Crime Categories	Hit Rates Test That Corrects for Police officer Selection of Crime Categories	Hausman
War on Drugs Crimes			
Black	-2.463***	-2.874***	0.0032
	(0.121)	(0.184)	
Major Economic Crimes			
Black	1.821***	1.446***	0.0000
	(0.072)	(0.113)	
Violent Crimes			
Black	-0.027	1.389***	0.0000
	(0.132)	(0.221)	
Minor Crimes			
Black	0.060	-0.632	0.0521
	(0.211)	(0.414)	

The dependent variable is the probability of being arrested conditional on being stopped and is multiplied by 100. *0.1 **0.05 ***0.01. Robust standard errors in parentheses for column 1, bootstrapped (1000 repetitions) and reweighted standard errors in parentheses for column 2. Specifications additionally include fixed effects for precincts and years as well as indicators for gender, youth, height, build and time of day. Column 3 presents the p-values from Hausman specification tests where the null hypothesis is that the estimated coefficient on black is the same across models from columns 1 and 2.

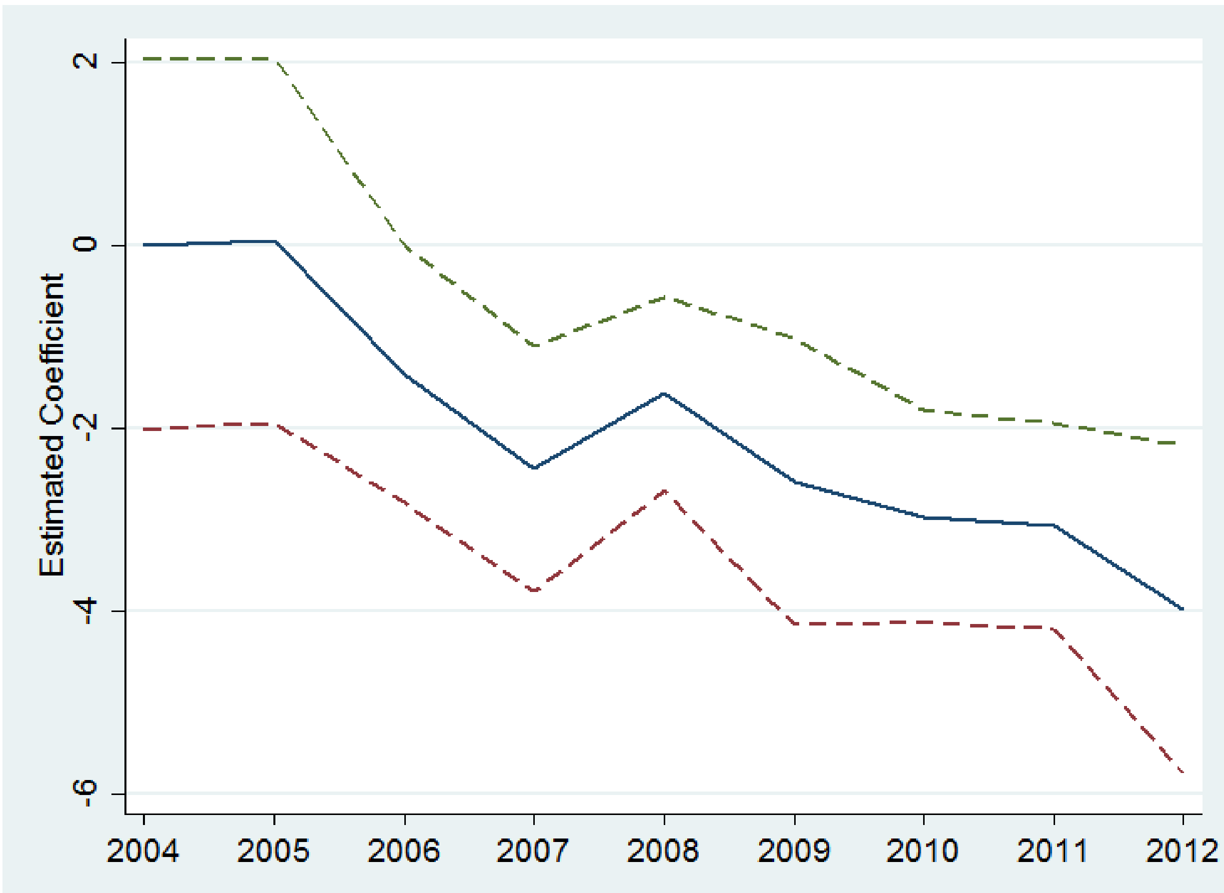
Table 5: Marginal Effects Estimates from Multinomial logit Estimation of First Stage Crime Classification Selection Equation

	War on Drugs	Major	Violent	Minor
Black	0.028***	-0.105***	0.097***	-0.020***
	(0.001)	(0.001)	(0.001)	(0.001)
Lag War on Drugs	0.139***	-0.093***	-0.038***	-0.008***
	(0.003)	(0.002)	(0.002)	(0.001)
Lag Other Economic	-0.083***	0.113***	-0.024***	-0.006***
	(0.003)	(0.002)	(0.002)	(0.001)
Lag Violent	-0.048***	-0.055***	0.112***	-0.008***
	(0.003)	(0.003)	(0.002)	(0.001)
Lag Minor	-0.014***	-0.002***	-0.009***	0.025***
	(0.004)	(0.004)	(0.003)	(0.001)
N. of Observations	2,399,717			
Pseudo R-squared	0.0706			
P-value H0: Exc. Restrictions = 0	0.0000			

The dependent variable is an aggregate of crime types which takes the value 1 for crimes related to the War on Drugs, 2 for major non-violent crimes, 3 for violent crimes and 4 for minor crimes. *0.1, **0.05 ***0.01. Standard errors in parentheses. Other Economic refers to economic crimes including trespassing, burglary, robbery, grand larceny and grand larceny auto. Violent refers to violent crimes including rape, murder and assault. Minor refers to minor crimes and includes petit larceny, graffiti and criminal misconduct. Lag crime variables are defined as the proportion of stops that involved crimes of that type in the day before the stop in the same precinct. The exclusion restrictions p-value refers to a joint test of significance for the four exclusion restrictions.

Appendix

Figure A1: Estimated Arrest Differential by Year, War on Drugs Crimes



The estimated coefficients correspond to the coefficient on black from a set of regressions of the probability of being arrested (multiplied by 100) conditional on being stopped on a dummy for black as well as precinct indicators estimated separately for each year. The standard errors are clustered at the precinct level. The outcome is multiplied by 100. The dashed lines represent the pointwise 95% confidence interval.

Table A1: Estimates of the Hit Rates Test on Summons, Overall and by Crime Type

Model	OLS	OLS	OLS	FE	FE	FE	FE
	1	2	3	4	5	6	7
Black							
All Crimes	0.056	0.0794*	0.0794	-1.760***	-1.746***	-1.746***	-1.545***
	(0.0427)	(0.0428)	(0.379)	(0.0527)	(0.0527)	(0.330)	(0.322)
War on Drugs	-0.922***	-0.886***	-0.886	-2.468***	-2.449***	-2.449***	-2.104***
	(0.103)	(0.103)	(0.619)	(0.122)	(0.122)	(0.498)	(0.499)
Other Economic	-0.922***	0.163***	0.163	-1.179***	-1.176***	-1.176***	-1.090***
	(0.103)	(0.0578)	(0.336)	(0.0744)	(0.0744)	(0.239)	(0.235)
Violent	-0.922***	-0.932***	-0.932**	-1.953***	-1.962***	-1.962***	-1.845***
	(0.103)	(0.0946)	(0.386)	(0.109)	(0.109)	(0.287)	(0.262)
Minor	-0.922***	-2.639***	-2.639***	-2.281***	-2.265***	-2.265***	-2.047***
	(0.103)	(0.132)	(0.900)	(0.148)	(0.148)	(0.423)	(0.371)
Clustered SE	no	no	yes	no	no	yes	yes
Time FE	no	yes	yes	no	yes	yes	yes
Precinct FE	no	no	no	yes	yes	yes	yes
Extra Controls	no	no	no	no	no	no	yes

Standard errors are presented in parentheses. *0.1 **0.05 ***0.01. The dependent variable is the probability of being issues a summons conditional on being stopped and is multiplied by 100. Extra controls refer to the inclusion of indicators for gender, youth, suspect height and build as well as time of day as defined in Table 1.

Table A2: Estimates of the Hit Rates Test on Arrests, by Crime Type and Borough

Model	OLS	OLS	OLS	FE	FE	FE	FE
	1	2	3	4	5	6	7
Black							
Manhattan	-4.559***	-4.521***	-4.521**	-3.544***	-3.471***	-3.471**	-3.299*
	(0.233)	(0.233)	(2.088)	(0.251)	(0.251)	(1.589)	(1.604)
Bronx	-1.985***	-2.039***	-2.039**	-1.905***	-1.944***	-1.944**	-1.574**
	(0.278)	(0.278)	(0.782)	(0.295)	(0.295)	(0.744)	(0.634)
Brooklyn	-3.109***	-3.115***	-3.115***	-1.582***	-1.631***	-1.631***	-1.493***
	(0.168)	(0.168)	(0.578)	(0.187)	(0.187)	(0.494)	(0.489)
Queens	-3.456***	-3.549***	-3.549***	-3.313***	-3.338***	-3.338***	-3.250***
	(0.254)	(0.255)	(0.997)	(0.307)	(0.308)	(0.956)	(0.934)
Staten Island	-1.204***	-0.918***	-0.918	-2.721***	-2.536***	-2.536**	-2.645**
	(0.283)	(0.288)	(1.154)	(0.373)	(0.373)	(0.583)	(0.383)
Clustered SE	no	no	yes	no	no	yes	yes
Time FE	no	yes	yes	no	yes	yes	yes
Precinct FE	no	no	no	yes	yes	yes	yes
Extra Controls	no	no	no	no	no	no	yes

Standard errors are presented in parentheses. *0.1 **0.05 ***0.01. The dependent variable is the probability of being arrested conditional on being stopped and is multiplied by 100. Extra controls refer to the inclusion of indicators for gender, youth, suspect height and build as well as time of day as defined in Table 1.

Table A3: Reason for Stop and Pedestrian Characteristics

	Overall	War on Drugs	Other Economic Crimes	Violent Crimes	Minor Offenses	Test of Equality
% Black	0.845	0.908	0.775	0.887	0.688	0.0000
	(0.361)	(0.289)	(0.418)	(0.316)	(0.463)	
% Male	0.929	0.945	0.905	0.945	0.896	0.0000
	(0.258)	(0.228)	(0.293)	(0.228)	(0.305)	
% Youth	0.541	0.575	0.455	0.634	0.503	0.0000
	(0.498)	(0.494)	(0.498)	(0.482)	(0.500)	
% Night	0.600	0.658	0.573	0.569	0.468	0.0000
	(0.490)	(0.474)	(0.495)	(0.495)	(0.499)	
% Tall	0.236	0.245	0.231	0.235	0.216	0.0000
	(0.425)	(0.430)	(0.421)	(0.424)	(0.412)	
% Heavy	0.085	0.086	0.089	0.078	0.084	0.0000
	(0.279)	(0.281)	(0.284)	(0.267)	(0.277)	

Standard deviations in parentheses. The last column presents the p-values from tests where the null hypothesis is that of equality in the probability of stop for the four crime categories.

Table A4: Estimates of Racial Discrimination for Summons by Crime Category that Both Ignore or Account for Selective Crime Categorization

Model	Hit Rates Test Exogenous Crime Categories	Hit Rates Test That Corrects for Police officer Selection of Crime Categories	Hausman
War on Drugs Crimes			
Black	-2.104***	-2.278***	0.0197
	(0.124)	(0.171)	
Major Economic Crimes			
Black	-1.090***	-1.660***	0.0000
	(0.076)	(0.369)	
Violent Crimes			
Black	-1.845***	-2.109***	0.0000
	(0.111)	(0.439)	
Minor Crimes			
Black	-2.047***	-2.252***	0.0053
	(0.150)	(0.728)	

The dependent variable is the probability of being issued a summons conditional on being stopped and is multiplied by 100. *0.1 **0.05 ***0.01. Robust standard errors in parentheses for column 1, bootstrapped (1000 repetitions) and reweighted standard errors in parentheses for column 2. Specifications additionally include fixed effects for precincts and years as well as indicators for gender, youth, height, build and time of day. Column 3 presents the p-values from Hausman specification tests where the null hypothesis is that the estimated coefficient on black is the same across models from columns 1 and 2.

Table A5: Robustness Checks for Sample Correction Estimates

Model	Exclusion restriction: 1 week	Exclusion restriction: 1 month	Correction Method: Lee (1983)	Correction Method: Dahl (2002)	First Stage: No Fixed Effects
War on Drugs Crimes					
Black	-3.011	-3.388	-2.293	-2.201	-2.305
	(0.169)	(0.226)	(0.104)	(0.188)	(0.138)
Major Economic Crimes					
Black	0.975	0.208	1.459	0.748	1.385
	(0.134)	(0.190)	(0.088)	(0.150)	(0.105)
Violent Crimes					
Black	1.319	0.691	1.266	1.529	0.930
	(0.198)	(0.323)	(0.154)	(0.204)	(0.155)
Minor Crimes					
Black	-0.331	0.183	-0.529	-1.531	-0.458
	(0.475)	(0.772)	(0.321)	(0.426)	(0.292)

The dependent variable is the probability of being arrested conditional on being stopped and is multiplied by 100. Robust standard errors in parentheses. Specifications additionally include fixed effects for precincts and years as well as indicators for gender, youth, height, build and time of day.