

An Economic Analysis of Black-White Disparities in the New York Police Department's Stop-and-Frisk Program

Decio Coviello and Nicola Persico

ABSTRACT

We introduce a model to explore the identification of two distinct sources of bias in the New York Police Department's stop-and-frisk program: the police officer making the stop decisions and the police chief allocating personnel across precincts. We analyze 10 years of data from the stop-and-frisk program in light of this theoretical framework. We find that white pedestrians are slightly less likely than African American pedestrians to be arrested conditional on being stopped. We interpret this finding as evidence that the officers making the stops are on average not biased against African Americans relative to whites, because the latter are stopped despite being a less productive stop for a police officer. We find suggestive evidence of police bias in the decision to frisk. Further research is needed.

1. INTRODUCTION

New York City's stop-and-frisk program is a police strategy whereby pedestrians are briefly stopped by police officers, engaged in conversation or questioning, and potentially searched. The procedure is perceived as demeaning for those who are stopped and/or frisked. The program also disproportionately impacts nonwhites. The racial impact of the program

DECIO COVIELLO is Associate Professor of Economics at HEC Montréal. NICOLA PERSICO is the John L. and Helen Kellogg Professor of Managerial Economics and Decision Sciences at Northwestern University. We thank Matthew Bloch for sharing with us the New York City electoral data and Tim Brophy for helping us in mapping the data onto police precincts. This paper is an update of Coviello and Persico (2013). We offer access to the replication material to academic faculty and graduate students. Please contact the first-named author for access.

[*Journal of Legal Studies*, vol. 44 (June 2015)]

© 2015 by The University of Chicago. All rights reserved. 0047-2530/2015/4402-0022\$10.00

has given rise to public protests¹ and widespread allegations of racial profiling.² Similar programs exist in many other jurisdictions.³

The stop-and-frisk program has been repeatedly challenged in court. In the most recent such lawsuit, US District Judge Shira Scheindlin noted that at issue was “the disproportionate number of African Americans and Latinos, as compared to whites, who become entangled in the criminal justice system” (Klasfeld 2012). The case, *Floyd v. City of New York* (959 F. Supp. 2d 540 [S.D.N.Y. 2013]), was decided against the New York Police Department (NYPD) on August 12, 2013, with the judge finding a violation of equal protection rights.⁴ The decision was front-page news in major US newspapers and generated a lively public debate, with Mayor Michael Bloomberg accusing the judge of blocking the chance for the city to have “a fair trial” and vowing to appeal the decision (Goldstein 2013) and the *New York Times* (2013) editorial board supporting the decision. This judicial decision is likely to have repercussions on police activity, and perhaps on crime,⁵ in other jurisdictions.

In this paper we examine the same stop-and-frisk data that were analyzed in the trial. Our aim is not to audit the judge’s opinion from a legal perspective. Rather, we inquire as to whether the stop-and-frisk program is administered in a racially biased way from the perspective of social science. From this perspective, legal tests and rules of evidence are not cen-

1. On March 16, 2012, several thousand people marched in New York City protesting the policy, which organizers of the demonstration assert “creates an atmosphere of martial law for the city’s black and Latino residents” (see Leland and Moynihan 2012).

2. According to the *New York Post*, “The stop-and-frisk policy has been under fire by vocal opponents, including many 2013 mayoral contenders, because the vast majority of people stopped are black or Hispanic” (Celona 2012). Sharpton (2012) writes, “When a majority of those targeted by police are young men of color and when the bulk of them are innocent, what else are we to conclude other than the fact that the NYPD has been implementing a policy of racial profiling and discrimination?”

3. Other major cities collect records of similar activities by their police. In Chicago police officers are required to file a contact card for every interview or pat down. These records include details about the person stopped and a description of the reason for the stop. In Los Angeles, police officers are required to complete field data reports when stopping a vehicle or a pedestrian.

4. The judge also found that the stop-and-frisk program gave rise to Fourth Amendment violations (unreasonable searches and seizures). We do not focus on Fourth Amendment issues.

5. According to William Bratton, the former New York City police commissioner and Los Angeles police chief, “[A]ny police department in America that tries to function without some form of stop and frisk, or whatever terminology they use, is doomed to fail” (Associated Press 2013).

tral; rather, what matters (or should matter) are precise definitions and the statistical analyses that speak to them. We believe that it is important for social scientists to offer their perspective when troubling allegations of racial bias are made. We also believe that it is important for this conversation to take place in academic journals.

In our view, the public debate on the practice of stop and frisk is missing a precise definition of what it means for the program to be racially biased. We propose one here. Following Becker (1957), we say that a program is biased if those who administer it are motivated in part by an (impermissible) preference for a specific racial group. Such a distortion in preferences can be present at two levels: at the level of the officers making the stops and at the level of the police chief allocating officers to precincts. (Both levels of bias are alleged in the public debate.) In Section 2 we lay out a game-theoretic model that incorporates these distortions, and we use it to determine which type of statistical evidence can be used to identify bias at either level, given the presence of confounding factors (unobservables). We conclude, consistent with previous literature, that bias at the officer level can be identified by looking at the success rates of stops. We reach a negative conclusion regarding bias at the level of the police chief: neither data on the frequency of stops nor data on their success rates can identify a latent distortion in a police chief's preferences. Interestingly, data on stop frequency are not helpful for identification at either level, which is at odds with the public focus on such data.

We then turn to the NYPD data. In Section 4 we analyze the prevalence of stops for different racial groups and find that, by any reasonable measure, African Americans are stopped much more frequently than whites as a proportion of their population. But, as mentioned above, we find it difficult to rule out unobservables, as opposed to officer bias, as potential explanations for this disparity.

In Section 5 we turn to arrest rates and find that the arrest rates of African Americans and whites who are stopped are essentially identical, at least on average across all precincts. We interpret this finding as inconsistent with the hypothesis that officers are biased in their stopping decisions, at least on average.

In Section 6 we comment on the results and discuss extensions. An important extension is to consider decisions to frisk in addition to decisions to stop. Using the same methodology, we find tentative evidence suggesting police bias in decisions to frisk, but further research is needed.

Two papers are closely related to ours. Gelman, Fagan, and Kiss

(2007) analyze New York City stop-and-frisk data for the years 1988–89. Most of their analysis focuses on documenting disparities in stops; using sophisticated statistical analysis, they conclude that “persons of African and Hispanic descent were stopped more frequently than whites, even after controlling for precinct variability and race-specific estimates of crime participation” (Gelman, Fagan, and Kiss 2007, p. 813). They also briefly address the disparity in arrest rates conditional on stops and tentatively conclude that police officers were indeed racially biased against African Americans. This tentative conclusion is based on the statistical fact that African Americans were less likely than whites to be arrested conditional on being stopped. We replicate their finding in our more recent and extensive data but show that it is overturned when we add precinct-level fixed effects. The second closely related paper is Ridgeway (2007), a RAND report sponsored by the New York City Police Foundation. Using data for 2006, the report finds that unjustified racial disparities in stopping rates are much smaller than those commonly reported in the literature. Key to this finding is the choice of benchmark for what level of disparity is justifiable.⁶ With regard to arrest rates, the report finds arrest rates that are overall higher for whites than for African Americans (Ridgeway 2007, tables 5.1 and 5.2). This is the opposite result of Gelman, Kagan, and Kiss (2007) and the same result we find using our more extensive data set. For this part of the analysis, Ridgeway (2007) uses a matching procedure that reweights observations so as to ensure an equal distribution of several characteristics of stops.⁷ If the matching procedure uses variables that are endogenous to the outcome (racial bias in our case), then the reweighting procedure might not be innocuous.⁸

6. Using census data on the fraction of residents of a given race is deemed inappropriate. This issue is well explained in Ridgeway (2007, p. 15).

7. These variables include the suspected crime, precinct, average age, time of day, location, month, sex, day of the week, type of identification (physical or verbal), “the x-y coordinates of the stop location, being reported by witness, being part of an ongoing investigation, being in a high-crime area, being at a high-crime time of day, being close to the scene of an incident, detecting sights and sounds of criminal activity, evasiveness, association with known criminals, changing direction at the sight of an officer, carrying a suspicious object, fitting a suspect description, appearing to be casing, acting as a lookout, wearing clothes consistent with those commonly used in crime, making furtive movements, acting in a manner consistent with a drug transaction or a violent crime, or having a suspicious bulge” (Ridgeway 2007, pp. 34–35).

8. A potential source of endogeneity might be the use of location as a matching variable. If the police express their bias by overpolicing a location that is mostly frequented by nonwhites, the matching procedure should cancel out this channel for bias.

In sum, a comparison of these two papers indicates that there is no consensus on the key question: is there any bias in the NYPD's stop-and-frisk program? We feel that our paper makes progress on this front. First, it organizes the evidence around a model. This is necessary to understand what features of the data can be indicative of bias and whose bias (the officers' or the chief's) is being identified. Second, it uses a data set that is more recent and more comprehensive in its time span (10 years as opposed to 2 years for Gelman, Kagan, and Kiss [2007] and 1 year for Ridgeway [2007]).

Knowles, Persico, and Todd (2001) introduce a version of the model that we present in Appendix A and derive in theorem 1. That paper does not feature the analysis in Section 2.2, which is an original contribution of this paper. Also of some relevance, Knowles, Persico, and Todd (2001) deal with a setting (highway searches) that does not feature geographic units, as do the precincts that feature prominently in our theory and empirical analysis.

A more broadly related literature is that of hit-rate analysis, which developed partly in reaction to Knowles, Persico, and Todd (2001). See Ayres (2002), Persico and Castleman (2005), Todd (2008), Whitney (2008), and Persico (2009) for reviews of this strategy. Ayres and Waldfogel (1994) use hit-rate analysis to look for racial bias in a judge's decision in setting bail. Hit-rate analysis is utilized in the policing context by Persico and Todd (2005, 2006, 2008), Sanga (2009), Hernández-Murillo and Knowles (2004), Childers (2012), Gershman (2000), and Persico (2002). Anwar and Fang (2006) offer an alternative approach to hit-rate analysis. Dharmapala and Ross (2004), Alpert, MacDonald, and Dunham (2005), and Smith, Makarios, and Alpert (2006) offer critical appraisals of this strategy.

2. THE MODEL

This model is meant to conceptualize what it means for policing to be biased and to give guidance as to where to look for bias in the data. Agents in our model are of three kinds: citizens of race $r \in \{A, W\}$ living in district i who choose whether to commit a crime that is detected through a stop-and-frisk instance, a mass P of police officers who stop and frisk citizens, and a police chief whose only action is to assign officers to precincts.

Following Becker (1957), we define bias as a taste for discrimination. By that we refer to a component of an agent's utility function that is dependent on the race of those with whom the agent interacts. A key modeling choice is whether an officer's possible taste for discrimination is innate or whether it reproduces the culture of the precinct to which the officer is assigned. In this model we opt for the second possibility and assume that officers who are assigned to precinct i inherit the bias of that precinct.⁹ Therefore, in this model it may be more appropriate to talk about precinct bias as opposed to officer bias.

The game has two stages. In the first stage, a police chief allocates police officers to precincts. In the second stage, a game in each precinct is played between officers and pedestrians. The second-stage games (one in each precinct) are assumed to reproduce that studied by Knowles, Persico, and Todd (2001) and Persico and Todd (2006). Because the model is somewhat standard, we do not reproduce its description here but relegate it to Appendix A. The equilibrium of each precinct-level game yields a function $K_i^r(P_i)$, which summarizes for each precinct the fraction of pedestrians of race r who commit a crime in district i when P_i officers are allocated to that district. The function $K_i^r(P_i)$ is derived in Appendix A.

Let us start with the first stage. What would be a permissible objective or motive for the chief that defines unbiased behavior? In the case of a police chief or other central authority allocating resources across districts, it seems reasonable to define this objective as the minimization of crime.¹⁰ If we make this assumption, then we may conceptualize the legally permissible version of the police chief's problem as follows:

$$\min_{P_i} \sum_i [N_i^A K_i^A(P_i) + N_i^W K_i^W(P_i)] \quad \text{subject to} \quad \sum_i P_i \leq P,$$

where P represents the total number of police officers available to the police chief and N_i^r denotes the number of pedestrians belonging to group

9. We make this assumption for two reasons: first, because we believe there is such a thing as a precinct culture and, second, because otherwise any observed differences in precinct bias would, in the model, result from the police chief's choice of which officers to allocate to which district. This does not seem to be a realistic assumption given that the personnel being allocated are patrol officers, whose bias is probably not known by the police chief.

10. Of note, this objective would be meaningless for the individual police officer in a district, because an individual officer probably has a negligible impact on aggregate crime. Put differently, it would be impractical to reward any police officer on the basis of total crime in New York City or even in her precinct, because that outcome depends only minimally on the officer's behavior.

r . According to this formulation, the legally permissible objective of the police chief is to minimize the sum of crimes across all precincts.

Deviations from this behavior can be classified as biased. What would be the impermissible, or biased, version of the police chief's objective function? Perhaps it would be one in which we allow the police chief to prioritize the crime rate of certain precincts, which may be objectionable, especially if these priorities turn out to be correlated with the race of the precinct's residents. This can be conceptualized by assigning welfare weights Γ_i to the crime rates of precinct i . In addition, the police chief may single out the crimes committed by pedestrians of a particular race. We can conceptualize this by adding race-specific weights γ^r . Then a potentially biased police chief would solve the following problem:

$$\min_{P_i} \sum_i \Gamma_i [\gamma^A N_i^A K_i^A(P_i) + \gamma^W N_i^W K_i^W(P_i)] \quad \text{subject to} \quad \sum_i P_i \leq P. \quad (1)$$

The parameters Γ_i , γ^A , and γ^W capture the police chief's bias.

2.1. Identifying Precinct-Level Bias

The precinct-level game is described and analyzed in Appendix A. The analysis reproduces that in Persico and Todd (2006) and points to the success rate of stops as a key indicator of officer bias. The logic, intuitively, is the following. Suppose that the success rate is lower for stops of African American pedestrians. An officer who is not biased against African Americans and is motivated by the prospect of making an arrest should reduce the number of less productive stops of African Americans and increase the more productive stops of whites. This arbitrage on the part of individual officers has aggregate effects: as officers shift to policing whites, the crime rate in the other group rises and the crime rate among whites decreases. This arbitrage will continue until, under a perfectly unbiased police force, arrest rates (hit rates) are equalized between the stops of white and African American pedestrians in the precinct. If the police force is biased, however, this arbitrage will stop earlier, at a point where the differential between the arrest rates of African Americans and whites is exactly offset by the officers' bias. This logic gives rise to the following result, which is proved in Appendix A.

Theorem 1: Positive Result for Identification of Police Officers' Bias. In the equilibrium of the precinct-level game, the arrest rate is the same across all subgroups in a race that are distinguishable by police. In

addition, if the police are unbiased, then the arrest rate is the same across races. If the police are biased against race r , the arrest rate is lower for race r than for the other race. Thus, officers' bias can be identified using arrest rates.

This theorem provides the justification for the hit-rate test applied in Section 2.2. It also identifies a significant advantage of the hit-rate test: under the behavioral assumptions stipulated in the model, the test is robust to omitted-variable bias. To see this, interpret the subgroups mentioned in the theorem as sets of pedestrians sharing certain characteristics that are observed by the officer but perhaps not observed by the econometrician. The econometrician, therefore, observes only the average arrest rate among all subgroups but not the arrest rate of each subgroup. In principle, this poses a problem because the arrest rate of a subgroup is what identifies bias against that subgroup. However, the theorem says that this is not a problem, because the equilibrium arrest rates must be the same across all subgroups within a race that are distinguishable by police, even if these subgroups are not distinguishable by the econometrician. Therefore, the econometrician's lack of ability to discern subgroups has no impact on his ability to infer bias from arrest rates.

2.2. Identifying Bias in Personnel Allocation across Precincts

Preliminary to the question of identifying the weights Γ_i , γ^A , and γ^W in problem (1), a conceptual ambiguity in the interpretation of these weights must be noted. One might be ambivalent about whether the configuration $\Gamma_i > \Gamma_j$ represents bias in favor of or against precinct i . On the one hand, $\Gamma_i > \Gamma_j$ means that precinct i 's crime rate is more salient than precinct j 's, and, accordingly, more resources will be proportionally devoted to precinct i , which results in a lower crime rate in that precinct. On the other hand, $\Gamma_i > \Gamma_j$ means that precinct i will be assigned more police officers, so there are more stops and more frisks, which some civil liberty advocates object to especially if police pressure correlates with the prevalence of minorities in the neighborhood. So it is conceptually and normatively ambiguous whether $\Gamma_i > \Gamma_j$ means that precinct i is favored or disfavored relative to precinct j .

Apart from the above conceptual and normative ambiguity, there is also an empirical difficulty in estimating the (unobserved) weights γ^r . To see the nature of this difficulty, we derive the equilibrium predictions that

would allow us to estimate the weights. The first-order conditions necessary for optimality in problem (1) are¹¹

$$\begin{aligned} & \Gamma_i \left[\gamma^A N_i^A \frac{\partial}{\partial P_i} K_i^A(P_i) + \gamma^W N_i^W \frac{\partial}{\partial P_i} K_i^W(P_i) \right]_{P_i=P_i^*} \\ &= \Gamma_j \left[\gamma^A N_j^A \frac{\partial}{\partial P_j} K_j^A(P_j) + \gamma^W N_j^W \frac{\partial}{\partial P_j} K_j^W(P_j) \right]_{P_j=P_j^*} \quad \text{for all } i, j, \end{aligned} \quad (2)$$

where P_i^* and P_j^* represent the optimal allocation. The conditions in expression (2) represent a system of equations, one for each precinct. In this system, N_i^r and P_i^* are known. The unknowns we seek to solve for are Γ_i , γ^A , and γ^W , which together are more than the number of equations. Clearly, identification is not possible from this system alone. Furthermore, even if we restrict attention to a subset of parameters (say, we somehow know all Γ_i and need to identify only all γ^r), we cannot identify them if we do not observe the elasticities of crime to policing, $\partial K_i^A(P_i)/\partial P_i$. This is proved in the next theorem.

Theorem 2: Negative Result for Identification of Police Chief Bias. The parameter Γ_i cannot be identified from system (2) without knowledge of the elasticity of crime to policing $\partial K_i^r(P_i)/\partial P_i$. The parameter γ^r cannot be identified from system (2) without knowledge of at least a pair (i, j) of the elasticities of crime to policing $\partial K_i^A(P_i)/\partial P_i$ and $\partial K_j^A(P_j)/\partial P_j$. Without knowledge of these elasticities, neither crime levels (equivalent to hit rates in the model) $K_i^r(P_i)$ nor stop intensities P_i^* are helpful in identifying any bias parameters in the police chief's problem (1).

Proof. Let us solve for Γ_1 . Write system (2) for $i = 1$. Even if the right-hand side is known, identification of Γ_1 separate from the term in brackets by which it is multiplied requires knowledge of the term in brackets. This in turn requires knowledge of $\partial K_i^A(P_i)/\partial P_i$ and $\partial K_i^W(P_i)/\partial P_i$. Thus, the first sentence in the theorem is proved. Now let us solve for γ^A . Manipulating system (2), we get

$$\begin{aligned} & \gamma^A \left[\Gamma_i N_i^A \frac{\partial}{\partial P_i} K_i^A(P_i) - \Gamma_i N_j^A \frac{\partial}{\partial P_j} K_j^A(P_j) \right] \\ &= \gamma^W \left[\Gamma_j N_j^W \frac{\partial}{\partial P_j} K_j^W(P_j) - \Gamma_i N_i^W \frac{\partial}{\partial P_i} K_i^W(P_i) \right]. \end{aligned}$$

11. It is convenient to assume that $C_i(\cdot)$ is a concave function. Under this assumption, which we maintain, the first-order conditions are also sufficient for optimality.

Even if the right-hand side is known, identification of γ^A separate from the term in brackets by which it is multiplied requires knowledge of the term in brackets. This in turn requires knowledge of $\partial K_i^A(P_i)/\partial P_i$ and $\partial K_j^A(P_j)/\partial P_j$ for at least one pair (i, j) . Thus, the second sentence in the theorem is proved. The third sentence follows because neither crime levels and hit rates nor stop intensities enter system (2) unmediated by the function $\partial K_i^r(\cdot)/\partial P_i$. Q.E.D.

This theorem proves an important point. Knowledge about the amount of policing directed to precincts or about hit rates cannot help identify the bias parameters in the police chief's problem. Instead, the key statistic is an elasticity. Unfortunately, it is generally difficult to get persuasive estimates of elasticities because they capture a counterfactual: what would happen to the crime rate if the police chief happened to perturb the allocation of personnel from its equilibrium level? Thus, estimating elasticities requires observing more than simply the level of crime at an equilibrium. This is an empirical challenge.¹²

The takeaway from this section is that identifying bias in the allocation of personnel across precincts is difficult for two reasons. The first difficulty is of a normative nature, and it has to do with what it means for an allocation to be biased against a precinct. The second difficulty is that it is difficult to obtain empirical estimates of the weights in problem (1).

3. THE DATA

We use data collected by the NYPD on individual stops, questionings, and frisks in New York City between 2003 and 2012 (NYPD 2003–12). This appears to be a slightly longer period than the one at issue in *Floyd*. The database contains information on whether the pedestrian was frisked, issued a summons, or arrested; the type of crime, which is recorded in our data as being suspected by the police making the stop; the race of the pedestrian; and the timing and location of the stop.

Unless explicitly mentioned, we restrict the sample to African American and white pedestrians, excluding Hispanics and other categories because the charge of racial bias seems to have special force with refer-

12. Typically, exogenous variation of police personnel is necessary to identify the elasticity of crime to policing. Levitt (1997) and McCrary (2002) discuss the difficulties of estimating such elasticity.

ence to the African American population.¹³ In this restricted sample of 2,947,865 stops, approximately 6 percent of the pedestrians who were stopped were arrested; 84 percent of the stops are of African American pedestrians and the rest are of whites. Most of the suspects' crimes recorded by the officers making stops fall into one of these categories: possession of a weapon (27 percent), robbery (17 percent), criminal trespass (12 percent), grand larceny auto (9.1 percent), and burglary (8.9 percent). Table 1 reports some descriptive statistics.

A possible caveat regarding these data is that NYPD officers are not required to record all interactions with private citizens. While NYPD policy requires officers to fill out a UF-250 form for every *Terry* stop,¹⁴ which is defined as a brief detention of a person based on a reasonable suspicion of crime, NYPD policy also specifically enumerates the circumstances requiring the completion of form UF-250. Those circumstances are when a person is stopped by use of force, a person who is stopped is frisked or searched, a person is arrested, or a person who is stopped refuses to identify himself (and is later identified by the officer) (New York State 1999, p. xv).¹⁵ It is possible, therefore, that recorded stops (the ones in our database) may be a selected sample of all stops. Judge Scheindlin noted the data limitations but ultimately accepted the data as a useful, albeit imperfect, tool to aid her decision in *Floyd*. We do the same in the main body of the paper.

It is tempting to address the selective-recording concern by restricting the sample to stops that are required by law to be recorded. In this sample, the problem of selective recording should not exist. The trouble with this strategy is that, at the time of choosing whom to stop, the officer cannot distinguish whether the stop will develop into one that has to be recorded. Conditioning our analysis on such *ex post* information would mean conditioning on information not possessed by the officer at the time of the stop. Put differently, the outcomes contained in this restricted data set cannot be said to fully portray the outcomes generated by any officer's stopping behavior. Thus, the hit-rate analysis cannot properly be applied to such a subsample. Nevertheless, for completeness, in Appendix C we construct a sample that we believe approximates the subsample with

13. We briefly extend our focus to Hispanics in Section 6.2.

14. "Prepare STOP, QUESTION AND FRISK REPORT WORKSHEET (PD344-151A) for EACH person stopped" (New York Police Department's Patrol Guide 212-11.6; emphasis in the original).

15. The outcome "refused to identify" is not recorded in the data. We proxy for it using the field "evasive response to questioning."

Table 1. Descriptive Statistics (%)

	Mean	SD
Arrest made	5.8	23
African American	84	37
Recorded crime:		
Possession of a weapon	27	44
Robbery	17	37
Criminal trespass	12	32
Grand larceny auto	9.1	29
Burglary	8.9	28
Grand larceny	4.3	20
Assault	4	20
Illegal possession of substances	3.6	19
Possession of marijuana	3.3	18
Illegal sales of substances	2.9	17
Petit larceny	2.5	16
Mischief	1.2	11
Graffiti	1.1	10
Other	4.3	20

Note. The crime categories represent 95 percent of the crimes recorded in the sample. Years 2003–5 have missing values for the recorded crimes. $N = 2,947,865$ observations and 2,496,267 recorded crimes.

mandated reports, and we replicate our analysis on that sample. That analysis identifies bias against African Americans.

A second, very important caveat must be raised regarding the suitability of arrests as an outcome for hit-rate analysis. The ideal outcome is a measure of productivity that the officer legitimately maximizes and that is itself objective, that is, not tainted by police bias. Arrests might not be objective because they might be subject to police discretion, and thus they may be tainted by police bias. For example, all else equal, the officer may be more likely to arrest an African American than a white pedestrian after having stopped either. This is a valid concern. The best evidence to address this concern would be the rates at which arrests, which are usually warrantless in our sample, are later upheld by a judge.¹⁶ Unfortunately, such data are not available to us. We return to this issue in Section 6.4.

16. Because the conversion is done by a judge, it is arguably an objective outcome in the sense that any judicial bias should be uncorrelated with the bias of the police officer making the arrest.

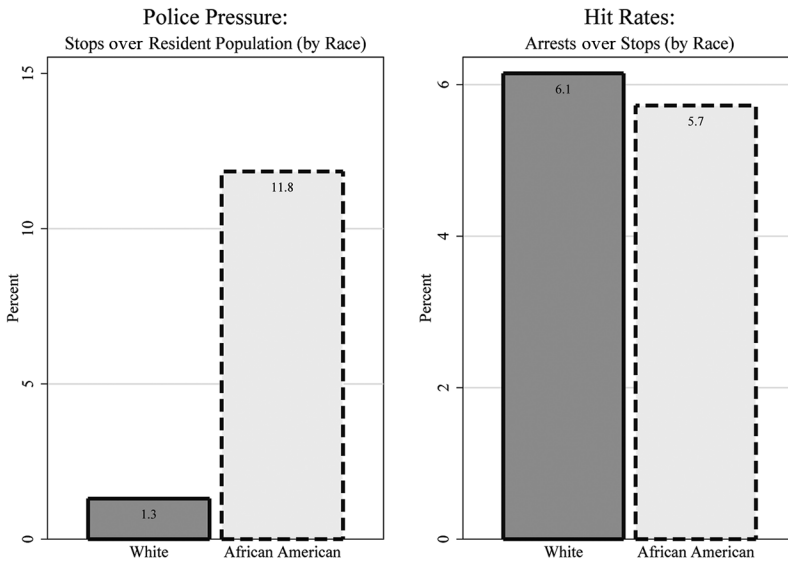


Figure 1. Average annual police pressure and hit rates in New York City, 2003–12

4. DISPARITIES IN POLICE PRESSURE

New York City's stop-and-frisk program disproportionately impacts minorities. The New York Civil Liberties Union (2012, p. 5) makes this point forcefully by documenting that, in 2011, 52.9 percent of stops were of African Americans and 33.7 percent were of Latinos, while whites accounted for only 9.3 percent of the stops. After restricting attention to African Americans and whites, Figure 1 summarizes this striking disparity.¹⁷ Police pressure is defined as the average number of stops of pedestrians of a race in a year divided by the total population of that race in New York City. In the full sample, pressure is about 10 times larger for African Americans than for whites.

This disparate impact is certainly problematic from a social viewpoint. However, disparate impact can reflect many factors, both observable and unobservable, that affect the stop-and-frisk process. Neither theorem 1 nor theorem 2 indicates that police pressure can help identify police bias. This lack of theoretical framework becomes problematic when, as in Table 2, conditioning on observables attenuates the disparity.

17. The resident population data in Figure 1 are from the 2010 census.

Table 2. Correlates of Relative Police Pressure in New York City

	(1)	(2)	(3)
% African American	-.222** (.073)	-.057 (.055)	-.066 (.049)
Income		.365** (.119)	.307** (.115)
Constant	23.118** (3.857)	-2.910 (6.473)	-10.874 (24.535)
Average relative police pressure			17
% African Americans in average precinct			26.78
Adjusted R^2	.083	.266	.467
Precinct controls	No	No	Yes
Year fixed effects	No	Yes	Yes

Note. Estimates are from ordinary least squares regressions on 75 precincts. The dependent variable is (relative) police pressure (arrests of American Americans/African American population)/(arrests of whites/white population). Column 3 includes the variable for the margin of Mayor Michael Bloomberg's victory. Missing years are computed using moving averages for the variables for the fraction of African Americans, income, age, fraction of females, fraction of college degrees, serious crime, graffiti, social capital, and African American commanding officers. Regressions with year fixed effects (nine dummies) control for possible time trends in the dependent variable and precinct-specific characteristics. Standard errors, in parentheses, are clustered at the precinct level. $N = 750$ observations.

** Significant at the 1% level.

Table 2 makes use of variability across precincts. The dependent variable is the ratio of police pressure for African Americans to that for whites.¹⁸ Column 1 shows that this disparity in police pressure is correlated with the precinct's racial makeup. From an atheoretical perspective, this dependence on race seems troubling. However, columns 2 and 3 show that conditioning on income takes away the effect of race (and eliminates the significance of the constant in the regression).¹⁹ Thus, it now appears that police personnel are allocated disproportionately to precincts with poorer residents—which is arguably less bad than allocating disproportionately to minority precincts. Is this encouraging news? Does this say anything about police bias? In our view, it is difficult to learn much from this atheoretical approach except at a descriptive level. For this reason, we regard the estimates in this section as suggestive of, but not dis-

18. Table 2 shows that, averaging within precincts first, then taking arithmetic averages of precinct-level means, the ratio of African American to white police pressure is 17. This result differs from the ratio (about 10) implied by Figure 1, where the average is computed at the citywide level.

19. The variables are described in Appendix D.

positive regarding, the presence of bias. We turn, therefore, to the analysis of arrest rates, which has its theoretical underpinning in theorem 1.

5. ANALYSIS OF ARREST RATES

Theorem 1 indicates that a comparison of arrest rates by race can identify bias in police officers' stop decisions. In this section we compute arrest rates and compare them across races.

We start by noting that, in the aggregate, the probability that a stop translates into an arrest is quite similar across races in our sample. This is shown by regressing an indicator variable coding whether the pedestrian who was stopped was arrested on another indicator variable coding the pedestrian's race. Figure 1 (right panel) shows the aggregate arrest rates for pedestrians by race. Clearly, the large disparity across races that is present in police pressure (left panel) is absent when we look at average arrest rates.

More detailed estimates are reported in Table 3. Depending on the specification, African American pedestrians who are stopped are between .42 percent and .44 percent less likely to be arrested than whites. Thus, the probability of a stop resulting in an arrest is about 6 percent for whites versus 5.6 percent for African Americans. Although the difference is very small, and perhaps unlikely to be perceived by an officer on the basis of his own experience alone, the difference is significant in two of three specifications. This pattern is similar to that found by Gelman, Fagan, and Kiss (2007) in their more limited sample.

This small difference in arrest rates between races changes sign, however, when we control for precincts. Precincts vary considerably in the likelihood that a stop translates into an arrest (see Figure 2).²⁰ Controlling for precincts is appropriate because precincts are, in effect, separate jurisdictions.²¹ It also prevents fallacy in aggregation if baseline arrest rates are correlated with race.

To understand the possible fallacy let us, for the sake of argument,

20. This is not surprising, given the heterogeneity among precincts. For institutional information about precincts, see Fyfe and Kane (2005).

21. The New York Police Department is organized into 76 precincts, each of which is responsible for a specific geographic area. An officer from one precinct cannot stop pedestrians in another precinct. According to New York state law (Crim. Proc. 140.50), "a police officer may stop a person in a public place located within the geographical area of such officer's employment."

Table 3. Arrests Made

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
African American	-.420** (.037)	-.437** (.037)	-.437 (.469)	.379** (.046)	.355** (.046)	.355+ (.207)	.340+ (.204)
Constant	6.140** (.034)						
Mean outcome							5.79
% African American							84
P-value				.001	.001	.001	.001
Year fixed effects	No	Yes	Yes	No	Yes	Yes	Yes
Precinct fixed effects	No	No	No	Yes	Yes	Yes	Yes
Year × precinct fixed effects	No	No	No	No	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of being arrested conditional on being stopped in New York City (in %). Regressions with year fixed effects (nine dummies) and precinct fixed effects on 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level in columns 3, 5, and 7. The *P*-value is for the joint test of all the precinct fixed effects equal to 0. *N* = 2,947,865.

+ Significant at the 10% level.

** Significant at the 1% level.

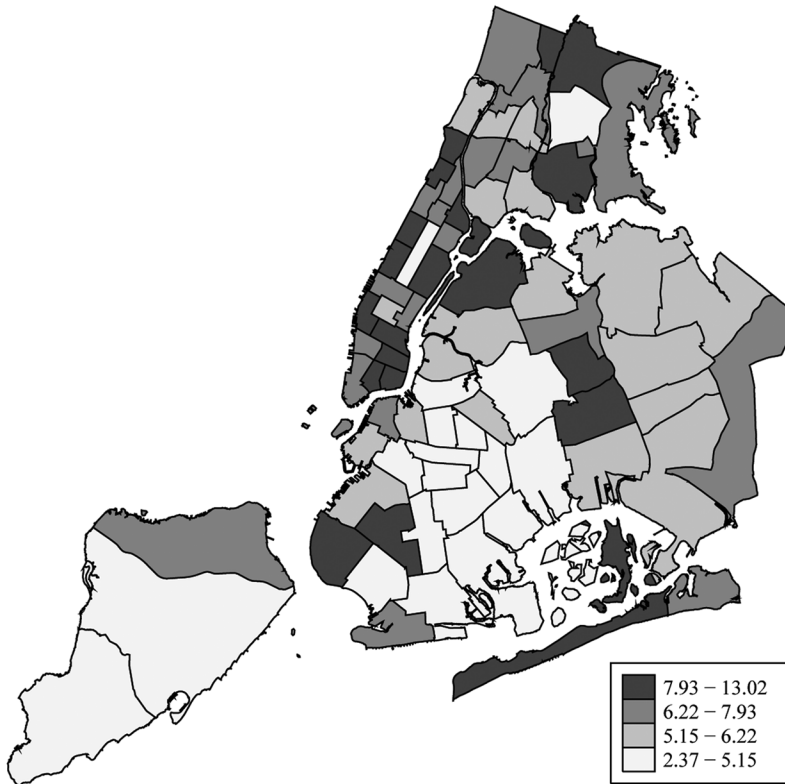


Figure 2. Probability of being arrested conditional on being stopped in New York City, 2003–12.

treat precincts as separate jurisdictions. If the police officers in each precinct are unbiased, then within each precinct the arrest rates of African American and white pedestrians should be the same conditional on being stopped. However, the levels of arrest rates need not be the same across precincts. For example, suppose hypothetically that, of all the African Americans and whites stopped in the Bronx, 3 percent were arrested and 6 percent of the African Americans and whites stopped in the Financial District were arrested. If we aggregate the data from the two precincts, we would mistakenly conclude that the police officers making the stops are biased against African Americans, because in the aggregate sample most African Americans are searched in the Bronx and have a 3 percent arrest rate, which is much lower than that of whites, most of whom are

searched in the Financial District. Thus, the hit-rate test carried out without controlling for precincts would be potentially biased or, more precisely, uninformative about the racial bias exhibited by police officers in each precinct.

A solution to this aggregation problem is to introduce precinct-level fixed effects in the statistical model that predicts arrest rates. In the above hypothetical example, introducing precinct-level fixed effects into the baseline specification allows the fixed effects to absorb the 3 percent and 6 percent baseline arrest rates, while the coefficient on African American would be estimated to equal 0. This coefficient of 0 would properly be interpreted as evidence that the police are not biased. Conversely, if the police are biased, we would observe lower arrest rates for African American searchees in many or all precincts, and this difference in arrest rates between African Americans and whites would be picked up by the coefficient on African American, after controlling for precinct-level fixed effects. Therefore, controlling for precinct fixed effects is necessary for the hit-rate test to function properly. Hence, precinct-level fixed-effects regressions represent our preferred specifications.

In columns 4–7 of Table 3 we present the same ordinary least squares regressions, this time with 76 precinct-level fixed effects. Notably, the coefficient on African American changes sign. Stopping an African American pedestrian results in a probability of arrest that is larger by .355 percent compared with stopping a white pedestrian. That is, after accounting for the fact that different precincts have different baseline rates of arrest conditional on search, African Americans are no longer less likely to be arrested conditional on being stopped.²² On the basis of theorem 1, we interpret this evidence as rejecting the hypothesis of a relative bias against African Americans in the officers' decisions to stop a pedestrian, at least on average across precincts. This is a key message from our analysis.²³ The evidence so far is consistent with the interpretation that, on average,

22. The precinct-level fixed effects jointly explain, in a statistical sense, the arrest rate (that is, the *P*-value is less than 5 percent for the joint test of all the precincts' fixed effects equal to 0; see Table 3).

23. For completeness, we repeated the analysis on the subsample of stops that are required by law to be reported (see New York State 1999, p. xv). In this subsample, the coefficient on African American does not become positive in the specification with precinct-level fixed effects. However, the estimates obtained controlling for the suspect's recorded type of crime and the analysis of police pressure are invariant to this sample selection. As mentioned before, this subsample cannot be a proper sample for a hit-rate analysis. Therefore, we disregard the results from this analysis.

there is no bias against African Americans in the police officers' decisions to stop a pedestrian.

It is important to point out that our main result is obtained without controlling for the suspected crime that is reported by the police officer. This is because theorem 1 requires equalization only across dimensions that are visible before the stop and hence can be used by the police to choose whom to stop. In contrast, the information contained in the variable for the suspected crime appears to reflect knowledge that can most plausibly be acquired during the stop. (For instance, the categories of recorded crime include "criminal possession of controlled substances," "criminal possession of marijuana," and "criminal possession of a weapon.") If post-stop information is included in the regression and this information is correlated with race, then the hit-rate test might be invalidated. To see this, imagine, hypothetically, that most African Americans who are stopped end up being arrested for substance possession, but whites who are stopped end up being arrested for all sorts of crimes. In this case introducing the variable for possession of a controlled substance (which is post-stop information) in the regression might absorb a lot of the predictive power that would otherwise accrue to the dummy variable African American and, therefore, potentially reduce the absolute magnitude of its estimated coefficient. So if the police are biased, the estimated coefficient on African American will not capture the full extent of police bias because of the presence of the inappropriate control for suspected crime. In any case, putting aside this theoretical argument, in Section 6.4 we repeat our analysis controlling for suspected crime and establish that our finding persists.

6. DISCUSSION AND EXTENSIONS

6.1. A Tale of Two Measures

Most of the public debate considers the wide racial disparity in police pressure (see Figure 1, left panel) as a strong clue, if not outright evidence, that the police act in a racially biased way. Our theoretical analysis (theorems 1 and 2) has led us to conclude that bias, to the extent that it can be identified with the data at hand, ought to be inferred from disparities in arrest rates (see Figure 1, right panel). One might ask whether these two measures, police pressure and arrest rates, are correlated across precincts. In other words, if we take arrest rate differentials as a good proxy of bias,

are they related to differentials in police pressure? Figure 3 suggests that this is not the case in our sample.²⁴ The horizontal axis reports the ratio, computed at the precinct level, of the values for police pressure in Figure 1. This is a measure of how much greater police pressure is on African American citizens compared with whites. The vertical axis represents our measure of bias (difference between the arrest rates of African Americans and whites, 10-year average, by precinct). Figure 3 suggests that there is no correlation between these two measures across precincts.²⁵ We interpret this observation as going against the presumption that large disparities in police pressure are necessarily correlates of police officers' bias.

6.2. Hispanics

So far we have restricted the hit-rate analysis to African American and white pedestrians. We now extend our analysis to a sample of 4,413,566 stops and frisks of African Americans, Hispanics (black and white), and whites. In this larger sample, African Americans and Hispanics represent 56.1 percent and 33.2 percent of the stops, respectively, and suspects are arrested, on average, 6 percent of the time.

In Table B1 we augment our baseline specification by regressing the indicator variable coding whether the pedestrian who was stopped was arrested on an indicator variable for African American pedestrians and on an indicator variable for Hispanic pedestrians. Depending on the specification, Hispanic pedestrians who are stopped are between .12 percent and .15 percent less likely to be arrested than whites. This small difference in arrest rates between races, however, vanishes if we control for precinct fixed effects.²⁶

6.3. Other Characteristics

So far we have focused on race and ethnicity. Other characteristics of a stop are available in our data such as gender, time of the stop, and so

24. Figure 3 plots the differential probability of an African American being arrested (that is, the estimated coefficient in the univariate regression of arrests on an indicator variable for African American pedestrians) against the natural logarithm of relative police pressure: (arrests of African Americans/African American population)/(arrests of whites/white population) in each of the New York City police precincts.

25. This intuition is supported by a regression analysis showing that relative police pressure is uncorrelated with our measure of bias. Results of these regressions are available on request.

26. In Table B2 we control for the suspect's recorded type of crime and find similar results.

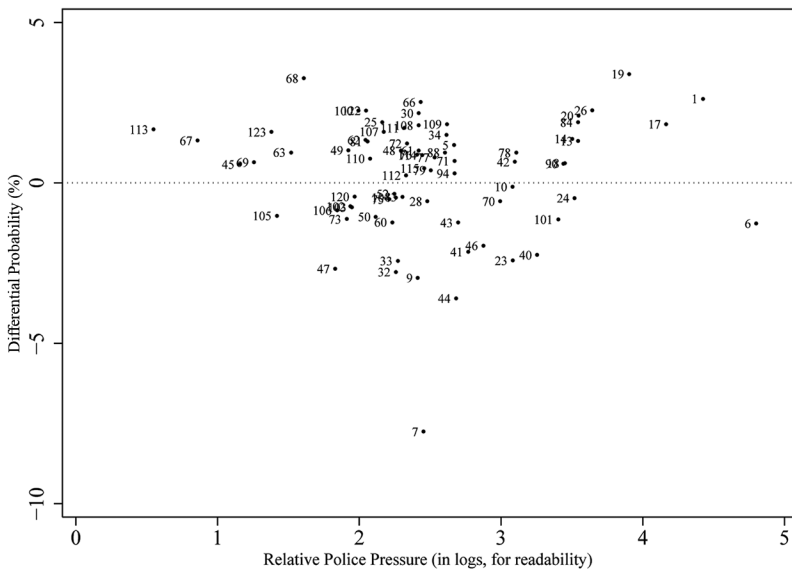


Figure 3. Differential probabilities of an African American being arrested across 76 precincts in New York City.

forth. To the extent that these are known to the officer before the stop, the theory predicts that the returns to these searches should be similar provided that the search cost is similar across characteristics. In Table 4 we compare arrest rates across these characteristics.

Although arrest rates are not equalized across characteristics, they are generally close in the sense that, for most characteristics, the estimates of the differential probability of being arrested are within 1 percent. Our favored interpretation for these coefficients is that such small statistical differences would be difficult for police officers to detect on the basis of their individual experiences. Moreover, the signs of some of the estimated coefficients might be compatible with the theoretical predictions discussed in Section 2.²⁷ For example, the negative coefficient (−.94 percent) for stops performed 7 p.m.–6 a.m. would arise in equilibrium if an arrest for a crime committed at night is perceived by the officer as more valuable than an arrest for a crime committed during the day (perhaps because the law tends to treat crimes committed at night more severely).

The exceptions are the coefficients for women (2.3 percent) and heav-

27. We thank the editor for making this observation.

Table 4. Other Pedestrian Characteristics

	All		African American		White	
	(1)	(2)	(3)	(4)	(5)	(6)
African American	-.367 (.469)	.423* (.202)				
Female	2.317** (.293)	2.308** (.260)	2.737** (.366)	2.729** (.323)	.838** (.295)	.759** (.266)
Above 6 feet	.553** (.068)	.434** (.047)	.477** (.070)	.354** (.052)	1.095** (.187)	.966** (.153)
Heavy build	1.224** (.116)	1.121** (.112)	1.176** (.116)	1.053** (.114)	1.420** (.238)	1.412** (.207)
Above 18 years old	.733** (.176)	.469** (.150)	.605** (.202)	.386* (.177)	1.548** (.285)	.938** (.167)
Stop 7 p.m.–6 a.m.	-.981** (.152)	-.939** (.160)	-1.160** (.167)	-1.104** (.178)	-.030 (.264)	.012 (.239)
Constant	5.289** (.474)	6.367** (.339)	7.640** (.689)	9.651** (.479)	4.826** (.533)	3.649** (.595)
Probability of arrest	5.855		5.786		6.217	
% African American	.840		1		0	
% Female	.0741		.0687		.102	
% Above 6 feet	.614		.618		.593	
% Heavy build	.0878		.0897		.0775	
% Above 18 years old	.847		.843		.868	
% Stops 7 p.m.–6 a.m.	.459		.460		.451	
Precinct fixed effects	No	Yes	No	Yes	No	Yes
N	2,853,320	2,853,320	2,397,038	2,397,038	456,282	456,282

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of being arrested conditional on being stopped in New York City (in %). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors clustered at the precinct level are in parentheses.

* Significant at the 5% level.

** Significant at the 1% level.

ily built pedestrians (1.2 percent); these higher arrest rates are indicative, in our framework, of a relative reluctance by officers to stop and frisk people in these categories. Why might officers refrain from stopping and frisking heavily built pedestrians and women? (Only 7.4 percent of stops are of women.) We conjecture that (mostly male) officers might shy away, at the margin, from searching women because such stops, although lawful, are even more controversial than those of male suspects (Ruderman 2012).²⁸ However, an analysis of this gender disparity is deferred to future work. The inferred reluctance to stop and frisk heavily built men may be due to a possibly higher risk of an aggressive reaction.

It is worth noting that conditioning on all these characteristics does not affect the main estimate of interest in this paper. Table 4 shows that even after controlling for these characteristics, the estimated coefficients on the African American indicator are parallel to those in Table 3.²⁹

6.4. Arrests as an Outcome for Hit-Rate Analysis

In this section we explore the concern (raised at the end of Section 3) that arrests might not be an objective outcome. To assess the objectivity of arrests, ideally one would want data on the fraction of arrests resulting from stops that are dismissed either by prosecutors before they get to court or by judges at a later proceeding. Unfortunately, we do not have access to such data.

We follow an (admittedly imperfect) alternative strategy by looking at an officer's behavior after the stop has been made. We check whether, given the suspect's type of crime as recorded by the police officer after the stop, the officer is more likely to arrest an African American than a white pedestrian. To see why this exercise speaks to the issue, imagine that out of 100 African Americans who are stopped, a police officer identifies two with weapons that he or she suspects (or believes) may be illegally held. Of 100 whites who are stopped, the police officer identifies four with such weapons. If the police officer arrests all African Americans with such weapons but only half of whites with such weapons, then our test of bias in the decision to stop will indicate an equal success rate of stops (2 percent in both races) and will therefore absolve the police officer

28. Ruderman (2012) indicates that in 2011 more than 80 percent of patrol officers were men and that it is deemed unsafe and often impractical for a male officer to summon a female officer to conduct a frisk of a female pedestrian.

29. The estimates in Table 4 differ slightly from those in Table 3 because of the smaller sample in the former, which is due to missing values for pedestrian characteristics.

of bias in the decisions to stop. However, in this example African American suspects are arrested more often than whites for the same suspected type of crime (as recorded by the officer), which indicates bias in the decision to arrest.

To check for bias in the decision to arrest, we check whether the race of the person stopped predicts the probability of arrest after conditioning on the suspect's recorded crime (as noted by the officer on form UF-250). This test will reveal whether the police show a race-based disparity in translating a suspect's crime into an arrest. An implicit assumption behind this test is that there is no discretion in an officer's recording of the suspect's crime. This test of bias in the decision to arrest is distinct from the main test in the paper, which looks for bias in the decision to stop.

Table B3 in Appendix B presents the results. When we control for type of crime, African Americans are slightly more likely to be arrested—.225 percent—but the effect is not statistically significant. We interpret this result as consistent with the hypothesis that, given a crime (as recorded by the officer) committed by a pedestrian of either race, officers are not using discretion in deciding whom to arrest or, at least, that any discretion they use is uncorrelated with race. Therefore, we do not find evidence against using arrest as the outcome in the hit-rate test.

We also break down the reasons for arrest by race of the pedestrian to explore whether some reasons might be more manipulable by biased officers. Table B4 reports the evidence. Unfortunately, we do not feel that there is a clear ranking of reasons in terms of manipulability. In the end, we believe that the objectivity of the decision to arrest is an assumption that cannot be fully tested in the data publicly released by the NYPD.

6.5. Alternative Unit of Analysis: Frisks

Until now the analysis has focused on bias in the decision to stop, ignoring whatever transpires between a stop and an arrest. Here we focus on frisks. About 53.7 percent of stops of African Americans develop into frisks, as opposed to 39.3 percent of stops of whites. Thus, stops of African American pedestrians are more likely to develop into a frisk. Frisks of African American pedestrians are also less likely to be productive than frisks of whites: about 9 percent of frisks of African Americans are associated with an arrest, compared with 13 percent of frisks of whites. This disparity in arrest rates is suggestive of bias in the decision to frisk. In Coviello and Persico (2013), we lay out a heuristic dynamic choice model in which the officer must first decide whom to stop (a stop

costs the officer t_1) and then, with the additional information that becomes available after the stop, whether to frisk (at an additional cost t_2).³⁰ We derive a modified hit-rate test for that model. When applied to our data, the test indicates that the 9 percent versus 13 percent disparity is inconsistent with the null hypothesis of no bias if $t_2/t_1 > 3/4$, that is, if the additional cost of a frisk is large relative to the cost of a stop. The intuition for this result is as follows. The officer will arbitrage frisks (do I frisk the pedestrian I have stopped already, or do I pass and possibly frisk the next pedestrian I stop?) in the same way that, in our main model, he arbitrages stops. However, note that arbitraging frisks (that is, opting to frisk the next pedestrian you stop as opposed to the one already stopped) requires paying an additional stop cost t_1 . If t_1 is large, arbitrage will be limited. Conversely, if the stop cost t_1 is small, we should expect arbitrage to work powerfully to equate the difference in success rates across races.

Some further insight can be obtained if we condition on a specific crime suspected, such as weapons possession, for which we think that a frisk is always warranted (indeed, the justification for the doctrine of the *Terry* stop is for the police to check a suspect for weapons) and will generate objective evidence either inculcating or exculpating the pedestrian. About one in four (27 percent) of all stops are associated with suspected weapons possession; of these, 93.5 percent are of African Americans (see Table B5). If the police officers claim that they suspected weapons possession, then a large disparity in frisk rates among those stops favoring whites might be prima facie evidence of racial bias. Similarly, differences in arrest rates for stops involving suspicion of weapons possession and a frisk could be interpreted through the lens of a hit-rate test.

Table B5 shows that, in the subsample of those who are suspected of carrying a weapon, whites are frisked 83 percent of the time, and African

30. The heuristic model is as follows. The police officer first decides whether to expend cost t_1 and stop a pedestrian. Then, after having interviewed the pedestrian and having learned some unobserved (to us) characteristic u , the officer is able to immediately make an arrest or may choose to frisk the pedestrian at an additional cost t_2 or to let the pedestrian go. In this model, the officer uses knowledge of u , which is gained after the stop, in his or her decision of whether to frisk. Only those pedestrians with u suggesting a sufficiently high likelihood of a crime will be frisked. Note that in the equilibrium of this model it is possible to have two characteristics u_1 and u_2 , both of which lead to a frisk, but with u_2 yielding a slightly higher probability of a successful frisk. Such small differences in the returns to frisks are not arbitrated away in equilibrium because officers need to pay t_1 in order to learn whether a pedestrian has characteristic u_1 or u_2 . The model yields precise bounds on the disparity in hit rates that can still be compatible with no bias for each given ratio t_1/t_2 .

Americans are frisked 86 percent of the time. This statistically significant difference (about 3 percent the size of the fixed-effects estimates) suggests, but does not establish, that blacks are somewhat overfrisked compared with whites. Let us look at the success rates of these frisks. There are two natural definitions of success for such a frisk: the pedestrian is arrested for illegal weapons possession or the pedestrian is arrested for any reason. Regardless of how success is defined, in our subsample frisks of African Americans are less likely to be successful than frisks of whites. Table B6 shows that in the subsample of those who are suspected of carrying a weapon and have been frisked, whites are arrested for weapons possession 2.6 percent of the time, and African Americans are arrested 1.3 percent less often. This difference is statistically significant. With regard to arrests for any reason, Table B7 indicates that whites are arrested 6.7 percent of the time, and African Americans are arrested 2.3 percent less often. This difference is also statistically significant.

Overall, these disparities in success rates suggest officer bias in the decision to frisk. This conclusion is subject to the caveat mentioned before: that the arbitrage of frisks is potentially costly because it requires the officer to perform a number of stops (about four, in our data) before finding another pedestrian who, after being stopped, is suspected of weapons possession. To the extent that this arbitrage is costly, small disparities in the success rate may not be indicative of bias. Further research is required on the question of bias in frisking behavior.

6.6. Alternative Outcome: Summonses

Summonses are notices of violation issued by police officers during stops of suspects. These notices often entail an order to appear in court. In our sample, 6 percent of the pedestrians who were stopped were issued a summons, and 84 percent of the summonses were issued to African Americans. Most of the summonses in our data fall into one of these categories: disorderly conduct, open container of alcohol, trespass, and consumption of alcohol on the street.

Presumably, issuing a summons is a lesser or secondary goal for a police officer compared with an arrest. Nevertheless, issuing a summons does make the stop to some extent successful, or productive. Therefore, in Table B8 we present the results of the hit-rate test on the outcome summons issued. The results are the opposite of those in Table 3: after controlling for precincts, the sign on African American switches and becomes

negative.³¹ The interpretation, according to the hit-rate analysis, would be that officers are biased against African Americans in their decision to stop them if officers care only about issuing a summons. But probably it is proper for officers to care regarding both issuing a summons and making an arrest.

This reasoning leads us to add a dimension to the model: the rate at which police officers trade off arrests and summonses. Mathematically, the officer's payoff from a stop can be conceptualized as follows:

$$\pi(\alpha) = \alpha \times I_{\text{Arrest}} + (1 - \alpha) \times I_{\text{Summons}},$$

where I_{Arrest} and I_{Summons} are indicators taking a value of one if the pedestrian is arrested or issued a summons, respectively. If α is close to 0, the payoff $\pi(\alpha)$ will closely mimic the variable Summons, and when α is close to 1, the payoff $\pi(\alpha)$ will be close to the variable Arrests.

Once the model is extended in this way, there are two unobserved parameters in the officer's objective function: one is the officer's possible bias against African Americans, and the other is the parameter α . For the purpose of this paper, both parameters are taken to reflect the police officer's tastes or values.³² Both parameters, arguably, have a normatively correct value: for bias this value is obviously 0, that is, no bias. We do not take a stand on the normatively correct value of α^* ; instead, we carry around this value as a free parameter α^* . If the parameters in the officer's utility function differ from their normatively correct values, then the officer's behavior will lead to a disparity in hit rates, where the hit rates are computed on the variable $\pi(\alpha^*)$. But it is difficult to determine whether the disparity is due to racial bias or to a normatively incorrect α (or both). Given this ambiguity, we next perform the following analysis. We stipulate as part of our null hypothesis that the police officer is unbiased, which implies that in equilibrium the hit rates on African Americans and whites need to be equalized on the variable $\pi(\alpha)$, whatever α is in the officer's mind (this prediction comes from the reasoning in Section 2). We then look for the value $\bar{\alpha}$, which equalizes the hit rates on the variable

31. Table B9 corresponds to Table B3 and has a similar result: given a certain crime committed by a pedestrian of either race, officers do not use discretion in favor of whites when deciding to whom to issue a summons. In fact, African Americans appear to be issued summonses less often than whites. Therefore, we see no evidence of police discretion that simultaneously affects the outcomes for summonses and is biased against African Americans.

32. In a broader treatment, one could argue that α might be determined by career incentives, for example.

$\pi(\alpha)$. The resulting $\bar{\alpha}$ is interpreted as the taste parameter that, given the null hypothesis of no bias, rationalizes the officer's observed stopping behavior. After this exercise, which is purely an exercise in the identification of unobserved taste parameters, we again ask the question of whether the estimated parameter $\bar{\alpha}$ is normatively acceptable. If it is not, we conclude that the police officer is behaving in a way that departs from the norm—even though we are not able to distinguish whether the departure is that the officer is racially biased, that the officer's α differs from the normatively acceptable α^* , or both.

Let us proceed. Fix any α , which we interpret as the known taste parameter in the police utility function. What happens if we perform the hit-rate test on the outcome variable $\pi(\alpha)$? This depends on the chosen value of α . If α is close to 0, then the payoff $\pi(\alpha)$ will closely mimic the variable *Summons*, and we know from Table B8 that the hit rates are not equalized. Conversely, if α is close to 1, then the payoff $\pi(\alpha)$ will closely replicate the variable *Arrests*, and we know from Table 3 that hit rates are not equalized either. We performed a search for the threshold value $\bar{\alpha}$ that equalizes the arrest rates on $\pi(\alpha)$ across races (not reported). This threshold value $\bar{\alpha}$ is around .8.³³ This means that the observed police behavior is consistent with that of a police force that is unbiased and uses an $\alpha \approx .8$. In addition, the observed police behavior is consistent with that of a police force that is not biased against African Americans and uses an $\alpha > .8$.³⁴ A value of .8 implies that the police value each arrest equal to about four summonses. We regard this conversion rate as normatively not unacceptable—although of course readers are free to make their own judgments. If that rate of 4:1 is found to be normatively acceptable, then we conclude that the police are behaving in a way that is observably equivalent to one whose tastes are in line with the normatively correct values; that is, no bias and a normatively acceptable $\alpha \approx .8$.

33. Tables B10 and B11 report the main results when the dependent variable is $\pi(.8)$. By construction, the estimated coefficients on African American in columns 6 and 7 of Table B10 are not significantly different from 0, which means that arrest rates are equalized.

34. This should be expected: if α is close to 1, then we are back to the analysis in Section 5.

7. CONCLUSIONS

New York City's stop-and-frisk program disproportionately impacts minorities. Yet former New York City police commissioner William Bratton has said, "Stop-and-frisk is not something that you can stop. It is an absolutely basic tool of American policing" (Feith 2013). If the stop-and-frisk process cannot be eliminated, then it becomes especially important to ensure that it is carried out in a racially unbiased way. To this end, we have argued that a strong theoretical foundation is needed that can rigorously identify two distinct sources of bias: the police officer making the decisions to stop a pedestrian and the police chief allocating personnel across precincts. Previous research offers positive identification results regarding officer bias; this paper adds a new, and negative, identification result for police chief bias.

We analyze 10 years of data from NYPD's stop-and-frisk program in light of this theoretical framework. After controlling for precinct-level fixed effects, we find that white pedestrians are slightly less likely than African American pedestrians to be arrested conditional on being stopped. We interpret this fact as evidence that the officers on average are choosing whom to stop in a manner consistent with no bias against African Americans, because whites are being stopped despite being less productive stops for police officers.

An analysis of the decision to frisk reveals that, after controlling for precinct-level fixed effects, African American pedestrians are less likely than white pedestrians to be arrested conditional on being frisked. We interpret this evidence as suggestive of bias against African Americans in the decision to frisk, although further research is needed on this point.

An important caveat: our analysis is based on the assumption that the decision to arrest (as opposed to the decision to stop and frisk) is not tainted by police officers' bias. We have tested this assumption to the extent possible with the data at hand and find no evidence pointing to its rejection. However, we feel that this assumption deserves further scrutiny.

Our results cannot be interpreted as proving that the stop-and-frisk program is lawful. It is possible that the program may be unlawful in other ways, for example, searches might not always arise from a reasonable suspicion.

APPENDIX A: THE THEORY OF PEDESTRIANS' AND OFFICERS' BEHAVIOR AND A TEST FOR OFFICERS' BIAS

Condition the analysis to precinct i , so that all variables in Appendix A are precinct specific. We want to allow for officer heterogeneity in the district and for additions (or subtractions) to the mass of police officers allocated to the precinct. This raises the question of how the differential officers should affect the pre-existing distribution of officer characteristics in the precinct. For simplicity, we assume that officers are blank slates and that they take on their characteristics (possible bias, cost of searching, and so on) by drawing from a precinct-specific distribution. This modeling device avoids the need for keeping track of changing characteristics of officers in a precinct as the precinct's personnel are changed.

Let r denote the race of the pedestrian, which is assumed to be observable by the police. Without loss of generality, assume that there are pedestrians of two races, either African American (A) or white (W). Other characteristics that are costlessly observable by the police are represented by $c \in \{1, \dots, C\}$. These characteristics might represent such things as the pedestrian's age or gender. From the officer's viewpoint, a pedestrian is characterized by two variables, r and c . Let N^{rc} denote the number of pedestrians belong to group (r, c) .

We assume that the police can distinguish between pedestrian groups (r, c) but cannot detect pedestrian heterogeneity within (r, c) groups. Two sources of unobserved pedestrian heterogeneity within groups are the payoffs to an individual of committing a crime and the costs of being detected. Let v represent the value of committing a crime. If the crime is detected, the payoff to the pedestrian is $v - j$, where j captures the cost of being detected. We allow v and j to vary across individuals within a (r, c) group and denote the joint conditional distribution of v and j by the cumulative distribution function $F_{rc}(v, j)$.

A pedestrian in group (r, c) makes a binary decision: commit a crime or not. Just as pedestrians may differ in their costs and benefits, we allow police officers to be heterogeneous in three respects: their search capacity, their per-search cost, and their racial bias. We assume that there is a mass P of police officers who, after being assigned to the precinct, draw a type p from a uniform distribution on $[0, 1]$. Each police officer p is endowed with a search capacity of S_p and a per-search cost t_p . If a search does not yield any evidence of crime, then we consider the search to be unsuccessful and assume that the police officer incurred the cost of search without any benefit. We introduce the potential for police bias by allowing the benefit that the officer derives from a successful search to depend on the race of the pedestrian. Suppose that the benefit to a police officer p of finding a criminal of race W is y_p^W and the benefit of finding criminal of race A is $y_p^A = y_p^W + B(p)$. We say that police are biased against African American pedestrians if $B(p) > 0$ for all p , against whites if $B(p) < 0$ for all p , and unbiased if $B(p) = 0$ for all p . If no search is conducted, there is a payoff of 0. As described, this accommodates police

heterogeneity in intensity of bias. However, we rule out environments in which $B(p)$ changes sign as p varies, that is, where some policemen are biased against whites and some are biased against African Americans. Below we propose a test for inferring the sign of $B(p)$.

A member of group (r, c) with given v, j , who commits a crime and expects σ members of his group to be searched receives an expected payoff

$$u_{r,c}(v, j, \sigma) = v - j \frac{\sigma}{N_{r,c}}.$$

When this payoff exceeds 0, the individual will choose to commit a crime. Let $K_{r,c}(v, j, \sigma)$ be an indicator function that equals one if the individual chooses to commit a crime. The fraction of pedestrians in each group (r, c) who commit a crime is given by

$$K^{r,c}(\sigma) = \int K_{r,c}(v, j, \sigma) dF_{r,c}(v, j).$$

The function $K^{r,c}(\sigma)$ summarizes the crime rate in group (r, c) when the police search that group with intensity σ . One can think of this function as a response function or as the supply of crime.

Denote by $S_p(r, c)$ the number of searches that officer p decides to devote to group (r, c) . The total number of searches of members of group (r, c) is obtained by aggregating the behavior of all police officers:

$$S(P, r, c) = P \int_0^1 S_p(r, c) dp.$$

Officer p 's expected payoff is the sum of the expected payoffs of all his searches, given by

$$\sum_{r,c} S_p(r, c) \{y_p^r K^{r,c}[S(P, r, c)] - t_p\}, \quad (\text{A1})$$

which depends on the officers perceived benefit from apprehending someone of race $r(y_p^r)$ as well as the officer's costs of search (t_p).

Persico and Todd (2006) show that a Nash equilibrium of this game exists and is generically unique. Let $[S^*(P, r, c)]_{r,c}$ be a vector denoting the search intensities at the Nash equilibrium. Suppose that groups (r, c) and (r', c') are searched in equilibrium. Then there must be a p and a p' such that

$$y_p^r K^{r,c}[S^*(P, r, c)] - t_p \geq y_{p'}^{r'} K^{r',c'}[S^*(P, r', c')] - t_{p'}$$

and

$$y_{p'}^{r'} K^{r',c'}[S^*(P, r, c)] - t_{p'} \leq y_{p'}^{r'} K^{r',c'}[S^*(P, r', c')] - t_{p'}.$$

If $r = r'$ or if the police are unbiased, then $y_p^r = y_{p'}^{r'}$ for all values of p , and so the two inequalities can be simultaneously satisfied only if

$$K^{r,c}[S^*(P, r, c)] = K^{r',c'}[S^*(P, r', c')]. \quad (\text{A2})$$

If the police are biased against race r , then $y_{p'}^r > y_{p'}^{r'}$, and so the second inequality can be satisfied only if the crime rates are such that

$$K^{r,c}[S^*(P, r, c)] < K^{r',c'}[S^*(P, r', c')].$$

Note that in our model the hit rate, that is, the likelihood that a search of group (r, c) yields a find, coincides with that group's crime rate $K^{r,c}$. Thus, the implications for the crime rate translate into testable implications for the hit rates. This observation yields theorem 1.

For use in Section 2, we now define the function

$$K_i^r(P) = K^{r,c}[S^*(P, r, c)].$$

Note that the left-hand side lacks a c argument, which makes sense because, by equation (A2), we have $K^{r,c}[S^*(P, r, c)] = K^{r,c'}[S^*(P, r, c')]$ for all c, c' . By the same argument, it is proper to omit reference to characteristics c in the model of Section 2.

We acknowledge that the test is based implicitly on the assumption that the only goal of stops is to make arrests. To the extent that stops also serve other functions (for example, deterrence), and this function is not achieved through an arrest, then the test might indicate discrimination where there is none.

APPENDIX B: ADDITIONAL TABLES AND FIGURES

Table B1. Arrest Made, Including Hispanics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
African American	-.420** (.037)	-.436** (.037)	-.436 (.469)	.279** (.043)	.247** (.043)	.247 (.187)	.235 (.185)
Hispanic	-.119** (.039)	-.149** (.039)	-.149 (.344)	-.004 (.043)	-.029 (.043)	-.029 (.164)	-.000 (.161)
Constant	6.140** (.034)						
Mean outcome (%)				5.86			
% African American				56.1			
% Hispanic				33.2			
P-value				.001	.001	.001	.001
Year fixed effects	No	Yes	Yes	No	Yes	Yes	Yes
Precinct fixed effects	No	No	No	Yes	Yes	Yes	Yes
Year × precinct fixed effects	No	No	No	No	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of being arrested conditional on being stopped in New York City (in %). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level in columns 3, 6, and 7. The *P*-value is for the joint test of all the precinct fixed effects equal to 0. *N* = 4,413,566 observations.

** Significant at the 1% level.

Table B2. Arrest Made, Including Hispanics and Controlling for the Suspect's Recorded Type of Crime

	(1)	(2)	(3)
African American	-.582 (.424)	.138 (.196)	.140 (.194)
Hispanic	-.095 (.322)	.046 (.172)	.069 (.169)
Mean outcome (%)		5.84	
% African American pedestrians		56.4	
% Hispanic pedestrians		33.1	
P-value		.001	.001
Precinct fixed effects	No	Yes	Yes
Crime fixed effects	Yes	Yes	Yes
Year \times precinct fixed effects	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of being arrested conditional on being stopped in New York City (in %). All regressions include year fixed effects and 13 indicators of the suspect's recorded type of crime representing 95% of the crimes (as recorded by the officer on Form UF-250). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level. The *P*-value is for the joint test of all the precincts fixed effects equal to 0. *N* = 3,733,833 observations.

Table B3. Arrest Made, Controlling for Suspect's Recorded Type of Crime

Model	(1)	(2)	(3)
African American	-.551 (.418)	.233 (.217)	.225 (.214)
Mean outcome (%)		5.76	
% African American pedestrians		84	
P-value		.001	.001
Precinct fixed effects	No	Yes	Yes
Crime fixed effects	Yes	Yes	Yes
Year \times precinct fixed effects	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of being arrested conditional on being stopped in New York City (in %). All regressions include year fixed effects and 13 indicators of the suspect's recorded type of crime representing 95% of the crimes (as recorded by the officer on Form UF-250). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level. The *P*-value is for the joint test of all the precinct fixed effects equal to 0. *N* = 2,496,267 observations.

Table B4. Reasons for Arrest: Sample Means (%)

	All Arrests (1)	African American (2)	White (3)
Possession of marijuana	11	11	10
Trespass	9.4	10	6.2
Possession of controlled substance	8.3	7.6	12
Possession of weapon	7.7	7.9	6.9
Assault	4.5	4.2	6.1
Robbery	4.3	4.7	2.2
Larceny	3.9	3.7	4.6
Burglary	3.3	3.4	2.9
Possession of stolen property	1.2	1.1	1.5
Possession of a forged instrument	.71	.74	.56
Menacing	.64	.61	.75
Other	46	46	47

Note. The sample consists of 170,595 arrests in New York City, 2003–12.

Table B5. Frisks, for Pedestrians Suspected of Weapons Possession

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
African American	4.900** (.164)	4.937** (.164)	4.938** (1.265)	2.972** (.184)	2.996** (.184)	2.996+ (1.643)	2.900+ (1.526)
Constant	83.058** (.158)						
Mean outcome (%)	87.64						
% African American	93.5						
P-value				.001	.001	.001	.001
Year fixed effects	No	Yes	Yes	No	Yes	Yes	Yes
Precinct fixed effects	No	No	No	Yes	Yes	Yes	Yes
Year × precincts fixed effects	No	No	No	No	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of being frisked in the sub-sample of stops on suspicion of weapons possession in New York City (in %). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level in columns 3, 6, and 7. The P-value is for the joint test of all the precinct fixed effects equal to 0.

+ Significant at the 10% level.

** Significant at the 1% level.

Table B6. Arrest for Weapons Possession, for Pedestrians Suspected of Weapons Possession and Frisked

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
African American	-2.027** (.069)	-2.023** (.069)	-2.023** (.218)	-1.259** (.079)	-1.263** (.079)	-1.263** (.218)	-1.254** (.215)
Constant	2.619** (.206)						
Mean outcome (%)				1.640			
% African American				93.8			
P-value				.001	.001	.001	.001
Year fixed effects	No	Yes	Yes	No	Yes	Yes	Yes
Precincts fixed effects	No	No	No	Yes	Yes	Yes	Yes
Year × precinct fixed effects	No	No	No	No	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of arrest for weapons possession, conditional on being stopped on suspicion of weapons possession, in the subsample of pedestrians who are stopped and frisked (in %). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level in columns 3, 6, and 7. The *P*-value is for the joint test of all the precinct fixed effects equal to 0. *N* = 581,763 observations.

** Significant at the 1% level.

Table B7. Arrest Made, for Pedestrians Suspected of Weapons Possession and Frisked

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
African American	-3.852** (.114)	-3.844** (.114)	-3.844** (.484)	-2.343** (.130)	-2.352** (.130)	-2.352** (.459)	-2.291** (.450)
Constant	6.728** (.340)						
Mean outcome (%)				4.626			
% African American				93.8			
P-value				.001	.001	.001	.001
Year fixed effects	No	Yes	Yes	No	Yes	Yes	Yes
Precinct fixed effects	No	No	No	Yes	Yes	Yes	Yes
Year × precinct fixed effects	No	No	No	No	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of arrest conditional on being stopped on suspicion of weapons possession in the subsample of pedestrians who are stopped and frisked (in %). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level in columns 3, 6, and 7. The *P*-value is for the joint test of all the precinct fixed effects equal to 0. *N* = 581,763 observations.

** Significant at the 1% level.

Table B8. Summons Issued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
African American	.070 ⁺ (.038)	.095* (.038)	.095 (.360)	-1.753** (.047)	-1.736** (.047)	-1.736** (.295)	-1.721** (.276)
Constant	6.122** (.035)						
Mean outcome (%)				6.18			
% African American				84			
P-value				.001	.001	.001	.001
Year fixed effects	No	Yes	Yes	No	Yes	Yes	Yes
Precinct fixed effects	No	No	No	Yes	Yes	Yes	Yes
Year × precinct fixed effects	No	No	No	No	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of a summons being issued conditional on being stopped (in %). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level in columns 3, 6, and 7. The *P*-value is for the joint test of all the precinct fixed effects equal to 0. *N* = 2,947,865 observations.

⁺ Significant at the 10% level.

* Significant at the 5% level.

** Significant at the 1% level.

Table B9. Summons Issued, Controlling for the Suspect's Recorded Type of Crime

	(1)	(2)	(3)
African American	-.254 (.311)	-1.464** (.293)	-1.487** (.276)
Mean outcome (%)	6.13		
% African American pedestrians		84	
P-value		.001	.001
Precinct fixed effects	No	Yes	Yes
Crime fixed effects	Yes	Yes	Yes
Year × precinct fixed effects	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of a summons being issued conditional on being stopped (in %). All regressions include year fixed effects and 13 indicators of the suspect's recorded type of crime representing 95% of the crimes (as recorded by the officer on Form UF-250). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level. The *P*-value is for the joint test of all the precinct fixed effects equal to 0. *N* = 2,496,267 observations.

** Significant at the 1% level.

Table B10. Arrest Made and Summons Issued: Weighted

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
African American	-.322** (.030)	-.331** (.030)	-.331 (.379)	-.048 (.037)	-.063+ (.037)	-.063 (.177)	-.072 (.171)
Constant	6.136** (.028)						
Mean outcome (%)				5.86			
% African American pedestrians				84			
P-value				.001	.001	.001	.001
Year fixed effects	No	Yes	Yes	No	Yes	Yes	Yes
Precinct fixed effects	No	No	No	Yes	Yes	Yes	Yes
Year × precinct fixed effects	No	No	No	No	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is $\pi(\alpha) = \alpha \times I_{\text{arrest}} + (1 - \alpha) \times I_{\text{summons}}$, the weighted sum of the probability of being arrested and the probability of a summons being issued conditional on being stopped (in %). The weights (α , $1 - \alpha$) are .8 and .2. Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level in columns 3, 6, and 7. The *P*-value is for the joint test of all the precinct fixed effects equal to 0. *N* = 2,947,867 observations.

+ Significant at the 10% level.

** Significant at the 1% level.

Table B11. Arrest Made and Summons Issued, Controlling for the Type of Crime: Weighted

	(1)	(2)	(3)
African American	-.492 (.333)	-.107 (.170)	-.118 (.166)
Mean outcome (%)		5.86	
% African American pedestrians		84	
P-value		.001	.001
Precinct fixed effects	No	Yes	Yes
Crime fixed effects	Yes	Yes	Yes
Year × precinct fixed effects	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is $\pi(\alpha) = \alpha \times I_{\text{arrest}} + (1 - \alpha) \times I_{\text{summons}}$, the weighted sum of the probability of being arrested and the probability of a summons being issued conditional on being stopped (in %). The weights (α , $1 - \alpha$) are .8 and .2. All regressions include year fixed effects and 13 indicators of the suspect's recorded type of crime representing 95% of the crimes (as recorded by the officer on Form UF-250). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level. The *P*-value is for the joint test of all the precinct fixed effects equal to 0. *N* = 2,496,267 observations.

APPENDIX C: REPLICATION WITH THE MANDATED-REPORTS SAMPLE

To identify mandated reports, we follow the definition in New York State (1999). These so-called mandated reports are stops that involve physical force, a frisk, an arrest, or the pedestrian's refusal to provide identification. The NYPD data, however, have a limitation: the outcome "refused to identify" is not recorded in the data. We proxy for it using the field "evasive response to questioning."

Table C1. Descriptive Statistics (%): Mandated Stops

	Mean	SD
Arrest made	.10	.30
African American	.87	.34
Recorded crime:		
Possession of a weapon	.4	.49
Robbery	.19	.39
Criminal trespass	.056	.23
Grand larceny auto	.065	.25
Burglary	.066	.25
Grand larceny	.037	.19
Assault	.032	.18
Illegal possession of substances	.031	.17
Possession of marijuana	.033	.18
Illegal sales of substances	.023	.15
Petit larceny	.017	.13
Mischief	.0085	.092
Graffiti	.0077	.088
Other	.032	.18

Note. The crime categories represent 95% of the crimes in the sample. Years 2003–2005 have missing values for the recorded crimes. $N = 1,681,600$ observations and 1,480,728 recorded crimes.

Table C2. Arrest Made: Mandated Stops

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
African American	-3.629** (.069)	-3.645** (.069)	-3.645** (.835)	-1.625** (.084)	-1.627** (.084)	-1.627** (.554)	-1.668** (.510)
Constant	13.30448** (.065)						
Mean outcome (%)				10.14			
% African American				87			
P-value				.001	.001	.001	.001
Year fixed effects	No	Yes	Yes	No	Yes	Yes	Yes
Precinct fixed effects	No	No	No	Yes	Yes	Yes	Yes
Year × precinct fixed effects	No	No	No	No	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of being arrested conditional on being stopped (in %). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level in columns 3, 6, and 7. The *P*-value is for the joint test of all the precinct fixed effects equal to 0. *N* = 1,681,600 observations.

** Significant at the 1% level.

Table C3. Arrest Made, Controlling for the Suspect's Recorded Type of Crime: Mandated Stops

	(1)	(2)	(3)
African American	-1.529** (.713)	-.524 (.427)	-.543 (.411)
Mean outcome (%)		9.705	
% African American pedestrians		87	
P-value		.001	.001
Precinct fixed effects	No	Yes	Yes
Crime fixed effects	Yes	Yes	Yes
Year × precinct fixed effects	No	No	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is the probability of being arrested conditional on being stopped (in %). All regressions include year fixed effects and 13 indicators of the suspect's recorded type of crime representing 95% of the crimes (as recorded by the officer on Form UF-250). Regressions with year fixed effects (nine dummies) and precinct fixed effects for 76 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level. The *P*-value is for the joint test of all the precinct fixed effects equal to 0. *N* = 1,480,728 observations.

** Significant at the 1% level.

Table C4. Correlates of Relative Police Pressure: Mandated Stops

	(1)	(2)	(3)
% African American	-.273** (.088)	-.061 (.072)	-.095 (.063)
Income		.469** (.141)	.435** (.135)
Constant	28.885** (4.703)	-5.201 (8.154)	-13.098 (30.915)
Average relative police pressure		21.58	
% African Americans in average precinct		26.78	
Adjusted R^2	.0793	.267	.470
Precinct controls	No	No	Yes
Year fixed effects	No	Yes	Yes

Note. Estimates are from ordinary least squares regressions. The dependent variable is (relative) police pressure (arrests of African Americans/African American population)/(arrests of whites/white population) in New York City. Column 3 includes a variable for the margin of Bloomberg's victory. Missing years are computed using moving averages for the variables for the fraction of African Americans, income, age, fraction of females, fraction of college degrees, serious crime, graffiti, social capital, and African American commanding officers. Regressions with year fixed effects (nine dummies) and precinct fixed effects for 75 precincts (75 dummies) control for a possible time trend in the dependent variable and precinct-specific characteristics, respectively. Standard errors, in parentheses, are clustered at the precinct level. $N = 750$ observations.

** Significant at the 1% level.

APPENDIX D

Table D1. Variables, Descriptions, and Sources

Variable	Description	Source
African American	Indicator variable coding whether the pedestrian is African American	NYPD (2003–12)
Hispanic	Indicator variable coding whether the pedestrian is Hispanic	NYPD (2003–12)
Relative Police Pressure	(Arrests of African Americans/African American population)/(arrests of whites/white population) in each of the New York City police precincts	NYPD (2003–12)
Margin of Bloomberg Victory	Difference in vote share between Mayor Michael Bloomberg and his closest opponent Mark Green, Fernando Ferrer, or Bill Thompson in the 2001, 2005, 2009 elections, respectively; missing years are computed using moving averages	NYC Board of Elections ^a
% African American	Percentage of the population that is African American in the precinct in 2010	US Census Bureau ^b
Income	Precinct average inflation-adjusted median income in 2010	US Census Bureau ^b
Age	Precinct average median age in 2010	US Census Bureau ^b
% Female	Precinct average percentage of the population that is female in 2010	US Census Bureau ^b
% College Degree	Precinct average percentage of the population ages 15–24 with a college degree in 2010	US Census Bureau ^b

Serious Crime	Number of annual crimes (murders, rapes, robberies, felony assaults, burglaries, grand larcenies, grand larcenies of autos) in each precinct divided by the precinct's population in 2010 (in 1,000 inhabitants) for the years 1998, 2001, 2012; missing years are computed using moving averages	New York Police Department crime statistics ^c
Graffiti	Number of sites with graffiti in each precinct in 2011 divided by the precinct's population in 2010 (in 1,000s of inhabitants)	NYC Open Data ^d
Social Capital	Number of annual civic initiatives (education, emergency preparedness, environment, helping neighbors in need, strengthening communities) in each precinct in 2011 divided by the precinct's population in 2010 (in 1,000s of inhabitants)	NYC Open Data ^e

^a The original data are electoral district data (for districts in 2001, 2005, 2009) matched with police precincts using the Stata routine `gpsmap.ado` (New York City Board of Elections, Election Results Summary 2001, 2005, 2009 [<http://vote.nyc.ny.us/html/results/results.shtml>]).

^b The original data are ZIP-code-level data matched with police precincts using the Stata routine `gpsmap.ado` (US Census Bureau, American Fact Finder [<http://factfinder2.census.gov>], 2007–2011 American Community Survey 5-Year Estimates, tables B02001, B19013, B01002, B01001, S1501; johnkeefe.net, Sharing NYC Police Precinct Data, NYC_Blocks_2010CensusData_Plus_Precincts [<https://www.google.com/fusiontables/DataSource?dsrclid=767562#rowsid=1>], blog entry by John Keefe, April 29, 2011).

^c The original data are data matched with police precincts by the New York Police Department (New York Police Department, Historical New York City Crime Data [http://www.nyc.gov/html/nypd/html/analysis_and_planning/historical_nyc_crime_data.shtml]).

^d The original data are data matched with police precincts by the New York Police Department (NYC Open Data, NYC Graffiti [<https://data.cityofnewyork.us/Social-Services/nyc-graffiti/8q69-4ke5f>]).

^e The original data are latitude-longitude geo-coded data matched with police precincts using the Stata routine `gpsmap.ado` (NYC Open Data, NYC Service Volunteer Opportunities [<https://data.cityofnewyork.us/Social-Services/NYC-Service-Volunteer-Opportunities/bquu-z2htl>]).

REFERENCES

- Alpert, Geoffrey P., John M. MacDonald, and Roger G. Dunham. 2005. Police Suspicion and Discretionary Decision Making during Citizen Stops. *Criminology* 43:407–34.
- Anwar, Shamena, and Hanming Fang. 2006. An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence. *American Economic Review* 96:127–51.
- Associated Press. 2013. California: Bratton Hired as Consultant. *New York Times*, January 23. <http://www.nytimes.com/2013/01/24/us/california-bratton-hired-as-consultant.html>.
- Ayres, Ian. 2002. Outcome Tests of Racial Disparities in Police Practices. *Justice Research and Policy* 4:131–42.
- Ayres, Ian, and Joel Waldfogel. 1994. A Market Test for Race Discrimination in Bail Setting. *Stanford Law Review* 46:987–1047.
- Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Celona, Larry. 2012. NYPD Issues Department-Wide Memo regarding Racial Profiling during “Stop-and-Frisks.” *New York Post*, May 17. <http://nypost.com/2012/05/17/nypd-issues-department-wide-memo-regarding-racial-profiling-during-stop-and-frisks>.
- Childers, Sean. 2012. Discrimination during Traffic Stops: How an Economic Account Justifying Racial Profiling Falls Short. *New York University Law Review* 87:1025–59.
- Coviello, Decio, and Nicola Persico. 2013. An Economic Analysis of Black-White Disparities in NYPD’s Stop and Frisk Program. Unpublished manuscript. HEC Montreal, Department of Applied Economics, Montreal.
- Dharmapala, Dhammika, and Stephen L. Ross. 2004. Racial Bias in Motor Vehicle Searches: Additional Theory and Evidence. *BE Journal of Economic Analysis and Policy* 3:1–23.
- Feith, David. 2013. William Bratton: The Real Cures for Gun Violence. *Wall Street Journal*, January 18. <http://www.wsj.com/articles/SB10001424127887323968304578246721614388346>.
- Fyfe, James J., and Robert Kane. 2005. Bad Cops: A Study of Career-Ending Misconduct among New York City Police Officers. Document No. 215795. Report to the Department of Justice, National Institute of Justice. <https://www.ncjrs.gov/pdffiles1/nij/grants/215795.pdf>.
- Gelman, Andrew, Jeffrey Fagan, and Alex Kiss. 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: 813–23.
- Gershman, Bennett L. 2000. Use of Race in “Stop-and-Frisk”: Stereotypical Beliefs Linger, but How Far Can the Police Go? *New York State Bar Association*

- Journal*, April, pp. 42–45.
- Goldstein, Joseph. 2013. Judge Rejects New York's Stop-and-Frisk Policy. *New York Times*, August 12. <http://www.nytimes.com/2013/08/13/nyregion/stop-and-frisk-practice-violated-rights-judge-rules.html>.
- Hernández-Murillo, Rubén, and John Knowles. 2004. Racial Profiling or Racist Policing? Testing in Aggregated Data. *International Economic Review* 45:959–89.
- Klasfeld, Adam. 2012. Court Strikes Challenge to Stop-and-Frisk Trial. *Courthouse News Service*, October 11. <http://www.courthousenews.com/2012/10/11/51188.htm>.
- Knowles, John, Nicola Persico, and Petra Todd. 2001. Racial Bias in Motor Vehicle Searches: Theory and Evidence. *Journal of Political Economy* 109:203–29.
- Leland, John, and Colin Moynihan. 2012. Thousands March Silently to Protest Stop-and-Frisk Policies. *New York Times*, June 17. <http://www.nytimes.com/2012/06/18/nyregion/thousands-march-silently-to-protest-stop-and-frisk-policies.html>.
- Levitt, Steven. 1997. Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime. *American Economic Review* 87:270–90.
- McCrary, Justin. 2002. Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment. *American Economic Review* 92:1236–43.
- New York Civil Liberties Union. 2012. Stop-and-Frisk 2011. New York Civil Liberties Union Briefing, May 9. New York Civil Liberties Union, New York. http://www.nyclu.org/files/publications/NYCLU_2011_Stop-and-Frisk_Report.pdf.
- New York State. Office of the Attorney General. 1999. Report on an Investigation into the New York Police Department's "Stop and Frisk" Program. New York State, Office of the Attorney General, New York. http://www.oag.state.ny.us/sites/default/files/pdfs/bureaus/civil_rights/stp_frsk.pdf.
- New York Times*. 2013. Racial Discrimination in Stop-and-Frisk. August 12. <http://www.nytimes.com/2013/08/13/opinion/racial-discrimination-in-stop-and-frisk.html>.
- NYPD (New York Police Department). 2003–12. Stop, Question, and Frisk Database. http://www.nyc.gov/html/nypd/html/analysis_and_planning/stop_question_and_frisk_report.shtml.
- Persico, Nicola. 2002. Racial Profiling, Fairness, and Effectiveness of Policing. *American Economic Review* 92:1472–97.
- . 2009. Racial Profiling? Detecting Bias Using Statistical Evidence. *Annual Review of Economics* 1:229–54.
- Persico, Nicola, and David A. Castleman. 2005. Detecting Bias: Using Statistical Evidence to Establish Intentional Discrimination in Racial Profiling Cases. *University of Chicago Legal Forum*, pp. 217–35.
- Persico, Nicola, and Petra E. Todd. 2005. Passenger Profiling, Imperfect Screening, and Airport Security. *American Economic Review: Papers and Proceed-*

- ings* 95:127–31.
- . 2006. Using Hit Rates to Test for Racial Bias in Law Enforcement: Vehicle Searches in Wichita. *Economic Journal* 116:F351–F367.
- . 2008. The Hit Rates Test for Racial Bias in Motor-Vehicle Searches. *Justice Quarterly* 25:37–53.
- Ridgeway, Greg. 2007. Analysis of Racial Disparities in the New York Police Department's Stop, Question, and Frisk Practices. Technical report. RAND Corporation, Santa Monica, CA. http://www.rand.org/pubs/technical_reports/TR534.
- Ruderman, Wendy. 2012. For Women in Street Stops, Deeper Humiliation. *New York Times*, August 6. <http://www.nytimes.com/2012/08/07/nyregion/for-women-in-street-stops-deeper-humiliation.html>.
- Sanga, Sarath. 2009. Reconsidering Racial Bias in Motor Vehicle Searches: Theory and Evidence. *Journal of Political Economy* 117:1155–59.
- Sharpton, Al. 2012. “Stop-and-Frisk” Is the New Racial Profiling. *Huff Post New York*, June 4. http://www.huffingtonpost.com/rev-al-sharpton/stop-frisk-is-the-new-rac_b_1569201.html.
- Smith, Michael R., Matthew Makarios, and Geoffrey P. Alpert. 2006. Differential Suspicion: Theory Specification and Gender Effects in the Traffic Stop Context. *Justice Quarterly* 23:271–95.
- Todd, Petra E. 2008. Racial Profiling. Pp. 1–13 *The New Palgrave Dictionary of Economics*, edited by Steven N. Durlauf and Lawrence E. Bloom. 2d ed. New York: Palgrave Macmillan.
- Whitney, Melissa. 2008. Statistical Evidence of Racial Profiling in Traffic Stops and Searches: Rethinking the Use of Statistics to Prove Discriminatory Intent. *Boston College Law Review* 49:263–99.