How Do NYPD Officers Respond to Terror Threats?

By Steven F. Lehrer† and Louis-Pierre Lepage‡

†*Queen's University and NBER* ‡*University of Michigan*

Final version received 29 October 2019.

Using data from the Stop-and-Frisk programme of the New York Police Department (NYPD), we evaluate the impact of a specific terrorist attack threat from Al Qaeda on policing behaviour in New York City. We find that after the Department of Homeland Security raised the alert level in response to this threat, people categorized as 'Other' by the NYPD, including Arabs, were significantly more likely to be frisked and have force used against them, yet were no more likely to be arrested. These individuals were in turn less likely to be frisked or have force used against them immediately after the alert level returned to its baseline level. Further, evidence suggests that these impacts were larger in magnitude in police precincts that have higher concentrations of mosques. Our results are consistent with profiling by police officers leading to low-productivity stops, but we cannot rule out that it constitutes efficient policing given important differences between deterrence of terrorism versus other crimes.

Introduction

Following the terrorist attacks on 11 September 2001 (known as '9/11'), law enforcement's focus has evolved to also include domestic terrorist threats. A large number of plots have been foiled, including two attempts to bomb the New York City (NYC) Subway and a car bombing attempt in Times Square on 1 May 2010.[1] Yet survey research shows that while most people are generally satisfied with the way in which the police perform their duties, beliefs have grown that individual police officers treat suspects of different races or ethnicities differently.[2] These potential racial differences have attracted substantial controversy, and in response many police departments introduced policies to deter racial profiling in order to reduce these beliefs and restore public support.

Being a police officer can be challenging, particularly when faced with the additional challenges posed by terrorism. Police officers occasionally have to make life-and-death split-second decisions that may leave room for unconscious racial bias to arise, regardless of training or department policies and procedures. Evidence of unconscious racial bias among police officers is summarized by Fridell (2008, 2016), who notes that while certain interventions can reduce associations between race or ethnicity and crime, they do not fully eliminate them.

More generally, theories of statistical discrimination often posit that racial differences in policing arise at least partly due to the effect of perceptions of identity.[3] A key implication from such models is that differentials in outcomes by race or ethnicity may be quite responsive to changes in perceptions. To empirically test this prediction requires a salient and exogenous event that could plausibly alter an individual police officer's racial perceptions of a group. In this paper, we argue that an announcement made by the Department of Homeland Security on 1 August 2004 provided such salient information to law enforcement officials. The announcement warned of an immediate terrorist threat to financial institutions in New York, Washington, DC, and New Jersey, and was accompanied by a change in the colour-coded Homeland Security Advisory System (HSAS) scale. The HSAS was designed to warn citizens and lead to additional security measures taken by various government institutions, generally including both federal agencies and law enforcement.

We use detailed data from the Stop-and-Frisk programme of NYC to examine whether this increase in the alert level from yellow (elevated risk) to orange (high risk) had an impact on police officer behaviour based on the race or ethnicity of alleged perpetrators. We conjecture that racial minorities from the Middle East and North Africa may have received additional scrutiny given the association between Al Qaeda and those regions.[4] This conjecture is motivated by the intuitive notion that terrorism is an area where individuals may be particularly prone to making assumptions regarding the ethnicity of perpetrators.[5] Similarly, any unconscious biases held by individual police officers may likely emerge in situations where threats and stakes are higher; and these situations may lessen the criteria of reasonable suspicion needed to stop and frisk individuals irrespective of their race or ethnicity.

Our empirical tests not only examine if the rate of stops leading to an arrest differed after the HSAS announcement,[6] but also explore if it altered police officer behaviour in terms of the number of stops as well as whether they frisked the suspects or used force as part of these stops. Complementary to Draca *et al.* (2011), who study how the July 2005 London terror attacks influenced police deployment and reported crime levels across districts, we focus on differential individual officer responses by suspect race or ethnicity. We motivate our empirical test with a model that extends the hit rates test from Knowles *et al.* (2001) to consider the addition of terrorism-related duties, how terrorist threats may affect perceptions of officers, and the possibility that deterring a terrorist attack offers larger benefits to a police officer than other crimes.

Our analysis uncovers that the 1 August 2004 increase in the alert level led to a substantial and statistically significant 13% increase in the probability of being frisked for people characterized as 'Other', including Arabs. This constitutes a 6% additional increase in frisking compared to other racial groups, and we also find evidence of a disproportionate 9% increase in the use of force, yet no higher odds of them being arrested or stopped. Further evidence suggests that these increases in frisking were particularly large during rush-hour times and in areas with higher concentrations of mosques. When the threat subsided, the relative probability of being frisked or have force used against them in turn decreased for this group. A natural question is whether it is possible to distinguish between statistical and taste- or bias-based discrimination in this setting. Although the additional frisks of Arabs led to no additional arrests, which points towards an irrational explanation, caution is warranted when interpreting our results. The results are consistent with the idea that making threats salient can bring out potentially latent discrimination, which appears inefficient in our application since there was no increase in arrests. Still, given important differences between terrorism and other crimes as well as the larger stakes at play, we cannot rule out that these changes in behaviour constitute efficient policing.

This paper has a natural link to two contentious literatures on law enforcement related to racial bias in the Stop-and-Frisk programme and the effects of Homeland Security terror alerts. On the latter, Shapiro and Cohen (2007) note that the HSAS was not viewed in a good light by citizens, who claimed that it was vague and uninformative regarding details of potential threats.[7] Changes in alert levels have been examined by Klick and Tabarrok (2005), who find that crime decreased in Washington, DC, during high-alert periods prior to July 2003 due to increased police presence, and by Omer *et al.* (2007), who find that changes in alert levels did not increase stress levels as proxied by calls to a law-enforcement peer-support hotline. On the former, despite front page headlines and criticism in the popular press,[8] Coviello and Persico (2015) do not find evidence of racial discrimination in the Stop-and-Frisk programme of the New York Police Department (NYPD) when using the hit rates test, but along with Goel *et al.* (2016), present evidence supportive of discrimination against African-Americans when restricting attention to stops relating to the possession of a concealed weapon. Lehrer and Lepage (2019) also use the hit rates test and additionally account for the fact that the category of stop reported reflects a behavioural choice of police officers. They report evidence of discrimination for crimes related to weapons and drugs. While this paper does not answer questions of how police officers should be monitored,[9] it provides relevant evidence on how their behaviour adapts to exogenous changes in information on criminality by racial or ethnic group.

The remainder of the paper is organized as follows. In Section I, we provide an overview of the data and research design employed to identify changes in policing. Our empirical results are presented in Section II, along with robustness checks and a discussion on interpreting our results. Finally, Section III summarizes our main findings and concludes.

I. Empirical Setting and Test

In this section, we first describe the event that provides an exogenous information shock to police officer perceptions. We then describe the dataset utilized, and provide an outline of the empirical strategy.

The terror alert of August 2004

The HSAS for terrorism was activated on 12 March 2002, and the threat level was raised six times from 'yellow' to 'orange', corresponding to a move from elevated to high threat.[10] The HSAS increase that took place on 1 August 2004 was accompanied by a specific warning that identified financial institutions in New York, Washington, DC, and New Jersey as being targeted by Al Qaeda.[11] This was the only increase of the six total alert increases that was specific about the potential targets and also explained the sources and quality of the intelligence on which the threat was based.[12] The increase from yellow to orange is also particularly important in our context since orange is the first level to explicitly require the coordination of security efforts with local law enforcement agencies, including the NYPD. Further, Morris (2003) reports that the NYPD commissioner mentioned that an orange alert level generally meant an 'orange plus' level for the city, which led to additional security measures such as increased security at public events and important locations, additional police presence in mass transit systems, and additional checkpoints at bridges, tunnels and other locations.

Since the alert had a marked emphasis on NYC, it may have been particularly salient to NYPD officers. It was also particularly long in that the orange status was maintained for nearly 100 days. The New York Police Commissioner Raymond Kelly publicly stated that that the NYPD would respond by providing 'significant security' at selected buildings and would step up both random and targeted searches of vehicles entering the city. Since the information concerning the threat suggested an attack by car or truck bomb, several streets near financial centres in midtown Manhattan were initially closed, and trucks were banned from entering bridges and tunnels leading to Wall Street.

The head of the Department of Homeland Security, Tom Ridge, mentioned being very confident in the information, some of which had been seized from an Al Qaeda operative as part of a CIA–Pakistani operation in Pakistan not long before and included detailed planning information as well as sketches of potential targets, including the Stock Exchange and Citigroup Center in New York. Further, information was found that suggested scouting by potential terrorists had recently been done to identify security in and around specific buildings, and midweek pedestrian traffic counts of the number of people per minute on each side of the street, identifying the best places for further reconnaissance, tips on how to make contact with employees who work in the buildings, general traffic patterns, and locations of hospitals and police departments.[13] The alert level was lowered to yellow on 10 November 2004 once permanent protective measures were put in place around specific locations throughout the financial services sector.

Even with these additional measures taken by the NYPD, it appears unlikely that increased targeting of Arabs during the period would be the product of a (likely illegal) top-down policy representing a conscious, concerted effort by the NYPD. While officers were certainly encouraged to be vigilant and may have been directed to increase stops and searching of suspicious individuals, we found no evidence of any explicit targeting, and NYC police commissioner Raymond Kelly has referred to racial profiling as a counter-terrorism measure as being 'just nuts'.[14]

Data

Our primary data are gathered from NYC Open Data and comprise all recorded stops from the Stop-and-Frisk programme between 2003 and 2012, which we restrict to stops from 2003 to the end of the HSAS in April 2011. For every stop, we are provided with detailed personal characteristics of the suspect, including age, gender, ethnicity, height, weight, hair and eye colour, plus the date and location of the stop as well as detailed information about the type of crime, weapons found and whether force was used.[15] In addition, we are provided with the officer's self-reported basis of search and reason for frisk as well as whether an arrest was made or summons issued to the suspect. The race/ethnicity categories used by the NYPD for the suspects are White, Black, White Hispanic, Black Hispanic, Asian, Unrecognized, American-Indian and Other.[16] The race/ethnicity category, along with other stop information, is reported by the police officer who performed the stop. We focus on 'Other' as the group of interest in our application, and pool Whites, Asians and Hispanics as the reference group. While it is not possible to specifically identify which racial groups enter in 'Other' or in what proportion, these are unlikely to change due to alert increases. Further, considering the different categories, it seems unlikely that estimated impacts of alert changes would be driven by racial subgroups other than Arabs and people from North Africa and the Middle East potentially contained in the 'Other' category. According to the 2010 Census, the New York Metro Area had the second largest Arab-American population in the USA, with 371,233 residents, an estimated 1.68% of the population, which could account for up to 40% of the group 'Other'. Following Coviello and Persico (2015) and other papers using the Stop-and-Frisk data, we consider these data representative of all stops even though police officers are not required to report those that do not involve the use of force or lead to a frisk, search, arrest or summons.[17] Our final sample restricted to the treatment and comparison groups of interest consists of 1,792,781 unique stops.

Table 1 presents summary statistics on the average number of daily stops and outcomes of interest. The average number of stops per day was 595 over the 3012 days studied in our sample, of which about 9.4% involved members of the racial group 'Other'. Throughout our analysis, we rescale the binary outcome variables to 100 or 0, therefore allowing coefficient estimates to be interpreted as direct percentage point effects. About half of the stops in the sample (48.42%) involved frisking for all racial groups combined, while the mean for suspects of the racial group 'Other' was slightly lower, at approximately 45.1%. The likelihood of an arrest is on average 5.82% overall and 4.2% for Others. Finally, unconditionally there is little difference in the likelihood of force being used by the NYPD between the two groups, the daily mean being around 23% for both.

Table 1 near here

Effect on racial profiling by police

In the Appendix, we provide a stylized model outlining how exogenous changes in terror alert levels could influence a variety of policing behaviours. It differs from the original hit rates model of Knowles *et al.* (2001) by allowing the behaviour of terrorists to differ from that of other criminals, mainly by considering the case where they may not be deterred by the threat of apprehension. The model generates clear testable predictions on how police officers respond to alert level increases by increasing the relative number of stops and the probability of frisking Arabs relative to suspects of other ethnicities. It provides a useful lens through which to interpret the results, but does not allow us to differentiate between statistical and taste- or bias-based discrimination.

Our model, along with the data structure and plausibly exogenous variation from the HSAS, suggests that a statistical comparison of time trends before and after the intervention will provide an unbiased estimate of the causal effect of the change in the HSAS level. Since the data consist of a time series, we collapse the data at the daily level and adopt an interrupted time series design that additionally accounts for secular trends. The simple logic of the time series experiment is analogous to a regression discontinuity (RD) design, in that if the graph of the dependent variable shows an abrupt shift in level or trend precisely at the point of the event, then the event is a cause of the effect on the dependent variable.[18]

Given the institutional details of the HSAS, we consider the following estimating equation to analyse the impact of the NYC-specific alert change on policing outcome Y_{it}:

$$\begin{aligned}(1)\qquad Y_{ipt} &= \beta_0 + \beta_1\, AI_t + \beta_2\, OTHER_i + \beta_3(AI_t * OTHER_i) + \beta_4\, AD_t \\ &+ \beta_5(AD_t * OTHER_i) + \beta_6 T_t + \beta_7(T_t * OTHER_i) + \beta_8 \sum_{l=1}^{L} Y_{i,t-l} \\ &+ \beta_9\, TIME_t + \beta_{10}\, PRECINCT_p + \beta_{11}(PRECINCT_p * YEAR_t) + \varepsilon_{ipt},\end{aligned}$$

where AI_t is an indicator variable corresponding to the period after the alert increase at time t, AD_t is an indicator variable corresponding to the period after the decrease in the alert, and the interaction terms $AI_t * OTHER_i$ and $AD_t * OTHER_i$ capture additional changes in response by police officers on those categorized as Others. Throughout, we control for a comprehensive set of calendar time and precinct fixed effects as well as their interaction. In our most general specifications, we also capture state dependence by conditioning on a set of lagged dependent variables $\sum_{l=1}^{L} Y_{i,t-l}$, which accounts for potential autocorrelation. The number of lags to include was selected using the most parsimonious specification suggested by either the Akaike information criterion or the Bayesian information criterion. The primary coefficients of interest are given by β_3 and β_5, which capture the additional jumps in the probability of an outcome for Others when compared to the comparison group from respectively increasing or decreasing the threat level. Equation (1) also allows us to test whether the behavioural responses of police officers offset as alert levels change in opposite directions.[19]

II. Results

We first provide graphical evidence of how alert levels influence dimensions of policing behaviour. We regressed each outcome variable on a linear trend, a set of month, year, precinct, day of week and year by precinct fixed effects. We then computed the residuals at the monthly level for those categorized as Others and the comparison group, plotted over time in Figure 1. Intuitively, it illustrates the portion of the variation in each outcome variable that could not have been predicted using predetermined characteristics. We would expect the relationship between our residualized outcomes and the calendar date to be flat, except for the potential impacts of alert level increases. Looking across the panels in Figure 1, we first observe that the NYC alert seems to have had little impact on policing for outcomes related to the number of stops and the arrest rate. In contrast, the remaining panels suggest that it was associated with disproportionate increases in the rate of frisk and force for Others compared to the comparison group. Finally, the panels in this figure also provide graphical evidence suggestive of the validity of the common trends assumption necessary for the validity of our design. All four residualized policing outcomes follow similar trends throughout the sample for the treatment and comparison groups outside of the higher alert periods.

Figure 1 near here

Table 2 presents estimates of equation (1) where the outcomes are the daily number of stops per precinct in panel A and the probability of being frisked in panel B. The columns differ on the basis of whether time and precinct fixed effects as well as lagged outcomes are included, and whether heteroscedasticity- and autocorrelation-consistent standard errors calculated using the test proposed in Cumby and Huizinga (1992) are reported. Lagged values of the dependent variable are included to account for path dependency, which captures the idea that police behaviour may be persistent over time depending on local conditions and not react immediately to a changing environment. Adding these lagged values often reduces the magnitude of other estimated coefficients, but helps to ensure that the remaining effect of alert changes is a contemporaneous response to the information shock. Adjusted standard errors account for potential serial correlation in the error term that may arise given the time series nature of our data.

Table 2 near here

In panel A of Table 2, we consider the daily number of stops per precinct as a proxy for policing activity since we do not have rich data on deployment like that in Draca *et al.* (2011). Across all four columns we observe that the NYC-specific alert increase leads to a decrease in the number of stops, which is consistent with the police shifting resources to other terrorism-prevention tasks. Further, we find that when considering only contemporaneous variables, the interaction term between 'Other' and the alert increase suggests a relative increase of around 4% in the number of stops for Others. The magnitude of the interaction term shrinks to a statistically insignificant 1% increase once lagged outcome values are included to capture state dependence in the number of stops.

Regarding frisking in panel B of Table 2, following the NYC-specific alert, all columns provide evidence of a statistically significant general increase as well as an additional statistically significant increase in the probability of being frisked for those classified as the racial group 'Other' relative to the comparison group.[20] Based on estimates in columns (3) and (4), all groups experienced on average a 2.76 percentage point increase (6% increase from the mean) in the probability of being frisked, and members of the group 'Other' experienced an additional 2.66 percentage point (or additional 6%) increase. Interestingly, we also observe a statistically significant differential decrease in being frisked for those in the group 'Other', which closes most of the estimated differential in frisking once the NYC-specific threat subsides and the alert level is decreased to yellow. We cannot reject the null hypothesis that the opposite alert changes offset each other, as indicated by the p-value in panel B of Table 2.

Table 3 presents estimates of equation (1) where the likelihood of an arrest and use of force are the dependent variables. In panel A, we observe no statistically significant effects on arrests in general, and no differential effect for Others. It is worth pointing out that the hit rates test introduced by Knowles *et al.* (2001) is nested within our specification, and in this case it would provide evidence of discrimination against members of the group 'Other' since there is a statistically significant lower baseline arrest rate (nearly 30% lower) for that group compared to the comparison group.

Table 3 near here

Turning to panel B of Table 3, we find that following the NYC-specific increase, there is little change in overall use of force but a statistically significant differential increase in the likelihood that force is used on the racial group 'Other' relative to the comparison group. The magnitude for the NYC-specific alert level corresponds to an additional increase of 9% for Others. Again, most of this differential dissipates after the alert level is decreased, and we cannot reject that both changes offset.

As a whole, both the graphical and regression evidence discussed above suggest that the NYC alert did not change the productivity of stops. We do not observe any substantial change in arrest rates even when officers disproportionately frisk Others and use force against them. The apparent decrease in the number of stops following the alert increase could mean that officers were more selective about which stops to make given their additional responsibilities, but this would in turn be expected to lead to an increase in the arrest rate, for which there is no evidence. Rather, the evidence is consistent with officers being no more selective but more 'thorough' or forceful in their stops, especially for Others.

If we additionally make the assumption that the impact of the NYC-specific alert was focused on Arabs, then the estimated differential impacts on this subgroup using the 2010 Census estimate for the Arab population of NYC would be of nearly 15% for frisks and 23% for force used. Thus the analysis presents robust evidence that the NYC-specific alert increase influenced policing behaviour differentially by racial group.

To shed additional light on the disparate increase in the probability of being frisked for Others after the NYC-specific alert, we present regression estimates by precinct and time of day to assess whether there was systematic geographic or temporal heterogeneity in our results.[21] On the former, we carried out an analysis at the precinct level, shown in Figure 2, which illustrates in darker shade the precincts that observed larger-than-average disparate increases in the likelihood that Others were frisked. Figure 2 also provides the locations of mosques in the city as a proxy for the concentration of Arab-Muslims.[22] The interpretation of this figure is *a priori* theoretically unclear. On the one hand, we may expect larger impacts in precincts where the Arab population is lower since police officers may be more likely to have unconscious biases against people with whom they interact less. On the other hand, we may expect that the impact would be larger where more Muslims live and congregate, and in precincts with more high-value targets. The former appears to be what is shown in Figure 2 since the impact on frisk appears to be primarily driven by areas with higher concentrations of mosques and in high-priority areas like lower- or mid-Manhattan. To shed further light on how the presence of mosques influences frisking by precinct, we expanded equation (1) to allow for differential responses by the number of mosques in the precinct. The results are shown in Subsection 1.1 of the Online Appendix, and confirm that differential increases in frisking for Others were larger in precincts with more mosques.

Figure 2 near here

Turning to temporal heterogeneity, Subsection 1.2 of the Online Appendix provides evidence that the increases in frisking occurred primarily during rush-hour (5am–10am, 5pm–7pm) and late in the evening (11pm–4am). The rush hour periods represent times where the risk of terrorism threat is greater and potentially deadlier, while activity late at night is more likely to appear suspicious to law enforcement.

Robustness of the main findings

To complement and explore the robustness of both the graphical evidence in Figure 1 and results in Tables 2 and 3, we considered a more general method to explore whether alert levels influenced policing. Specifically, we estimated a restricted version of equation (1) that does not include indicators for alert changes or their interactions with other covariates. In their place, we ran the model repeatedly with a single indicator for each two-week or one-month interval in the sample, along with an interaction of this indicator with the racial category 'Other'. This exercise has many uses; it allows us to observe the timing of any jump and see whether it corresponds to an actual alert level change, further examines pre-alert trends, investigates the length of policing responses, and conducts a placebo test over the entire sample.

Figure 3 provides the results from this exercise at the two-week level where each data point presents the estimate for that specific two-week window.[23] Notice that there is a sustained spike in the frisk differential around the NYC-specific threat. Statistically significant estimates are denoted by numbers above each panel. In total, only four (two) two-week periods are estimated to have had a statistically significant impact on the frisk (force) differential for the entire sample period, including one during the NYC-specific threat.[24] For both outcomes, the increase associated with the alert increase of interest also appears to be much more sustained.

Figure 3 near here

We also considered an RD design relaxing the continuity requirements of the running variable (time) and extrapolating from the discretization around the cut-off given by the dates of alert increases. From a conceptual standpoint, both an interrupted time series analysis and RD are similar in the region of interest as they approximate the conditional expectation on both sides of the cut-off. A graphical analysis as well as the corresponding estimates for the NYC-specific alert are presented in Subsection 1.4 of the Online Appendix.[25] We do not focus on the RD approach given the time series nature of our data, but it serves as a robustness exercise that does not rely on a comparison group for validity. The RD design yields similar conclusions to our main specification for arrests and frisks. For the likelihood of frisk, the estimates for the group 'Other' are positive, large in magnitude (around 13%) and statistically significant, while those for the comparison group are small and statistically insignificant. For arrests, there was little change for either group as the estimates are small and statistically insignificant. For the outcomes of force used and the number of stops, the impacts are also estimated to be small and statistically non-significant for both groups. This contrasts with our main specification, which finds a decrease in overall stops as well as an increase of force used on Others.

We conducted three additional robustness checks. First, we estimated equation (1) at the hourly level with hour of day fixed effects rather than at the day level as shown in Subsection 1.5 of the Online Appendix. We prefer the daily specification given the uncertainty about which time of day the different alerts were announced, but the main results are similar. Second, in Subsection 1.6 of the Online Appendix, we present results from equation (1) where we condition the outcomes on the number of stops. That is, we calculate the daily number of frisks, arrests and events involving force per stop to partial out the effects of the change in stops from the effects of other police behaviour. Our main findings are robust to this empirical strategy. Third, we conducted the analysis using only African-Americans as the comparison group in Subsection 1.7 of the Online Appendix, and the main results are quantitatively similar except for the estimated differential change in the probability of force used for Others, which is smaller and becomes statistically insignificant.

Interpretation of the main results

Following the 9/11 attacks and the 2003 invasion of Iraq, there was much discussion of threats from terrorist organizations such as Al Qaeda, and the HSAS level was often referred to as the 'terror alert level'. Intuitively, the HSAS can be seen as introducing a signal into the policing environment that could influence both police officers and citizens since it affects the tasks and focus of the police force, which may in turn lead to strategic responses from criminals. Additional efforts by police officers in specific areas could deter criminals or have them change the location of their crime.[26] This is less likely to arise in the case of terrorism since potential terrorists may be unlikely to be deterred by the threat of apprehension or change their target after months of planning. As such, unlike for common crimes, allocating more police officers alone is unlikely to prevent terrorism unless it is accompanied by active measures such as frisking. New information regarding the source of threats may in turn lead police officers to revise their beliefs about the likelihood that individuals from different backgrounds will yield a successful stop for a given type of crime. If police officers associated Al Qaeda threats with Arabs or Muslims, then they may naturally have targeted them more after the alert increase, either consciously or subconsciously.

This stresses that one should be cautious when interpreting our main results since changes in policing by ethnic group or race may not represent racial prejudice. These policing responses indicate increased targeting or profiling of a certain group, but result from information regarding an increased likelihood of a threat with an associated increase in the stakes regarding the success of stops.

A model encompassing these considerations and motivating our empirical analysis is presented in the Appendix. The model modifies the hit rates test by focusing on the impact of an exogenous information shock on policing behaviour, rather than looking at static outcome differentials. It incorporates the additional duties of law enforcement officials related to terrorism, and allows officers to update their beliefs about the likelihood that Arabs may engage in terrorist acts during periods of high alert levels. The model generates the following key predictions: periods of high alerts should be associated with disproportionate increases in frisking, the relative quantity of stops (holding resources allocated to the Stop-and-Frisk programme constant), and potentially the use of force for Arabs or Muslims if officers believe them to be relatively more likely to engage in terrorist acts. This is consistent with our empirical findings apart from that regarding the number of stops, which could be explained by manpower re-allocations and task changes in response to the threat.

While our model and related tests cannot distinguish between statistical and taste- or bias-based discrimination, using intuition similar to Abaluck *et al.* (2016) provides additional insight. Our results indicate that following the alert level increase, police officers frisked and used force against relatively more members of the group 'Other' *without* an increase in the arrest rate. This suggests that the return to additional frisks was low, consistent with a non-rational behavioural reaction. Nevertheless, this is again complicated by the fact that police officer stops may not have a deterrent effect on potential terrorists as in the standard hit rates test. Successful stops or arrests regarding terrorism are also rare events with very high payoffs to law enforcement such that even veteran officers may not have much experience in this context. They may in turn less accurately assess potential risks, resorting to biases and broader profiling of specific groups as a result. Thus an explanation for our findings is that limited cognitive capacity could have led to profiling given the strong associations between the Middle East, 9/11 and other terrorist threats during that period, as well as evidence from Lichtbau and Drew (2004) suggesting that changes in the HSAS may place undue stress on police officers, particularly when the threat is more salient.[27] Since stress has also been shown to influence decision-making,[28] the judgment of officers may have changed following terrorist threats, and less attention was paid to being unbiased and fair.

Our results thus appear particularly consistent with a broad attempt at statistical discrimination influenced by group stereotypes, although we cannot reject that terrorist threats led to increased taste for discrimination against Arabs or Muslims. Similarly, while the response appears unwarranted due to the lack of successful stops, we cannot rule out that additional frisking was efficient given that terrorists may not otherwise be deterred. The interpretation of our results thus hinges on a careful consideration of the daily duties of police officers regarding both terrorism and common crimes, and of differences in behaviour between the two types of criminals. In times of a terrorist threat, the responsibilities of police officers partly shift away from minimizing crime towards prevention and protection.

III. Conclusion

It is well established that the duties of police officers in cities like New York changed greatly after 11 September 2001, due to increased terrorism concerns. On 1 August 2004, a specific terrorist threat led the US Department of Homeland Security to warn the country that Al Qaeda had apparently studied financial institutions in three cities, including New York, and were possibly planning an imminent attack. This warning led to an increase in the terror alert level from yellow to orange associated with additional security measures taken by the NYPD. The news of this terrorist threat also plausibly provided salient information to law enforcement about changes in the risk of an attack, potentially affecting the day-to-day decisions of individual officers.

In this paper, we examine if this change in the HSAS level led to differential changes in policing as part of the Stop-and-Frisk programme for suspects classified as 'Other' racial group by the NYPD, including Arabs. Our empirical results provide robust evidence that the alert level increase led to a disproportionate increase in the likelihood of being frisked or having force used against them for this group. Yet there was no corresponding increase in arrests or the likelihood of being stopped. These outcome differentials in turn reverted back to normal after the alert was lowered to yellow.

Racial profiling and police accountability have been at the centre of important ongoing public debates. Policing is an area where biases may be particularly likely to emerge, either consciously or unconsciously, in response to changing information about potential threats from organizations commonly associated with certain ethnic groups. Bias may be unconscious and discrimination latent, but the consequences in New York City for Arab citizens are very real, as illustrated by disproportionate increases in frisking and force used in periods when the alert level was heightened. One interpretation of our findings is that latent discrimination emerged when alert levels were more salient to police officers. This is consistent with racial profiling that did not lead to additional productive stops, as proxied by arrest rates. Nevertheless, we cannot isolate the source of discrimination or rule out that this policing response to the higher alert level was efficient. This is due to both deterrence considerations and the important distinction between traditional criminal activity and preventing terrorist attacks, as shown in the theoretical model.

These considerations also provide a potential rationale for the low test–retest reliability and predictive validity of psychological tests designed to measure implicit bias if both evolving social context and current events affect views of groups such as Arabs.[29] Negative views that individual officers hold about a group are more likely to emerge and affect their decision-making following salient events associated with the group.[30] In summary, our analysis suggests that to inform debates on racial profiling and police accountability, future work is needed to develop more general tests of discrimination in settings that consider the full set of duties faced by police officers that go well beyond traditional crime minimization.

APPENDIX: ECONOMIC MODEL

The Knowles *et al.* (2001) hit rates test remains the major avenue through which researchers examine police racial bias. This test was developed in the context of motor vehicle searches and is based on a simple model where police officers decide which vehicles to subject to searches and motorists decide whether to break the law by carrying drugs or illegal weapons. Motorists are assumed to have *ex ante* taken into account the probability of being searched and the penalty if they are caught. Racial bias is introduced as a preference parameter that reduces the perceived cost of searching vehicles of minority groups. This framework stipulates that in the absence of racial bias, each officer would pursue a monitoring strategy that maximizes the number of successful search outcomes given a cost of search, where a successful search is defined as one that uncovers some contraband.

Our model has several parallels, but is adapted to consider that individual officers must allocate their efforts to conducting stops s relating to either potential terrorist threats t or other criminal activity o. To ease exposition, pedestrians are considered to be in one of two racial groups: Arabs (A) and an aggregated group comprising of all other races (G). The key element of the model for our empirical analyses relates to the possibility of there being two terrorism threat levels l: baseline b and high h. By exploiting changes in terror alert levels as exogenous information shocks, we are able to overcome standard issues of conditioning on *ex post* information when performing a hit rates test on frisking suspects of different races. We next describe the behaviour of each actor in the model, the characteristics of an equilibrium, and theoretical predictions from a change in the HSAS level on policing outcomes, in turn.

Criminals

Individuals can be either type of criminal (t or o), and police officers correctly identify their type. We use $r \in \{A, G\}$ to denote the race of the individual, and $c \in \{1, \ldots, C\}$ to represent other characteristics that are observed by the police (but potentially unobserved by the econometrician) at zero cost and assumed to be independent of threat levels. The number of individuals in group (r, c) is expressed by $N^{r,c}$. An individual of type o receives a benefit v_o from committing a crime, and faces a cost h_o if subsequently stopped by the police.[31] Since criminals may be aware of potential changes in policing activities in response to changes in the likelihood of a terrorist attack, we allow the expected number of daily stops made to target group (r, c) to depend on the terrorism alert level (l), and define this value as σ_l.

An individual who engages in criminal activity of type o has an expected payoff of

$$u_{r,c,o}(v_o, h_o, \sigma_l) = v_o - h_o \frac{\sigma_l}{N^{r,c}},$$

and commits a crime if $u_{r,c,o}(v_o, h_o, \sigma_l) \geq 0$. Defining $F_{r,c}(v_o, h_o)$ to represent the group-specific joint conditional distribution of v_o and h_o, the crime o rate for group (r, c) denoted by $K_o^{r,c}(\sigma_l)$ is given by

$$P\left(v_o - h_o \frac{\sigma_l}{N^{r,c}} \geq 0\right) = K_o^{r,c}(v_o, h_o, \sigma_l),$$
$$K_o^{r,c}(\sigma_l) \equiv \int K_o^{r,c}(v_o, h_o, \sigma_l)\, \mathrm{d}F_{r,c}(v_o, h_o).$$

Similarly, an individual of type t faces benefits and costs respectively denoted by v_t and h_t. We define $M_{r,c}(v_t, h_t)$ as the joint conditional distribution of v_t and h_t. We assume that criminals of type t have a simplified payoff function of the form

$$u_{r,c,t}(v_t, h_t, \sigma) = v_t - h_t,$$

since they are assumed to not attach any weight to the likelihood of being stopped or arrested. In other words, this assumption implies that potential terrorists, unlike other criminals, cannot be deterred by the threat of judicial sanctions.[32] Thus a terrorist activity will be undertaken if and only if $v_t - h_t \geq 0$, the probability of which is given by $M(v_t - h_t)$. We do not require this assumption for our empirical tests to be valid, but it influences the interpretation. Rather, we aim to show that the model can be extended to motivate empirical tests even in the case where terrorists cannot be deterred. If terrorists behave as common criminals do, then the standard hit rates test applies and we can interpret outcome differentials accordingly.

Police officers

Assume that there is a mass P of police officers who, after having exogenously been assigned to a given precinct, draw a type p from a uniform distribution on $[0,1]$.[33] Given each officer's search capacity S_p, the total number of stops is given by

$$S(r,c,s,l) = P\int_0^1 S_p(r,c,s,l)\,\mathrm{d}p.$$

Since each officer allocates their efforts between both types of crimes, the following condition must be satisfied:

$$S_p^o(r,c,l) + S_p^t(r,c,l) = S_p(r,c,s,l) = \overline{S}_p.$$

Under the NYPD Stop-and-Frisk programme, each police officer first chooses whether or not to stop a suspect and then decides whether to additionally frisk that suspect. We define $y^r_{p,s,l}$ as the benefit that an officer receives from a successful stop (arrest). Officers are assumed to face a cost δ to stop a suspect, and an additional cost ρ to frisk the suspect. For stops related to crimes o we define

$$W_o(r,c,l) = P(\text{Guilty of crime } o \mid r,c,l) = K_o^{r,c}(S_p(r,c,o,l)),$$

and for stops related to crimes t we define[34]

$$W_t(r,c,l) = P(\text{Guilty of crime } t \mid r,c,l) = F_{r,c}(M,l),$$

where W_t is the subjective probability that an officer assigns to an individual being guilty of a terrorism crime, and $F_r : [0,1]\times\{b,h\} \longrightarrow [0,1]$ is a distortion function that can vary by racial group and depends on the true probability of guilt and on the alert level. In general, we would expect that $y^r_{p,t,l} > y^r_{p,o,l}$ and $W_t < W_o$, since (i) the benefits of stopping a terrorist attack likely outweigh those from arresting a suspect for a common crime, and (ii) the crime rate for terror activities is lower than that for other crimes. To understand police officer behaviour in equilibrium, we use backwards induction and solve a two-stage decision process, outlined below.[35]

Second-stage equilibrium Conditional on having made a stop, a police officer decides whether or not to frisk the suspect. Presumably, the decision to frisk the suspect relies on a different information set than that available when stopping the suspect, since the officer is likely to have gathered additional information during the stop. Denote by η the unobserved (to the econometrician) additional signal of guilt. The officer, conditional on W_s,[36] optimizes

$$\pi_2 = \max\left\{0, H(W_s,\eta,y^r_{p,s,l}) - \rho\right\},$$

where H is a function representing the benefits of frisking, which is increasing in W_s, η and $y^r_{p,s,l}$.

The optimal choice is given by

$$\begin{cases} 1 & \text{if } H(W_s,\eta,y^r_{p,s,l}) - \rho \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Intuitively, the officer chooses to frisk a suspect if the value of the additional information that they gain during the stop exceeds the cost of undertaking the frisk.[37]

Further discussion regarding η is warranted. It is well recognized in the traditional hit rates model that η precludes the use of standard regression analysis since a disparity in the rate of frisk cannot be interpreted as evidence of profiling or discrimination given that the decision to frisk relies on unobserved private information acquired by the officer during the stop. By imposing an additional assumption on η, we can overcome this limitation since our focus is on not the disparity in the rate of frisk but rather the additional *difference* in the disparity in the rate of frisk caused by an exogenous information change. Changes in the terror alert level are uncorrelated with η. Therefore our tests require that η is stable across sudden changes in the terrorism threat level such that the quality of the unobserved information acquired by police officers during stops is the same whether the HSAS level is yellow or orange, ruling out the possibility that the difference in frisking hinges on η.

First-stage equilibrium In the first stage, the officer decides whether or not to stop a suspect taking into account both the costs and expected benefits if they subsequently frisk the suspect. The officer simultaneously allocates their total number of stops between the two types of crimes.[38] Formally, each officer maximizes

$$\pi_1 = \sum_{r,c,l} S_p^o \pi_1^o + \sum_{r,c,l} S_p^t \pi_1^t = \sum_{r,c,s,l} S_p^s \left\{ W_s y_{p,s,l}^r - \delta + E[\pi_2] \right\}.$$

Ignoring capacity constraints, an officer will stop an individual of group (r, c) for a crime of type s in a period of alert level l if

$$W_s(r, c, l) y_{p,s,l}^r - \delta + E[\pi_2] > 0.$$

Without imposing additional restrictions, this model cannot provide a prediction on the fraction of stops that are allocated to each type of crime or racial group.[39] However, under some mild additional assumptions, we can expect an interior solution in which individuals are stopped for both types of crime. First, consider a scenario where police officers allocate their entire effort to terrorism crimes.[40] In this case, criminals of type o would rationally expect $\sigma_l = 0$, and they would not be searched in equilibrium. This would naturally lead to $W_o(r, c, l) = 1$, and it would therefore be optimal for an officer to start stopping suspects for other crimes provided that the payoff from a *guaranteed* arrest outweighs the cost of stopping the suspect *plus* the foregone expected benefit of stopping an additional terrorist.

While this remains an assumption, we believe that it is fairly innocuous and evidently satisfied in the data. It states that even in periods associated with high terrorist threats, police officers would continue focusing on their full slate of duties and not devote their entire attention to preventing terrorism. However, officers may change the emphasis that they place on each of their duties as the risk of a terrorist attack changes. Conversely, a parallel assumption is needed to guarantee that police officers do not completely ignore terrorism threats, no matter the terror alert level. This assumption is also fairly mild and consistent with NYPD protocols as described in Section I.

Predicted effects from an exogenous change in the alert level

Having described the general model, we use it to study the impact of changes in the terrorism alert level on policing, and derive testable implications. We assume the following:

(a) $$F_{r,c}(M, h) \geq F_{r,c}(M, l),$$

(b) $$F_{A,c}(M, h) \geq F_{G,c}(M, h),$$

(c) $$F_{A,c}(M, l) = F_{G,c}(M, l),$$

(d) $$y_{p,t,h}^r \geq y_{p,t,l}^r \quad \text{and} \quad y_{p,t,h}^A \geq y_{p,t,h}^G.$$

Condition (a) suggests that officers believe a terrorist attack to be more likely during periods of elevated threat, and condition (b) suggests that officers may update their beliefs and consider Arabs relatively more likely to engage in terrorist acts during periods of high alert levels. Condition (c) is a simplifying assumption that rules out baseline discrimination to focus on the impacts of alert levels.[41] Condition (d) suggests that any payoff to thwarting a terrorist attack may be higher during periods of increased threat, even though stopping a terrorist attack from any group may yield an identical benefit. In the previous subsection, we implicitly assumed that the payoff and cost of successfully carrying out a terrorist attack (i.e. $G(v_t - h_t)$) did not change with the alert level, such that changes in alert levels simply reflect changes in the information set of law enforcement officers.

The difference in the likelihood of carrying out a frisk between a period of high alert and a period of baseline alert level is

$$H\big(F_{r,c}(M, h), \eta, y_{p,s,h}^r\big) - H\big(F_{r,c}(M, l), \eta, y_{p,s,l}^r\big),$$

and given the previous assumptions,

$$\begin{aligned} &H\big(F_{A,c}(M, h), \eta, y_{p,s,h}^A\big) - H\big(F_{A,c}(M, l), \eta, y_{p,s,l}^A\big) \\ &\geq H\big(F_{G,c}(M, h), \eta, y_{p,s,h}^G\big) - H\big(F_{G,c}(M, l), \eta, y_{p,s,l}^G\big). \end{aligned}$$

Our model thus predicts that following an increase in the alert level, Arabs would be frisked more in equilibrium, and that this increase is larger than for other racial groups. A racial differential in the probability of frisking a suspect in this setting can be explained either by an asymmetric increase in the benefits of preventing an attack from Arabs or from changes in beliefs about guilt. If we rule out changes in beliefs, then potential bias is introduced only through the payoff modifier as in the standard hit rates test, and a difference in frisking could be interpreted as taste-based discrimination. In the presence of both channels, we cannot distinguish between taste-based and statistical discrimination based on correct or incorrect beliefs.

Nevertheless, we can assess whether these additional frisks were productive, reflecting true changes in the probability of terrorism crimes. Statistical discrimination implies that officers use race as a proxy for the likelihood of committing a certain crime, and conduct stops based on productivity. Therefore an increase in the probability of frisking should be accompanied by an increase in the arrest rate. Otherwise, if police officers frisk more suspects but make no more arrests, then this suggests a low return on the additional frisks performed and thus makes an explanation based on stereotypes, biases or taste more likely.

Now consider the impact of an alert level increase on the first stage. The model has an unambiguous prediction on the direction of expected changes in the allocation of stops. An increase in alert level under the previous assumptions increases the value of stops related to terrorism. Therefore we would expect a relative increase in the share of stops that are related to terrorism. The model does not provide a closed-form solution for the allocation of stops, and we cannot identify terrorism-related stops (or any reason for the stop or type of crime suspected), but we may expect this to be reflected by an increase in the number of stops of Arabs. In the presence of decreasing returns to stopping members of a given group, the differential impact of alert levels on Arabs may be smaller, but the same logic and conclusions apply. Since terrorism stops are relatively more important in periods of high alert, more Arabs will be stopped if police officers are more likely to associate that group with the terrorist threats. Since we assume that η is constant before and after alert changes, this would also translate into a higher probability of frisk.

ACKNOWLEDGMENTS

We are grateful to the editor, one anonymous reviewer, seminar participants at the 2016 CLEF/CEA meetings, John Bound, Charles Brown, Michael Mueller-Smith and Jeffrey Smith for helpful comments and suggestions on this project. NYC Stop-and-Frisk data are publicly available at https://nycopendata.socrata.com (accessed 14 November 2019) as well as at the ICPSR website at the University of Michigan. Lehrer thanks SSHRC for research support. We are responsible for all errors.

NOTES

1. A complete list of foiled attempts since the terrorist attacks of 9/11 is provided at https://en.wikipedia.org/wiki/List_of_unsuccessful_terrorist_plots_in_the_United_States_post-9/11 (accessed 16 November 2019).

2. For example, see Tankebe (2013), Mazerolle *et al.* (2013) and Schuck *et al.* (2008) as well as surveys of the literature on perceptions of the police. While much research focuses on black and white differences, Sun and Wu (2015) survey Arab Americans and find that the majority have a great deal or a lot of confidence in the local police. They also find no statistically significant difference in confidence between Muslims, who accounted for 40% of their sample, and those of other religious affiliations.

3. See Fang and Moro (2011) for a recent survey of the literature on statistical discrimination. Within this literature, perceptions of identity are defined as stereotypes that relate to situations where a police officer observes the group identity of the suspect and assigns them a perceived value from a latent index of criminality based on these stereotypes.

4. Consistent with this hypothesis, Jobard *et al.* (2012) find that Arab passengers were more than seven times likelier to be stopped and searched on the Paris metro than Whites, and cases of racially discriminatory police stops in France and Spain have recently been brought to the European Court of Human Rights (Seydi and others v. France, Muhammed v. Spain).

5. Legewie (2013) suggests that terrorist attacks can have a profound short-term effect on individual perceptions, while Legewie (2016) argues that local events related to police shootings can influence police behaviour related to force.

6. During high-alert times, there is a change in how officers are deployed, and additional patrols are allocated to high-profile locations across cities. This change in deployment is well documented in the popular press and used for identification in both Klick and Tabarrok (2005) and Draca *et al.* (2011).

7. See also Fischhoff (2002) regarding the vagueness of the HSAS more generally.

8. See, for example, New York Civil Liberties Union, Stop-and-Frisk Data, at www.nyclu.org/content/stop-and-frisk-data (accessed 13 November 2019).

9. Shi (2009) provides compelling evidence that officers reduce policing intensity when faced with greater expected penalties by exploiting events surrounding the April 2001 riot in Cincinnati.

10. Threats are first transmitted through governmental channels to federal agencies and other government bodies before being transmitted to governors, state authorities, major law enforcement agencies and mayors. They are then shared with selected private sector entities before being publicly announced through a press conference.

11. See Eggen *et al.* (2004), Aguirre (2004), Jehl and Johnston (2004) and Taylor (2005) for details regarding the 1 August 2004 threat discussed here.

12. The dates of other alert changes are provided in the Online Appendix, along with an analysis of their effects on policing outcomes.

13. The media later reported that there may also have been 'test runs' in days prior to the alert increase, although this information was never corroborated by the Department of Homeland Security.

14. Social psychologists also provide evidence that modern-day bias among law enforcement personnel is most likely to be in the form of implicit bias, while Harcourt (2007) makes it clear that we have no evidence or theory to suggest whether racial profiling may be an effective counter-terrorism measure or instead be completely inadequate.

15. Force is defined by the NYPD as a stop that involves one or more of the following: physically manhandling the suspect with hands, physically having to put the suspect on the ground or against a wall, drawing and/or pointing their service weapon, using a baton, handcuffs or pepper spray, or other scenarios that involve a specific physical threat or action against the suspect. Approximately 35% of stops involving force did not lead to frisking.

16. We do not include African-Americans in the primary analysis since they constitute a large fraction of practising Muslims in the USA and may therefore be inappropriate as a comparison group. See, for example, Bagby (2012). The main results are robust if we include African-Americans in the comparison group, and Subsection 1.7 of the Online Appendix presents results contrasting Others to African-Americans directly.

17. In the sample, 35% of all stops were not stops that officers had to report by law, hinting at some possible incentive scheme in which police officers want to convey that they are making efforts. The sample cannot be restricted to only stops that have to be reported due to conditioning on *ex post* information. The external validity of the results relies on the sample being representative of all stops in the city, an assumption that is untestable with our data but seems plausible given the high percentage of stops that did not have to be reported. In addition, we may worry that police officers specifically underreport racially sensitive stops to avoid outside criticism, or overreport them to signal efforts in preventing terrorism, making the net effect on reporting ambiguous.

18. The major distinction is that time series data are discretized and generally not independent and identically distributed (i.i.d.), whereas RDs conceptually require that the assignment variable be continuous with no selective heaping on either side of the discontinuity.

19. While the increases received greater media attention and results concerning alert decreases may yield less clear results if there are lags between the end of a threat and the lowering of the HSAS level, decreases are important to consider if alerts impacted day-to-day NYPD activities and led to additional tasks and measures.

20. General disparities in the probability of being frisked should not be interpreted as evidence of racial profiling. The decision to frisk having stopped a suspect relies on private information obtained by the officer before and during the stop. Persico and Todd (2008) stress that the hit rates test cannot be performed in this case and that this alone does not constitute evidence of profiling. This problem does not apply when considering whether the alert level changes may create profiling because the change in the probability of being frisked is caused by an exogenous information shock.

21. Geographical coordinates of stops are not available in the data until 2006, preventing more precise heterogeneity analyses around the NYC alert increase.

22. The list of mosques was obtained from http://hirr.hartsem.edu/mosque/database.html (accessed 14 November 2019).

23. Performing the analysis at the monthly level leads to similar conclusions, as shown in Subsection 1.3 of the Online Appendix.

24. There is suggestive evidence that some of the placebo periods associated with a higher frisk differential were linked to other terrorist plots such as the arrest in November 2003 of a terrorist suspected to have links with another terrorist who plotted an attack on the Brooklyn Bridge, and the arrest in December 2005 of a man suspected of plotting an attack on oil and natural gas targets in NYC and New Jersey.

25. We follow the bias-corrected estimation procedure with robust inference in Calonico *et al.* (2014) with MSE-optimal bandwidth selection when undertaking this analysis.

26. We may expect this to be particularly true for traditional criminals from racial groups perceived as likely to be targeted by the NYPD as part of counter-terrorism measures. This would go in the opposite direction of what we find and lead to relative decreases in stops, frisks and uses of force for those groups.

27. Dowling *et al.* (2006) present evidence that the events of 9/11 led to the manifestation of psychological issues in NYPD officers.

28. See Starcke and Brand (2012) or Morgado *et al.* (2015) for recent interdisciplinary reviews of this literature.

29. Psychologists have designed tests to detect an individual's ability to engage in discriminatory behaviour by either considering differences in subject response times (Correll *et al.* 2007) to images or tracking a physiological response such as one's heart rate.

30. In addition, our results also point to the importance of changing contexts to detect potentially latent discrimination, an issue that deserves further study not only for policing behaviour but also in considerations of algorithmic fairness, since decisions are increasingly made by automated statistical or machine-learning predictive models.

31. This modelling for crime type o has a one-to-one correspondence to the hit rates model of Knowles *et al.* (2001), to ease comparisons.

32. This does not mean that would-be terrorists are non-responsive to any police measure, but we assume that they will not abandon their plan of committing an attack due to the likelihood of being caught or the threat of being sanctioned.

33. We abstract from precinct assignment considerations in our setting by assuming that it is exogenous of terrorism considerations and unaffected by changes in the terrorism threat level.

34. Police officers' main tasks and formation primarily relate to non-terrorist crimes for which they likely have better information. Terrorism crimes are such that it would be difficult to credibly estimate a rate of crime by ethnic group, and may thus be more likely to be affected by stereotypes. It also seems that threats of attacks may be more likely to affect perceptions of guilt of certain minorities than give police officers additional utility from stopping those minorities. Having discrimination enter as in the usual hit rates test would ignore the possibility that alert levels hold salient information about the threat of terrorism from the Middle East, whether founded or not. This explicitly highlights the difficulty in distinguishing between statistical and taste-based discrimination in this setting.

35. We suppress the more straightforward arguments of functions in the following subsubsections for ease of notation.

36. At the time of the decision, the officer has already formed his beliefs regarding the probability of a suspect's guilt given (r, c), therefore W_s is *not* a random variable but rather the realization of one.

37. This value could be negative if the officer believes that the suspect is less likely to be guilty after having performed the stop, or have a value near 0 if the officer either did not gain valuable information or is already highly confident of the suspect's guilt. In each of these cases, the suspect would not be frisked.

38. Realistically, we may expect the marginal benefit of stopping a given group for a specific type of crime to decrease with the amount of stops if police officers first target the most 'suspicious' individuals. We abstract from this possibility in the main model for simplicity, but discuss the implications of allowing this feature below.

39. In other words, there is no closed-form solution for the allocation of stops.

40. This could occur if the (presumably) much higher benefit of a successful terrorism-related arrest outweighs the (presumably) much lower probability of arresting a terrorist.

41. This assumption is not necessary for our empirical tests to be valid, but simplifies the exposition.

REFERENCES

Abaluck, J., Agha, L., Kabrhel, C., Raja, A. and Venkatesh, A. (2016). The determinants of productivity in medical testing: intensity and allocation of care. *American Economic Review*, **106**(12), 3730–64.

Aguirre, B. E. (2004). Homeland Security warnings: lessons learned and unlearned. *International Journal of Mass Emergencies and Disasters*, **22**(2), 103–15.

Bagby, I. (2012). *The American Mosque 2011.* Washington, DC: Council on American–Islamic Relations.

Calonico, S., Cattaneo, M. D. and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression discontinuity designs. *Econometrica*, **82**(6), 2295–326.

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S. and Keesee, T. (2007). Across the thin blue line: police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology*, **92**(6), 1006–23.

Coviello, D. and Persico, N. (2015). An economic analysis of black-white disparities in the New York Police Department's Stop-and-Frisk program. *Journal of Legal Studies*, **44**(2), 315–60.

Cumby, R. E. and Huizinga, J. (1992). Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica*, **60**(1), 185–95.

Dowling, F. G., Moynihan, G., Genet, B. and Lewis, J. (2006). A peer-based assistance program for officers with the New York City Police Department: report of the effects of Sept. 11, 2001. *American Journal of Psychiatry*, **163**(1), 151–3.

Draca, M., Machin, S. and Witt, R. (2011). Panic on the streets of London: police, crime, and the July 2005 terror attacks. *American Economic Review*, **101**(5), 2157–81.

Eggen, D., Mintz, J. and Priest, D. (2004). Financial sector is put on high alert. *Washington Post*, 8 August; available online at https://www.washingtonpost.com/archive/politics/2004/08/08/financial-sector-is-put-on-high-alert/c216bcd4-675c-480c-bd73-4cb2951ebddc/?utm_term=.7f6f1c4b77c6 (accessed 13 November 2019).

Fang, H. and Moro, A. (2011). Theories of statistical discrimination and affirmative action: a survey. In J. Benhabib, M. O. Jackson and A. Bisin (eds), *Handbook of Social Economics*, Vol. 1A. Amsterdam: Elsevier, pp. 33–200.

Fischhoff, B. (2002). Assessing and communicating the risks of terrorism. In A. H. Teich, S. D. Nelson and S. J. Lita (eds), *Science and Technology in a Vulnerable World.* Washington, DC: AAAS, pp. 51–64.

Fridell, L. A. (2008). Racially biased policing: the law enforcement response to the implicit black–crime association. In M. Lynch, E. B. Patterson and K. Childs (eds), *Racial Divide: Race, Ethnicity and Criminal Justice.* Monsey, NY: Criminal Justice Press, pp. 39–59.

—— (2016). Racial aspects of police shootings. *Criminology & Public Policy*, **2**(15), 481–9.

Goel, S., Rao, J. M. and Shroff, R. (2016). Precinct or prejudice? Understanding racial disparities in New York City's stop-and-frisk policy. *Annals of Applied Statistics*, **10**(1), 365–94.

Harcourt, B. (2007). Muslim profiles post 9-11: is racial profiling an effective counter-terrorist measure and does it violate the right to be free from discrimination? In B. J. Goold and L. Lazarus (eds), *Security and Human Rights*. Oxford: Hart Publishing, pp. 73–98.

Jehl, D. and Johnston, D. (2004). Threats and responses: the overview; reports that led to terror alert were years old, officials say. *New York Times*, 3 August; available online at https://www.nytimes.com/2004/08/03/us/threats-responses-overview-reports-that-led-terror-alert-were-years-old.html (accessed 14 November 2019).

Jobard, F., Lévy, R., Lamberth, J., Névanen, S. and Wiles-Portier, E. (2012). Measuring appearance-based discrimination: an analysis of identity checks in Paris. *Population*, **67**(3), 349–75.

Klick, J. and Tabarrok, A. (2005). Using terror alert levels to estimate the effect of police on crime. *Journal of Law and Economics*, **48**(1), 267–79.

Knowles, J., Persico, N. and Todd, P. (2001). Racial bias in motor vehicle searches: theory and evidence. *Journal of Political Economy*, **109**(1), 203–29.

Legewie, J. (2013). Terrorist events and attitudes toward immigrants: a natural experiment 1. *American Journal of Sociology*, **118**(5), 1199–245.

—— (2016). Racial profiling and use of force in police stops: how local events trigger periods of increased discrimination. *American Journal of Sociology*, **122**(2), 379–424.

Lehrer, S. F., and Lepage, L. (2019). Does selective crime reporting influence our ability to detect racial discrimination in the NYPD's stop-and-frisk program? In K. P. Huynh, D. T. Jacho-Chavez and G. Tripathi (eds), *The Econometrics of Complex Survey Data: Theory and Applications*, Advances in Econometrics Vol. 39. Bingley: Emerald Press, pp. 259–86.

Lichtblau, E. and Drew, C. (2004). Ridge's record: color alerts and mixed security reviews. *New York Times*, 1 December; available online at https://www.nytimes.com/2004/12/01/politics/ridges-record-color-alerts-and-mixed-security-reviews.html (accessed 14 November 2019).

Mazerolle, L., Bennett, S., Davis, J., Sargeant, E. and Manning, M. (2013). Legitimacy in policing: a systematic review. *Campbell Systematic Reviews*, **9**(1), i–147.

Morgado, P., Sousa, N. and Cerqueira, J. J. (2015). The impact of stress in decision making in the context of uncertainty. *Journal of Neuroscience Research*, **93**(6), 839–47.

Morris, V. (2003). Edgy city moves to orange+. New York Post, 23 December; available online at https://nypost.com/2003/12/23/edgy-city-moves-to-orange (accessed 14 November 2019).

Omer, S. B., Barnett, D. J., Castellano, C., Wierzba, R. K., Hiremath, G. S., Balicer, R. D. and Everly, G. S. Jr (2007). Impact of Homeland Security alert level on calls to a law enforcement peer support hotline. *International Journal of Emergency Mental Health*, **9**(4), 253–57.

Persico, N. and Todd, P. E. (2008). The hit rates test for racial bias in motor-vehicle searches. *Justice Quarterly*, **25**(1), 37–53.

Schuck, A. M., Rosenbaum, D. P. and Hawkins, D. F. (2008). The influence of race/ethnicity, social class, and neighborhood context on residents' attitudes toward the police. *Police Quarterly*, **11**(4), 496–519.

Shapiro, J. N. and Cohen, D. K. (2007). Color bind: lessons from the failed Homeland Security Advisory System. *International Security*, **32**(2), 121–54.

Shi, L. (2009). The limit of oversight in policing: evidence from the 2001 Cincinnati riot. *Journal of Public Economics*, **93**(1–2), 99–113.

Starcke, K. and Brand, M. (2012). Decision making under stress: a selective review. *Neuroscience & Biobehavioral Reviews*, **36**(4), 1228–48.

Sun, I. Y. and Wu, Y. (2015). Arab Americans' confidence in police. *Crime & Delinquency*, **61**(4), 483–508.

Tankebe, J. (2013). Viewing things differently: the dimensions of public perceptions of police legitimacy. *Criminology*, **51**(1), 103–35.

Taylor, A. (2005). The government's response to terrorism. Independent Institute, 27 December; available online at www.independent.org/publications/policy_reports/detail.asp?id=9260 (accessed 14 November 2019).

SUPPORTING INFORMATION

Additional supporting information may be found in the Online Appendix for this paper:

Additional Results and Robustness Checks

1.1 Estimates by Number of Mosques per Precinct
1.2 Estimates by Time of Day
1.3 Placebo Test at the Monthly Level
1.4 Regression Discontinuity Plots and Estimates
1.5 Analysis at the Hourly Level
1.6 Estimates Conditional on the Number of Stops per Precinct
1.7 African-Americans as the Comparison Group
1.8 Other Alert Changes

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than about missing material) should be directed to the corresponding author for the paper.

FIGURE CAPTIONS

FIGURE 1. Residualized policing outcomes in periods of orange alert level.

Notes: The series represent the monthly mean residual outcomes for Others and the comparison group. We use residuals from a regression of the raw data on a linear trend and month, year, precinct, precinct $\times$ year and day of week fixed effects.

FIGURE 2. Impact of the NYC alert change on the probability of frisk, by precinct.

FIGURE 3. Bi-weekly estimates of the impact of changes in the HSAS alert level on policing outcomes.

Notes: Numbers above panels denote statistical significance at the 5% level.

TABLE 1

SUMMARY STATISTICS

	Full sample		Others	
Variable	Mean	S.D.	Mean	S.D.
	(1)	(2)	(3)	(4)
Outcomes				
Frisk	48.43	(49.98)	45.13	(49.76)
Arrest	5.82	(23.41)	4.20	(20.06)
Force	23.41	(42.34)	23.45	(42.37)
Daily number of stops	595.21	(252.31)	54.55	(29.61)
Demographics				
Male	92.43	(26.45)	90.90	(28.76)
Age	28.08	(11.90)	28.48	(12.92)
Youth	54.06	(49.84)	53.98	(49.84)
Time				
Night	59.34	(49.12)	60.79	(48.82)
Friday–Sunday	43.22	(49.54)	45.60	(49.81)
Winter	27.32	(44.56)	27.17	(44.48)
Spring	28.34	(45.06)	28.15	(44.97)
Summer	21.52	(41.09)	22.04	(41.45)
Fall	22.83	(41.97)	22.65	(41.86)
Race				
White	20.99	(40.72)		
Hispanic	63.46	(48.16)		
Asian	6.19	(24.10)		
Other	9.37	(29.14)		
Observations	1,792,781		167,920	

Notes

Standard deviations in parentheses. Columns (1) and (2) refer to the sample of all observations, while columns (3) and (4) refer to the subsample composed of observations pertaining to the racial group ‘Other’. Youth refers to the fraction of suspects aged below 25. Night refers to the fraction of stops performed between 7pm and 6am.

TABLE 2

IMPACT OF THE NYC ALERT CHANGE ON THE NUMBER OF STOPS AND ON THE PROBABILITY OF FRISK

	(1)	(2)	(3)	(4)
Panel A: Number of stops				
Other	−0.567***	−0.646***	−0.0297	−0.0297
	(0.032)	(0.031)	(0.024)	(0.025)
Alert increase	−0.411***	−0.655***	−0.104***	−0.104***
	(0.039)	(0.048)	(0.038)	(0.039)
Alert increase × Other	0.191***	0.163***	0.0397	0.0397
	(0.056)	(0.059)	(0.048)	(0.049)
Alert decrease	1.061***	0.473***	0.109**	0.109**
	(0.040)	(0.055)	(0.047)	(0.045)
Alert decrease × Other	−0.674***	−0.524***	−0.0605	−0.0605
	(0.054)	(0.061)	(0.048)	(0.048)
Number of observations	415,341	415,341	415,341	415,341
p-value H_0: I = D	0.000	0.000	0.257	0.259
p-value H_0: I = −D	0.000	0.000	0.580	0.585
Outcome mean	4.316	4.316	4.316	4.316
Panel B: Frisk				
Other	2.949***	0.963**	1.862***	1.862***
	(0.425)	(0.416)	(0.417)	(0.465)
Alert increase	−0.915*	2.962***	2.764***	2.764***
	(0.479)	(0.508)	(0.506)	(0.544)
Alert increase × Other	3.702***	3.207***	2.659***	2.659**
	(1.026)	(0.984)	(0.983)	(1.073)
Alert decrease	−3.951***	−1.255*	−1.208*	−1.208
	(0.463)	(0.698)	(0.696)	(0.740)
Alert decrease × Other	−3.211***	−1.993**	−2.192**	−2.192**
	(1.005)	(0.971)	(0.969)	(1.059)
Number of observations	415,341	415,341	415,341	415,341
p-value H_0: I = D	0.000	0.005	0.009	0.017
p-value H_0: I = −D	0.418	0.041	0.431	0.483
Outcome mean	44.925	44.925	44.925	44.925
Time and precinct fixed effects	N	Y	Y	Y
Dependent variable lags	N	N	Y	Y

HAC standard errors	N	N	N	Y

Notes

Robust or HAC standard errors in parentheses. The binary outcome variable Frisk is rescaled to 100 or 0. Additional covariates include a linear trend, an interaction between the trend and the treatment group, fixed effects for month, year, day of week, precinct and year × precinct. The p-values refer to tests of the differential impact of the alert increase for Others being equal, and equal but of opposite sign to that of the decrease.

*, **, *** indicate $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

TABLE 3

IMPACT OF THE NYC ALERT CHANGE ON THE NUMBER OF STOPS AND ON THE PROBABILITY OF ARREST AND FORCE

	(1)	(2)	(3)	(4)
Panel A: Arrest				
Other	−2.696***	−1.997***	−1.913***	−1.913***
	(0.206)	(0.209)	(0.209)	(0.210)
Alert increase	−1.433***	0.252	0.252	0.252
	(0.248)	(0.270)	(0.270)	(0.271)
Alert increase × Other	0.472	0.456	0.422	0.422
	(0.460)	(0.459)	(0.459)	(0.457)
Alert decrease	−1.559***	0.189	0.179	0.179
	(0.235)	(0.363)	(0.363)	(0.364)
Alert decrease × Other	0.481	0.324	0.319	0.319
	(0.442)	(0.446)	(0.446)	(0.443)
Number of observations	415,341	415,341	415,341	415,341
p-value H_0: I = D	0.991	0.878	0.904	0.9037
p-value H_0: I = −D	0.001	0.005	0.008	0.0085
Outcome mean	6.680	6.680	6.680	6.680
Panel B: Force				
Other	−0.776**	−0.893**	0.113	0.113
	(0.390)	(0.390)	(0.390)	(0.445)
Alert increase	−2.912***	−0.018	0.108	0.108
	(0.427)	(0.457)	(0.456)	(0.470)
Alert increase × Other	2.503***	2.495***	2.088**	2.088**
	(0.919)	(0.901)	(0.899)	(0.961)
Alert decrease	−3.746***	0.411	0.287	0.287
	(0.406)	(0.633)	(0.631)	(0.654)
Alert decrease × Other	−1.608*	−1.469*	−1.496*	−1.496
	(0.890)	(0.877)	(0.875)	(0.924)
Number of observations	415,341	415,341	415,341	415,341
p-value H_0: I = D	0.017	0.020	0.034	0.046
p-value H_0: I = −D	0.093	0.053	0.264	0.304
Outcome mean	23.013	23.013	23.013	23.013
Time and precinct fixed effects	N	Y	Y	Y
Dependent variable lags	N	N	Y	Y

HAC standard errors	N	N	N	Y

Notes

Robust or HAC standard errors in parentheses. The binary outcome variables Arrest and Force are rescaled to 100 or 0. Additional covariates include a linear trend, an interaction between the trend and the treatment group, fixed effects for month, year, day of week, precinct and year × precinct. The p-values refer to tests of the differential impact of the alert increase for Others being equal, and equal but of opposite sign to that of the decrease.

*, **, *** indicate $p < 0.1$, $p < 0.05$, $p < 0.01$, respectively.

Figure 1

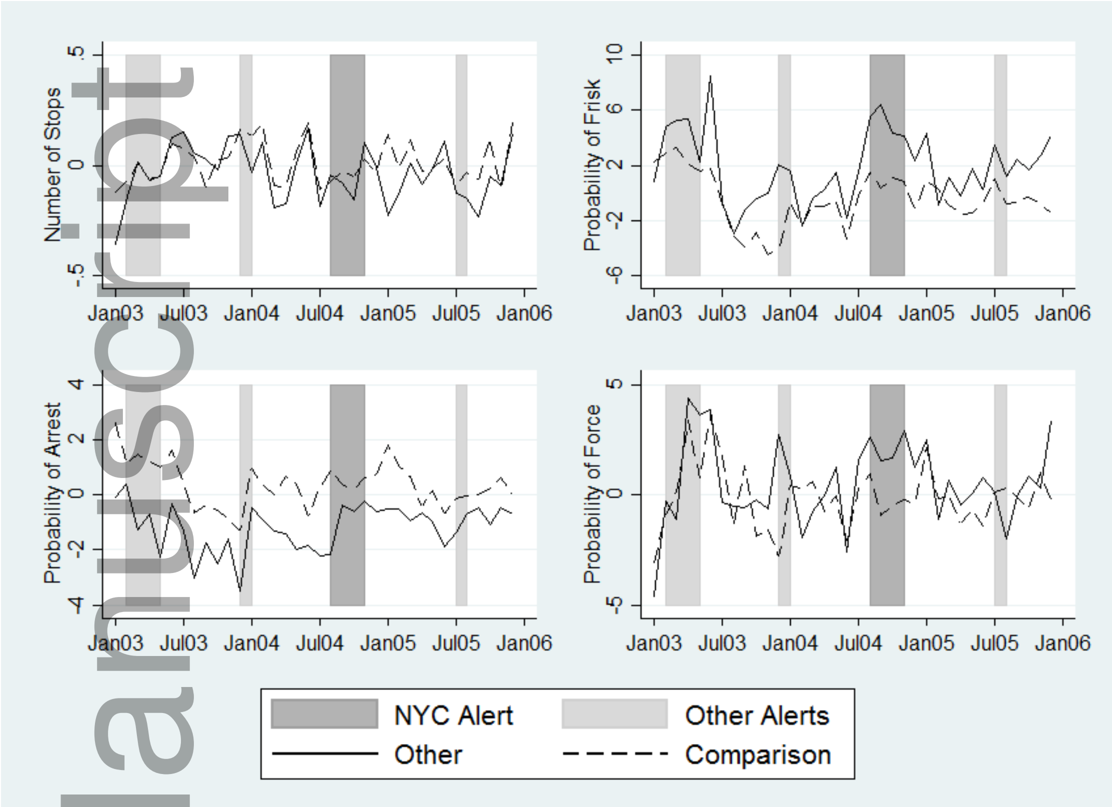

Number of Stops
Probability of Frisk
Probability of Arrest
Probability of Force
Jan03 Jul03 Jan04 Jul04 Jan05 Jul05 Jan06
NYC Alert
Other Alerts
Other
Comparison

Figure 2

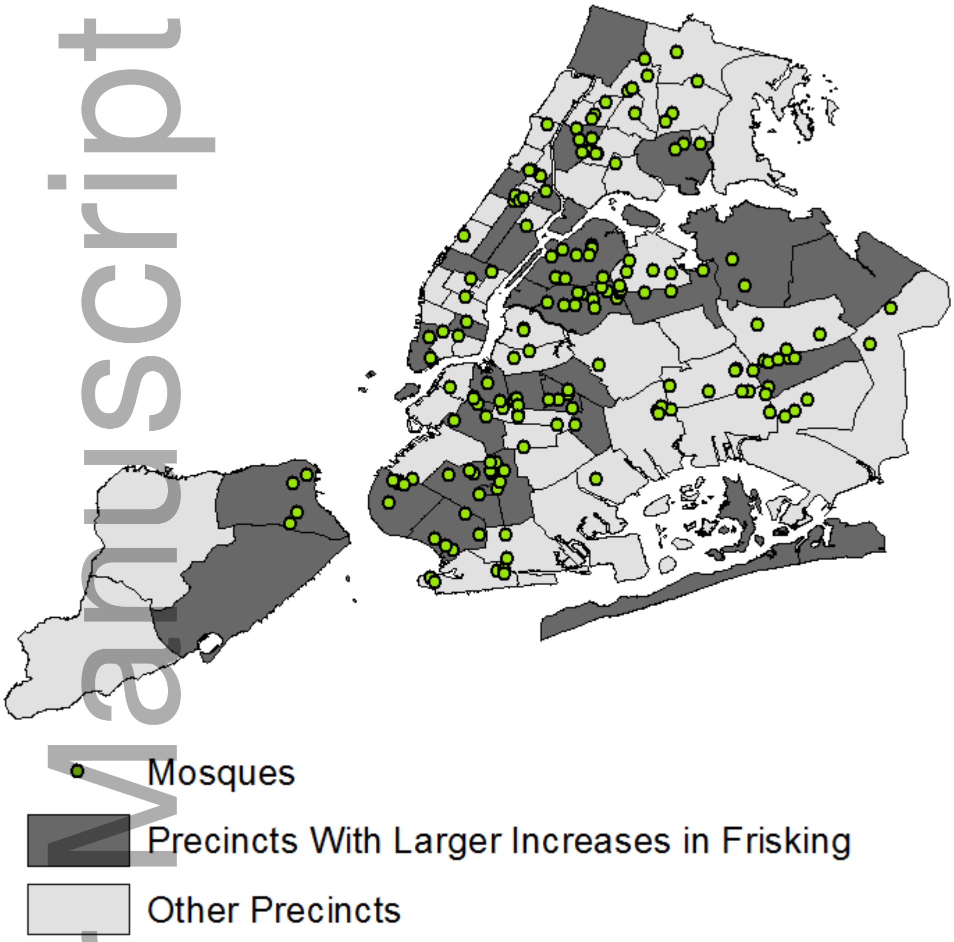
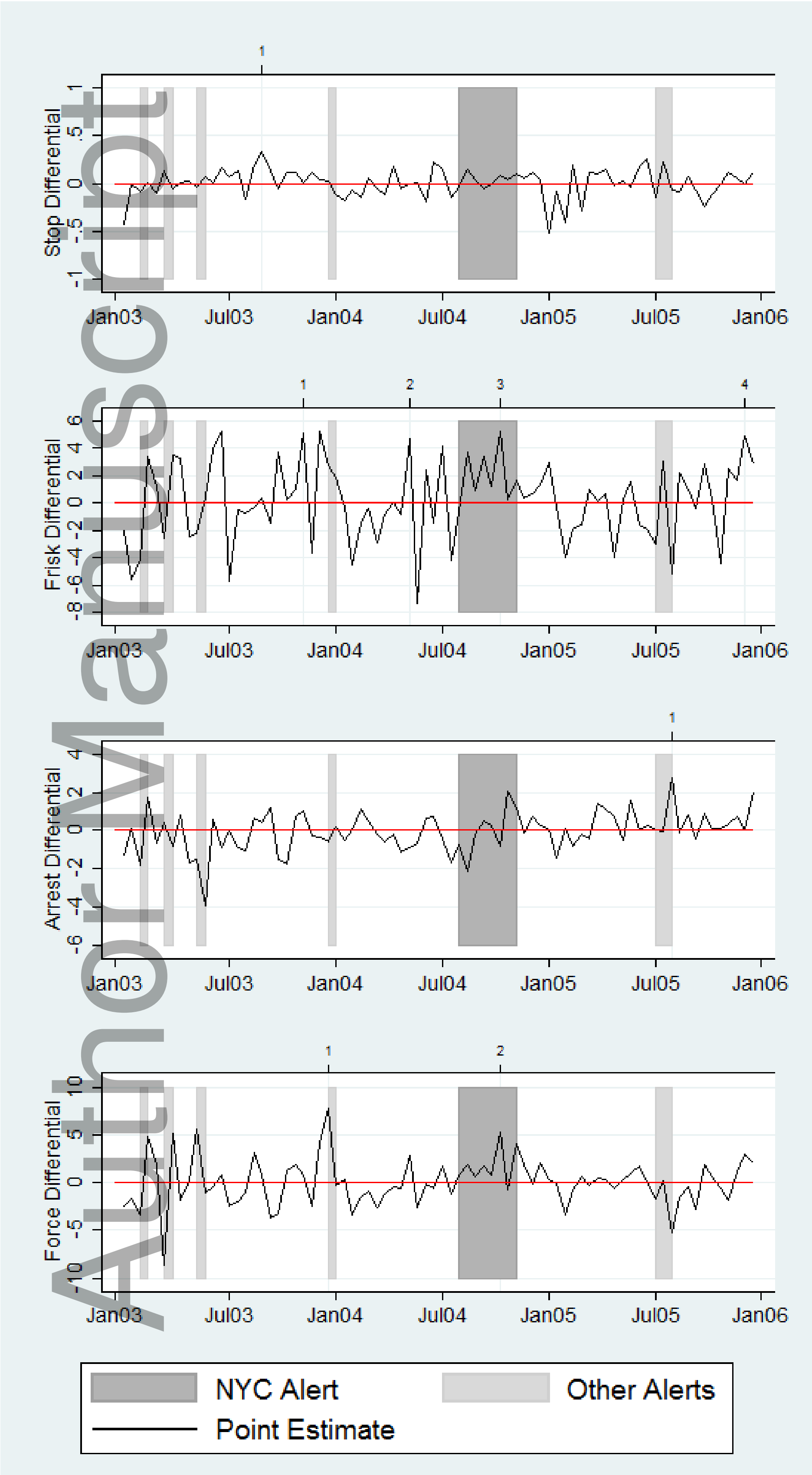

Mosques
Precincts With Larger Increases in Frisking
Other Precincts

Figure 3



Stop Differential
Frisk Differential
Arrest Differential
Force Differential
Jan03
Jul03
Jan04
Jul04
Jan05
Jul05
Jan06
NYC Alert
Other Alerts
Point Estimate