Racial Bias in Stop, Question, and Frisk Implementation: Raging Racism, or Statistical Discrimination

By CALEB DAME, ZACHARY HANCOCK, AND MITCHELL WILSON*

By limiting information of each police interaction from the Stop, Question, and Frisk Program in NYC, regression analysis was performed after Double-LASSO Regression was used in covariate selection. The regression found which possible factors were most predictive of the officer finding contraband or weapons on the subject given they were chosen by an officer to be stopped. It was found that being a minority lowered the chance of having illicit materials, illustrating that the balance of that data skewed too heavily minority, with officers finding fewer weapons or contraband per stopped minority compared to their non-minority counterparts after controlling for other factors relevant to the decision of being stopped. Young minority males were found to be stopped with either a gun or contraband 35-45% less often than comparable non-minorities.

* Brigham Young University, 1040 E Campus Drive, Provo, Utah 84604

I. Introduction

In light of the current racial climate and discrimination being at the forefront of the public conversation, this study delves further into the debate surrounding the New York City (NYC) Stop, Question, and Frisk Policy. There is widespread analysis and some clear data that suggest that racial discrimination is a determining factor in the police officers' choices of which NYC residents to stop, search, and frisk (Milner, George, Allison 2016). This study, however, is focused on evaluating the policy's effectiveness in stopping people that were breaking the law by holding weapons or contraband at the time of the stop, including profiling race as a proxy for criminal propensity. While the ethics of taking one's race into account when deciding to frisk or search them is questionable at best, this paper will evaluate whether or not the increased number of minority stops was justified by an increased incidence of weapons or contraband possession. This can lead to determining if any disparity in frisking rates by race may simply be the result of statistical discrimination, i.e taking into account basic group statistics to make more informed decisions when there is only access to limited information. Otherwise, if the minority subjects were less likely to be holding this incriminating evidence given the police decided to stop them at higher rates, the implication would be that any type of stereotyping or predictive heuristic used by officers was both ineffective and biased.

II. Background

In 2002, the New York Police Department (NYPD) instituted a new proactive policing policy called the "Stop, Question, and Frisk" (SQF) program. Under this program NYPD officers were encouraged to stop suspicious individuals and search for contraband and weapons in an effort to stop crimes before they occur. During the implementation of this NYPD program, the police department was also engaged in "broken windows" policing, where police focused their efforts more on preventing the frequent, smaller crimes in order to prevent larger, more serious crimes from ever occurring. These proactive objectives to reduce crime required officers on the beat to question individuals, perform stops, temporarily detain individuals, frisk individuals by patting down the outsides of their clothing, or perform searches where officers check the insides of clothing or any closed compartment to confiscate contraband and weapons. The type of interaction that the officer has with an individual is a response to the level of suspicion that the officer has regarding that person and their behavior. The SQF program has resulted in over five million stops and at the height of the program, NYPD officers made 685,724 stops in the 2011 calendar year.

The SQF program has been highly controversial, with critics claiming that the program is unconstitutional and discriminatory. In 1968, the United States Supreme Court determined in the case of *Terry v. Ohio* that police officers are justified in frisking or searching any individual, without a warrant, on the basis of suspicious behavior and not solely on probable cause. The rules and regulations regarding the SQF program have their foundations in this case. Critics have claimed that the NYPD's SQF program was discriminatory against Black and Latino New Yorkers. While only accounting for about 50% of the city's population, Blacks and Latinos made up over 80% of all stops that occurred under the SQF program. Defenders of the SQF program are quick to point out that nearly 80% of all crime and 90% of all violent crime are committed by Blacks and Latinos. These criticisms against the SQF program culminated with the case of *Floyd v. City of New York* where the program was deemed to be unconstitutional for violating the right “against unreasonable searches and seizures” and for being discriminatory against minorities. As a result of this case it was ordered that the SQF program be discontinued as it was then being practiced. In response to this decision, the SQF program was overhauled with court-appointed remedies and an independent monitor to oversee the change of the program. Since this decision, the number of stops made as a result of the program have dramatically decreased to the point where only 13,459 stops were made in 2019, less than 2% of the stops made in 2011.

While a lower court determined in *Floyd v. City of New York* that the SQF program was discriminatory against Blacks and Latinos, other research has both confirmed and disputed these findings. A RAND corporation study claimed that suggestions of overwhelming racial discrimination from the NYPD as a whole were unsubstantiated; however; this study was able to identify a small pool of 15 police officers who were discriminating against Blacks and Latinos (Ridgeway, 2007). The study developed benchmarks from propensity scores to determine whether discrimination occurred. In another study, the investigation on racial biases found that large, tall, and heavy Black and Latino men were disproportionately frisked, searched, and subjected to police use of force (Milner, George, Allison 2016). Additional evidence suggesting bias and stereotyping occuring in the NYPD was determined in studies that address differences in “hit rates" or the proportion of stops that result in arrests. Whites were found to have a significantly higher hit rate than Blacks and Latinos (Gelman, Fagan, Kiss 2012), which provides additional evidence to suggest that too many Blacks and Latinos were being pulled over. Following the decision of the *Floyd* case, there was a significant decrease in stops made through the SQF program. The resulting effect was a decreased number of stops occurring in predominantly Black and Latino communities (MacDonald, Braga 2019).

Our research is focused on determining whether the New York Police Department has been inaccurately stereotyping racial minorities through the duration of the Stop, Question, and Frisk Program. We specifically investigate whether the police’s efforts to stop those in possession of contraband or weapons was done in a targeted fashion to simply stop those who were most likely breaking the law, or if the stopping strategy was discriminatory and influenced by race. Past studies have focused on this very question, and we hope to provide additional evidence to determine the extent of racial bias in NYPD police officers' proactive policing measures. As cries for police reform become more and more prevalent, our results are meant to provide some insight into whether policing policies in America disproportionately affect minority groups, and for what reason.

III. Data Collection, Cleaning, and Criteria

Data collected throughout the duration of the SQF Program was submitted by the officers after each encounter on UF250 forms. Several internal audits and incentive programs ensured that the data would be filled out consistently and completely so that no reports would be left unreported or without sufficient data. The data has been made available by the NYPD and is published in raw form on their website by year.

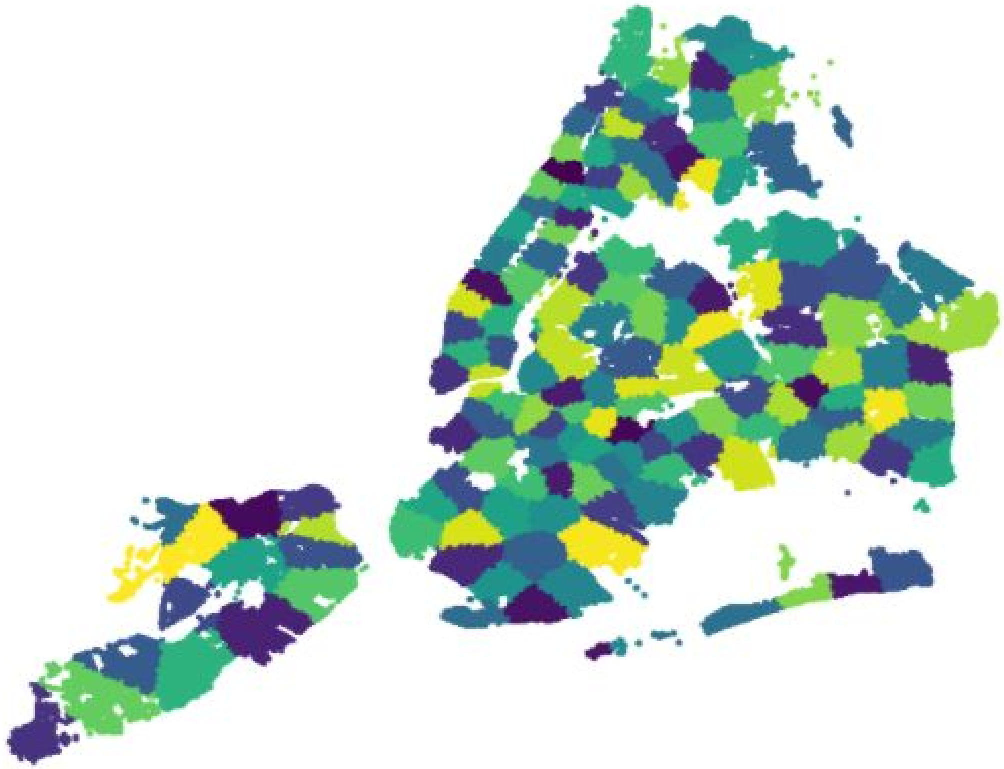
Recorded on each UF250 are the appearance of the one stopped, various descriptors of the officer, and the location of the stop. Physical traits such as race, age (verified by ID if present on suspect), eye color, hair color, and height (also verified by ID if present) are catalogued by the officer. The information regarding the officer includes, among much more, complaints filed against them personally, their assigned precinct, if they were in uniform, and if they presented ID while not in uniform. In addition to details around those involved, the form asks for the GPS coordinates of the stop, the responsible precinct, the NYC district postal code, which beat number the officer was on, the street name and adjacent streets, the time of day, and much more. Including 126 features, with one thereof being an index corresponding to the reason the officer thought the stop was necessary, the data is very detailed and descriptive.

Once found and combined between various annually reported spreadsheets published by the NYPD, many features were ill-suited for any linear model as many of them simply include non-unique text filled by the officer, such as street name or miscellaneous features that the officer wished to include on the report but for which there was no official space. The reason-for-stop index mentioned above is left off of the analysis as it could easily be a biased variable that the officer may manipulate. Such manipulation would render the estimated effect of race unreliable if certain reported suspicions were serially reserved for minorities, perhaps as a justification for harsher treatment. For many features, including those listed above, the original entries were too messy or consisted of unorganized text, making it difficult to systematically parse and integrate into the model as usable data. Additionally, certain data such as police number had too many possible values to one-hot encode, and would have required much more RAM than was available to our researchers. Otherwise, initial feature engineering created variables for the subject having facial hair, visible scars, or visible tattoos as they were among the most common appearances in the miscellaneous features data. Time of day was separated into fixed effects by hour due to how time enters non-linearly into a function of crime. Year and month were entered as fixed effects to capture longitudinal and seasonal trends.

Much of the written and street-specific data were not suited for a linear model, but fixed effects by postal code, city, beat number, and precinct were included to control for variation in crime statistics throughout New York. Additionally, GPS data for each stop was used to train a K-Means clustering algorithm to provide another method of controlling for variation in crime by location and provided 100 cluster fixed effects. These efforts to control for location are essential in understanding the difference between discrimination against minorities more genereally and independent of neighborhood and potential crime environment. If not controlled for, the effect of being in the "wrong" neighborhood where there are higher crime and stopping rates would bias our model. 100 points were chosen in place of more clusters in hopes of cluster predictive power being significant enough to be left in the final OLS analysis after the double-selection LASSOs. Also, the argument against fewer clusters would be that the number of neighboring precincts, postal codes, and different beats already clustered the data into 30-50 different groups, so the logical use of GPS coordinates would be to cluster the data into finer subgroups that are not captured by other categorical variables.

Many features in the original data, such as if an arrest was made or if the officer used handcuffs, were not included in the final analysis since they are unlikely to have been the cause for the stop. In our efforts to control for different reasons an officer would stereotype a suspect and frisk them, these would not enter into play. One such feature was if the person

Figure 1. Clustered Map of GPS Coordinates



had a criminal record as the officer would not necessarily know the record status of just anyone on the street; fortunately, in this case the data also lists if the officer knew the person from past encounters, so we can include this factor as it is something that can be known before the stop that helps the officer build a predictive model. The removal of such data was in an effort to isolate all information known to the officer before the subject was frisked, therefore including all environmental, physical, and pre-frisk data and not any data that is itself known as a consequence of the subject being frisked. In addition to discarding post-frisk variables, hair and eye color were dispensed with due to the highly correlative nature of those covariates with minority status, potentially causing a problem for parameter selection via double-selection LASSOs.

IV. Method

In order to assess the efficiency of the NYPD's targeted stops in terms of finding any weapon or contraband, we must determine which racial appartenance, if any, are predictive of having a weapon or contraband, given that the individuals were stopped. If the police officers were doing their best to target weapon or drug carriers and stopped every person that was most likely to be in possession of illicit objects, one would expect a disparity in racial composition of those who were frisked that is identical to the racial composition of the set of those carrying weapons. This would cause the person's race to not be causally related to holding weapons or drugs; therefore, in the case of simple statistical discrimination, a linear regression's coefficient for being a minority would be indistinguishable from zero. However, if there is excess discrimination evidenced by stopping too many minorities without finding additional weapons or drugs, then the predictive effect of being a minority on weapon or drug possession would be negative. This would show that frisked minorities turned up weapons or drugs at a significantly lower rate than their non-minority counterparts

To find a relationship between the minority variable and outcome variables, it is crucial that we control for any variables which may be correlated with treatment and outcome. Omitting these important covariates may bias the estimated causal effect of discrimination on weapon or contraband retreival rates. The dataset contains, after feature engineering and cleaning, there was still an unwieldy amount of variables—well over 500. So, we look to accomplish two tasks: reduce the dimensionality of our dataset and control for the most relevant covariates. These goals may both be accomplished nicely by employing the method of post-double selection LASSO to determine the most relevant controls given the large set of possibilities.

In order to validate this method, we require the assumption that, with only a limited amount of cleverly chosen covariates, we may control for any confounding variables to a great degree. This assumption has been termed "approximately sparse" (Belloni, Chernozhukov, and Hansen, 2014). With it, we conduct the three following steps to comprise the post-double selection LASSO. First, we find the covariates which effectively predict treatment using LASSO with a cross-validated penalty parameter. (The variables with nonzero coefficients are the ones that qualify.) Then, we find the covariates in a LASSO model that predict outcome and define the set of final controls as the union of the covariates with non-zeros coefficients from the LASSO regressions. They have been included precisely because they predict treatment (race affiliation) and/or outcomes, and therefore would introduce bias if left out of the model. Finally, we conduct a multiple linear regression of the outcomes on treatment with the chosen covariates, where the coefficient in front of race is our estimate of the marginal effect of race on likelihood of carrying a weapon or drugs.

The table presents the results of the least squares regressions on a minority dummy variable (Minority) and chosen covariates. We adopt a trial-and-error approach with regard to choosing the penalty parameter used in determining the LASSO covariate selections. The table organizes the estimating equations by penalty parameter. It is not surprising that many of these chosen covariates are the same. First, the variables that predict the treatment (minority status) are identical for both models (because the treatment is the same for both). Further, many of the variables that predict whether or not the stopped individual carried a weapon are understandably similar to those which predict whether or not they were in possession of drugs.

V. Results

Upon regression of weapons or contraband possession on the chosen covariates, found in Table 1, we find that for all estimating equations, the sign of the coefficient on minority is negative and statistically significant with well over 99% confidence. This indicates that police were incorrectly stereotyping minorities as carrying contraband or weapons more often than their non-minority counterparts. The negative sign is evidence that, given an individual was stopped and due to the racial balance of the dataset, being a minority on average lowered the probability of carrying illicit materials once confounding factors were controlled for. While NYC crime statistics do show that minorities commit the majority of gun and drug offenses, this is evidence that the police were still much too thorough in their proactive policing, stopping many more minorities than could have been attributed to statistical discrimination alone.

Using the coefficients associated with LASSO's first penalisation parameter, $\boldsymbol{\alpha} = 0.01$, we can estimate the differential in probability of weapon or drug possession. For a 25 year old, non-minority man, weighing 200 lbs and measuring 72", the estimated probabilities for weapon and drug possession are .03614 and .0515 respectively. Once the negative effects for minority status are added, the probabilities are reduced to .02114 and .0315 for non-whites, implying that such minorities sampled were roughly 40% less likely to be carrying drugs or weapons when frisked, compared to equivalent non-minorities. Statistical discrimination would extend to police having a bias toward those who commit crime, skewing minority due to crime statistics; however, the analysis indicates that even more minorities than would be justified by statistical discrimination alone were stopped under SFQ practice.

It is interesting to consider the r-squared values of the regressions which included the most variables: 0.041 and 0.037 for possession of weapons and drugs, respectively. Ideally, the variables associated to the race, gender and physical build of the person stopped would be zero. This would signify that the police were perfectly discriminating statistically and not incorrectly profiling people based on observable immutable characteristics. For example, at first glance, it may appear that police stop people in public housing neighborhoods too often to justify. However, they may be justified by the high contraband possession rates found there. One must be careful not to forget that this model doesn't allow conclusions to be drawn about the causality of these variables on finding drugs or weapons more generally. It simply shows whether police have been focused too much on one location, indicator, or stereotype to the detriment of finding increased crime.

VI. Conclusion and Future Inquiry

While it was no surprise that different numbers of people of varying race were being stopped during the SQF program, the analysis shows that police officers oversampled the minority population beyond any preexisting propensity to carry contraband. While the claim of statistical discrimination may be very much weakened, it may still not be a clear case of taste-based discrimination either, and should be investigated.

One possible reason for such a bias could be the proactive focus and time spent by officers in minority communities where there may not have been nearly as many viable suspects who were non-minorities. Officers do not choose their beat and their assignment—the police bureaucracy and leaders do. Thereby, working in a minority neighborhood would not have been their choice to make. Also, small, concentrated areas where minorities may be treated differently would have been unobservable in our model as we lacked the computational power to include minority-location interaction terms, which would have enabled our research to have judged if the bias in treatment was universal and widespread, or concentrated in specific neighborhoods or sectors. Additionally, studies have provided evidence that police in New York were using contemporaneous models at the height of SQF to predict zones where crimes would be committed while continuously updating using arrest data (Ensign et. al.). Therefore, by trying to allocate police labor effectively, the resulting bias may not be taste-based, but a result of orders from a black box in the office, influencing heterogeneous treatment of minorities based on location.

In any case, it is clear that minorities were differentially treated by the SQF policy implementation beyond the treatment that may be justified by the racial makeup of crime statistics. Fortunately, the SQF has begun a decline in practice in recent years and, while it is still on the books, has been mostly defanged by public scrutiny and the results of litigation through the court. Hopefully, further conclusions can be drawn after testing for heterogeneous effects by neighborhood. This would imply that the general issue of discrimination may not have been as universal or widespread as the entirety of the SQF and grant a more nuanced lens through which we can judge the program. After which, one can better evaluate officers and precincts individually and gage the extent to which the minority communities may have been hurt the most, where the most injustice and discrimination were performed, and how policing procedure in those areas can change in the future to correct past injustices.

VII. References

An Analysis of the New York City Police Department’s ”Stop-and-Frisk” Policy in the Context of Claims of Racial Bias. Gelman, A., Fagan, J., Kiss, A., 2012.

Analysis of Racial Disparities in the New York Police Department’s Stop, Question, and Frisk Practices. Santa Monica, CA: RAND Corporation, 2007.

Black and Hispanic Men Perceived to Be Large Are at Increased Risk for Police Frisk, Search, and Force. Milner, A. N., George, B. J., & Allison, D. B., 2016.

Did Post-Floyd et al. Reforms Reduce Racial Disparities in NYPD Stop, Question, and Frisk Practices? An Exploratory Analysis Using External and Internal Benchmarks. MacDonald, J., Braga, A., 2019.

High-Dimensional Methods and Inference on Structural and Treatment Effects. Belloni, A., Chernozhukov, V., Hansen, C., 2014.

Publications, Reports - New York Police Department. 2020. Retrieved from URL: https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page

Runaway Feedback Loops in Predictive Policing. Ensign, D., Friedler, S., Neville, S., Scheidegger, C., Venkatasubramanian, S., 2018.

Weapons Incidence Estimates

Weapons Incidence Variables	α = 0.01	α = 0.005	α = 0.0025
Minority	**-0.015**	**-0.0147**	**-0.0127**
	(0)	(0)	(0)
Age	**0.0002**	**.0001**	**-0.0000121** *
	(0.0000079)	(0.00000792)	(0.0000079)
Weight	**0.000053**	**0.0000387**	**0.0000435**
	(0.00000313)	(0.00000313)	(0.00000308)
Height	**-0.00008**	**-0.0001**	**-0.0000736**
	(0.0000242)	(0.0000242)	(0.0000238)
Casing the Area	-	**-0.0183**	**-0.0109**
		(0)	(0)
Presented Photo ID	-	**0.0106**	**0.0099**
		(0)	(0)
Alerted by Radio	-	**-0.0026**	**-0.00190**
		(0)	(0)
Staten Island	-	**-0.0072**	**-0.0043**
		(0.001)	(0.001)
Officer in Uniform	-	**-0.0058**	**-0.0072**
		(0)	(0)
Public Housing	-	**-0.01**	**-0.0112**
		(0)	(0)
Suspicious Object	-	-	**0.129**
			(0.001)
Group	-	-	**-0.0035**
			(0)
Suspicious Bulge	-	-	**0.0253**
			(0)
Public Housing	-	-	**-0.0112**
			(0)
Queens	-	-	**0.0057**
			(0)
Close to Active Crime Scene	-	-	-0.0006
			(0)
Suspicious of Weapon	-	-	0.0397
			(0)
Committing Violent Crime	-	-	**0.0022**
			(0)
Constant	**0.0263**	**0.0357**	**0.0292**
	(0.002)	**(0.002)**	(0)

Contraband Incidence Estimates

Contraband Incidence Variables	α = 0.01	α = 0.005	α = 0.025
Minority	**-0.02**	**-0.019**	**-0.0167**
	(0)	(0)	(0)
Age	**0.0006**	**0.0005**	**0.0005**
	(0.00000963)	(0.00000958)	(0.00000966)
Weight	**0.000022**	**-0.00000193** *	**0.0000006** *
	(0.00000382)	(0.00000378)	(0.00000376)
Day of Month	**-0.0002**	**-0.0001**	**-0.0001**
	(0.0000122)	(0.000012)	(0.000012)
Height	**-0.0001**	**-0.0002**	**-0.0001**
	(0.0000296)	(0.0000293)	(0,0000291)
Officer In Uniform	-	**-0.0127**	**-0.0122**
		(0)	(0)
Casing the Area	-	**-0.0153**	**-0.0125**
		(0)	(0)
Staten Island	-	**0.0021**	**0.0054**
		(0.001)	(0)
Public Housing	-	**0.0203**	**0.0185**
		(0)	(0)
Alerted By Radio	-	**-0.0073** *	**-0.0097**
		(0)	(0)
Appears to be a Drug Transaction	-	**0.0845**	**0.0842**
		(0)	(0)
Presented Photo ID	-	**0.0117**	**0.0108**
		(0)	(0)
Group	-	-	**0.0024**
			(0)
Committing Violent Crime	-	-	**-0.0093**
			(0)
Suspiscion of Weapons	-	-	**0.052**
			(0)
Close to Active Crime Scene	-	-	**-0.0013**
			(0)
Suspiscious Bulge	-	-	**-0.0085**
			(0)
Queens	-	-	**0.0067**
			(0)
Constant	**0.0393**	**0.049**	**.373**
	(0.002)	(0.002)	(0.002)

* $p > 0.01$