Evidence from the NYPD's Stop-and-Frisk Program:

Economic Analyses of the Impact of Social Capital on Crime and of Racial Discrimination Heterogeneity in Policing Behavior

By

Louis-Pierre Lepage

An essay submitted to the Department of Economics

in partial fulfillment of the requirements for

the degree of Master of Arts

Queen's University

Kingston, Ontario, Canada

August 2013

Acknowledgements

I gratefully acknowledge guidance and essential comments from Steven Lehrer. I also thank Decio Coviello for providing data as well as insightful institutional information. Lastly, I thank Andre Lepage, José Guerra and George Kontoleon for their useful comments. I am responsible for all errors.

1-General Introduction

The numerous costs generated by crime can have a critical impact on individual and social welfare. Accordingly, economists have long been interested in studying the determinants of crime, starting with early contributions such as Becker (1968) and Ehrlich (1973) on modeling criminal participation using economic tools. This literature has gained in popularity in recent years with papers such as Grogger (1998), Kelly (2000), Imrohoroglu, Merlo and Rupert (2001) and Lochner and Moretti (2001)[1]. The study of crime also provides several additional research opportunities with important policy implications. It has moved beyond understanding crime participation and is now assessing further components such as the economic motivators or deterrents of crime and the evaluation of crime reduction strategies.

This paper contributes to the literature by using data from the New York Police Department's (NYPD) Stop-and-Frisk program to examine two ways to reduce the costs of crime to society. Specifically, in the first part of the paper, we investigate whether social capital, as measured through the incidence of volunteerism, has a statistically significant negative impact on crime. Although social capital has been studied in the sociology literature since the creation of Social Disorganization Theory, few empirical papers can credibly establish a causal link between social capital and crime. This paper uses crime measures which are most likely to be influenced by economic factors and is the first study to focus on a single city. To identify the effect of social capital on crime, we use an instrumental variables (IV) estimator. Our main finding is that an increase of one standard deviation in volunteerism leads to a statistically significant decrease of 13.71% in arrests and 11.82% in court summons issued by the NYPD.

In the second part of the paper, we extend the analysis performed in Coviello and Persico (2013) on racial discrimination in the Stop-and-Frisk program. This is a controversial program that has

[1] Buonanno (2003) presents an overview of the main contributions.

.

attracted substantial recent attention in the popular press. Using the "hit rates" statistical test developed by Knowles, Persico and Todd (2001) to detect the presence of racial discrimination, we analyze the universe of stops made by police officers in New York City over a 10 year period. This can provide evidence of whether there is any racial bias against African-Americans. Given the possibility that studying mean impacts across all sets of crime may fail to identify discrimination, we extend earlier research by examining whether there is heterogeneity in the relationship across various subgroups including temporal, geographic and on the basis of the type of crime. Our evidence indicates that it is important to consider different types of crime independently when investigating racial discrimination. Specifically, our replication of Coviello and Persico (2013) continues to find limited evidence of racial discrimination when considering the aggregate of all crime types, but we find strong and robust evidence of discrimination when only crimes related to the War on Drugs are considered. To the best of our knowledge, these results constitute the only scientific evidence of discrimination against African-Americans in the New York City Stop-and-Frisk program.

This paper is organized as follows. Following a detailed description of the data, we conduct two empirical investigations independently. That is, we discuss each research question separately in a self-contained manner. The relationship between social capital and crime is presented in section 3. Our investigation of racial discrimination in the Stop-and-Frisk program is presented in section 4. In our concluding section, we discuss the implications of our main findings for both the economics literature and policy community.

.

2-Data

The primary data used in this paper is gathered from NYC Open Data[2] and comprises all recorded stops from the Stop-and-Frisk program between 2003 and 2012. For every stop, we are provided with the outcome, personal characteristics such as age, sex and ethnicity, the date and location of the stop as well as detailed information about the type of crime, weapons found and whether force was used. Following Coviello and Persico (2013), we consider the data representative of all stops even though police officers are not required to report those which do not involve the use of force or lead to a frisk, search, arrest or summons[3]. The overall sample is comprised of 4,791,153 stops. The analysis of social capital is first restricted to a subsample of 685,724 observations which represent all stops performed in 2011. Summary statistics for this subsample are presented in Table 1. Notice that over 93% of the suspects are male and 58% percent of the suspects are black or Black-Hispanic. Approximately 6% of the stops led to an arrest while 5.8% of the stops led to a court summons.

Data on volunteering opportunities in 2011 is gathered from the NYC Open Data initiative. This source includes all volunteering opportunities managed through NYC Service, a "citywide initiative tasked with setting a new standard for how cities can tap the power of their people to tackle their most pressing challenges"[4], launched in 2009. Observations fall into 6 categories: Emergency Preparedness, Health, Education, Helping Neighbors in Need, Community Strengthening and Environment.

[2] Available at *https://nycopendata.socrata.com/*

[3] In the sample, 35% of all stops were not stops which 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 which have to be reported due to conditioning on ex-post information. The external validity of the results rely on the sample being somewhat representative of all stops in the city, an assumption which is untestable with our data but seems plausible given the high percentage of stops which did not have to be reported.

[4] Information can be found at *http://www.nycservice.org/about*

.

Table 1: Summary Statistics for Social Capital

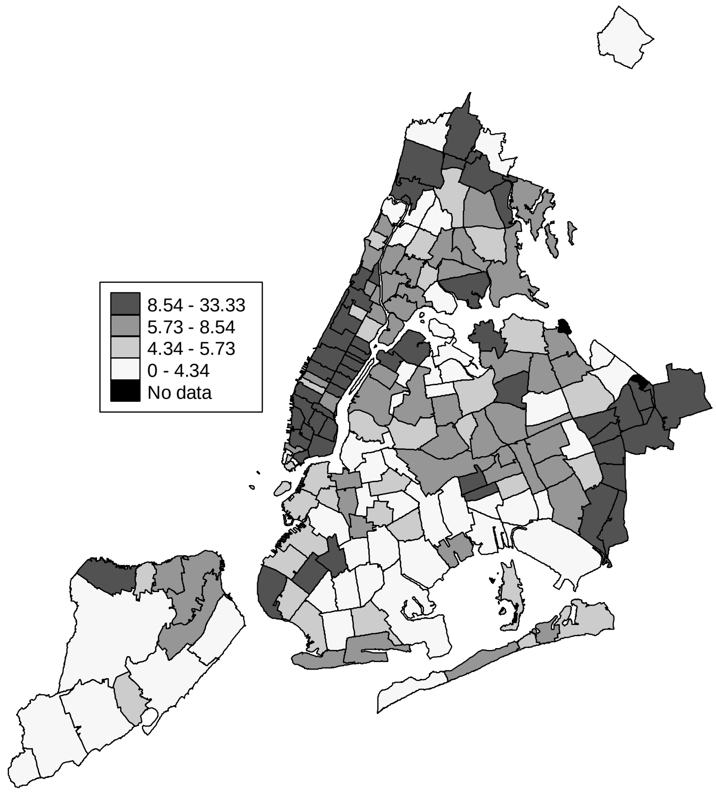
	Obs.	Mean	Std. Dev.	Min	Max
Outcome:					
Arrest (%)	65,992	6.04	23.82	0	100
Summons Issued (%)	65,992	5.84	23.45	0	100
Suspect Characteristics:					
Male	64,821	0.93	0.25	0	1
Age	65,644	27.92	11.59	8	90
Social Capital (per thousand people):					
Volunteering Opportunities	65,788	0.09	0.46	0	9.22
Obstructed Driveway Complaints	65,788	5.99	4.14	0	21.05
Illegal Parking	65,788	3.64	2.50	0	75.47
Vehicle Noise	65,788	1.67	1.22	0	8.01
Smoking Violations	65,788	0.25	0.63	0	37.74
Chronic Speeding	65,788	0.06	0.19	0	18.87
Cultural Associations	65,788	0.23	0.8	0	12.67
Controls:					
Income (1000$)	65,788	25.77	11.87	12.42	137.4
High School Degree (%)	65,788	26.82	8.08	0.4	100
Structures with More than 20 Units (%)	65,782	46.62	26.04	0	100
Population Aged 15 to 24 (%)	65,788	15.45	2.88	0	22.7
Unemployment (%)	65,788	6.76	2.04	1.7	30.2
Median Age	65,788	33.94	4.13	27.5	49.3
Poverty (%)	65,788	23.97	10.04	1.6	71.7
Less than High School (%)	65,788	24.99	10.74	0	49.2
Single Mother Headed Households (%)	65,782	13.45	7.48	0	32.1
Borough:					
Manhattan	65,992	0.2	0.4	0	1
Bronx	65,992	0.2	0.4	0	1
Brooklyn	65,992	0.34	0.47	0	1
Queens	65,992	0.22	0.42	0	1
Staten Island	65,992	0.04	0.19	0	1

Given that the goals of several of these categories are highly correlated, we merge them in order to facilitate the analysis. That is, Health and Education are merged together in a "Policy" category, while Helping Neighbors in Need and Community Strengthening are merged into a

.

"Community" category. In order to control for differences in population between zip codes, volunteerism is normalized by the total population of each zip code.

Figure 1: Probability of Arrest when Stopped, 2011, by Zip Code



In our analysis, we also need to account for other variables which could influence the relationship between volunteering and crime. These variables are gathered from the United States Census Bureau of the US Department of Commerce and include 2011 projections from the American Community Survey 5-year estimates on inflation adjusted median income, percentage of the population living under the poverty line, high school completion rates, median age, proportion of the population aged between the ages of 15 and 24, unemployment, proportion of housing structures with more than 20 housing units and proportion of single-mother headed households.

.

The analysis for section 3 is conducted at the zip code level, which is the unit of aggregation for social capital data as well as for controls. To obtain the zip code in which every stop was performed, reverse geocoding methods were implemented by using geographical coordinates. This process involves using longitude-latitude coordinates, which are included in the data, to pinpoint the location of the stop and retrieve the zip code using *Google Maps*. More information on geocoding and computational methods is presented in Appendix 1. New York City exhibits strong spatial heterogeneity in both crime and social capital, as can be seen from Figures 1 to 3. The data provides significant variation across zip codes, which can be exploited to establish a link between volunteering and crime.

In order to make things computationally feasible, a randomly selected 10% subsample of all stops performed in 2011 was gathered[5]. In this subsample of 65,995 observations, we utilize 70% of the zip codes located in the city[6]. Examples of crimes which led to arrests include assault, larceny, robbery, possession of a concealed weapon, rape, narcotics and prostitution. Examples of crimes which led to a court summons include drinking in public, disorderly conduct, trespassing and driving without a license. More information on crime types is presented in Table 2.

[5] Reverse geocoding all observations could take up to three months.

[6] Some zip codes are unusable since they are of an administrative nature and are not attached to any territory (or attached to a single building). Others cover territory in which there were no stops in 2011, such as small islands around Manhattan. Lastly, some control variables are not available for all zip codes.

Table 2: Crimes Leading to Arrests or Court Summons (%)

	Arrests	Summons
Alcohol in Public		32.58
Disorderly Conduct		30.1
Marijuana	27.75	4.24
Trespassing	19.46	7.5
Weapon	14.69	
Robbery	11.78	
Bicycle on Sidewalk		10.52
Other		10.5
Controlled Substance	9.69	
Assault	8.37	
Larceny	6.91	
Public Urination		4.54
Forged Instruments	1.32	

Notes: All values are percentages. Crimes presented are a subsample of all those committed. Crime codes or descriptions are not coded in a standardized way in the data, leading to many redundant categories. Crime titles with less than 25 occurrences for arrests and 10 occurrences for summons were removed from the sample.

Figure 2: Probability of Being Issued a Summons when Stopped, 2011, by Zip Code

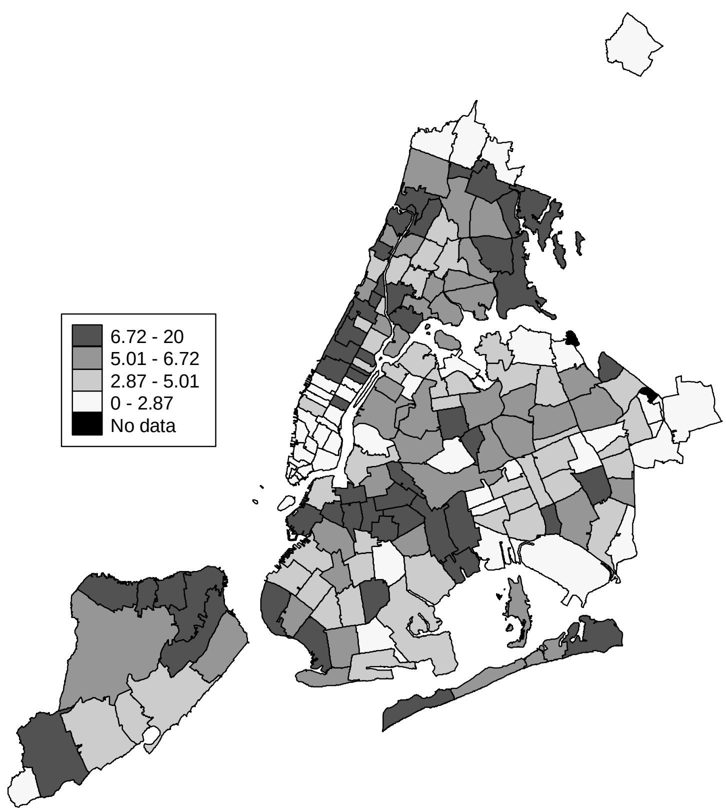
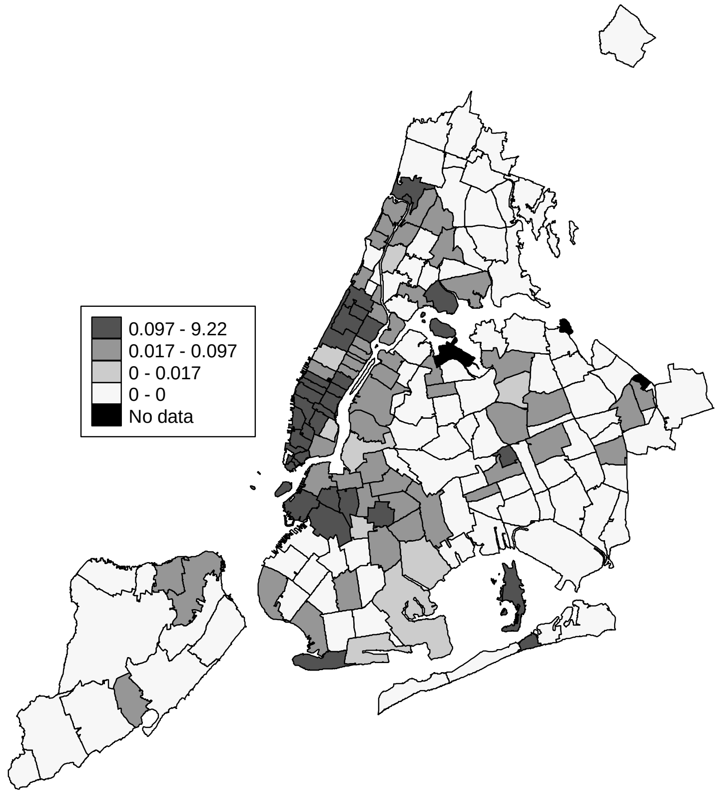


Figure 3: Volunteering Opportunities per 1000 Inhabitants, 2011, by Zip Code



In section 4, we turn our attention to testing for whether there is evidence of discrimination against African-Americans when compared to whites in the Stop and Frisk program. We follow Coviello and Persico (2013), who place a restriction to include only those two demographic groups from the full sample of the entire 2003-2012 period. Over 84% of this subsample of 2,947,867 observations is composed of African-American suspects. Summary statistics and information about the types of crimes which represent over 95% of all recorded stops is presented in Table 3. As a robustness check, we also take a larger subsample of 4,413,568 observations which includes White-Hispanics along with whites and Black-Hispanics along with blacks. In this subsample, of all suspects stopped by the NYPD (of which 62% were black or Black-Hispanic), 5.87% were arrested and 6.3% were issued a court summons.

Table 3: Summary Statistics for Discrimination

	Mean	St.Dev	N
Outcome			
Arrest	5.79	23.35	2,497,865
Summons	6.18	24.08	2,497,865
Race of the pedestrian			
Black	84	37	2,497,865
Crimes			
Possession of a Weapon	27	44	2,496,267
Robbery	17	37	2,496,267
Criminal Trespass	12	32	2,496,267
Grand Larceny Auto	9.1	29	2,496,267
Burglary	8.9	28	2,496,267
Grand Larceny	4.3	20	2,496,267
Illegal Possession of Substances	3.6	19	2,496,267
Assault	4	20	2,496,267
Marihuana	3.3	18	2,496,267
Illegal Sales of Substances	2.9	17	2,496,267
Petit Larceny	2.5	16	2,496,267
Mischief	1.2	11	2,496,267
Graffiti	1.1	10	2,496,267
Other Crimes	4.3	20	2,496,267

Mean is in percent. The 13 categories of crimes represent over 95% of all recorded crimes. Some values for the type of crime are missing for the years 2003-2005

3-Social Capital

3.1-Introduction

While the role of individual characteristics which lead to criminal behavior has received a lot of attention among economists, the literature offers little empirical insight into broader social characteristics. Specifically, the way in which members of a community interact and the feelings which they entertain towards one another may influence the decision to engage in criminal behavior[7].The idea that various intangible elements such as civic norms, peer pressure, trust and cooperation may influence crime is intuitively sound and has been a long staying concept in the criminology and sociology literatures since the development of Social Disorganization Theory.[8] A community in which such elements are present may be enacting a form of informal social control on criminal behavior by increasing the cost of criminal participation through the creation of a sense of guilt or betrayal. These unobservable assets of a community are referred to as social capital.

This section provides insight into the relationship between social capital and crime by using a research design to establish a causal relationship. Specifically, we investigate whether there is a causal link between social capital, measured through volunteerism, and crime. That is, do stronger civic norms, measured through the extent of volunteering from NYC Service in a zip code area of New York City, lead to decreases in crimes targeted by the NYPD's Stop-and-Frisk program in the area?

It adds to the limited empirical evidence on the topic by using data on a single city, New York City, which is well suited for such an analysis since it is diverse enough to exhibit variation in social capital. Using data from one city reduces the chance of the results being significantly biased by confounding factors. Indeed, some areas of New York City vary substantially along

[7] For example, see Calvó-Armengol and Zenou, (2004)

[8] Some major contributions include Shaw and McKay (1942), Bursik (1986, 1988), Sampson and Groves (1989), Wilson (1990, 1996), Sampson, Morenoff and Earls (1999), Sampson, Raudenbush and Earls (1997), Baumer, Messner, and Rosenfeld (2004) and Salmi and Kivivuori (2006).

sociocultural and socioeconomic characteristics while being subject to similar laws and policies. We exploit variation in social capital at the zip code level to identify its effect on crime. We study a single branch of social capital, civic norms, measured by the extent of volunteerism in an area.

In addition, we focus mainly on crimes which are more likely to be the outcome of a rational decision, such as property crimes, since they are presumably more influenced by societal norms than violent crimes or crimes of passion. Our measures of crime are the probability of a suspect being arrested or issued a court summons when stopped by the police. We posit that these probabilities accurately reflect – albeit indirectly - the incidence of crime in the area.

One common concern in the literature is establishing a causal link, since social capital could be negatively linked to crime if more disadvantaged neighborhoods also exhibit lower communal development. In addition, crime may also deter accruement of social capital, as trust and cooperation are less likely to thrive in high crime areas. Thus, Ordinary Least Squares (OLS) estimates may be negative regardless of causality. Contrastingly, there could also be a positive relationship if volunteering is predominantly found in less developed neighborhoods due to a bigger need for these services.

To overcome the inherent endogeneity problem, we use instrumental variables (IV) estimation with several instruments which reflect the amount of social capital present in a community. We also control for several variables known to influence criminal participation, and include month and borough dummies given that the five regions of the city may differ along unobserved characteristics.

Our results show that social capital, through volunteering, is significantly related to crime. An increase in volunteering leads to a decrease in the probability of a suspect being arrested or summoned to court once stopped. The results are robust to considering additional controls and specifications. Correcting for endogeneity is also critical, as the coefficient estimated by OLS is less than half that of the IV and we can reject the null hypothesis that volunteering opportunities can be treated as exogenous in the analysis. In our baseline IV model, an increase of one standard deviation in the number of volunteering opportunities per thousand people would decrease arrests by 13.71% and the number of summons issued by 11.82%.

In the remainder of this section, we begin with a brief review of prior research examining the links between crime and measures of social capital. In Section 3.3, we discuss the economic model that underlies our empirical analysis and introduce both the empirical and identification strategy. The main results are presented and discussed in section 3.4. Section 3.5 includes several robustness checks. A concluding section summarizes the main findings of the overall section.

3.2-Literature Review

Extensive theoretical research has been conducted on the role of social capital in public policy and other social phenomenon like crime. Central to understanding the roles of social capital are the works of Putnam (1993, 2000), in which he defines it as:

"*social capital refers to connections among individuals – social networks and the norms of reciprocity and trustworthiness that arise from them. In that sense social capital is closely related to what some have called "civic virtue." The difference is that "social capital" calls attention to the fact that civic virtue is most powerful when embedded in a sense network of reciprocal social relations*."

There is an overall consensus in the literature from other fields that there exists a negative relationship between crime and social capital. Those conclusions are reached in studies such as Rosenfeld, Messner, and Baumer (2001, 2004), along with Salmi and Kivivuori (2006). Although sociologists and criminologists have long been interested in the link between crime and social capital, little research has been done by economists. This is somewhat surprising given the vast amount of research on understanding the foundations of crime[9]. When rational behaviour models for criminal participation are extended to include community elements, it gives rise to what is

[9] See Buonanno (2003) for an overview of the main contributions

referred in the literature as “multiple equilibria” caused by the different aspects of social capital. Theoretical models from studies such as Calvo´-Armengol and Zenou (2004) posit that stronger associational networks could foster crime by easing communication between criminals while others, such as Weibull and Villa (2005), find that strong civic norms or networks can act as an informal control mechanism which discourages crime through the increase of intangible costs such as loss of reputation, guilt and shaming.

Of the few papers using economic tools, some such as Lederman, Loayza, and Mene´ndez (2002) focus on cross-country data and restrict their analysis to violent crimes. Cross-country studies are more likely to suffer from omitted variable bias, whereas studies focusing on violent crimes are likely to understate the importance of social capital. Indeed, New York City has similar policies and institutions in all areas, while differences in crime in cross-country studies may come from differences in the quality of institutions or policy-making. These are likely positively correlated with social capital and negatively correlated with crime, which would lead to a downwards bias of the effect of social capital. Additionally, cross-country studies must often settle for fewer, broader controls and therefore have lower internal validity.

The papers closest to ours are Akcomak and ter Weel (2008), which focuses on social capital at municipality level in Denmark, and Buonnano, Montolio and Vanin (2009), which analyses the impact of provincial social capital on crime in Italy. Following the latter, and other works such as Knack and Keefer (1997) and Bjørnskov (2006), we separate social capital in three categories: generalized trust, civic norms and associational networks. This paper focuses on civic norms, measured through the incidence of volunteerism.

3.3-Empirical Strategy

In this section, we develop the basic framework from which most of the results are obtained. The role of social capital in explaining spatial crime heterogeneity can be illustrated using an extremely simple framework in which social capital is considered as an informal control mechanism. As in Akçomak and ter Weel (2008), we assume social capital to be increasing in the degree of participation in civic life, altruism and care, security and trust as well as informal controls, contacts and acquaintances.

Given an individual *i* considering committing a crime *c* in a neighborhood *z*, the agent gains Y_{iz} from the crime and incurs an institutional punishment cost of T_{iz} if detected (probability Θ_{iz}). Then, the value of committing the crime is simply given by:

$$V_{iz} = Y_{iz} - \Theta_{iz}(n_z,p_z,x_i,c)\, T_{iz}(p_z,x_i,c) \quad (1)$$

In which x_i, p_z and n_z are respectively vectors of personal characteristics, policing and judicial characteristics and social capital. The remarkable aspect of the latter is that, unlike costs such as incarceration or social rejection, it enters the decision to commit the crime regardless of Θ_{iz} or T_{iz}. Consider $Y_{iz} = U_{iz}(x_i,c) + S_{iz}(n_z,x_i,c)$, where U_{iz} is the personal benefit derived from the crime. The impact of social capital on the value of the crime $S_{iz}(n_z,x_i,c)$, unlike T_{iz}, enters in the decision regardless of the outcome. A functional form is preferred since the impact of social capital, particularly of trust and associational networks, may not be linear. Furthermore, trust could provide more opportunities for crime while associational networks could facilitate either cooperation between criminals or their apprehension by the authorities[10]. Thus, we also allow community characteristics to influence the probability of being caught. For simplicity, n_z is restricted to impact S_{iz} and Θ_{iz} in opposite directions. That is, we rule out cases in which community characteristics lower the gains from committing a crime while also lowering the probability of being caught, or vice versa.

Ceteris paribus, if $S_{iz}(n_z,x_i,c)$ is negative and decreasing (more negative) in n_z, then social capital decreases the value of engaging in criminal activity for the individual. Neighborhoods with more

[10] See Calvo´-Armengol and Zenou (2004).

social capital would exhibit lower crime rates, partially explaining the spatial heterogeneity of crime. The goal of our analysis is to study and measure the relationship between n_z and crime. We do not distinguish between the impact on the probability of being caught and the impact on the gains derived from the crime.

Our data is not rich enough to directly estimate the above model, so we consider the following strategy to identify the reduced form relationship between volunteering opportunities and crime. Even though OLS estimates are likely biased due to endogeneity, they remain useful as a reference point. We consider the basic OLS model with covariates:

$$Y_{in} = \alpha_1 + \beta V_n + \eta A_i + \rho G_i + \delta X_n + \varepsilon_{in} \qquad (2)$$

Where Y_{in} is a binary outcome (0-100) for arrests or summons issued, V_n is volunteering opportunities per thousand people, A_i is the age of the suspect, G_i is the gender of the suspect, and X_n is a vector of zip code socio-economic characteristics.

The estimate of β given by (2) is likely to be biased and may not uncover a causal relationship. One reason is that a relationship likely exists regardless of causality. Intuitively, a neighborhood with higher human development may have less need for volunteering, or perhaps high crime areas discourage the presence of volunteering. Specifically, the bias is likely due to the endogeneity of volunteering. Volunteerism could reduce crime, or there may be less volunteerism in high crime areas to begin with, or an increase in volunteering may follow an increase in crime if the need and demand for assistance is increased.

This problem is similar to a well-known 2002 paper on police and crime by Steven Levitt. When studying the impact of the size of the police force on crime, the author faced the same issue since higher crime rates could lead to increases in police hiring. He highlights the importance of finding an instrument (number of firefighters) related to the size of the police force but unrelated to the crime rate. He also highlights the importance of controlling for other factors which can affect both the instruments and crime, without which the exogeneity assumption is violated. His results show the importance of correcting for endogeneity, since his OLS estimates are positive while his IV estimation yields the expected negative estimates. Our empirical strategy is similar and we use a wide set of diverse instruments to overcome the aforementioned problems.

We use a Two Stage Least Squares (2SLS) procedure, which accounts for endogeneity, given by the following equations:

$$1^{st} \text{ Stage: } V_n = \alpha_2 + \beta Z_n + \eta A_i + \rho G_i + \delta X_n + u_{in} \tag{3}$$

$$2^{nd} \text{ Stage: } Y_{in} = \alpha_3 + \beta \tilde{V}_n + \eta A_i + \rho G_i + \delta X_n + \varepsilon_{in} \tag{4}$$

Where Z_n is a vector of five instruments for volunteering and $\tilde{V}_n$ is the predicted value from the first stage regression[11]. The instruments were also gathered from NYC Open Data and they include information on 311 complaint calls to the NYPD. The selected infractions are presumably associated with low social capital and have little impact on more serious crimes. They are: blocked driveway complaints, illegal parking infractions, vehicle noise complaints, smoking infractions and chronic traffic speeding violations. They are also normalized by population.

The coefficient of interest, β, is an estimate of the relationship between volunteering and crime after controlling for variables which are presumably correlated with crime and social capital. The quality of the estimate relies on the assumption that there remains no unexplained variation which is common to each area and stable through time. This assumption, similar to the one used in Buonanno, Montolio & Vanin (2009), is necessary given that a standard panel data solution is unfeasible due to data restrictions and the slow moving nature of social capital.

[11] The real values of volunteering are used to calculate the standard errors

We control for variables which are likely correlated with both crime and social capital. Specifically, more income and less poverty should increase the opportunity cost of crime. Education also raises the opportunity cost of crime and may also have a “civilizing” externality (Fajnzylber, Lederman and Loayza, 2002). Suspects who have not completed high school are likely to be less employable and therefore likelier to engage in crime. All else constant, neighborhoods with more youth may be more criminally active (Freeman 1991; Grogger 1998). Unemployment reflects the inability to earn income through the labor market and reduces the cost of partaking in crime. The number of housing units per structure is included to distinguish between the denser areas with large apartment blocks and less densely populated areas. Lastly, the proportion of single mothers captures both social and economic hardships which are likely to lead to higher crime.

To capture unobserved heterogeneity, we include dummies for the five boroughs of New-York City, as they are likely to be structurally different. Their social, demographic and economic characteristics are likely to vary significantly, while policies and institutions are homogenous. For instance, all boroughs, except Staten Island in the past 12 years, have elected democratic district attorneys and borough presidents for over two decades. The city council is also composed of 46 democrats and 4 republicans; only Staten Island has a republican majority (2 out of 3). In the 2012 presidential elections, votes in Staten Island were approximately evenly split between both parties while democrats received over 79% of the votes in other boroughs. We also include month dummies, since stops may vary depending on the seasons, and include interactions between month and borough dummies to capture any heterogeneity across regions. Robust standard errors are used whenever possible to account for heteroskedasticity.

As seen in Table 4, correlations between the instruments and the outcome variables are low, ranging from -2.7% to 1.5% for arrests and -2.1% to 0.6% for summons, while correlations with volunteering are stronger, ranging from -12% to 60%. One concern with those instruments is that they may be similar to infractions deserving of a court summons, like drinking in public or driving without a license. We eventually address this issue by comparing the results with those obtained using different instruments. The five instruments allow for overidentification of the model and for testing conditional exogeneity. Furthermore, including several instruments improves the efficiency of the estimator. We present additional tests on the validity and performance of the estimator in the results section. Another noteworthy aspect is the low correlation between volunteering and both outcome variables. This indicates that volunteering may not explain a large proportion of the variation in arrests or summons. Zip codes may be too large to capture the intricacies of community spirit, since some of them include over 60 000 people. Another explanation is that the decision to partake in criminal activity remains primarily motivated by individual characteristics.

The main issue with our instruments is that, while structurally different from a Stop-and-Frisk related crime, they are similar to infractions which could lead to a court summons. To investigate this issue, we compare our results with those obtained using a presumably more exogenous instrument. We use the normalized number of registered cultural associations per zip code, which we show to be fully exogenous with regards to arrests. The idea to use cultural capital as a predictor for social capital is borrowed from the sociology literature, in which both forms of capital have been modeled as interlinked since the publication of *The Forms of Capital* by Pierre Bourdieu in 1986.

Introduced primarily in Bourdieu (1986), cultural capital refers to non-financial capital which the individual equips himself with in order to further his success and social mobility beyond economic means. It is similar to a soft-skill version of human capital, focusing less on technical knowledge and more on concepts like intellect, way of speaking and appearance. Explained further in Grenfell (2008, 2011), the concept is modeled as a step in between economic capital and social capital. That is, one uses economic capital to obtain cultural capital which is then used to accumulate social capital. To our knowledge, there is very little empirical research on cultural capital in economics, especially related to crime.

Table 4: Correlations

	Arrest	Summons	Volunteering	Blocked Driveway	Illegal Parking	Vehicle Noise	Smoking	Chronic Speeding
Arrest	1							
Summons	-0.046	1						
Volunteering	0.01	-0.018	1					
Blocked Driveway	-0.027	-0.013	-0.118	1				
Illegal Parking	0.01	-0.017	0.399	0.288	1			
Vehicle Noise	0.015	0.006	0.251	-0.363	0.161	1		
Smoking	0.006	-0.021	0.598	-0.054	0.593	0.209	1	
Chronic Speeding	-0.002	-0.007	0.174	-0.002	0.445	0.037	0.73	1

Notes: Arrests is the probability of being arrested conditional on being frisked and Summons is the probability of being issued a summons conditional on being frisked. All other variables are normalized by 1000 inhabitants for each zip code.

From the theory, it would follow that areas with higher cultural capital should also have higher social capital. In addition, this relationship should be independent from crime given that we control for economic and human capital. We first use cultural capital as a regressor in lieu of volunteering and then as an instrument, both on its own and along with the other five instruments. Correlations are presented in Table 5.

Table 5: Cultural Capital Correlations

	Arrest	Summons	Volunteering	Cultural Associations
Arrest	1			
Summons	-0.046	1		
Volunteering	0.010	-0.018	1	
Cultural Associations	0.013	-0.018	0.734	1

Arrests is the probability of being arrested conditional on being stopped, Summons is the probability of being issued a summons conditional on being stopped. Cultural Associations and Volunteering are expressed by 1000 people.

To compare the performance of OLS and IV, we conduct a Hausman test for the endogeneity of volunteering. Under the null hypothesis, there is no endogeneity, both the OLS and IV estimators are consistent, while OLS is efficient. Under the alternative, only the IV estimate is consistent. To improve the performance of the test, we first compute the test statistic by basing both covariance matrices on the disturbance variance estimate from the efficient estimator. Then, given the restrictive nature of the standard test and our use of robust standard errors, we conduct a bootstrapped Hausman test following Cameron and Trivedi (2010) to obtain more robust estimates which may address concerns presented in papers such as Hahn, Ham and Moon (2011).

One issue is that data restrictions, coupled with the slow moving nature of social capital, prevents us from using fixed effects (FE) and can lead to issues when clustering. Further information on clustering can be found in the robustness checks section. Additionally, FE relies on strict exogeneity, which is violated given the endogeneity of volunteering. This restriction is commonplace in the literature, including the paper closest to ours, Buonnano, Montolio and Vanin (2009). To address these issues, we include dummies for the 5 boroughs of New York City as well as for the 12 months of 2011.

Another issue is that the interpretation of the results may not be straightforward, as our indicators are not a direct measure of crime. The exactitude of the interpretation relies on the assumptions that volunteering has no other impact on the Stop-and-Frisk program or on the ability of criminals to avoid suspicion or detection. It seems unlikely that volunteerism could have an impact on these other than through social capital.

Another consideration is that the Stop-and-Frisk program targets specific crimes which may not be representative of all crimes in the city. The program emphasizes prevention and has a lesser effect on crimes of passion. It is therefore better suited to analyze crimes of opportunity, which is a beneficial distinction given that social capital presumably has a larger impact on economic crimes. Our estimates may therefore better reflect the importance of social capital than some found in the literature.

Additionally, we extend the baseline models in two ways. We allow for nonlinearity by including a squared term for volunteering and estimate the relationship between crime and the four categories of volunteering. The first extension is presented in the robustness checks section while the second extension is presented in Appendix 2.

3.4-Results

3.4.1-Baseline Model

Table 6 presents OLS estimates from equation 2. The first three columns are for arrests and the last three for summons. The columns differ based on whether we control for borough and month effects, as indicated at the bottom of the table. The estimated coefficient for volunteering is statistically significant and negative across all specifications and for both outcome variables. For arrests, the coefficient on volunteering is statistically significant at the 5% level when controlling for time and borough effects. The results predict that a one standard deviation (0.46) increase in volunteering would lead to a 0.24 percent decrease in the probability of being arrested when stopped. Considering the mean of 6.04% for arrests, this constitutes a decrease of 3.97%. The results for summons are similar. The coefficients for volunteering are all negative and statistically significant at the 1% level. Based on the estimate of approximately -0.5%, a one standard deviation increase in volunteering would lead to a 0.23% decrease in the probability of being issued a court summons – an overall decrease of 3.94% given the mean of 5.84%. It is interesting to note that most estimates are consistent across specifications except for youth, which has no statistically significant effect when controlling for borough and month effects, and for poverty and the proportion of people with less than a high school degree, which become statistically significant when the additional controls are added.

Table 7 presents estimates from equation 4. In the IV model, the estimated coefficient for volunteering is larger and statistically significant at the 1% level across all specifications. For arrests, based on the coefficient of around -1.8%, an increase of one standard deviation in volunteering opportunities per 1000 people would lead to a 0.83% decrease in the probability of being arrested when stopped. This constitutes a 13.71% decrease in all arrests of the Stop-and-Frisk program. Additionally, the IV model appears well specified and superior to OLS. The Kleibergen-Paap Wald rk F statistic against weak identification is above 1200.28 for all specifications, well above the standard cut-off of 10 for weak instruments. The p-value for underidentification is 0.0000 across all specifications, while the null hypothesis of the Sargan-Hansen test for overidentification is not rejected at the 10% level for any specification. This indicates that the instruments are not correlated with the error term of the outcome equation. Lastly, we can reject that the OLS and the IV estimates are the same based on the Hausman test values; OLS suffers from endogeneity and is inconsistent. Additionally, it appears as though OLS estimates are biased downwards and understate the impact of social capital on crime.

For summons; the IV coefficients are more than three times larger than those from OLS. Based on the estimate of -1.5%, an increase of one standard deviation in normalized volunteering opportunities would lead to a 0.69% decrease in the probability that a suspect be issued a court summons. This amounts to an overall reduction of 11.82% in all court summons. Volunteering is therefore predicted to have a similar impact on both crime indicators. The first stage statistics are the same as for arrests, the underidentification p-value is 0.0000 and we cannot reject the null of the Sargan-Hansen Test at the 10% level when controlling for borough and month effects. The smaller p-values for the test of overidentification restrictions and for the Hausman test may be a consequence of the similarity between the instruments and crimes which lead to court summons. Estimates from OLS, due to endogeneity, once again lead to an underestimation of the impact of social capital. It may be interesting to note that the proportion of single mother headed households is a useful predictor for both outcome variables when all controls are included.

An additional concern with the OLS estimates is that there may be two channels through which volunteering impacts crime. It may foster the development of networks and norms which act as a deterrent on crime. That is the social capital channel. On the other hand, volunteers provide goods and services to the more disadvantaged members of their community, which reduces the need to commit economic crimes. This is an indirect transfer mechanism. While it is not possible to distinguish between channels for OLS, IV estimates solely reflect the first channel. This is due to the IV estimator allowing us to recover a Local Average Treatment Effect (LATE), an Average Treatment Effect (ATE) only for individuals whose outcome is affected by changes in the instruments. That is, the estimates which we recover represent the impact of social capital on crime for those whose crime participation is dictated by the strength of civic norms in their area.

Table 8 presents the first stage estimates for the IV model. The coefficient on the instruments is highly statistically significant in all specifications, as are most controls. The controls which are less significant are age and gender, which is expected given that they are individual variables.

Lastly, we present the reduced form estimates for the IV model in Table 9. For arrests, the coefficients on the instruments are generally negative and not statistically significant once controlling for time and borough effects. For summons, coefficients for vehicle noise and smoking violations are both statistically significant, though of opposite signs. The larger overall negative sign, coupled with the mostly positive first stage coefficients, accurately predicts a negative second stage estimate.

Table 6: Determinants of Arrests and Summons, OLS

	Arrests			Summons		
Model	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Volunteering	-0.458*	-0.516**	-0.513**	-0.511***	-0.495***	-0.508***
	(0.25)	(0.25)	(0.251)	(0.163)	(0.329)	(0.329)
Age	0.047***	0.0467***	0.047***	0.118***	0.117***	0.116***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Gender	-2.999***	-2.993***	-2.953***	-0.827**	-0.758*	-0.756*
	(0.448)	(0.45)	(0.45)	(0.395)	(0.447)	(0.447)
Income	0.010	0.007	0.008	-0.112***	-0.129***	-0.128***
	(0.025)	(0.026)	(0.026)	(0.022)	(0.026)	(0.026)
Poverty	-0.046	0.087***	0.088***	0.105***	0.083***	0.083***
	(0.029)	(0.033)	(0.033)	(0.032)	(0.033)	(0.033)
High School	-0.118***	-0.058**	-0.056**	-0.092***	-0.088***	-0.087***
	(0.022)	(0.023)	(0.023)	(0.024)	(0.023)	(0.023)
Less than High School	0.023	-0.047**	-0.046**	-0.147***	-0.142***	-0.143***
	(0.02)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Median Age	0.187***	0.141**	0.143***	0.037	-0.008	-0.012
	(0.054)	(0.056)	(0.056)	(0.051)	(0.056)	(0.056)
Youth	0.202***	0.048	0.051	-0.163***	-0.243***	-0.245***
	(0.066)	(0.071)	(0.071)	(0.063)	(0.073)	(0.073)
Housing Density	0.018***	-0.017***	-0.01**	-0.027***	-0.029***	-0.029***
	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Unemployment	-0.194***	-0.145**	-0.143**	-0.14**	-0.133**	-0.131*
	(0.068)	(0.068)	(0.068)	(0.067)	(0.068)	(0.068)
Single Mothers	0.039	-0.055	-0.059	0.128***	0.144***	0.145***
	(0.032)	(0.037)	(0.037)	(0.035)	(0.037)	(0.037)
Constant	1.535	4.008	6.241	11.731***	15.492***	13.406***
	(3.589)	(3.814)	(3.907)	(3.536)	(3.848)	(3.936)
Observations	64281	64281	64281	64281	64281	64281
Number of Zip Codes	182	182	182	182	182	182
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

Notes: The dependent variables are the probability of being arrested or issued a court summons conditional on being stopped. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units.

Table 7: Determinants of Arrests and Summons, IV

	Arrests			Summons		
Model	IV (1)	IV (2)	IV (3)	IV (4)	IV (5)	IV (6)
Volunteering	-0.859*** (0.33)	-1.182*** (0.329)	-1.176*** (0.329)	-1.446*** (0.239)	-1.524*** (0.241)	-1.553*** (0.242)
Age	0.048*** (0.009)	0.047*** (0.009)	0.048*** (0.009)	0.118*** (0.009)	0.118*** (0.009)	0.117*** (0.009)
Gender	-2.995*** (0.448)	-2.987*** (0.447)	-2.947*** (0.447)	-0.818** (0.395)	-0.749* (0.395)	-0.747* (0.395)
Income	0.021 (0.026)	0.026 (0.026)	0.027 (0.026)	-0.086*** (0.022)	-0.1*** (0.022)	-0.099*** (0.022)
Poverty	-0.042 (0.029)	0.094*** (0.033)	0.095*** (0.033)	0.112*** (0.028)	0.093*** (0.032)	0.093*** (0.032)
High School	-0.114*** (0.022)	-0.05** (0.023)	-0.048** (0.023)	-0.083*** (0.022)	-0.076*** (0.024)	-0.075*** (0.024)
Less than High School	0.025 (0.02)	-0.043** (0.021)	-0.042** (0.021)	-0.141*** (0.02)	-0.136*** (0.021)	-0.137*** (0.021)
Median Age	0.188*** (0.054)	0.143*** (0.056)	0.145*** (0.056)	0.041 (0.05)	-0.005 (0.051)	-0.009 (0.051)
Youth	0.229*** (0.068)	0.092 (0.073)	0.095 (0.073)	-0.099 (0.061)	-0.174*** (0.063)	-0.175*** (0.063)
Housing Density	0.02*** (0.005)	-0.016** (0.007)	-0.016** (0.007)	-0.025*** (0.006)	-0.027*** (0.007)	-0.027*** (0.007)
Unemployment	-0.184*** (0.068)	-0.128* (0.068)	-0.126* (0.068)	-0.116* (0.066)	-0.106 (0.067)	-0.104 (0.067)
Single Mothers	0.03 (0.032)	-0.072* (0.037)	-0.075** (0.037)	0.106*** (0.029)	0.119*** (0.035)	0.119*** (0.035)
Constant	0.536 (3.634)	2.332 (3.848)	4.652 (3.936)	9.399*** (3.378)	12.9*** (3.517)	10.9*** (3.544)
Observations	64281	64281	64281	64281	64281	64281
Number of Zip Codes	182	182	182	182	182	182
First Stage F-Test (dof =)	1200.28	1448.81	1445.85	1200.28	1448.81	1445.85
Sargan-Hansen Test (p-value)	0.153	0.672	0.717	0.003	0.103	0.11
Hausman Test (p-value)	0.097	0.005	0.005	0.0001	0.0000	0.0000
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

Notes The dependent variables are the probability of being arrested or issued a court summons conditional on being stopped.. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units. First Stage F-Test values are adjusted for heteroskedasticity. Hausman test was performed by basing both covariance matrices on the disturbance variance estimate from the efficient estimator.

Table 8: First Stage Estimates

Model	Five Instruments			Cultural Capital		
	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Obstructed Driveway Complaints	-0.009***	-0.017***	-0.017***			
	(0.0004)	(0.0005)	(0.001)			
Illegal Parking	0.013***	0.020***	0.020***			
	(0.0012)	(0.0013)	(0.001)			
Vehicle Noise	0.007***	0.010***	0.010***			
	(0.0012)	(0.0013)	(0.001)			
Smoking Violations	0.588***	0.588***	0.588***			
	(0.0139)	(0.0149)	(0.015)			
Chronic Speeding	-0.255***	-0.276***	-0.275***			
	(0.0535)	(0.0671)	(0.067)			
Cultural Capital				0.387***	0.390***	0.390***
				(0.005)	(0.005)	(0.005)
Age	0.0002*	0.0002*	0.0002*	0.0001	0.0001	0.0001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Gender	0.014**	0.009**	0.009**	0.015***	0.013***	0.013***
	(0.0046)	(0.0046)	(0.0046)	(0.005)	(0.004)	(0.004)
Income	0.018***	0.019***	0.019***	0.021***	0.022***	0.022***
	(0.0013)	(0.0014)	(0.0014)	(0.002)	(0.002)	(0.002)
Poverty	-0.003***	-0.005***	-0.005***	0.002***	0.002**	0.002**
	(0.0004)	(0.0005)	(0.0005)	(0.001)	(0.001)	(0.001)
High School	0.011***	0.012***	0.012***	0.015***	0.014***	0.014***
	(0.0009)	(0.0009)	(0.0009)	(0.001)	(0.001)	(0.001)
Less than High School	0.008***	0.009***	0.009***	0.011***	0.011***	0.011***
	(0.0007)	(0.0007)	(0.0007)	(0.001)	(0.001)	(0.001)
Median Age	-0.001*	0.001	0.0004	0.005***	0.007***	0.007***
	(0.0006)	(0.0006)	(0.0006)	(0.0004)	(0.0005)	(0.0005)
Youth	0.047***	0.054***	0.054***	0.039***	0.044***	0.044***
	(0.0013)	(0.0016)	(0.0016)	(0.001)	(0.001)	(0.001)
Housing Density	0.001***	0.002***	0.002***	0.002***	0.003***	0.003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Unemployment	0.015***	0.013***	0.013***	0.022***	0.020***	0.020***
	(0.0018)	(0.0018)	(0.0018)	(0.002)	(0.002)	(0.002)
Single Mothers	-0.006***	-0.006***	-0.006***	-0.012***	-0.010***	-0.010***
	(0.0006)	(0.0006)	(0.0006)	(0.001)	(0.001)	(0.001)
Constant	-1.702***	-2.014***	-1.989***	-2.127***	-2.365***	-2.403***
	(0.0996)	(0.1106)	(0.119)	(0.128)	(0.137)	(0.142)
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

The dependent variable is the number of volunteering opportunities per 1000 people. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units. Sample of 64281 observations.

Table 9: Reduced Form Estimates

Model	Arrests			Summons		
	IV	IV	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Obstructed Driveway Complaints	-0.02	0.054	0.052	-0.069**	-0.005	-0.007
	(0.037)	(0.043)	(0.043)	(0.034)	(0.04)	(0.04)
Illegal Parking	0.048	-0.042	-0.038	0.00004	-0.053	-0.046
	(0.067)	(0.069)	(0.069)	(0.062)	(0.064)	(0.064)
Vehicle Noise	-0.186*	-0.046	-0.043	0.234**	0.249**	0.243**
	(0.097)	(0.11)	(0.111)	(0.096)	(0.105)	(0.106)
Smoking Violations	-0.27	-0.45	-0.464	-0.975***	-0.910***	-0.949***
	(0.288)	(0.304)	(0.305)	(0.254)	(0.267)	(0.267)
Chronic Speeding	-3.172*	-1.923	-1.82	0.74	0.334	0.421
	(1.871)	(2.112)	(2.115)	(1.79)	(1.913)	(1.916)
Age	0.047***	0.047***	0.048***	0.118***	0.118***	0.117***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Gender	-3.002***	-2.997***	-2.957***	-0.826**	-0.766*	-0.764*
	(0.448)	(0.447)	(0.447)	(0.395)	(0.395)	(0.395)
Income	0.001	0.004	0.006	-0.120***	-0.129***	-0.129***
	(0.024)	(0.025)	(0.025)	(0.021)	(0.021)	(0.021)
Poverty	-0.044	0.105***	0.106***	0.090***	0.094***	0.093***
	(0.03)	(0.035)	(0.035)	(0.029)	(0.033)	(0.033)
High School	-0.133***	-0.076***	-0.073***	-0.075***	-0.083***	-0.082***
	(0.023)	(0.025)	(0.025)	(0.023)	(0.025)	(0.025)
Less than High School	0.016	-0.057***	-0.055***	-0.144***	-0.147***	-0.148***
	(0.02)	(0.021)	(0.021)	(0.02)	(0.022)	(0.022)
Median Age	0.173***	0.151***	0.153***	-0.005	-0.019	-0.023
	(0.057)	(0.058)	(0.058)	(0.051)	(0.052)	(0.052)
Youth	0.196***	0.032	0.035	-0.195***	-0.259***	-0.261***
	(0.066)	(0.071)	(0.071)	(0.059)	(0.062)	(0.062)
Housing Density	0.017***	-0.019***	-0.018**	-0.029***	-0.031***	-0.031***
	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)
Unemployment	-0.188***	-0.146**	-0.143**	-0.118*	-0.120*	-0.118*
	(0.068)	(0.069)	(0.069)	(0.066)	(0.067)	(0.067)
Single Mothers	0.032	-0.061	-0.064*	0.103***	0.118***	0.119***
	(0.032)	(0.038)	(0.038)	(0.029)	(0.036)	(0.036)
Constant	3.479	4.934	6.895*	14.493***	16.437***	14.536***
	(3.745)	(3.908)	(4.043)	(3.463)	(3.555)	(3.58)
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

The dependent variables are the probability of being arrested or issued a court summons conditional on being stopped. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units. Sample of 64281 observations.

3.4.2-Cultural Capital

Given that our instruments are similar to crimes deserving of a summons, as previously discussed, we also use cultural capital as an instrument for social capital. The quantity of cultural capital in an area is likely to be a predictor of social capital, while also being conditionally exogenous. We consider cultural associations as an instrument for an exactly identified model and then together with the five other instruments. We also use cultural capital as a regressor in place of volunteering and show that it is exogenous with regards to arrests. Those results are presented in Appendix 3. Table 10 shows the estimates of the exactly identified model for both outcome variables. The estimated coefficients for volunteering are extremely similar to the ones estimated previously with the five instruments. The difference in coefficients when using the five instruments together or cultural capital alone is 0.004. Additionally, we show cultural capital to be fully exogenous with regards to crime in Appendix 3. Our results for arrests are therefore robust to the use of different instruments and the potential issue with the set of five instruments has no impact on the results. For summons, the difference between the estimates is larger (about 18%), but estimates from any set of instruments predict that volunteering has a strong negative impact on the probability of a court summons being issued to a suspect. The precision of the estimation is lesser than for arrests, but we can conclude that there exists a negative relationship. Appendix 4 presents estimates from a model estimated with all six available instruments, cultural capital and the five previous instruments. The results are very similar overall across all specifications, strengthening our confidence in the estimates. Additionally, for summons, the vehicle noise instrument is omitted from the last model in order to satisfy the overidentification restrictions.

The reduced form estimates for the exactly identified model are presented in Table 11. The coefficient for cultural capital is negative and statistically significant in all specifications. Therefore, if the exclusion restriction assumptions are valid, the reduced form estimates, coupled with the positive first stage coefficients (presented in Table 8), confirm the negative IV estimates.

Table 10: Determinants of Arrests and of Summons, Cultural Capital as an Instrument

	Arrests			Summons		
Model	IV (1)	IV (2)	IV (3)	IV (4)	IV (5)	IV (6)
Volunteering	-0.9***	-1.180***	-1.180***	-1.196***	-1.25***	-1.28***
	(0.334)	(0.334)	(0.334)	(0.228)	(0.23)	(0.23)
Age	0.048***	0.047***	0.048***	0.118***	0.118***	0.117***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Gender	-2.995***	-2.987***	-2.947***	-0.818**	-0.749*	-0.747*
	(0.448)	(0.447)	(0.447)	(0.395)	(0.395)	(0.395)
Income	0.021	0.026	0.027	-0.086***	-0.1***	-0.099***
	(0.026)	(0.026)	(0.026)	(0.021)	(0.022)	(0.022)
Poverty	-0.042	0.094***	0.095***	0.112***	0.093***	0.093***
	(0.029)	(0.033)	(0.033)	(0.028)	(0.032)	(0.032)
High School	-0.114***	-0.05**	-0.048**	-0.083***	-0.076***	-0.075***
	(0.022)	(0.023)	(0.023)	(0.022)	(0.024)	(0.024)
Less than High School	0.025	-0.043**	-0.042**	-0.141***	-0.136***	-0.137***
	(0.02)	(0.021)	(0.021)	(0.02)	(0.021)	(0.021)
Median Age	0.188***	0.143***	0.145***	0.041	-0.005	-0.009
	(0.054)	(0.056)	(0.056)	(0.05)	(0.051)	(0.051)
Youth	0.229***	0.092	0.095	-0.099	-0.174***	-0.175***
	(0.068)	(0.073)	(0.073)	(0.061)	(0.063)	(0.063)
Housing Density	0.02***	-0.016**	-0.016**	-0.025***	-0.027***	-0.027***
	(0.005)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)
Unemployment	-0.184***	-0.128*	-0.126*	-0.116*	-0.106	-0.104
	(0.068)	(0.068)	(0.068)	(0.066)	(0.067)	(0.067)
Single Mothers	0.03	-0.072*	-0.075**	0.106***	0.119***	0.119***
	(0.032)	(0.037)	(0.037)	(0.029)	(0.035)	(0.035)
Constant	0.536	2.332	4.652	9.399***	12.9***	10.9***
	(3.634)	(3.848)	(3.936)	(3.378)	(3.517)	(3.544)
First Stage F-Test (dof =)	1200.28	1448.81	1445.85	1200.28	1448.81	1445.85
Sargan-Hansen Test (p-value)	0.153	0.672	0.717	0.003	0.103	0.11
Hausman Test (p-value)	0.097	0.005	0.005	0.0001	0.0000	0.0000
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

Notes: The dependent variable is the probability of being arrested conditional on being stopped. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units. First Stage F-Test values are adjusted for heteroskedasticity. Hausman test was performed by basing both covariance matrices on the disturbance variance estimate from the efficient estimator. IV+ refers to estimation using all six instruments while the first three columns only use cultural capital. Sample of 64281 observations.

Table 11: Reduced Form Estimates, Cultural Capital as an Instrument

Model	Arrests			Summons		
	IV	IV	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Cultural Capital	-0.349***	-0.468***	-0.461***	-0.463***	-0.488***	-0.500***
	0.129	0.130	0.130	0.088	0.090	0.090
Age	0.047***	0.047***	0.048***	0.118***	0.118***	0.117***
	0.009	0.009	0.009	0.009	0.009	0.009
Gender	-3.008***	-3.002***	-2.962***	-0.839**	-0.767*	-0.766*
	0.448	0.447	0.447	0.395	0.395	0.395
Income	0.003	-0.0001	0.001	-0.118***	-0.135***	-0.134***
	0.024	0.025	0.025	0.020	0.021	0.021
Poverty	-0.044	0.092***	0.093***	0.108***	0.089***	0.088***
	0.029	0.033	0.033	0.028	0.032	0.032
High School	-0.128***	-0.067***	-0.065***	-0.104***	-0.096***	-0.096***
	0.022	0.023	0.023	0.022	0.024	0.024
Less than High School	0.016	-0.057***	-0.055***	-0.156***	-0.152***	-0.153***
	0.019	0.021	0.021	0.020	0.022	0.022
Median Age	0.184***	0.134**	0.136**	0.034	-0.015	-0.019
	0.054	0.056	0.056	0.050	0.051	0.051
Youth	0.197***	0.041	0.043	-0.163***	-0.247***	-0.249***
	0.065	0.070	0.070	0.058	0.061	0.061
Housing Density	0.018***	-0.019***	-0.019***	-0.028***	-0.031***	-0.031***
	0.005	0.007	0.007	0.006	0.007	0.007
Unemployment	-0.202***	-0.151**	-0.149*	-0.148**	-0.138**	-0.137**
	0.067	0.068	0.068	0.066	0.067	0.067
Single Mothers	0.039	-0.061*	-0.064***	0.126***	0.138***	0.138***
	0.031	0.036	0.036	0.028	0.034	0.035
Constant	2.347	5.124	7.478*	12.567***	16.547***	14.631***
	3.527	3.750	3.853	3.307	3.451	3.488
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

The dependent variables are the probability of being arrested conditional on being stopped and the probability of being issued a summons conditional on being stopped. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units. Sample of 64281 observations.

3.4.3-Bootstrapped Hausman Test

Hausman tests for all models are performed by basing both covariance matrices on the disturbance variance estimate from the efficient estimator in order to successfully compute the test statistics. While this procedure is generally acceptable for the type of endogeneity tests that we perform, it may cast doubt on the validity of the standard Hausman test results given its strict assumptions. In addition, the mere fact that robust standard errors are used throughout (and do impact some results) violates the assumption that one of the two estimators must be fully efficient under the null. It is not uncommon in general to not require either estimator to be fully efficient, but this may lead to misleading conclusions. To account for this eventuality, we conduct a bootstrapped version of the Hausman test, following Cameron and Trivedi (2010).

Formally, the test statistic of the test can be expressed in the quadratic form:

$$H = (b_1 - b_0)' (V(b_1 - b_0))^{-1} (b_1 - b_0) \qquad (6)$$

Where b_1 is the efficient estimator under the null and b_0 is consistent under both the null and the alternative. This test follows asymptotically the chi-squared distribution with degrees of freedom h, where h is the rank of the matrix of the difference in variances. When performed in common statistical software, it is assumed that b_1 is fully efficient, which eventually leads to a simplification in the denominator leading to the simplified test statistic:

$$H = (b_1 - b_0)' (Var(b_0) - Var(b_1))^{-1} (b_1 - b_0) \qquad (7)$$

This simplification, performed during the standard computation of the test, is inaccurate in our case at the least because of heteroskedasticity. We therefore use bootstrapping to estimate $V(b_1 - b_0)$, avoiding the need for the problematic assumption. We bootstrap with N = 400 repetitions in order to obtain 400 estimates of both b_1 and b_0 and therefore compute their difference. The estimation of $V(b_1 - b_0)$ then becomes:

$$(N\text{-}1)^{-1} \textstyle\sum_b ((b_{1b} - b_{0b} - (N \sum_b (b_{1b} - b_{0b}))^{-1}) ((b_{1b} - b_{0b} - (N \sum_b (b_{1b} - b_{0b}))^{-1})' \qquad (8)$$

This process allows us to compute the less restrictive test statistic and better assess the extent of endogeneity. The results, shown in Table 12, are very similar to the ones from the standard test.

The only noticeable difference is that we cannot reject the null at the 10% level for arrests in the baseline model, if we do not control for month and borough effects.

For all other specifications, the difference is small and does not alter the results in any way. Regardless of heteroskedasticity and of our decision to base both covariance matrices on the disturbance variance estimate from the efficient estimator, the results remain unchanged.

Table 12: Bootstrapped Hausman P-Values, Selected Models

	Hausman	Bootstrapped Hausman	Borough FE	Month FE	Borough X Month
Basic Model					
Arrests	0.097	0.115	no	no	no
	0.005	0.007	yes	yes	no
	0.005	0.008	yes	yes	yes
Summons	0.0001	0.0000	no	no	no
	0.0000	0.0000	yes	yes	no
	0.0000	0.0000	yes	yes	yes
IV with Cultural Capital					
Arrests	0.0723	0.0704	no	no	no
	0.0055	0.0053	yes	yes	no
	0.0067	0.0064	yes	yes	yes
Summons	0.0000	0.0000	no	no	no
	0.0000	0.0000	yes	yes	no
	0.0000	0.0000	yes	yes	yes
IV with Six Instruments					
Arrests	0.0837	0.0856	no	no	no
	0.0045	0.0052	yes	yes	no
	0.0051	0.0058	yes	yes	yes
Summons	0.0000	0.0000	no	no	no
	0.0000	0.0000	yes	yes	no
	0.0000	0.0000	yes	yes	yes

Notes: Numbers are all p-values. The dependent variable Arrests is the probability of being arrested conditional on being stopped and the dependent variable Summons is the probability of being issued a summons conditional on being stopped. Bootstrapping was performed with 400 repetitions. Estimation was performed with robust standard errors. Sample of 64281 observations.

3.5-Robustness Checks

3.5.1-Additional Controls

Since the validity of the results relies on there being no omitted variable bias, we test the robustness of the results to the inclusion of additional controls. Our baseline IV results are robust to controlling for the proportion of people who have obtained a bachelor's or graduate degree, the proportion of foreign-born residents in the area, a squared income term, the normalized number of graffiti in the area and the race of the suspect (restricted to white and white Hispanic or black and Black Hispanic)[12].

Further, the results are not sensitive to our aggregation unit. We also estimate a negative though smaller relationship between social capital and crime for every specification of the baseline models when aggregating variables at the police precinct level. This may be due to precincts being larger than zip code areas and therefore capturing social capital less precisely. Methods for aggregating at the precinct level are presented in Appendix 1.

3.5.2-Clustering

In addition to the previous robustness checks, the statistical significance of the IV results is also robust to clustering the standard errors at either the borough or zip code level. We do not present the results with these corrections given potential issues inherent to clustering in our sample. That is, clustering at the borough level would yield only five clusters, while inference asymptotics are based on the number of clusters approaching infinity[13]. Clustering at the zip code level leads to vastly imbalanced clusters and it relies on the implicit assumption that there is no intra-correlation

[12] Results are omitted due to spacing concerns but are available upon request.
[13] Bertrand, Duflo and Mullainathan (2004), Cameron and Miller (2010) and Bell, Morgan, Kromrey and Ferron (2010)

in the standard errors between boroughs[14]. These approaches can lead to "the cure being worse than the disease", see Austin and Schaffer (2007). Since the statistical significance is invariant to the level at which the standard errors are corrected for the residuals departing from the i.i.d. assumption, our confidence in the findings is increased.

3.5.3-Nonlinearity

Volunteering is negatively linked to both crime outcomes, but the models above restrict the relationship to a linear one. It may not be as simple; volunteering may have a different effect depending on the quantity provided. Specifically, some volunteering could help the community at first, but a lot more could reflect that something in the area creates a vast need for volunteering while also encouraging crime. For instance, a natural disaster would create an immediate need for volunteering and lead to more crimes. Another example would be an area with an exceptionally high level of drug addiction.

To capture this potential relationship, we also estimate the models with a squared term added for volunteering. The results are presented in Appendix 5. All five instruments, including their squares, are included in the estimation for arrests, while Vehicle Noise and its square are excluded from the estimation for summons. We find that the higher order term is not statistically significant in any IV specification for either dependent variable, while the lower order terms retain similar values as before. We can also reject that the IV and OLS estimates are the same; therefore IV is the only consistent estimator. We find no evidence of nonlinearity in the data and conclude that the relationship between social capital and crime is linear.

[14] Austin and Schaffer (2007), Bertrand, Duflo and Mullainathan (2004), Cameron and Miller (2010)

3.6-Conclusion

While the impact of social capital on crime has been present in the literature for several years, little credible evidence has been provided. In this section, we use IV estimators to present conclusive evidence that civic norms, measured through volunteerism, have a statistically significant negative impact on crimes targeted by the Stop-and-Frisk program. The relationship between volunteering and crime is linear and is robust to controlling for several social and economic factors as well as correcting for endogeneity and additional unobserved heterogeneity. Specifically, an increase in volunteering would bring a statistically significant decrease in both the probability of a suspect being arrested and the probability of a suspect being issued a court summons. There appears to be no other channel through which volunteering could affect the Stop-and-Frisk program or the decision of individuals to commit crimes. We also show that OLS results are likely to be biased due to endogeneity. Civic norms do help reduce crime, but areas with heavy volunteering may be more vulnerable in the first place, leading to an understatement of the effect of social capital.

There may remain concerns with the results, as they are obtained from a randomly selected 10% subsample of all stops for the year 2011 due to computational limits with reverse geocoding using *Google Maps*. It would be interesting to test how the results translate to the whole sample given that they do change slightly when altering the sample to control for ethnicity. If anything, given random sampling and a large sample size, we would expect the significance of the results to strengthen.

An important limitation is that the definition of volunteering used may be restrictive. The volunteering data only comprises opportunities from the NYC Service initiative, which are likely to represent only a fraction of all volunteerism. These opportunities may have a larger impact since they are recognized municipally and operate through an official channel, but many other types of volunteering presumably help develop social capital in the city. Further, we investigate one out of the three types of social capital generally recognized in the literature. To understand the overall importance of social capital for a community, one would also need to consider trust and associational networks.

The results also have significant policy implications, a topic we return to in the general concluding section.

4-Discrimination

4.1-Introduction

The New York Police Department's (NYPD) Stop-and-Frisk program, in which an officer of the law can routinely stop, question and frisk any citizen suspected of criminal intent, has come under fire from various advocacy groups[15]. While defenders of the Stop-and-Frisk program claim it has saved over 7000 lives[16] and played a key role in New York City's decrease in crime over the past years, some civil rights groups claim it constitutes a violation of freedom. Further, others have accused it of racial profiling given the overwhelming majority of minorities targeted by the program (consistently around 85% of all stops in each year)[17]. The program has been targeted by legislative action, including high profile cases and class action lawsuits such as *Floyd, et al. v. City of New York, et al.*[18].

Beyond legal and philosophical implications, the social desirability of a program such as Stop-and-Frisk relies on its ability to efficiently deter crime. This in turn relies on police officers stopping suspects in a productive manner by targeting legitimate criminal activity. Testing for discrimination is challenging since an analysis of disparate impact alone does not constitute evidence of discrimination. After all, police officers who stop more members of a group are not biased provided that the stops are productive and lead to arrests or summons. Furthermore, if the majority of the members of a group are concentrated into higher or lower crime neighborhoods, disparate impact and even more sophisticated techniques may lead to misleading conclusions. Therefore, when testing if police officers are discriminating against blacks, we must investigate

---

[15] New York Civil Liberties Union, Stop-and-Frisk Data, www.nyclu.org/content/stop-and-frisk-data, Consulted August 10th 2013

[16] Burke, Kathy, NYPD's Ray Kelly: "Stop-and-Frisk Saved 7,383 Lives", Newsmax, http://www.newsmax.com/Newsfront/kelly-stop-frisk-saved/2013/07/22/id/516422, July 22nd 2013

[17] Bump, Philip, Why Racism in Numbers Will Bring Down the NYPD in the Stop-and-Frisk Trial, The Atlantic Wire, http://www.theatlanticwire.com/national/2013/05/nypd-stop-and-frisk-numbers/65561/, May 24th 2013

[18] Center for Constitutional Rights, *Floyd, et al. v. City of New York, et al*, /ccrjustice.org/floyd, Consulted August 10th 2013

whether the stops which involve blacks are as productive as the ones involving whites. If an officer stops many blacks without just cause, then those stops will not lead to arrests or summons and the probability of these two outcomes will therefore be lower for this subgroup.

One way to test for discrimination empirically is the hit rates test, which goes beyond disparate impact. This test was developed in Knowles, Persico and Todd (2001), an important contribution which has been the foundation of several papers on discrimination. In this section of the paper, we build on results from Coviello and Persico (2013), which uses the hit rates test and finds no evidence of racial discrimination when considering the whole city of New York over the 2003-2012 period. While their overall result is telling, racial inequity, unless institutionalized, may be restricted to certain subgroups of the population and vary with time, location and crime type. We therefore investigate subgroup treatment effect heterogeneity.

Specifically, we conduct the hit rates test for each borough of the city individually, then for every year separately and lastly for every borough-year combination. Further, we also look for gender inequity as well as discrimination of black men specifically, using the richer data from the sample of the social capital section. In addition, we conduct nonlinear Oaxaca decompositions (Yun, 2004) as another tool to detect bias. Our analysis uncovers plausible evidence of racial discrimination in police stops in the boroughs of Manhattan, the Bronx and Brooklyn, when considering all the data collected from 2003 to 2012. Further, when considering individual years for the entire city, we find some indication of bias for the entire year of 2011. Including individual of Hispanic origin to the analysis reduces the difference in the arrest rates between white and white Hispanics and black and Black Hispanics, lowering the likelihood that these tests can detect racial bias.

We uncover large and surprising differences in the presence of discrimination across categories of crime. Drug crimes and possession of a weapon are examples of crimes which we hypothesize have a higher likelihood of being tainted by discrimination. These categories of crimes have received a special emphasis since they are related to the War on Drugs, which has also been accused by many advocacy organizations of being discriminatory[19]. We perform the test on all stops for which the suspected crime was related to drugs or weapons and find robust, conclusive evidence of racial inequity against African-Americans. After controlling for the year effects, adding precinct fixed effects and clustering by precinct, blacks are more than 1.89% less likely to be arrested when stopped. Considering the mean of 5.98% for arrests, this is a difference of approximately 30%, which reflects that drug related stops are conducted in an unproductive and racially discriminatory manner. Further, African-Americans are 2.6% less likely to be issued a summons (mean of 8.36%).

Unlike our investigation of discrimination of all crimes, we find robust evidence of discrimination in every borough of the city for these two specific crime categories. Our analysis uncovers a worrisome trend, as inequity in these categories has increased markedly each year since 2009. In contrast, we did not find any evidence of discrimination when conducting the analysis on data collected prior to 2009, and this is unlikely due to there being missing records, which would only affect the data quality from 2003-2005. Whether this change in the trend in stops is due to the changing economic environment or differences in policing after The Great Recession begun is beyond the scope of this section and a topic for further research.

In the remainder of this section, we begin with a brief review of the discrimination literature. In section 4.3, we discuss the economic model that underlies our empirical analysis and introduce both the empirical and identification strategy. The main results are presented and discussed in section 4.4.Robustness checks are presented in section 4.5. Section 4.6 then summarizes the main findings of the entire section.

---

[19] Human Rights Watch, United States – Punishment and Prejudice: Racial Disparities in the War on Drugs, http://www.hrw.org/legacy/reports/2000/usa/Rcedrg00-05.htm, Consulted on August 10th 2013.

4.2-Literature Review

The foundation of the model we use can be found in the literature on optimal auditing, in which papers such as Becker (1968) modeled the decision to participate in criminal activity given a certain probability of being audited. The idea that discrimination would lead to lower profits for those who engage in it comes originally from Becker (1957). Intuitively, agents who discriminate are restricting their behavior based on irrational criterion or beliefs which do not translate to any practical benefits. They should therefore not operate as efficiently as agents who optimize their behavior based on true costs and benefits.

Similar tests have found many applications in income reporting Scotchmer (1987), mortgage lending Van Order and Zorn (1995) or academic publishing Smart and Waldfogel (1996). One earlier contribution which uses a similar empirical strategy is Ayres and Waldfogel (1994), which finds evidence of discrimination against minorities in the setting of bail bonds by judges.

The hit rates test for discrimination that we use was developed by Knowles, Persico and Todd (2001), and underwent some minor extensions in Persico and Todd (2006) and Coviello and Persico (2013). In each of these papers, the authors use the test to investigate whether police officers stop or search more African-Americans based on reasonable evidence or out of racial bias. For example, Coviello and Persico (2013) use the same data set as this paper and find that the difference in the arrest rates between the two racial groups is not statistically significant; therefore blacks are not significantly less likely to be arrested when stopped or searched. They conclude their analysis as providing evidence that suggests police officers operate without bias and perform stops or searches based on their productivity.

This section of the paper extends Coviello and Persico (2013) by considering some important distinctions in the application of the test and has parallels to procedures considered by Dharmapala and Ross (2003). These authors argue that the results of the hit rates test may not be robust to allowing for variations in the probability of an audit or offenses of varying degrees of severity across types of crime. We consider heterogeneity across temporal, geographic and crime dimensions.

This section of the paper extends Coviello and Persico (2013) by considering some important distinctions in the application of the test and has parallels to procedures considered by Dharmapala and Ross (2003). These authors argue that the results of the hit rates test may not be robust to allowing for variations in the probability of an audit or offenses of varying degrees of severity across types of crime. We consider heterogeneity across temporal, geographic and crime dimensions.

4.3-Empirical Strategy

We use the model of pedestrian and police behaviour from Coviello and Persico (2013), itself adapted from Persico and Todd (2006). The model provides a framework which demonstrates the strength of the hit rates test in detecting the presence of discrimination. Those interested in an in-depth look at the theory should consult the aforementioned papers as we only provide a summary overview.

The main feature of the model is that it "incorporates potential police heterogeneity in intensity of racial bias and in costs of searching as well as pedestrian heterogeneity in the benefits and costs from committing a crime.", Coviello and Persico (2013). The intuition behind the model is that, if officers are not biased, then the arrest rate should be equal across races. If it is not, it implies that police officers stop people belonging to one group even though it is less productive to do so. Given a pedestrian of race r with r = White (W) or African-American (A) and with a set of additional characteristics c effortlessly observable by the police. An officer can distinguish without cost between groups (r,c) but not across individuals belonging to the same group.

Let P denote the number of officers *p* who each have a capacity to search given by S_p and a cost per search given by t_p. Discrimination is incorporated by allowing the gains from a successful search to vary depending on the race of the suspect. Specifically, the gains of an officer *p* from arresting a pedestrian of race *A* is given by $Y_{pA} = Y_{pW} + B(p)$, where B(p) denotes discrimination and is assumed to be of the same sign across all officers. Then, with $S_p(r,c)$ denoting the sum of the searches that an officer *p* performs on a group *(r,c)* and $K^{r,\,c}$ denoting the crime rate of group *(r,c)* , the expected payoff of an officer is:

$$\sum_{r,c} S_p(r,c)\,[y_{pr}\,K^{r,\,c}\,(S(r,c)) - t_p] \tag{5}$$

Persico and Todd (2006) shows that a generically unique Nash equilibrium exists for this game, which ultimately leads to the following theorem:

Theorem 1 *(Persico and Todd (2006): foundation for hit rates test):*

In the equilibrium, the hit rate is the same across all subgroups within a race that are distinguishable by police. Also, if the police are unbiased, then the hit rate is the same across races. If the police are biased against race r, the hit rate is lower in race r than in the other race.

This theorem implies that the hit rates test is *robust to omitted variable bias*, since it is the same across all groups within a race which are distinguishable by a police officer.

We first apply the test individually by borough and by year. Following Coviello and Persico (2013), we consider both OLS (which has been shown to be inadequate by the authors) and precinct fixed effects in order to account for cross-regional heterogeneity in crime rates. Additionally, we include clustering at the precinct level to capture any remaining heterogeneity across precincts, which may lead to artificially low standard errors.

Specifically, our results come from the following equation:

$$Y_{in} = \alpha_1 + \beta B_i + \Theta_n + \varepsilon_{in} \quad (6)$$

Where Y_{in} is the probability of being arrested or issued a summons, B_i is a dummy variable which has value 1 if the suspect is black or 0 if the suspect is white and Θ_n is a vector of 76 dummy variables for each precinct of the city. If the coefficient β is negative and statistically significant, then blacks have a lower probability of being arrested or issued a summons and, since the test is robust to omitted variables bias by construct (which we also test later), this is evidence of discrimination since the difference in probabilities is due to race alone. It reflects that blacks are stopped too often; if they were not, then the productivity of the stops would be the same across races and the coefficient on black would be non-significant.

We can obtain results by estimating the model imposing the restriction $\Theta_n = 0$. This will be referred to as the OLS analysis and assumes that there is no unobserved heterogeneity at the precinct level which is constant through time and correlated with regressors. As shown in Coviello and Persico (2013), in our case, not controlling for unobserved heterogeneity common to each precinct of the city would lead to misleading results since the coefficient for black could reflect a difference across neighborhood quality or crime rates. If more members of a group are stopped in precincts which have higher crime rates in general, then a higher probability of being arrested or summoned to court would be mistakenly attached to that group.

In contrast, we can estimate the model while relaxing the previous restriction; this is referred as the FE analysis. It assumes that there are time-invariant characteristics within each precinct which may affect the outcome variables. It also assumes that these characteristics are uncorrelated across precincts; the regressors must be exogenous and therefore uncorrelated to past, present and future shocks. This procedure imposes additional assumptions but allows us to control for crime heterogeneity across precincts which are likely to affect arrest or summons rates.

To conduct statistical inference, we cluster the standard errors at the precinct level. A potential concern is that the number of clusters when performing the analysis by borough does not meet the usual minimum of approximately 50 found in the literature. Therefore, when studying boroughs individually, clustering may lead to artificially inflated standard errors and the second FE specification may provide more reliable results. Fortunately, in our analysis, clustering did not lead to any difference in the statistical significance of the results except for those using data from the Bronx alone. In all of the tables that follow which explore differences across boroughs, we present the results with both fixed effects and clustering, but our preferred estimates are contained in the fifth column of each table.

Similar to Dharmapala and Ross (2003), we investigate heterogeneity over various subgroups of crime, since the hit rates test may be better suited to detect discrimination for certain types of crime than others. Specifically, the model above considers that every stop comes after an officer has observed a suspect and established reasonable suspicion. This is not the case for crimes which are in progress or which may require an immediate intervention. Stopping these crimes provides an immediately observable benefit to the officer at a search cost t_p of 0. As an extreme example, while racial bias may affect the prosecution and sentencing of a pedestrian who is caught assaulting another, it is unlikely that B(p) would be so large as to dictate whether the officer would stop the assault. It follows that the hit rates test may be better suited to detect discrimination for crimes which leave room for judgement because B(p) implicitly enters in the establishment of a reasonable suspicion. Similarly, even for crimes which do not require immediate intervention, the cost of searching may differ for different types of crime.

More importantly, the perception of an officer who discriminates against a group is likely to vary with the type of crime. A biased officer may unjustly identify certain types of crimes with African-Americans or whites and therefore stop too many members of the group for those crimes. For instance, a biased officer who suspects too many blacks of carrying drugs does not necessarily have the same inaccurate perceptions about blacks and disorderly conduct. We therefore perform the test on subsamples of different types of crime.

Those subsamples are suitable for hit rates analysis since they include stops which the officers were not obligated to report (no use of force, frisking, arrest, summons or refusal to identify). Additionally, the samples are selected using the type of crime which was suspected by the officer at the time of the decision to perform the stop. This means that we do not use any ex-post information; we do not use a sample of stops involving a type of crime but a sample of stops involving all *suspicions* related to a type of crime. Since the entire Stop-and-Frisk program relies on an officer only stopping pedestrians conditional on suspecting them of a crime, it follows that officers should always be able to provide the type of crime for which a suspect was stopped.

We also put an emphasis on crimes related to drugs or possession of a weapon. One interesting aspect of these crimes is that, in addition to representing a large fraction of all stops, they are central to the controversial and long-lasting nationwide War on Drugs[20] which has, along with the Stop-and-Frisk program, been accused several times of discrimination against African-Americans, who constitute the majority of suspects arrested for drug related crimes (Nunn, 2002).

[20] Popularized in 1971 by Nixon, the War on Drugs has since been identified as one of the leading causes for the particularly high incarceration rate in the United States - more than 1% of the adult population. In particular, African-American males are 6 times as likely to be incarcerated as white males and three times as likely as Hispanic males.

The PEW Center on the States, Pew Report Finds More than One in 100 Adults are Behind Bars, http://web.archive.org/web/20080303025427/http://www.pewcenteronthestates.org/news_room_detail.aspx?id=35912, Consulted August 10th 2013
Human Rights Watch, United States – Punishment and Prejudice: Racial Disparities in the War on Drugs, http://www.hrw.org/legacy/reports/2000/usa/Rcedrg00-05.htm, Consulted on August 10th 2013.
The Sentencing Project, New Incarceration Figures: Thirty-Three Consecutive Years of Growth, http://www.sentencingproject.org/doc/publications/inc_newfigures.pdf, Consulted August 10th 2013

4.4-Results

4.4.1-Subgroup Heterogeneity for all Crimes

First, when considering arrests by borough, as presented in Table 13, we find weak evidence of inequity in the Bronx, since the estimate from column 5 is negative but only of approximately half a percent. The other estimated coefficients are all positive or statistically non-significant and therefore point towards there being no racial bias.

When considering discrimination in arrests from year to year, presented in Table 14, we find no proof of racial bias, as blacks are not significantly less likely to be arrested in any year of the sample. It appears as though officers of the NYPD should even have stopped more African-Americans between 2005 and 2008.

When considering arrests for both years and boroughs together, as presented in Appendix 6, there may be indications of discrimination only for 2011 in the Bronx, but it is an isolated result which on its own does not make a strong case. We revisit this result later when considering both arrests and summons together.

For summons, presented in Table 15 by borough and Table 16 by years, all the coefficients are negative and statistically significant, which would be evidence of discrimination if the police *only* considered issuing summons. Therefore, as in previous literature, the results for summons do not imply that there was any inequity. Most likely, the police consider both arrests and summons together and both outcomes must therefore be combined to investigate overall discrimination. In the model presented earlier, it is logical to assume that police officers want to minimize crime but that they put a larger weight on crimes deserving of an arrest since they are more severe and cause more damage. Accordingly, Coviello and Persico (2013) have shown that a weight of 4 to 1 between arrests and summons would lead to no evidence of discrimination for their overall results.

We believe that this assumption is not unlikely to hold overall, but it is not sufficient to rule out discrimination in the more detailed results. For our results, the assumption is enough to rule out racial bias in Queens and Staten Island, but a six or seven to one ratio would have to be considered for Manhattan and Brooklyn while even a 9:1 ratio would leave evidence of discrimination for the Bronx. The plausibility of such ratios is left to the judgement of the reader, but we believe this is a first hint that there may be some inequity in the program and that further investigation is warranted. For year to year estimates, a 4:1 weighting scheme is sufficient to rule out discrimination in all years except for 2011, for which a 9:1 ratio is necessary. Again, the likeliness of the assumption is subjective, but we believe it is most likely indication of bias and that it appears stronger in the later years of the sample. These two results also point towards discrimination being at its highest in the Bronx in 2011, which is consistent with the earlier results for arrests by year and borough and lends credence to our previous estimates.

In Appendix 7, we also perform the test with Hispanics added in the sample. This addition leads to a decrease in the difference in arrest rates and, though the overall signs do not change, a 4:1 weighting scheme between arrests and summons is sufficient to erase any sign of racial bias. Discrimination appears to be stronger when considering African-Americans versus whites.

Table 13: Arrest Made, Overall and by Borough

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-0.420***	-0.437***	-0.437	0.379***	0.355***	0.355*
	(0.037)	(0.037)	(0.469)	(0.046)	(0.046)	(0.207)
Constant	6.140***					
	(0.034)					
P-value of H0 : ui = 0				0.0000	0.0000	0.0000
Black						
Manhattan	-0.572***	-0.571***	-0.571	0.219**	0.212**	0.212
	(0.097)	(0.097)	(0.830)	(0.104)	(0.104)	(0.612)
Bronx	-0.415***	-0.329**	-0.329	-0.593***	-0.547***	-0.547
	(0.157)	(0.157)	(0.647)	(0.167)	(0.167)	(0.495)
Brooklyn	-1.164***	-1.216***	-1.216***	0.584***	0.531***	0.531*
	(0.053)	(0.053)	(0.465)	(0.069)	(0.069)	(0.292)
Queens	0.254***	0.258***	0.258	0.551***	0.535***	0.535*
	(0.082)	(0.082)	(0.450)	(0.099)	(0.099)	(0.288)
Staten Island	1.740***	1.803***	1.803**	0.583***	0.594***	0.594
	(0.113)	(0.113)	(0.763)	(0.143)	(0.143)	(0.765)
Constant						
Manhattan	8.621***					
	(0.089)					
Bronx	7.023***					
	(0.151)					
Brooklyn	4.998***					
	(0.050)					
Queens	6.607***					
	(0.072)					
Staten Island	4.346***					
	(0.080)					
P-value of H0 : ui = 0						
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0000	0.0000	0.0000
Clustered SE	no	no	yes	No	no	yes
Time FE	no	yes	yes	No	yes	yes
Precinct FE	no	no	no	Yes	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis.

Table 14: Arrest Made, by Year

Model	OLS	OLS	FE	FE
	(1)	(2)	(3)	(4)
Black				
2003	-0.776***	-0.776	0.850***	0.85
	(0.224)	(0.862)	(0.271)	(0.518)
2004	-1.082***	-1.082	0.754	0.754
	(0.143)	(0.601)	(0.174)	(0.417)
2005	-0.367***	-0.367	1.189***	1.189***
	(0.118)	(0.475)	(0.147)	(0.283)
2006	-0.752***	-0.752*	0.280**	0.28
	(0.094)	(0.449)	(0.117)	(0.295)
2007	-0.447***	-0.447	0.177	0.177
	(0.112)	(0.564)	(0.140)	(0.364)
2008	0.558***	0.558	0.613***	0.613**
	(0.108)	(0.653)	(0.138)	(0.262)
2009	-0.244**	-0.244	0.178	0.178
	(0.111)	(0.675)	(0.138)	(0.354)
2010	0.086	0.086	0.410**	0.41
	(0.117)	(0.578)	(0.141)	(0.253)
2011	-0.868***	-0.868*	-0.156	-0.156
	(0.104)	(0.509)	(0.125)	(0.243)
2012	-1.033***	-1.033	-0.038	-0.038
	(0.113)	(0.659)	(0.136)	(0.403)
P-value of H0 : ui = 0				
2003			0.0000	0.0000
2004			0.0000	0.0000
2005			0.0000	0.0000
2006			0.0000	0.0049
2007			0.0000	0.2319
2008			0.0000	0.0000
2009			0.0000	0.0000
2010			0.0000	0.0000
2011			0.0000	0.0000
2012			0.0000	0.0000
Clustered SE	no	yes	no	yes
Time FE	no	no	no	no
Precinct FE	no	no	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis.

Table 15: Summons Issued, Overall and by Borough

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.070*	0.095**	0.095	-1.753***	-1.736***	-1.736***
	(0.038)	(0.038)	(0.360)	(0.047)	(0.047)	(0.295)
Constant	6.122***					
	(0.035)					
P-value of H0 : ui = 0				0.0000	0.0000	0.0000
Black						
Manhattan	-0.13	-0.12	-0.12	-2.359***	-2.350***	-2.350***
	(0.082)	(0.082)	(1.043)	(0.087)	(0.087)	(0.369)
Bronx	-1.477***	-1.353***	-1.353**	-1.637***	-1.495***	-1.495***
	(0.160)	(0.160)	(0.546)	(0.171)	(0.171)	(0.235)
Brooklyn	0.175**	0.149**	0.149	-2.191***	-2.221***	-2.221***
	(0.068)	(0.068)	(0.730)	(0.088)	(0.088)	(0.608)
Queens	-0.069	-0.038	-0.038	-0.672***	-0.635***	-0.635
	(0.074)	(0.074)	(0.677)	(0.089)	(0.089)	(0.694)
Staten Island	-1.521***	-1.522***	-1.522***	-1.592***	-1.586***	-1.586***
	(0.113)	(0.113)	(0.359)	(0.143)	(0.143)	(0.449)
Constant						
Manhattan	5.733***					
	(0.075)					
Bronx	8.338***					
	(0.155)					
Brooklyn	6.458***					
	(0.064)					
Queens	5.589***					
	(0.066)					
Staten Island	5.987***					
	(0.080)					
P-value of H0 : ui = 0				0.0000	0.0000	0.0000
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0000	0.0000	0.0106
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variables is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis.

Table 16: Summons Issued, by Year

Model	OLS (1)	OLS (2)	FE (3)	FE (4)
Black				
2003	-0.793***	-0.793	-1.488***	-1.488***
	(0.187)	(0.538)	(0.228)	(0.421)
2004	0.434***	0.434	-1.521***	-1.521***
	(0.155)	(0.513)	(0.189)	(0.403)
2005	0.612***	0.612	-1.478***	-1.478***
	(0.141)	(0.811)	(0.175)	(0.474)
2006	0.085	0.085	-0.771***	-0.771
	(0.110)	(0.615)	(0.136)	(0.591)
2007	0.534***	0.534	-1.207***	-1.207**
	(0.125)	(0.805)	(0.155)	(0.530)
2008	0.602***	0.602	-1.810***	-1.810***
	(0.111)	(0.500)	(0.141)	(0.364)
2009	0.283**	0.283	-2.230***	-2.230***
	(0.112)	(0.467)	(0.139)	(0.461)
2010	-0.184**	-0.184	-2.656***	-2.656***
	(0.117)	(0.497)	(0.142)	(0.449)
2011	-0.702***	-0.702	-2.135***	-2.135***
	(0.102)	(0.488)	(0.122)	(0.295)
2012	-0.126	-0.126	-1.388***	-1.388***
	(0.104)	(0.415)	(0.125)	(0.224)
P-value of H0 : ui = 0			0.0000	0.0000
2003			0.0000	0.0000
2004			0.0000	0.0000
2005			0.0000	0.0000
2006			0.0000	0.6492
2007			0.0000	0.0000
2008			0.0000	0.0000
2009			0.0000	0.0000
2010			0.0000	0.3017
2011			0.0000	0.0000
2012			0.0000	0.0000
Clustered SE	no	yes	no	yes
Time FE	no	no	no	no
Precinct FE	no	no	yes	yes

The dependent variables are the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis.

We also conduct the test by gender for the restricted 2011 sample, the results are presented in Appendix 8. For arrests, we find strong evidence of discrimination against men in every borough, though most notably in Manhattan, Brooklyn and Staten Island. Police officers consistently stop too many men even if the stops are less productive. The bias of -2.874% is very large, as it constitutes a difference of over 47% in arrests when considering the mean of 6.04%. This may be due to the fact that men commit a higher proportion of crimes in general, so police officers may be quicker to mistakenly suspect them. The NYPD should stop more women as part of the program in order to maximize the productivity of their stops. For summons, there is little sign of discrimination except for the Bronx, as most coefficients for boroughs are positive or statistically non-significant, perhaps because police officers are more likely to suspect women of smaller crimes. Overall, if we consider any weight which is not heavily skewed towards summons, we find statistically significant signs of inequity against men.

We also present Oaxaca decompositions in Appendix 9 as another method to investigate differences in arrests and summons between groups. We present decompositions computed using both IV and OLS, given the previously identified endogeneity of volunteering. Additionally, given the binary nature of our outcome variables, the decompositions are nonlinear and follow the method proposed in Yun (2004). These decompositions lead to the same conclusions as the hit rates test once controlling for precinct fixed effects. There is little evidence of city-wide racial discrimination once controlling for precinct FE, though we find strong signs of gender discrimination.

4.4.3-Discrimination by Crime Type: The War on Drugs

We proceed with the same test, now restricted to drug and weapon related crimes. Approximately 71% of the sample is related to weapon crimes and 29% to drugs. We include possession and sale of both Marihuana as well as other controlled substances. Statistics for the sample are presented in Table 17. Additionally, we include dummy variables for the different types of crime in all specifications. Data for 2003 is unavailable and observations for years 2004-2006 are almost exclusively for Brooklyn.

Table 17: Summary Statistics, War on Drugs Related Crimes

	Mean	St.Dev	N
Outcome			
Arrest	5.98	23.72	932,918
Summons	8.36	27.69	932,918
Race of the pedestrian			
Black	90.77	28.95	932,918
Crimes			
Possession of a Weapon	71.15	45.3	932,918
Marihuana	11.48	31.88	932,918
Possession of Substances	9.74	29.64	932,918
Illegal Sales of Substances	7.63	26.54	932,918
Mandated Stops			
Mandated	77.8	41.56	932,918

Mean is in percent. Mandated Stops represents the proportion of stops which must be reported by law.

The overall results for all years and boroughs, presented in the first section of Table 18, yield strong, consistent evidence of discrimination in all specifications, even after the inclusion of precinct FE and clustering. It appears as though stops related to the War on Drugs, which have larger implications for the African-American community, have been conducted in a discriminatory manner by the NYPD between the years 2004 and 2012. For arrests, considering the estimated coefficient of -1.89% and the mean of 5.98%, this constitutes an overall difference of over 32%.

As can also be seen from Table 18, the coefficient on black is statistically significant at the 1% level and negative in all five boroughs. Discrimination has been city-wide, though most prominent in Manhattan, Queens and Staten Island. It may be interesting to note that those boroughs have the lowest proportion of African-Americans in the sample, respectively 89%, 87% and 64% (95% for the Bronx and 94% for Brooklyn).

Looking at the results from year to year, shown in Table 19, the difference in arrest rates has grown every year since 2009, an increase of over 66% from -1.95% to -2.94% in 2012. It may be interesting to note that the large increase in media coverage documented in Coviello and Persico (2013) started in early 2011, after the program had been operating in a discriminatory manner for around 2 years. While this increase in coverage provides no evidence of discrimination, the program started to be heavily criticized at the time where we do find signs of inequity.

The estimates for summons, displayed in Table 20 by borough and Table 21 by year, provide even stronger evidence of discrimination and they are statistically significant at the 1% level. The important conclusion implied from these estimates is that evidence of racial bias remains no matter the weight which is assigned to either outcome. In fact, if the police consider summons at all, overall discrimination will be larger than 1.89%. For instance, with a weight scheme of 4:1 as used earlier, overall discrimination would be of 2.03%. There are also important disparities between boroughs, the strongest indication of inequity coming from Brooklyn, the Bronx and Manhattan, while estimates for Queens are less conclusive.

While the strongest indication of racial bias for summons is for the years 2004-2005, this is likely due to most observations being for Brooklyn, which was identified as the borough with the most evidence of inequity. There is no obvious trend as for arrests; discrimination has increased from 2007 to 2010 but has since been decreasing.

In Appendix 12, we also present results by year and borough, again dismissing the results which include clustering. Consistent with our previous findings, there is strong and consistent indication of discrimination in all boroughs, particularly concentrated in the later years. Some coefficients are very large and make up for over 50% of the mean of arrests. This provides overwhelming evidence that there has been discrimination in the program, at least in recent years.

Table 18: Arrest Made, Overall and by Borough, War on Drugs Crimes

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-2.509***	-2.461***	-2.461***	-1.889***	-1.888***	-1.888***
	(0.085)	(0.085)	(0.641)	(0.100)	(0.100)	(0.450)
Constant	7.695***					
	(0.113)					
P-value of H0 : ui = 0						
Black						
Manhattan	-3.391***	-3.347***	-3.347*	-2.781***	-2.711***	-2.711*
	(0.195)	(0.195)	(1.906)	(0.215)	(0.215)	(1.416)
Bronx	-0.957***	-1.002***	-1.002	-1.320***	-1.349***	-1.349**
	(0.239)	(0.239)	(0.755)	(0.252)	(0.252)	(0.666)
Brooklyn	-2.356***	-2.322***	-2.322***	-0.978***	-1.003***	-1.003**
	(0.136)	(0.135)	(0.536)	(0.159)	(0.159)	(0.453)
Queens	-1.578***	-1.606***	-1.606	-2.183***	-2.169***	-2.169***
	(0.227)	(0.227)	(1.036)	(0.273)	(0.273)	(0.793)
Staten Island	-0.343	-0.089	-0.089	-2.233***	-2.076***	-2.076***
	(0.284)	(0.286)	(1.033)	(0.345)	(0.345)	(0.463)
Constant						
Manhattan	20.614***					
	(0.676)					
Bronx	4.579***					
	(0.325)					
Brooklyn	13.751***					
	(0.379)					
Queens	21.717***					
	(0.734)					
Staten Island	18.048***					
	(1.031)					
P-value of H0 : ui = 0						
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0000	0.0000	0.0046
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis. All estimations include dummy variables for each crime type in the sample.

Table 19: Arrest Made, by Year, War on Drugs Crimes

Model	OLS	OLS	FE	FE
	(1)	(2)	(3)	(4)
Black				
2004	-0.563	-0.563	-0.061	-0.061
	1.027	0.986	1.061	0.929
2005	-1.478***	-1.478	0.095	0.095
	0.519	1.017	0.624	0.962
2006	-2.070***	-2.070***	-0.936***	-0.936
	0.201	0.631	0.234	0.662
2007	-3.008***	-3.008***	-1.902***	-1.902***
	0.226	0.747	0.266	0.660
2008	-1.111***	-1.111	-0.773***	-0.773
	0.229	0.787	0.274	0.533
2009	-2.469***	-2.469***	-1.846***	-1.846**
	0.225	0.793	0.265	0.739
2010	-2.147***	-2.147**	-2.098***	-2.098***
	0.241	0.922	0.283	0.555
2011	-2.964***	-2.964***	-2.315***	-2.315***
	0.219	0.835	0.254	0.527
2012	-2.990***	-2.990**	-2.942***	-2.942***
	0.248	1.133	0.288	0.824
P-value of H0 : ui = 0			0.0000	0.0000
2004			0.0000	0.0000
2005			0.0947	0.0000
2006			0.0000	0.0000
2007			0.0000	0.0000
2008			0.0000	0.0000
2009			0.0000	0.0000
2010			0.0000	0.0000
2011			0.0000	0.0000
2012			0.0000	0.0000
Clustered SE	no	yes	no	yes
Time FE	no	no	no	no
Precinct FE	no	no	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0: ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis. All estimations include dummy variables for each crime type in the sample.

Table 20: Summons Issued, Overall and by Borough, War on Drugs Crimes

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-1.212***	-1.219***	-1.219**	-2.615***	-2.611***	-2.611***
	(0.100)	(0.100)	(0.544)	(0.117)	(0.117)	(0.524)
Constant	7.276***					
	(0.134)					
P-value of H0 : ui = 0						
Black						
Manhattan	0.473**	0.495**	0.495	-2.507***	-2.492***	-2.492***
	(0.216)	(0.215)	(1.463)	(0.236)	(0.236)	(0.663)
Bronx	-2.919***	-2.776***	-2.776***	-2.927***	-2.791***	-2.791***
	(0.307)	(0.307)	(0.724)	(0.324)	(0.323)	(0.524)
Brooklyn	-2.388***	-2.546***	-2.546**	-3.828***	-3.900***	-3.900***
	(0.179)	(0.179)	(1.169)	(0.210)	(0.209)	(1.184)
Queens	-0.760***	-0.596***	-0.596	-0.674**	-0.535**	-0.535
	(0.222)	(0.222)	(0.925)	(0.267)	(0.267)	(1.106)
Staten Island	-2.969***	-2.858***	-2.858***	-2.316***	-2.234***	-2.234*
	(0.263)	(0.264)	(1.049)	(0.319)	(0.319)	(1.278)
Constant						
Manhattan	6.941***					
	(0.748)					
Bronx	8.106***					
	(0.418)					
Brooklyn	8.502***					
	(0.502)					
Queens	5.013***					
	(0.717)					
Staten Island	7.451***					
	(0.954)					
P-value of H0 : ui = 0						
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0000	0.0000	0.0003
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0: ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis. All estimations include dummy variables for each crime type in the sample.

Table 21: Summons Issued, by Year, War on Drugs Crimes

Model	OLS	OLS	FE	FE
	(1)	(2)	(3)	(4)
Black				
2004	-3.658**	-3.658	-5.201***	-5.201**
	(1.657)	(2.313)	(1.709)	(2.448)
2005	-0.150	-0.150	-4.373***	-4.373***
	(0.951)	(1.620)	(1.133)	(1.598)
2006	-0.757***	-0.757	-0.756**	-0.756
	(0.272)	(1.108)	(0.314)	(0.916)
2007	-1.209***	-1.209	-1.786***	-1.786**
	(0.296)	(1.233)	(0.345)	(0.869)
2008	-0.905***	-0.905	-2.938***	-2.938***
	(0.278)	(0.775)	(0.333)	(0.781)
2009	-1.263***	-1.263	-3.649***	-3.649***
	(0.268)	(0.886)	(0.316)	(1.177)
2010	-1.819***	-1.819**	-4.294***	-4.294***
	(0.273)	(0.805)	(0.321)	(0.887)
2011	-1.347***	-1.347**	-2.142***	-2.142***
	(0.233)	(0.636)	(0.268)	(0.539)
2012	-1.018***	-1.018*	-2.016***	-2.016***
	(0.250)	(0.572)	(0.293)	(0.451)
P-value of H0 : ui = 0				
2004			0.0000	0.0000
2005			0.0000	0.0000
2006			0.0000	0.0000
2007			0.0000	0.0000
2008			0.0000	0.0000
2009			0.0000	0.0000
2010			0.0000	0.0000
2011			0.0000	0.0000
2012			0.0000	0.0000
Clustered SE	no	yes	no	yes
Time FE	no	no	no	no
Precinct FE	no	no	yes	yes

The dependent variable is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0: ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis. All estimations include dummy variables for each crime type in the sample.

We next perform the hit rates test on other crimes, which make up approximately 38% of the remaining sample. These are crimes which we identify as being perhaps less likely to exhibit discrimination, allowing us to test our previous intuition and the sensitivity of the hit rates test results to the inclusion of different types of crime. Statistics for this subsample are presented in Table 22.

Table 22: Summary Statistics, Other Selected Crimes

	Mean	St.Dev	N
Outcome			
Arrest	6.6	24.83	774,372
Summons	5.33	22.46	774,372
Race of the pedestrian			
Black	82.44	38.05	774,372
Crimes			
Criminal Trespass	37.54	48.42	774,372
Grand Larceny Auto	29.39	45.55	774,372
Grand Larceny	13.9	34.6	774,372
Assault	10.59	30.77	774,372
Petit Larceny	8.12	27.32	774,372
Murder and Rape	0.5	6.8	774,372
Mandatory Stops			
Mandatory	39.76	48.94	774,372

Mean is in percent. Mandated Stops represents the proportion of stops which must be reported by law. Some values for the type of crime are missing for the years 2003-2005.

All results for this sample are presented in Appendix 13. In accordance with our previous intuition, we find no sign of discrimination. Indeed, for arrests, the coefficients are positive, slightly larger than the ones obtained from the test on all crimes and the discrepancy in the arrest rates is significantly smaller than for drug-related crimes. The results by borough yield the same conclusions. We find no evidence of discrimination, and it appears as though whites may in fact be stopped too often for these crimes, particularly in Manhattan and Staten Island. The results by year are also consistent; we find no evidence of discrimination, as the difference in arrest rates between both races is mostly statistically non-significant.

The results for summons are similar to those obtained from the test on all crimes; the coefficients are statistically significant and negative. Still, even with equal weights given to both crime outcomes, we find no evidence of discrimination overall. The boroughs with the largest coefficients for arrests are also the boroughs with the smallest estimates for summons, which leads to a smaller discrepancy between boroughs if we consider both outcomes. The conclusion is the same for individual years; there does not appear to be any discrimination.

The results are in accordance with our intuition regarding the use of the model. Discrimination is likely to vary depending on the crime and it is therefore necessary to conduct separate analyses for different subsets of crimes. We cannot say whether the racial bias we identified may be due to drug and weapons related crimes leaving more room for discrimination, to structural differences from being associated with the War on Drugs or biased beliefs of police officers regarding those crimes. Regardless, we do find statistically significant evidence of racial bias for the subset of crimes which we had identified as likely to give rise to discrimination and find no indication for crimes which we had identified as less likely to give rise to discrimination.

4.5-Robustness Checks

The results of our analysis are robust to controlling for the percentage of stops for which police officers were obligated to report as well as to omitting the crime indicator variables from all regressions.

Additionally, while the hit rates test is robust to omitted variable bias by assumption, reverse geocoding methods allow us to directly test this assumption for the first time in the literature.

Using the methods described in Appendix 1, we are able to conduct robustness checks using the randomly selected subsample from the year 2011 which was used in the social capital section. Specifically, we find strong evidence of discrimination for crimes related to the War on Drugs whether we use FE by precincts or zip codes. The estimated coefficient on the indicator for being African-American when we include zip code FE is approximately three percent lower for arrests and two percent lower for summons, which is consistent with our previous finding that 2011 was one of the years with the most discrimination.

Interestingly, including additional control variables to the estimating equation has an impact on the estimated coefficient for black but not on the overall conclusion. This is surprising, since the creators of the testing procedure have argued that it should be robust to omitted variable bias, yet this appears to be a questionable assumption - though one with little impact in our case. Indeed, evidence of discrimination remains when controlling for inflation adjusted median income, percentage of the population living under the poverty line, high school completion rates, bachelor and graduate degrees completion rate, median age, proportion of the population aged between the ages of 15 and 24, unemployment, proportion of housing structures with more than 20 housing units, proportion of single-mother headed households and the proportion of African-Americans living in each zip code. When controlling for these variables, the coefficient on black for arrests is -2.92 and is statistically significant at the 1% level. The coefficient for black for summons is -1.43 but is no longer statistically significant. Therefore, the controls do have an impact on the results of the hit rates test, though these do not alter the overall conclusion that there is discrimination when considering both outcomes. It appears as though the assumption about the robustness of the hit rates test to omitted variable bias may not always be empirically valid, though further investigation would be necessary.

Lastly, our results may be impacted by the problem of multiplicity, as some of our tests include many hypotheses. The tests are based on rejecting the null hypothesis when it is unlikely to be true based on the observed data. Multiplicity arises when we perform many such tests, since the probability of committing a type I error (rejecting the null when it is true) increases with the number of hypotheses which are tested. Formally, the familywise error rate (FWER) is “the probability of making one or more type I errors among all the single hypotheses when performing multiple pairwise tests on families of hypotheses that are similar in purpose”[21]. For example, the chance of committing at least one type I error when testing an effect on six different outcomes (5% significance level, two-sided tests) is 15.9%.[22]

The importance of accounting for multiple level testing has been demonstrated in several contributions in economics, including Ding and Lehrer (2011). The authors show that the estimated benefits of students being assigned to smaller sized classes vary whether the p-values are corrected for the multiplicity of outcomes (effect on reading score, mathematics score, listening score, etc.). The correction can have important policy implications. In our case, not applying it could lead to the conclusion that police officers are discriminating against African-Americans when the estimated effect is in fact random.

To account for this issue, we apply the Bonferroni correction, a procedure regarded by many researchers as very conservative, to preserve the familywise error rate. The overwhelming majority of our important results (column 5 for results by borough, column 4 for results by year) remain statistically significant at the 5% level. Fortunately, the design of our results provides a very simple way to see how the results change with the correction. Throughout all tables presented previously, all results which were not statistically significant at the 1% level are not significant when applying the correction. Of the results which were statistically significant at the 1% level, all are significant at the 5% level when applying the correction, except for three results which become significant at the 10% level. These are: the coefficient for arrests in 2008 in Table

[21] Extract from Ding and Lehrer (2011)
[22] Calculations performed in Ding and Lehrer (2011)

14 (all crimes), the coefficient for summons in 2005 in Table 19 (War on Drugs related crimes) and the coefficient for arrests in 2012 in Table 52 (selected other crimes). Therefore, none of our overall conclusions are affected by the correction.

4.6-Conclusion

We extend the literature on discrimination in the Stop-and-Frisk program by considering whether the impacts differ across boroughs and time as well as studying racial bias for different types of crime. When considering arrests over all crimes, African-Americans are not less likely to be arrested when precinct fixed effects are included. The coefficients for summons are statistically significant and negative. When the objective function of a police officer comprises both arrests and summons, the assumption used in Coviello and Persico (2013) is insufficient to rule our discrimination in Manhattan, the Bronx and Brooklyn. They showed that, conditional on the police prioritizing arrests to summons at a rate of at least four to one, there would be no evidence of racial bias against blacks. We find that a rate of approximately ten to one would be necessary to rule out discrimination in the Bronx. When considering different years individually, 2011 is the year which provides the strongest evidence of discrimination, though the plausibility of the results relies on the assumption made about policing priorities.

Adding Hispanics to the sample reduces the discrepancy in arrest rates. It does not change the overall results, but a lower ratio between arrests and summons is required to rule out discrimination. We identify large inequity against men in 2011, overall and in every borough. It appears as though police officers overweight the propensity of men to commit crimes, leading to suboptimal crime reduction. Additionally, the Oaxaca decompositions also substantiate this result.

We also considered crimes related to the War on Drugs specifically, both because they have been the target of racism allegations and because, as argued previously, they are crimes for which the hit rates test may be more likely to uncover racial bias. For arrests, we find strong evidence of discrimination across all specifications, statistically significant at the 1% level. This observed inequity is larger in Manhattan, Queens and Staten Island, in which it is estimated to be approximately equal to a third of all arrests. We also identify a surprising trend when performing the hit rates test on individual years; discrimination has been growing every year since 2009, even with the additional attention and criticism aimed at the program.

Additionally, the results for summons provide stronger evidence of bias, implying that overall discrimination is larger than that identified for arrests. Regardless of the assumption about policing priorities, there has been significant discrimination in the past for drug related crimes.

In line with our intuition, we find no indication of bias when performing the hit rates test on a sample of crimes including trespassing, larceny and assault. This may be evidence that police officers who discriminate against blacks for drug related crimes may in turn be proportionally less likely to stop them for other crimes.

We highlight the importance of considering multiple specifications and performing analyses of varying degrees of precision when researching discrimination. Considering smaller subsets allows more precision, which can make a difference when looking for clues of something that is inherently subjective and biased.

To the best of our knowledge, this constitutes the first definite scientific evidence of racial discrimination by the NYPD in the Stop-and-Frisk program. It could give weight and credence to the criticism of Stop-and-Frisk in recent years. While we cannot speak as to the legality of the program, it has not been operating in a socially optimal way given that African-Americans are needlessly stopped too often. Police officers should shift their attention to whites, at least for drug related crimes, which would increase the productivity of their stops and lead to more crime reduction at no additional cost or effort.

5-General Conclusion

This paper examined two questions related to deterring crime, using data from NYC. Our results have several implications, both for the economics literature and for policy-makers.

In the social capital section, we present robust evidence that civic norms, measured through the incidence of volunteerism, have a statistically significant negative impact on crime, even after controlling for a variety of possible other determinants Our results are consistent with the prior literature on civic norms and crime in finding a decrease of approximately 14% in the probability of being arrested, while Buonnano, Montolio and Vanin (2009) obtain estimates of approximately 15%. Our study is one of the only ones to estimate causal effects via an instrumental variables estimator. Since we only consider two aggregates of crime, it would be interesting to conduct the analysis on the entire Stop-and-Frisk sample for 2011 and estimate the impacts on precise types of crime. Our results show that an expansion of the NYC Service program and of volunteerism in general could significantly reduce crime. This could be an especially beneficial endeavor compared to law enforcement, as crime reduction remains only a side effect of the program, therefore volunteering could offer additional social benefits as well as positive externalities associated with social capital. The policy implications of our results are fairly straightforward. Our analysis is conducted using data from a single program in a single city, but there is substantial variation across the zip codes in the city and we are the first paper in the literature to have access to data at this level. On the other hand, external validity may be limited since, while there is no indication that the relationship would not hold for other cities, especially in the same country, it would be interesting to conduct additional research using similar microeconomic data.

In the discrimination section, we present evidence of racial bias against African-Americans, at the least for crimes related to the War on Drugs, and plausibly for all crimes in certain boroughs in later years of the sample. Our results come at a time where the Stop-and-Frisk program is severely critiqued for its alleged disproportionate impact on minorities. Our evidence supports this claim, since the current program discriminates heavily against blacks, which may not only be socially undesirable in itself, but also reflects inefficient policing and a waste of resources. Officers of the NYPD are not decreasing crime as productively as they could if they targeted more white suspects. This is not to say that the program has not contributed to decreasing crime, but that both opposing parties would gain from the NYPD targeting blacks less disproportionately, since it would lead to even less crime while increasing the respect of civil rights. Our results, similarly to Dharmapala and Ross (2003), also emphasize some critical distinctions regarding the use of the hit rates test as originally proposed. The results vary critically depending on the types of crime included in the sample. This could be due to several reasons such as differences in the types of criminal, the propensity of certain crime types to give rise to discrimination or simply diverging policing methods between crime types. Additionally, given the nature of racial bias, it is likely to not be stable through time or common to all geographical regions. When considering these shortcomings of the results previously established in the literature, we reach the opposite conclusion and find conclusive evidence of discrimination. Additionally, unlike Dharmapala and Ross (2003), we also correct for multiple testing using the Bonferroni correction to adjust for the probability of committing a type I error. All our results are therefore robust to the issue of multiplicity in testing.

For both sections, it would also be interesting to consider different cities, perhaps in which there is no Stop-and-Frisk program, to investigate whether part of the effect is specific to New York City.

This paper also illustrates several examples of the potential benefits of combining rich administrative data with geographical information and geocoding methods to obtain additional data which allows deeper analyses with stronger, more credible results.

6-Bibliography

Akçomak, I. Semih and Bas ter Weel, 2008, The Impact of Social Capital on Crime: Evidence from the Netherlands, *IZA DP No. 3603*.

Baumer, Eric, Steven F. Messner, and Richard Rosenfeld, 2004, “Dimensions of Social Capital and Rates of Criminal Homicide.” *American Sociological Review* 69:882-903.

Becker, G.S., 1968, “Crime and Punishment: An Economic Approach”, *Journal of Political Economy*, 76 (2), pp. 169-217.

Bell, B, G.B Morgan, J. D. Kromney and J. M. Ferron, 2010, The Impact of Small Cluster Size on Multilevel Models: A Monte Carlo Examination of Two-Level Models with Binary and Continuous Predictors, Section on Survey Research Methods – JSM 2010, p.457-467

Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan, 2004, How Much Should We Trust Differences-in Differences Estimates?, *Quarterly Journal of Economics*, 2004, *119*(1), pp. 249-75.

Bjørnskov, Christian. 2006. The Multiple Facets of Social Capital. *European Journal of Political Economy* 22:22–40.

Bourdieu, Pierre, 1986, The Forms of Capital, In J. Richardson (Ed.) Handbook of Theory and Research for the Sociology of Education (New York, Greenwood), 241-258.

Buonanno, Paolo, 2003, The Socioeconomic Determinants of Crime. A Review of the Literature, Working Paper Series, No. 63, University of Milan - Bicocca

Bursik, R.J.,1986, Ecological stability and the dynamics of delinquency. In A. J. Reiss and M. Tonry (Eds.), *Communities and Crime* (pp. 35−66). Chicago: University of Chicago Press.

Bursik, R.J. , 1988, Social disorganization and theories of crime and delinquency: Problems and prospects. *Criminology*, 26, 519−551.

Buonanno, Paolo, Daniel Montolio and Paolo Vanin, 2009, *Does Social Capital Reduce crime?*, *The Journal of Law & Economics*, vol.52, p.145-170.

Calvo´ -Armengol, Antoni, and Yves Zenou, 2004, Social Networks and Crime Decisions: The Role of Social Structure in Facilitating Delinquent Behavior. *International Economic Review* 45:939–58.

.

Calvo´ -Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, 2005, Peer Effects and Social Networks in Education and Crime. Working Paper No. 645. Research Institute of Industrial Economics, Stockholm.

Calvo´ -Armengol, Antoni, Thierry Verdier, and Yves Zenou, 2007, Strong and Weak Ties in Employment and Crime. *Journal of Public Economics* 91:203–33.

Cameron, A. Colin and Douglas L. Miller, 2011, Robust Inference with Clustered Data, A. Ullah and D. E. Giles eds., *Handbook of Empirical Economics and Finance*, CRC Press, pp.1-28.

Cameron, A.Colin and Pravin K. Trivedi, 2010, Microeconometrics Using Stata, *Stata Press – Revised Edition*, 706 pages.

Coviello, Decio and Nicola Persico, 2013, An Economic Analysis of Black-White Disparities in NYPD's Stop-and-Frisk Program.

Ding, Weili and Steven F. Lehrer, 2011, Experimental Estimates of the Impacts of Class Size on Test Scores: Robustness and Heterogeneity, *Education Economics*, *Taylor and Francis Journals*, vol. 19(3), pages 229-252.

Dharmapala D, Ross SL, 2003, Racial bias in motor vehicle searches: additional theory and evidence. Work. Pap. No. 2003–12R

Ehrlich, Isaac, 1973, Participation in Illegitimate Activities -- A Theoretical and Empirical Investigation, *Journal of Political Economy*, Vol. 81 (3)

Fajnzylber, Pablo, Daniel Lederman, and Norman Loayza, 2002, Inequality and Violent Crime. *Journal of Law and Economics* 45:1–40.

Freeman, Richard B, 1991, Crime and the Employment of Disadvantaged Youths.Working Paper No. 3875. National Bureau of Economic Research, Cambridge, Mass.

Grenfell, M., 2008, Pierre Bourdieu: Key Concepts, Acumen Publishing, 256 pages

Grenfell, M, 2011, Bourdieu, Language and Linguistics, Continuum, 272 pages

Grogger, Jeffrey T., 1998, Market Wages and Youth Crime. *Journal of Labor Economics* 16: 756–91.

Hahn, J, J. Ham and H. Moon, 2011, The Hausman Test and Weak Instruments, *Journal of Econometrics*

Heaton, Paul, 2006, Does Religion Really Reduce Crime? *Journal of Law and Economics*, 49:147–72.

Imrohoroglu, A., Merlo, A. and P. Rupert , 2001, “What Accounts for the Decline in Crime?”, Federal Reserve Bank of Cleveland, wp 0008.

Kelly, M., 2000, “Inequality and Crime”, *The Review of Economics and Statistics*, 82 (4), pp. 530-539.

Knack, Stephen, and Philip Keefer, 1997, Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics* 112:1251–88.

Knowles, J, Nicola Persico and Petra Todd, 2001, "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy*, University of Chicago Press, vol. 109(1), pages 203-232

Lederman, Daniel, Norman Loayza, and Ana Marı´a Mene´ndez, 2002, Violent Crime: Does Social Capital Matter? *Economic Development and Cultural Change* 50:509–39.

Levitt, Steven D., 2002, Using Electoral Cycles in Police Hiring to Estimate the Effects of Police on Crime: Reply, *The American Economic Review*, Vol. 92, No. 4, pp. 1244-1250

Lochner, L. and E. Moretti, 2001, "The effect of education on crime: evidence from prison inmates, arrests, and self-reports", NBER Working Paper no. 8605.

Messner, Steven F., Eric Baumer, and Richard Rosenfeld, 2004, Dimensions of Social Capital and Rates of Criminal Homicide. *American Sociological Review* 69:882–903.

Moulton, B. R., 1990, An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units, *The Review of Economics and Statistics*, 72, 334-338.

Nichols, Austin and Mark E Schaffer, 2007, Clustered standard errors in Stata, United Kingdom Stata Users' Group Meetings 2007, Stata Users Group.

Nunn, Kenneth B., 2002, Race, Crime and the Pool of Surplus Criminality: Or Why the War on Drugs Was a War on Blacks, *Gender, Race & Justice* (6): 381

Persico, Nicola, 2002, Racial Profiling, Fairness, and Effectiveness of Policing" American Economic Review 92(5): 1472-97.

Persico, Nicola, 2008, Racial Profiling? Detecting Bias Using Statistical Evidence.

Persico, Nicola and Petra Todd, 2006, Using Hit Rates to Test for Racial Bias in Law Enforcement: Vehicle Searches in Wichita." The Economic Journal 116: F351-F367.

Putnam, Robert D, 1993, *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, N.J.: Princeton University Press.

———, 2000, *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Rainsford, H. F., 1955, Long Geodesics on the Ellipsoid, *Bull. Géod.* 37 (1): 12–22

Rosenfeld, Richard, Steven F. Messner, and Eric P. Baumer, 2001, Social Capital and Homicide. *Social Forces* 80:283–310.

Rubio, Mauricio, 1997, Perverse Social Capital—Some Evidence from Colombia. *Journal of Economic*

Salmi, V and J. Kivivuori, 2006, The Association between Social Capital and Juvenlie Crime: The Role of Individual and Structural Factors, *European Journal of Criminology*, 3, 2, 123-148.

Sampson, R.J. and W.B. Groves, 1989, Community structure and crime: Testing social disorganization theory. *American Journal of Sociology*, 94, 774−802.

Sampson, R.J., J.D. Morenoff and F. Earls, 1999, Beyond social capital: Spatial dynamics of collective efficacy of children. *American Sociological Review*, 64, 633−660.

Sampson, R.J., S.W. Raudenbush, and F. Earls, 1997, Neighbourhoods and violent crime: A multi level study of collective efficacy. *Science*, 227, 916−924.

Shaw, C. R. and H.D. McKay, 1942, Juvenile delinquency and urban areas; A study of rates of delinquents in relation to differential characteristics of local communities in American cities. Chicago: University of Chicago Press.

Vincenty, T., 1975, Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations, *Survey Review*, XXIII (176): 88–93.

Weibull, Jorgen, and Edgar Villa, 2005, Crime, Punishment and Social Norms. SSE/EFI Working Paper 610. Stockholm School of Economics, Stockholm.

Wilson, W.J., 1990, The Truly Disadvantaged. Chicago: University of Chicago Press.

Wilson, W.J., 1996, When Work Disappears. New York: Vintage Books.

Yun, Myeong-Su, 2004, Decomposing differences in the first moment, Economics Letters, 82: 275-280.

Yun, Myeong-Su, 2005, A Simple Solution to the Identification Problem in Detailed Wage Decompositions, Economic Inquiry, 43: 766-772.

Appendix 1 – Geocoding Methods for Data Aggregation

Microeconomic data may sometimes include coordinates or part of an address with each observation. If this information is used efficiently, it can provide the researcher with much deeper and more detailed data to exploit. Since these geographical methods are seldom used in economics, we aim to provide an example of successful aggregation of economic data using different denominations, which provides deeper research opportunities and more robust results. In particular, it allows the researcher to convert a list of disparate observations into panel data and allows for a wider range of estimation techniques.

From Addresses to NYC Precincts and Zip Codes

We use NYC data on volunteerism as an example to illustrate the methods. The process of gathering longitude-latitude coordinates from addresses (which may not include zip codes) is referred to as geocoding and can be done by simply entering an address into *Google Maps* and retrieving the coordinates of the location. Note that *Google Maps* imposes a limit of 2500 observations per 24 hour period for every IP address.

For aggregation by zip code, it becomes trivial to match every volunteering opportunity to a zip code, as the coordinates can be used to reverse geocode every observation (enter the coordinates in *Google Maps* to pinpoint the address) to obtain the full address, in which zip codes are included. Note that more complete addresses lead to higher success rates in general; city names as well as regional or state names should be included to avoid confusion. Directly associating zip codes to precincts may prove much more arduous since there is no straightforward way to establish the center point of a zip code to minimize distances. In those instances, publicly available geographic "*shapefile*" maps can be used to overlap polygons representing both precincts and zip codes for an overall idea of the locations. It is then theoretically feasible to calculate the proportion of each zip code polygon which is contained in each precinct polygon and associate the two by maximizing this proportion over all zip codes. A less precise but much simpler approach is to select some addresses or sets of coordinates for each zip code and pinpoint their location on a map over which precinct polygons are overlapped, in order to identify the precinct in which they are located.

For aggregation by precincts, the goal is to associate every observation to the closest precinct. The addresses of every precinct headquarters being publicly available, we can in turn find the coordinates of every headquarters by geocoding each address individually. The next step is to minimize the geographical distance between the two sets of coordinates: those for volunteering opportunities and those for precinct headquarters. While it may seem attractive to use Pythagoras on the absolute value of the differences in order to calculate the Euclidian distance, this would be incorrect as it implies that the Earth is a flat two dimensional Cartesian plane. More sophisticated methods are necessary and, while there exist many ways to minimize a geographical distance, we present only two: the Haversine and Vincenty formulas. The first is used to calculate great circle distances between two points on a sphere using latitude and longitude. It is simpler as it relies purely on the use of trigonometry. The Vincenty calculations are iterative methods to calculate distances on the surface of a spheroid and require more sophisticated computations.

The Haversine formula can be expressed as:

$$\text{Haversin}(d / r) = \text{Haversin}(\Phi_2 - \Phi_1) + \cos(\Phi_1)\cos(\Phi_2)\,\text{Haversin}(L_2 - L_1) \tag{1}$$

Where:

$$\text{Haversin}(\Theta) = \sin^2(\Theta/2) \tag{2}$$

And d is the distance between two points located on a sphere, r is the radius of the sphere, Φ_1 and Φ_2 are the latitudes of the two points while L_1 and L_2 are the longitudes. By using the radius of the Earth (6371 kilometers) and the coordinates from every volunteering opportunity N and every precinct M, it is possible to minimize the distance between each N and the 76 different M by solving for *d* in the Haversine formula.

Note that this requires the inverse Haversine function or the arcsine function given by:

$$d = r\ \mathrm{Haversin}^{-1}(h) = 2r \arcsin (h^{1/2}) \qquad (3)$$

or equivalently:

$$= 2r \arcsin ((\sin^2((\Phi_2 - \Phi_1)/2) + \cos(\Phi_1) \cos(\Phi_2) \sin^2((L_2 - L_1)/2))^{1/2}) \qquad (4)$$

We can then identify which precinct headquarters is closest to every volunteering opportunity and aggregate the data accordingly.

Another more sophisticated method was developed in Vincenty (1975). It assumes the figure of the Earth to be that of an oblate spheroid, which leads to accuracy gains over longer distances. We propose that the researcher consider both approaches when distances are small, as the Haversine calculations may be more accurate for distances which tend to 0. This comes from the simple fact that, for small values of Θ, $\sin(\Theta)$ is approximately linear in Θ and therefore leads to increased accuracy as the distance approaches 0.

The Vincenty formula also uses the geographical coordinates of two distinct points to find the distance *d*. Note that it is an iterative method and requires an ellipsoid model of the earth; therefore it cannot be computed mechanically. The method first requires the calculation of $L = L_2 - L_1$ and of $X_i = \arctan((1-f)\tan(\Phi_i))$, $i=1,2$ where f is the flattening of the ellipsoid used for the calculations. We then set L as the initial value for λ and evaluate the following equations by iteration until λ converges:

$$\mathrm{Sin}\Theta = ((\cos X_2 \sin\lambda)^2 + (\cos X_1 \sin X_2 - \sin X_1 \cos X_2 \cos\lambda)^2)^{1/2} \qquad (5)$$

$$\cos\Theta = \sin X_1 \sin X_2 + \cos X_1 \cos X_2 \cos\lambda \quad (6)$$

$$\Theta = \arctan(\sin\Theta/\cos\Theta) \quad (7)$$

$$\sin a = \cos X_1 \cos X_2 \sin\lambda / \sin\Theta \quad (8)$$

$$\cos^2 a = 1 - \sin^2 a \quad (9)$$

$$C = (f/16)\cos^2 a\,(4 + f(4-3\cos^2 a)) \quad (10)$$

$$\cos(2\Theta_n) = \cos\Theta - 2\sin X_1 \sin X_2 / \cos^2 a \quad (11)$$

$$\lambda = L + (1-C)\, f \sin a\,(\Theta + C\sin\Theta(\cos(2\Theta_n) + C\cos\Theta(-1+2\cos(2\Theta_n)))) \quad (12)$$

After having achieved converge at the desired level of accuracy, the distance can be obtained from:

$$Y^2 = \cos^2 a\,(e^2 - p^2 / p^2) \quad (13)$$

$$A = 1 + (Y^2 / 16\,384)(4096 + Y^2(-768 + Y^2(320-175Y^2))) \quad (14)$$

$$B = (Y^2 / 1024)(256 + Y^2(-128 + Y^2(74 - 47Y^2))) \quad (15)$$

$$\Delta\Theta = B\sin\Theta(\cos(2\Theta_n) + 0.25B(\cos\Theta(-1 + 2\cos^2(2\Theta_n)) - (1/6)B\cos(2\Theta_n)(-3+4\sin^2\Theta)(-3+42\cos^2(2\Theta_n)))) \quad (16)$$

$$d = pA\,(\Theta - \Delta\Theta) \quad (17)$$

Where *a* is the azimuth at the equator, *e* is the radius at the equator, *p* is the radius at the poles and Θ is the arc length between two points on the auxiliary sphere. Those interested in more details on the method should consult Rainsford (1955) and Vincenty (1975).

For our purpose, the solution of (17) is the desired distance between one volunteering opportunity and one precinct. Aggregation can then be achieved by associating every volunteering opportunity to the precinct with the closest headquarters.

A word on conversion between types of coordinates

The former methods require the use of geographical coordinates in degrees, which are then converted into radians. Many sets of coordinates in the US, such as the Stop-and-Frisk data, may come in the form of State Plane Coordinates (SPC). In order to use reverse geocoding and associate each stop to an address and a zip code, coordinates must first be converted from SPC to latitude-longitude. An important distinction is that converters often use SPC data reported in US Survey Feet to convert coordinates, while raw data may be reported in International feet or meters. While the difference between both is quite small, it can lead to vast differences when using SPC, since we may be calculating distances of many hundreds of feet or more. It is therefore critical to convert the coordinates into US Survey Feet, which are defined exactly as 1200/3937 of a meter. Failure to do so will lead to a large loss of accuracy and render reverse geocoding extremely imprecise. Once the SPC data is adjusted, the conversion between sets of coordinates requires the use of specialized geographic information system software such as *ArcGIS*, since is not implemented in most software programs used in economics (to our knowledge).

Appendix 2 – Volunteering in Categories

Another interesting question is whether different types of volunteering have divergent impacts on crime. It may be that volunteerism reinforces civic norms regardless of the objective or that some types of volunteering develop social capital more efficiently. For instance, volunteering may have a larger impact in situations of crisis such as floods or hurricanes, since the community may rally together and form a stronger bond.

To investigate the question, we use the six reported types of volunteering in the data to create four distinct categories with different objectives. We merge education and health into a public policy category while strengthening communities and helping neighbors in need are grouped into a general community service category. Opportunities regarding the environment and emergency preparedness are left as two distinct categories. Together, these four categories amount to over 92% of all volunteering. One difficulty is that some categories of volunteering are only present in a few areas, which leads to sample issues

Using the baseline models with the four independent variables leads to inconclusive results overall. For arrests, emergency preparedness is statistically significant at the 5% level for OLS, though any inference from this estimate is difficult as this category in particular is concentrated in few areas. It seems plausible that being prepared for emergencies may lessen their impacts but it is not clear that the effect would be through social capital.

The effect is no longer statistically significant when considering IV, but we cannot reject the null hypothesis of the Hausman test at the 5% level and should therefore use the efficient OLS estimates.

Table 23: Determinants of Arrests, Volunteering in Categories

Model	OLS	OLS	OLS	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Community and Neighbors	1.036	1.112	1.147	-2.664	-10.263	-9.652
	(0.943)	(0.944)	(0.944)	(5.996)	(9.862)	(9.848)
Health and Education	-3.17	-4.052	-4.179*	13.686	21.343	19.966
	(2.468)	(2.475)	(2.479)	(13.45)	(20.012)	(20.007)
Environment	-9.803	-7.126	-7.429	-97.61	164.673	154.114
	(6.865)	(6.885)	(6.893)	(137.616)	(196.895)	(196.992)
Emergency Preparedness	-5.379**	-5.92**	-5.768**	-3.308	1.77	1.345
	(2.734)	(2.737)	(2.746)	(11.212)	(12.848)	(12.788)
Age	0.047***	0.047***	0.047***	0.047***	0.047***	0.047***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Gender	-3.003***	-3.001***	-2.961***	-2.927***	-3.059***	-3.016***
	(0.448)	(0.447)	(0.447)	(0.458)	(0.457)	(0.457)
Income	0.01	0.005	0.007	0.139	-0.159	-0.147
	(0.026)	(0.026)	(0.026)	(0.163)	(0.217)	(0.217)
Poverty	-0.049*	0.086***	0.087***	-0.037	0.077**	0.079**
	(0.029)	(0.033)	(0.033)	(0.033)	(0.035)	(0.035)
High School	-0.116***	-0.058**	-0.055**	-0.055	-0.121	-0.115
	(0.022)	(0.023)	(0.023)	(0.082)	(0.086)	(0.086)
Less than High School	0.023	-0.05**	-0.049**	0.114	-0.137	-0.13
	(0.02)	(0.021)	(0.021)	(0.094)	(0.126)	(0.126)
Median Age	0.041	0.132**	0.133**	0.21***	0.152**	0.153**
	(0.032)	(0.056)	(0.056)	(0.062)	(0.061)	(0.061)
Youth	1.908	0.048	0.051	0.205*	-0.068	-0.056
	(3.61)	(0.072)	(0.072)	(0.119)	(0.167)	(0.167)
Housing Density	0.02***	-0.016**	-0.016**	0.025	-0.045	-0.043
	(0.005)	(0.007)	(0.007)	(0.017)	(0.033)	(0.033)
Unemployment	-0.2***	-0.153**	-0.149**	-0.15	-0.265*	-0.255*
	(0.068)	(0.069)	(0.069)	(0.14)	(0.145)	(0.146)
Single Mothers	0.041	-0.052	-0.056	0.023	-0.037	-0.042
	(0.032)	(0.037)	(0.037)	(0.06)	(0.047)	(0.047)
Constant	1.908	4.438	6.664*	-7.212	15.041	17.545
	(3.61)	(3.836)	(3.931)	(11.371)	(15.315)	(16.647)
First Stage F-Test C&N (dof= 5)				1004.94	1143.92	1140.66
First Stage F-Test H&E (dof= 5)				1186.22	1806.11	1784.93
First Stage F-Test Env (dof= 5)				77.29	72.69	70.25
First Stage F-Test Emer (dof= 5)				801.53	916	917.32
Sargan-Hansen Test (p-value)				0.672	0.734	0.77
Hausman Test (p-value)				0.101	0.075	0.083
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

Notes: The dependent variable is the probability of being arrested conditional on being stopped. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units. First Stage F-Test values are adjusted for heteroskedasticity. Hausman test was performed by basing both covariance matrices on the disturbance variance estimate from the efficient estimator. Sample is 64281 observations in 182 zip codes.

The estimated impact is very large; a one standard deviation increase in emergency preparedness volunteering would lead to a 43% decrease in all arrests. This result is not likely to apply to the entire city, but the effect might indeed be large for a few areas vulnerable to disasters. Regardless, we dismiss these results due to the issues with the sample.

In addition, the magnitude of some coefficients such as environment is very large, to a point of concern, and the IV model seems ill suited to estimate the relationship with precision. It is not evident that instruments meant to capture the strength of social capital in an area would also capture specific efforts to prepare for disasters or protect the environment.

The issues are similar for the estimation for summons; most categories are not statistically significant when controlling for boroughs and time effects and the estimates are very imprecise. We dismiss the results due to the large standard errors for OLS along with the counterintuitive result that environmental volunteering would have a very large positive impact on the probability of being issued a court summons when stopped. We are unable to draw any conclusion as to whether the effect of volunteering depends on the objective of the program.

Table 24: Determinants of Summons Issued, Volunteering in Categories

Model	OLS	OLS	OLS	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Community and Neighbors	-0.564	-0.575	-0.597	19.288***	7.674	8.503
	(0.505)	(0.509)	(0.512)	(5.913)	(9.148)	(9.15)
Health and Education	-2.457*	-2.527*	-2.449*	-52.465***	-29.14	-30.755*
	(1.452)	(1.46)	(1.468)	(13.318)	(18.424)	(18.46)
Environment	10.061**	10.851**	10.645**	-275.442**	-10.595	-30.711
	(4.271)	(4.331)	(4.352)	(131.801)	(184.366)	(184.629)
Emergency Preparedness	-0.54	-0.322	-0.388	-20.45*	-13.298	-13.985
	(1.739)	(1.742)	(1.747)	(11.345)	(11.589)	(11.566)
Age	0.118***	0.117***	0.116***	0.118***	0.117***	0.116***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Gender	-0.833**	-0.763*	-0.762*	-0.712*	-0.789*	-0.778*
	(0.395)	(0.395)	(0.395)	(0.414)	(0.403)	(0.404)
Income	-0.12***	-0.138***	-0.137***	0.152	-0.136	-0.114
	(0.022)	(0.023)	(0.023)	(0.158)	(0.203)	(0.204)
Poverty	0.105***	0.085***	0.085***	0.115***	0.094***	0.095***
	(0.028)	(0.032)	(0.032)	(0.034)	(0.035)	(0.035)
High School	-0.097***	-0.09***	-0.09***	0.047	-0.086	-0.077
	(0.022)	(0.024)	(0.024)	(0.08)	(0.081)	(0.081)
Less than High School	-0.156***	-0.152***	-0.153***	-0.026	-0.181	-0.169
	(0.02)	(0.022)	(0.022)	(0.091)	(0.118)	(0.118)
Median Age	0.035	-0.012	-0.016	-0.065	-0.063	-0.066
	(0.05)	(0.051)	(0.051)	(0.057)	(0.056)	(0.056)
Youth	-0.147**	-0.23***	-0.232***	0.07	-0.153	-0.139
	(0.06)	(0.063)	(0.063)	(0.115)	(0.152)	(0.152)
Housing Density	-0.027***	-0.03***	-0.03***	0.013	-0.022	-0.019
	(0.006)	(0.007)	(0.007)	(0.016)	(0.031)	(0.031)
Unemployment	-0.14**	-0.131*	-0.13*	0.099	-0.103	-0.088
	(0.067)	(0.068)	(0.068)	(0.139)	(0.139)	(0.139)
Single Mothers	0.124***	0.138***	0.139***	0.021	0.117***	0.116**
	(0.029)	(0.035)	(0.035)	(0.058)	(0.045)	(0.045)
Constant	12.129***	16.028***	14.086***	-3.62	16.773	13.48
	(3.392)	(3.535)	(3.571)	(11.17)	(14.412)	(15.704)
First Stage F-Test C&N (dof= 5)				1004.94	1143.92	1140.66
First Stage F-Test H&E (dof= 5)				1186.22	1806.11	1784.93
First Stage F-Test Env (dof= 5)				77.29	72.69	70.25
First Stage F-Test Emer (dof= 5)				801.53	916	917.32
Sargan-Hansen Test (p-value)				0.75	0.899	0.965
Hausman Test (p-value)				0.0000	0.0002	0.0002
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

Notes: The dependent variable is the probability of being issued a summons to appear in court conditional on being stopped. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units. First Stage F-Test values are adjusted for heteroskedasticity. Hausman test was performed by basing both covariance matrices on the disturbance variance estimate from the efficient estimator. Sample is 64281 observations in 182 zip codes.

Appendix 3 – Cultural Capital as a Regressor

We use the amount of cultural capital per thousand inhabitants as a regressor in lieu of volunteering to measure its own impact on crime. These results constitute the only empirical evidence on the topic and provide a way to test the exogeneity of cultural associations before using the variable as an instrument.

The results for arrests are presented in Table 25. For both OLS and IV, cultural associations are negatively linked to the probability of being arrested when stopped, and the estimates are similar for the two estimators. In fact, we cannot reject that they are the same according to a Hausman test, which shows that cultural associations are exogenous with regards to arrests. The OLS estimates predict that a one standard deviation increase in cultural associations would lead to a 0.37% decrease in the probability of being arrested when stopped. This constitutes an overall decrease of 6.32%.

For summons, as presented in Table 26, the estimated coefficient is also statistically significant and negative, though IV estimates are slightly larger. In fact, we can now reject the null hypothesis that both estimators are the same. Cultural associations appear endogenous with regards to summons, but this may be due once again to issues with the instruments. According to the IV estimates, a one standard deviation increase in cultural capital would lead to a 0.54% decrease in the probability of a suspect being issued a court summons - an overall decrease of 8.93%. It follows from the theory that cultural capital should have a negative though smaller impact on crime than social capital, since it is one of its inputs.

Table 25: Determinants of Arrests, Cultural Capital as a Regressor

Model	OLS	OLS	OLS	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Cultural Associations	-0.349***	-0.468***	-0.461**	-0.371***	-0.505***	-0.503***
	(0.13)	(0.251)	(0.130)	(0.139)	(0.14)	(0.14)
Age	0.047***	0.047***	0.047***	0.048***	0.047***	0.048***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Gender	-3.008***	-3.002***	-2.962***	-3.009***	-3.003***	-2.963***
	(0.448)	(0.447)	(0.447)	(0.448)	(0.447)	(0.447)
Income	0.003	-0.0001	0.001	0.004	0.001	0.002
	(0.024)	(0.025)	(0.0246)	(0.024)	(0.025)	(0.025)
Poverty	-0.044	0.092***	0.093***	-0.044	0.093***	0.094***
	(0.029)	(0.033)	(0.033)	(0.029)	(0.033)	(0.033)
High School	-0.128***	-0.067***	-0.065**	-0.128***	-0.067***	-0.065***
	(0.022)	(0.023)	(0.023)	(0.022)	(0.023)	(0.023)
Less than High School	0.016	-0.057***	-0.056**	0.016	-0.058**	-0.056***
	(0.022)	(0.021)	(0.021)	(0.02)	(0.021)	(0.021)
Median Age	0.185***	0.134**	0.136**	0.184***	0.134**	0.135***
	(0.054)	(0.056)	(0.056)	(0.054)	(0.056)	(0.056)
Youth	0.197***	0.041	0.043	0.198***	0.043	0.046
	(0.065)	(0.071)	(0.07)	(0.065)	(0.07)	(0.07)
Housing Density	0.018***	-0.019***	-0.019**	0.018***	-0.019***	-0.019***
	(0.005)	(0.007)	(0.007)	(0.005)	(0.007)	(0.007)
Unemployment	-0.202***	-0.151**	-0.149**	-0.202***	-0.151**	-0.148**
	(0.068)	(0.068)	(0.068)	(0.067)	(0.068)	(0.068)
Single Mothers	0.039	-0.061*	-0.064*	0.039	-0.062*	-0.065*
	(0.032)	(0.036)	(0.036)	(0.031)	(0.036)	(0.037)
Constant	2.347	5.125	7.478	2.326	5.11	7.479
	(3.527)	(3.750)	(3.854)	(3.528)	(3.75)	(3.851)
Observations	64281	64281	64281	64281	64281	64281
Number of Zip Codes	182	182	182	182	182	182
First Stage F-Test (dof = 5)				4515.67	4462.30	4468.66
Sargan-Hansen Test (p-value)				0.1709	0.6719	0.72
Hausman Test (p-value)				0.6932	0.4652	0.4144
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

Notes: The dependent variable is the probability of being arrested conditional on being stopped. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units. First Stage F-Test values are adjusted for heteroskedasticity. Hausman test was performed by basing both covariance matrices on the disturbance variance estimate from the efficient estimator. All five instruments were used to calculate the IV estimates.

Table 26: Determinants of Summons, Cultural Capital as a Regressor

Model	OLS	OLS	OLS	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Cultural Capital	-0.463***	-0.488***	-0.5***	-0.605***	-0.657***	-0.67***
	(0.88)	(0.09)	(0.09)	(0.1)	(0.101)	(0.102)
Age	0.118***	0.118***	0.117***	0.118***	0.118***	0.117***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Gender	-0.839**	-0.758*	-0.766*	-0.841**	-0.767*	-0.768*
	(0.395)	(0.395)	(0.395)	(0.395)	(0.395)	(0.395)
Income	-0.118***	-0.135***	-0.134***	-0.115***	-0.132***	-0.132***
	(0.02)	(0.021)	(0.022)	(0.02)	(0.021)	(0.021)
Poverty	0.108***	0.089***	0.088***	0.11***	0.092***	0.092***
	(0.028)	(0.032)	(0.032)	(0.028)	(0.032)	(0.033)
High School	-0.104***	-0.097***	-0.096***	-0.106***	-0.098***	-0.097***
	(0.022)	(0.024)	(0.024)	(0.022)	(0.024)	(0.024)
Less than High School	-0.156***	-0.152***	-0.153***	-0.157***	-0.155***	-0.156***
	(0.02)	(0.022)	(0.022)	(0.02)	(0.022)	(0.022)
Median Age	0.034	-0.015	-0.019	0.034	-0.017	-0.021
	(0.058)	(0.051)	(0.051)	(0.05)	(0.051)	(0.051)
Youth	-0.163***	-0.247***	-0.249***	-0.152	-0.237***	-0.239***
	(0.06)	(0.061)	(0.061)	(0.058)	(0.061)	(0.061)
Housing Density	-0.028***	-0.031***	-0.031***	-0.027***	-0.031***	-0.031***
	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)
Unemployment	-0.149**	-0.138**	-0.137*	-0.147*	-0.136	-0.134
	(0.066)	(0.067)	(0.067)	(0.066)	(0.067)	(0.067)
Single Mothers	0.126***	0.138***	0.138***	0.122***	0.131***	0.131***
	(0.028)	(0.035)	(0.035)	(0.028)	(0.035)	(0.035)
Constant	12.567***	16.547***	14.631***	12.433***	16.481***	14.633***
	(3.307)	(3.506)	(3.488)	(3.311)	(3.452)	(3.486)
Observations	64281	64281	64281	64281	64281	64281
Number of Zip Codes	182	182	182	182	182	182
First Stage F-Test (dof =)				4515.67	4462.30	4468.66
Sargan-Hansen Test (p-value)				0.0022	0.1214	0.1328
Hausman Test (p-value)				0.0015	0.0001	0.0001
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

Notes: The dependent variable is the probability of being issued a summons to appear in court conditional on being stopped. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units. First Stage F-Test values are adjusted for heteroskedasticity. Hausman test was performed by basing both covariance matrices on the disturbance variance estimate from the efficient estimator.

Appendix 4 – IV Using All Six Instruments

Table 27: Determinants of Arrests and of Summons, All Six Instruments

	Arrests			Summons		
Model	IV+	IV+	IV+	IV+	IV+	IV+
	(1)	(2)	(3)	(4)	(5)	(6)
Volunteering	-0.867***	-1.186***	-1.176***	-1.366***	-1.445***	-1.474***
	(0.326)	(0.325)	(0.325)	(0.239)	(0.231)	(0.232)
Age	0.048***	0.047***	0.048***	0.118***	0.118***	0.117***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Gender	-2.995***	-2.987***	-2.947***	-0.818**	-0.749*	-0.747*
	(0.448)	(0.447)	(0.447)	(0.395)	(0.395)	(0.395)
Income	0.021	0.026	0.027	-0.086***	-0.1***	-0.099***
	(0.026)	(0.026)	(0.026)	(0.022)	(0.022)	(0.022)
Poverty	-0.042	0.094***	0.095***	0.112***	0.093***	0.093***
	(0.029)	(0.033)	(0.033)	(0.028)	(0.032)	(0.032)
High School	-0.114***	-0.05**	-0.048**	-0.083***	-0.076***	-0.075***
	(0.022)	(0.023)	(0.023)	(0.022)	(0.024)	(0.024)
Less than High School	0.025	-0.043**	-0.042**	-0.141***	-0.136***	-0.137***
	(0.02)	(0.021)	(0.021)	(0.02)	(0.021)	(0.021)
Median Age	0.188***	0.143***	0.145***	0.041	-0.005	-0.009
	(0.054)	(0.056)	(0.056)	(0.05)	(0.051)	(0.051)
Youth	0.229***	0.092	0.095	-0.099	-0.174***	-0.175***
	(0.068)	(0.073)	(0.073)	(0.061)	(0.063)	(0.063)
Housing Density	0.02***	-0.016**	-0.016**	-0.025***	-0.027***	-0.027***
	(0.005)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)
Unemployment	-0.184***	-0.128*	-0.126*	-0.116*	-0.106	-0.104
	(0.068)	(0.068)	(0.068)	(0.066)	(0.067)	(0.067)
Single Mothers	0.03	-0.072*	-0.075**	0.106***	0.119***	0.119***
	(0.032)	(0.037)	(0.037)	(0.029)	(0.035)	(0.035)
Constant	0.536	2.332	4.652	9.399***	12.9***	10.9***
	(3.634)	(3.848)	(3.936)	(3.378)	(3.517)	(3.544)
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

The dependent variable is the probability of being arrested or issued a court summons conditional on being stopped. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units. Estimates obtained using cultural capital plus all other instruments except vehicle noise.

Appendix 5 – Nonlinearity in Volunteering

Table 28: Determinants of Arrest and Summons with Nonlinearity in Volunteering, OLS

	Arrests			Summons		
Model	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)
Volunteering	-0.612 (0.58)	-0.617 (0.368)	-0.572 (0.37)	-1.843*** (0.371)	-1.864*** (0.74)	-1.874*** (0.745)
Volunteering Squared	0.026 (0.082)	0.017 (0.051)	0.01 (0.051)	0.226*** (0.051)	0.232*** (0.11)	0.231*** (0.11)
Age	0.047*** (0.009)	0.047*** (0.009)	0.047*** (0.009)	0.118*** (0.022)	0.118*** (0.009)	-0.135*** (0.009)
Gender	-3*** (0.448)	-2.994*** (0.395)	-2.953*** (0.395)	-0.838** (0.024)	-0.77* (0.449)	-0.097*** (0.447)
Income	0.009 (0.025)	0.006 (0.021)	0.008 (0.022)	-0.118*** (0.064)	-0.135*** (0.026)	-0.216*** (0.026)
Poverty	-0.046 (0.029)	0.087*** (0.028)	0.088*** (0.032)	0.103*** (0.007)	0.079** (0.033)	-0.029*** (0.033)
High School	-0.119*** (0.022)	-0.059** (0.023)	-0.056** (0.024)	-0.103*** (0.067)	-0.097*** (0.024)	-0.147** (0.024)
Less than High School	0.022 (0.02)	-0.048** (0.02)	-0.046** (0.022)	-0.155*** (0.395)	-0.149*** (0.021)	-0.768* (0.021)
Median Age	0.188*** (0.054)	0.141** (0.05)	0.143** (0.051)	0.042 (0.051)	-0.005 (0.057)	-0.009 (0.057)
Youth	0.205*** (0.067)	0.05 (0.061)	0.052 (0.064)	-0.133** (0.032)	-0.214*** (0.074)	0.078** (0.074)
Housing Density	0.019*** (0.005)	-0.017** (0.006)	-0.01** (0.007)	-0.026*** (0.035)	-0.029*** (0.007)	0.145*** (0.007)
Unemployment	-0.196*** (0.068)	-0.146** (0.066)	-0.143** (0.067)	-0.155** (0.022)	-0.148** (0.069)	-0.15*** (0.069)
Single Mothers	0.039 (0.032)	-0.055 (0.029)	-0.059 (0.035)	0.128*** (0.009)	0.144*** (0.037)	0.117*** (0.037)
Constant	1.551 (3.588)	4.027 (3.364)	6.257 (3.508)	11.874*** (3.539)	15.752*** (3.847)	13.789*** (3.937)
Observations	64281	64281	64281	64281	64281	64281
Number of Zip Codes	182	182	182	182	182	182
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

The dependent variables are the probability of being arrested and the probability of being issued a summons conditional on being stopped. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged between the ages of 15 and 24 for each zip code. Housing Density refers to the proportion of housing structures with 20 or more units.

Table 29: Determinants of Arrest and Summons with Nonlinearity in Volunteering, IV

	Arrests			Summons		
Model	IV 1	IV 2	IV 3	IV 4	IV 5	IV 6
Volunteering	-1.692** (0.088)	-1.627** (0.74)	-1.572** (0.745)	-1.729*** (0.597)	-2.13*** (0.59)	-2.195*** (0.59)
Volunteering Squared	-0.084 (0.109)	0.074 (0.11)	0.066 (0.11)	0.068 (0.089)	0.132 (0.087)	0.128 (0.87)
Age	0.047*** (0.009)	0.048*** (0.009)	0.048*** (0.009)	0.118*** (0.009)	0.119*** (0.009)	0.117*** (0.009)
Gender	-2.99*** (0.448)	-3.018*** (0.449)	-2.95*** (0.447)	-0.822** (0.395)	-0.757* (0.4)	-0.76* (0.395)
Income	0.028 (0.026)	0.024 (0.026)	0.025 (0.026)	-0.091*** (0.021)	-0.108*** (0.022)	-0.106*** (0.022)
Poverty	-0.041 (0.029)	0.093*** (0.033)	0.094*** (0.033)	0.111*** (0.028)	0.089*** (0.032)	0.089*** (0.032)
High School	-0.115*** (0.023)	-0.053** (0.024)	-0.051** (0.024)	-0.087*** (0.023)	-0.083*** (0.024)	-0.082*** (0.024)
Less than High School	0.024 (0.02)	-0.046** (0.021)	-0.044** (0.021)	-0.144*** (0.02)	-0.141*** (0.021)	-0.142*** (0.022)
Median Age	0.191*** (0.054)	0.148** (0.057)	0.146*** (0.057)	0.042 (0.05)	-0.004 (0.051)	-0.008 (0.051)
Youth	0.22*** (0.125)	0.101 (0.074)	0.104 (0.074)	-0.098** (0.63)	-0.17*** (0.065)	-0.167** (0.065)
Housing Density	0.263*** (0.006)	-0.016** (0.007)	-0.016** (0.007)	-0.025*** (0.006)	-0.027*** (0.007)	-0.027*** (0.007)
Unemployment	-0.181** (0.068)	-0.133* (0.069)	-0.13* (0.069)	-0.123* (0.068)	-0.119* (0.068)	-0.116* (0.098)
Single Mothers	0.022 (0.031)	-0.072* (0.037)	-0.075** (0.037)	0.11*** (0.031)	0.123*** (0.035)	0.122*** (0.068)
Constant	-0.196 (3.625)	2.403 (3.847)	4.747 (3.937)	9.737*** (3.371)	13.486*** (3.505)	11.385*** (3.525)
Observations	64281	64281	64281	64281	64281	64281
Number of Zip Codes	182	182	182	182	182	182
First Stage F-Test (dof= 5)	3599.91	3793.93	3773.68	3899.97	4084.12	4459.69
First Stage F-Test Sq (dof= 5)	1001.43	1034.88	1028.71	1078.49	1116.08	1194.02
Sargan-Hansen Test (p-value)	0.0008	0.6395	0.6554	0	0	0
Hausman Test (p-value)	0.0046	0.01	0.0104	0	0	0
Borough FE	no	yes	yes	no	yes	yes
Month FE	no	yes	yes	no	yes	yes
Borough X Month	no	no	yes	no	no	yes

The dependent variables are the probability of being arrested and being issued a summons when stopped. Robust standard errors are in parentheses.* denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Youth refers to the proportion of inhabitants aged 15-24. Housing Density refers to the proportion of housing structures with 20 or more units. First Stage F-Test values are adjusted for heteroskedasticity. Hausman test was performed by basing both covariance matrices on the disturbance variance estimate from the efficient estimator. Vehicle noise is excluded from the estimation.

Appendix 6 – Discrimination for All Crimes, Year and Borough

Table 30: Arrest Made, by Year and Borough

Model			OLS					OLS					FE					FE		
			(1)					(2)					(3)					(4)		
	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.
Black																				
2003	-0.683	0.337	-2.986***	1.163**	1.130*	-0.683	0.337	-2.986**	1.163	1.130	0.616	-0.239	-0.380	3.044***	-1.000	0.616	-0.239	-0.380	3.044***	-1.000
	(0.629)	(0.889)	(0.337)	(0.454)	(0.609)	(0.994)	(1.098)	(1.416)	(1.624)	(1.569)	(0.659)	(0.978)	(0.432)	(0.532)	(0.794)	(0.729)	(0.691)	(0.606)	(0.887)	(2.197)
2004	0.357	-1.378**	-1.826***	1.015***	-0.845	0.357	-1.378	-1.826*	1.015	-0.845	0.396	-0.798	0.541**	1.935***	-0.253	0.396	-0.798	0.541	1.935***	-0.253
	(0.395)	(0.692)	(0.195)	(0.304)	(0.553)	(1.000)	(1.546)	(1.040)	(0.888)	(2.327)	(0.423)	(0.742)	(0.244)	(0.368)	(0.686)	(0.905)	(1.397)	(0.566)	(0.554)	(3.014)
2005	0.221	0.071	-0.822***	0.187	1.366**	0.221***	0.071***	-0.822	0.187***	1.366*	1.087***	-0.598	1.573***	1.264***	0.986	1.087*	-0.598	1.573***	1.264**	0.986
	(0.312)	(0.623)	(0.145)	(0.270)	(0.581)	(0.837)	(1.227)	(0.473)	(1.036)	(0.821)	(0.334)	(0.661)	(0.193)	(0.333)	(0.734)	(0.657)	(1.152)	(0.421)	(0.600)	(0.490)
2006	-1.837***	0.328	-1.096***	-0.067	3.295***	-1.837	0.328***	-1.096	-0.067***	3.295	-0.657**	-0.194	0.789***	0.212	2.083***	-0.657	-0.194	0.789***	0.212	2.083*
	(0.235)	(0.440)	(0.124)	(0.218)	(0.369)	(0.785)	(0.906)	(0.535)	(1.029)	(0.844)	(0.250)	(0.467)	(0.163)	(0.263)	(0.461)	(0.678)	(0.483)	(0.240)	(0.706)	(1.205)
2007	-1.328***	1.144**	-1.110***	-0.200	2.495***	-1.328**	1.144	-1.110**	-0.200	2.495***	-0.694**	0.225	0.843***	-0.199	1.823***	-0.694	0.225	0.843*	-0.199	1.823***
	(0.284)	(0.465)	(0.168)	(0.247)	(0.347)	(1.034)	(0.795)	(0.572)	(0.921)	(0.074)	(0.302)	(0.500)	(0.227)	(0.298)	(0.412)	(0.801)	(0.676)	(0.480)	(0.818)	(0.223)
2008	0.656**	0.397	-0.791***	0.972***	2.635***	0.656	0.397	-0.791*	0.972	2.635***	0.362	0.263	0.788***	0.669**	0.870**	0.362	0.263	0.788*	0.669	0.870**
	(0.291)	(0.493)	(0.168)	(0.241)	(0.267)	(0.945)	(0.770)	(0.892)	(0.509)	(0.542)	(0.314)	(0.527)	(0.224)	(0.290)	(0.350)	(0.710)	(0.451)	(0.405)	(0.548)	(0.437)
2009	0.354	-0.978**	-1.370***	0.655***	0.485***	0.354	-0.978	-1.370	0.655*	0.485***	0.835**	-0.913**	0.594***	-0.413	-0.122	0.835	-0.913	0.594	-0.413	-0.122
	(0.316)	(0.435)	(0.169)	(0.235)	(0.266)	(0.892)	(1.203)	(1.282)	(0.857)	(0.824)	(0.338)	(0.463)	(0.222)	(0.290)	(0.339)	(0.897)	(1.133)	(0.460)	(0.778)	(0.637)
2010	0.055	-0.363**	-1.183***	0.791***	1.677*	0.055	-0.363	-1.183	0.791	1.677	0.632*	-0.889*	0.646***	0.340	0.463	0.632	-0.889	0.646	0.340	0.463
	(0.302)	(0.488)	(0.177)	(0.239)	(0.312)	(0.839)	(1.184)	(0.775)	(0.527)	(1.301)	(0.323)	(0.514)	(0.224)	(0.288)	(0.398)	(0.694)	(0.706)	(0.425)	(0.278)	(1.121)
2011	-1.205	-1.155	-1.215***	-0.578***	2.208***	-1.205	-1.155	-1.215**	-0.578	2.208***	-0.164	-1.118***	-0.169	0.101	0.244	-0.164	-1.118**	-0.169	0.101	0.244
	(0.259)	(0.389)	(0.158)	(0.235)	(0.334)	(1.161)	(0.853)	(0.577)	(0.699)	(0.622)	(0.277)	(0.411)	(0.195)	(0.276)	(0.403)	(0.688)	(0.569)	(0.388)	(0.380)	(0.231)
2012	-1.520***	-1.048**	-1.274***	-0.778***	1.627***	-1.520	-1.048	-1.274	-0.778	1.627	0.281	-0.674	0.060	-0.264	-0.160	0.281	-0.674	0.060	-0.264	-0.160
	(0.290)	(0.439)	(0.151)	(0.283)	(0.367)	(1.631)	(0.730)	0.658*	(1.139)	0.695**	(0.307)	(0.462)	(0.191)	(0.333)	(0.480)	(1.118)	(0.862)	(0.495)	(0.918)	(0.686)
H0 : ui = 0											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2003											0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000	0.1371
2004											0.0000	0.0000	0.0000	0.0000	0.0015	0.0000	0.0009	0.1107	0.2928	0.2918
2005											0.0000	0.0000	0.0000	0.0000	0.0000	0.4887	0.0000	0.0000	0.0033	0.1732
2006											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1405	0.0773
2007											0.0000	0.0000	0.0000	0.0000	0.0061	0.0000	0.0273	0.0098	0.5527	0.0142
2008											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0255
2009											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0403	0.8066
2010											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5691
2011											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0038
2012											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0614	0.0000	0.8993	0.0371
Clustered SE	no	no	no	no	no	yes	yes	yes	yes	yes	no	no	no	no	no	yes	yes	yes	yes	yes
Time FE	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Precinct FE	no	no	no	no	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero.

Table 31: Summons Issued, by Year and Borough

Model			OLS					OLS					FE					FE		
			(1)					(2)					(3)					(4)		
	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.
Black																				
2003	-1.182***	0.492	0.415	-1.808***	-4.970***	-1.182	0.492	0.415	-1.808*	-4.970***	-2.026***	0.567	-0.719	-2.034***	-2.841***	-2.026**	0.567	-0.719	-2.034**	-2.841***
	(0.422)	(0.759)	(0.347)	(0.327)	(0.530)	(0.749)	(1.067)	(0.771)	(0.943)	(0.882)	(0.445)	(0.840)	(0.446)	(0.385)	(0.690)	(0.767)	(0.772)	(0.824)	(0.814)	(0.868)
2004	-0.488	-2.756***	1.091***	-0.853***	-1.706***	-0.488	-2.756	1.091	-0.853	-1.706**	-1.391***	-1.778**	-2.212***	-1.065***	-0.593	-1.391***	-1.778	-2.212**	-1.065**	-0.593***
	(0.356)	(0.670)	(0.286)	(0.250)	(0.415)	(0.593)	(1.861)	(1.064)	(0.806)	(0.773)	(0.381)	(0.720)	(0.357)	(0.303)	(0.514)	(0.508)	(1.628)	(1.017)	(0.403)	(0.162)
2005	0.754**	-1.461**	1.416***	0.193	-4.990***	0.754	-1.461	1.416	0.193	-4.990**	-1.878***	-1.130	-1.729***	-0.540*	-2.338***	-1.878***	-1.130	-1.729*	-0.540	-2.338***
	(0.303)	(0.649)	(0.225)	(0.267)	(0.601)	(0.933)	(1.237)	(1.399)	(1.086)	(2.169)	(0.320)	(0.689)	(0.298)	(0.328)	(0.757)	(0.674)	(0.946)	(0.943)	(1.002)	(0.849)
2006	-0.297	0.315	-0.836***	0.975***	-1.394***	-0.297	0.315	-0.836	0.975	-1.394	-1.855***	-1.456**	-1.066***	0.969***	-0.808**	-1.855***	-1.456**	-1.066**	0.969	-0.808
	(0.239)	(0.538)	(0.177)	(0.235)	(0.322)	(1.218)	(1.418)	(1.070)	(1.381)	(1.792)	(0.253)	(0.569)	(0.233)	(0.284)	(0.400)	(0.475)	(0.741)	(0.414)	(1.980)	(1.541)
2007	-0.065	1.094*	0.816***	0.510**	-4.203***	-0.065	1.094	0.816	0.510	-4.203	-2.259***	-0.783	-1.258***	0.390	-2.402***	-2.259***	-0.783	-1.258*	0.390	-2.402
	(0.243)	(0.568)	(0.226)	(0.258)	(0.370)	(1.594)	(1.440)	(1.027)	(1.011)	(3.747)	(0.255)	(0.605)	(0.306)	(0.310)	(0.434)	(0.469)	(0.548)	(0.680)	(1.349)	(3.529)
2008	0.040	-0.122	0.785***	-1.028***	-2.134***	0.040	-0.122	0.785	-1.028**	-2.134***	-2.332***	-1.025	-1.954***	-1.427***	-1.540***	-2.332***	-1.025	-1.954*	-1.427***	-1.540**
	(0.233)	(0.541)	(0.215)	(0.210)	(0.243)	(0.956)	(1.242)	(1.242)	(0.457)	(0.644)	(0.250)	(0.576)	(0.286)	(0.253)	(0.318)	(0.550)	(1.094)	(1.008)	(0.523)	(0.719)
2009	-0.596**	-3.070***	0.939***	0.250	-2.482***	-0.596	-3.070***	0.939	0.250	-2.482***	-2.616***	-2.700***	-2.930***	-0.913***	-2.249***	-2.616***	-2.700***	-2.930**	-0.913**	-2.249***
	(0.243)	(0.441)	(0.213)	(0.203)	(0.275)	(0.890)	(0.576)	(1.149)	(1.275)	(0.580)	(0.259)	(0.471)	(0.279)	(0.250)	(0.350)	(0.613)	(0.488)	(1.417)	(0.383)	(0.333)
2010	-0.768***	-3.180***	-0.020	0.699***	-2.819***	-0.768	-3.180**	-0.020	0.699	-2.819***	-3.854***	-2.993***	-3.053***	-0.755***	-2.866***	-3.854***	-2.993***	-3.053***	-0.755	-2.866***
	(0.255)	(0.455)	(0.215)	(0.223)	(0.317)	(1.207)	(1.298)	(0.977)	(1.169)	(0.607)	(0.270)	(0.480)	(0.272)	(0.268)	(0.405)	(0.961)	(0.670)	(0.998)	(0.516)	(0.636)
2011	0.354	-3.228***	-1.081***	-1.034***	1.899***	0.354	-3.228***	-1.081	-1.034	1.899***	-2.337***	-2.406***	-2.900***	-1.451***	-0.478	-2.337***	-2.406***	-2.900***	-1.451**	-0.478***
	(0.218)	(0.364)	(0.184)	(0.199)	(0.360)	(1.638)	(0.789)	(0.825)	(0.651)	(0.510)	(0.230)	(0.385)	(0.227)	(0.235)	(0.435)	(0.442)	(0.436)	(0.563)	(0.614)	(0.144)
2012	0.191	-0.111	-1.037***	0.511**	3.533***	0.191	-0.111	-1.037*	0.511	3.533**	-1.843	-0.343	-1.823	-0.788	-0.871	-1.843***	-0.343	-1.823***	-0.788**	-0.871
	(0.226)	(0.423)	(0.159)	(0.232)	(0.413)	(1.083)	(0.416)	(0.536)	(0.734)	(1.514)	0.239***	(0.445)	0.201***	0.273***	(0.539)	(0.338)	(0.523)	(0.431)	(0.389)	(1.379)
H0 : ui = 0											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1910	0.9212	0.1866	0.0232
2003											0.0000	0.0000	0.0000	0.0000	0.0000	0.0255	0.0026	0.0000	0.0000	0.0010
2004											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1932	0.0000	0.0005	0.0067
2005											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2370
2006											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6111
2007											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.8179
2008											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.7760	0.0000	0.0000
2009											0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0002	0.0008	0.0201
2010											0.0000	0.0000	0.0000	0.0000	0.0014	0.0000	0.0000	0.0000	0.0000	0.0882
2011											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0058	0.0004
2012											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1479	0.0533	0.0217
Clustered SE	no	no	no	no	no	yes	yes	yes	yes	yes	no	no	no	no	no	yes	yes	yes	yes	yes
Time FE	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Precinct FE	no	no	no	no	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

The dependent variable is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero.

Appendix 7 – Discrimination including Hispanics, All Crimes

Table 32: Arrest Made, Overall and by Borough

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black and Black-Hispanic	-0.188***	-0.187***	-0.187	0.356***	0.339***	0.339***
	(0.023)	(0.023)	(0.310)	(0.026)	(0.026)	(0.102)
Constant	5.983***					
	(0.018)					
P-value of H0 : ui = 0				0.000	0.000	0.000
Black and Black-Hispanic						
Manhattan	0.031	0.01	0.01	0.397***	0.360***	0.360***
	(0.058)	(0.058)	(0.447)	(0.060)	(0.060)	(0.267)
Bronx	0.118**	0.130**	0.13	-0.03	-0.022	-0.022
	(0.059)	(0.059)	(0.171)	(0.060)	(0.060)	(0.130)
Brooklyn	-0.802***	-0.826***	-0.826**	0.262***	0.230***	0.23
	(0.035)	(0.035)	(0.325)	(0.042)	(0.041)	(0.152)
Queen's	1.023***	1.023***	1.023***	0.856***	0.854***	0.854***
	(0.050)	(0.050)	(0.371)	(0.061)	(0.061)	(0.148)
Staten Island	1.298***	1.330***	1.33	0.359***	0.358***	0.358
	(0.102)	(0.102)	(0.832)	(0.115)	(0.115)	(0.532)
Constant						
Manhattan	7.930***					
	(0.046)					
Bronx	6.509***					
	(0.046)					
Brooklyn	4.691***					
	(0.030)					
Queens	5.774***					
	(0.036)					
Staten Island	4.734***					
	(0.067)					
P-value of H0 : ui = 0						
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0000	0.0000	0.0000
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero. White includes whites and White-Hispanics.

Table 33: Arrest Made, by Year

Model	OLS	OLS	FE	FE
	(1)	(2)	(3)	(4)
Black and Black-Hispanic				
2003	-0.515***	-0.515	0.724***	0.724**
	(0.149)	(0.505)	(0.168)	(0.310)
2004	-0.424***	-0.424	0.968***	0.968***
	(0.089)	(0.393)	(0.100)	(0.244)
2005	-0.127*	-0.127	0.916***	0.916***
	(0.075)	(0.341)	(0.085)	(0.181)
2006	-0.404***	-0.404	0.345***	0.345**
	(0.061)	(0.337)	(0.069)	(0.144)
2007	-0.333***	-0.333	0.313***	0.313*
	(0.073)	(0.397)	(0.083)	(0.174)
2008	0.284***	0.284	0.268***	0.268*
	(0.069)	(0.437)	(0.079)	(0.136)
2009	-0.102	-0.102	0.152*	0.152
	(0.067)	(0.447)	(0.077)	(0.180)
2010	0.157**	0.157	0.224**	0.224
	(0.070)	(0.399)	(0.080)	(0.149)
2011	-0.121**	-0.121	0.114*	0.114
	(0.062)	(0.305)	(0.069)	(0.120)
2012	-0.723***	-0.723*	-0.035	-0.035
	(0.070)	(0.399)	(0.079)	(0.220)
P-value of H0 : ui = 0				
2003			0.0000	0.0000
2004			0.0000	0.0000
2005			0.0000	0.0000
2006			0.0000	0.0000
2007			0.0000	0.0000
2008			0.0000	0.0000
2009			0.0000	0.0000
2010			0.0000	0.0000
2011			0.0000	0.0000
2012			0.0000	0.0000
Clustered SE	no	yes	no	yes
Time FE	no	no	no	no
Precinct FE	no	no	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero. White includes whites and White-Hispanics.

Table 34: Summons Issued, Overall and by Borough

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black and Black-Hispanic	-0.196***	-0.185***	-0.185	-1.311***	-1.299***	-1.299***
	(0.024)	(0.024)	(0.286)	(0.027)	(0.027)	(0.156)
Constant	6.423***					
	(0.019)					
P-value of H0 : ui = 0				0.0000	0.0000	0.0000
Black and Black-Hispanic						
Manhattan	-0.568***	-0.573***	-0.573	-1.493***	-1.491***	-1.491***
	(0.051)	(0.051)	(0.872)	(0.053)	(0.053)	(0.246)
Bronx	-0.827***	-0.792***	-0.792***	-0.938***	-0.892***	-0.892***
	(0.062)	(0.062)	(0.224)	(0.063)	(0.063)	(0.169)
Brooklyn	0.108**	0.106**	0.106	-1.434***	-1.423***	-1.423***
	(0.044)	(0.044)	(0.522)	(0.052)	(0.052)	(0.261)
Queen's	-0.149***	-0.154***	-0.154	-1.128***	-1.132***	-1.132**
	(0.047)	(0.047)	(0.566)	(0.058)	(0.058)	(0.530)
Staten Island	-1.747***	-1.761***	-1.761***	-2.046***	-2.046***	-2.046***
	(0.103)	(0.103)	(0.212)	(0.116)	(0.116)	(0.105)
Constant						
Manhattan	6.355***					
	(0.041)					
Bronx	7.801***					
	(0.049)					
Brooklyn	6.511***					
	(0.037)					
Queens	5.594***					
	(0.034)					
Staten Island	6.197***					
	(0.067)					
P-value of H0 : ui = 0						
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0000	0.0000	0.1239
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero. White includes whites and White-Hispanics.

Table 35: Summons Issued, by Year

Model	OLS (1)	OLS (2)	FE (3)	FE (4)
Black and Black-Hispanic				
2003	-0.491***	-0.491	-0.929***	-0.929***
	(0.125)	(0.381)	(0.142)	(0.260)
2004	0.172*	0.172	-0.818***	-0.818***
	(0.096)	(0.374)	(0.108)	(0.189)
2005	0.522***	0.522	-0.942***	-0.942***
	(0.089)	(0.561)	(0.100)	(0.207)
2006	-0.164**	-0.164	-0.718***	-0.718**
	(0.071)	(0.423)	(0.081)	(0.317)
2007	0.236***	0.236	-1.103***	-1.103***
	(0.081)	(0.459)	(0.092)	(0.286)
2008	0.085	0.085	-1.334***	-1.334***
	(0.071)	(0.400)	(0.082)	(0.165)
2009	-0.160**	-0.16	-1.810***	-1.810***
	(0.068)	(0.428)	(0.079)	(0.207)
2010	-0.599***	-0.599	-2.012***	-2.012***
	(0.071)	(0.445)	(0.081)	(0.343)
2011	-0.542***	-0.542	-1.302***	-1.302***
	(0.061)	(0.399)	(0.069)	(0.162)
2012	-0.578***	-0.578*	-1.181***	-1.181***
	(0.065)	(0.322)	(0.073)	(0.187)
P-value of H0 : ui = 0				
2003			0.0000	0.0000
2004			0.0000	0.0000
2005			0.0000	0.0000
2006			0.0000	0.0655
2007			0.0000	0.0000
2008			0.0000	0.0000
2009			0.0000	0.0000
2010			0.0000	0.0000
2011			0.0000	0.0000
2012			0.0000	0.0000
Clustered SE	no	yes	no	yes
Time FE	no	no	no	no
Precinct FE	no	no	yes	yes

The dependent variable is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero. White includes whites and White-Hispanics.

Appendix 8 – Gender Discrimination, All Crimes

Table 37: Arrest Made, Overall and by Borough, by Gender

Model	OLS (1)	OLS (2)	OLS (3)	FE (4)	FE (5)	FE (6)
Male	-3.078***	-3.099***	-3.099***	-2.853***	-2.874***	-2.874***
	(0.376)	(0.376)	(0.491)	(0.376)	(0.375)	(0.479)
Constant	8.952***					
	(0.364)					
P-value of H0 : ui = 0				0.0000	0.0000	0.0000
Male						
Manhattan	-3.815***	-3.747***	-3.747**	-3.475***	-3.410***	-3.410**
	(1.157)	(1.158)	(1.702)	(1.153)	(1.155)	(1.571)
Bronx	-2.999**	-2.919**	-2.919	-2.905**	-2.822**	-2.822
	(1.288)	(1.285)	(2.160)	(1.285)	(1.282)	(2.141)
Brooklyn	-2.022***	-2.051***	-2.051**	-1.896***	-1.918***	-1.918**
	(0.658)	(0.658)	(0.892)	(0.656)	(0.656)	(0.880)
Queens	-2.223*	-2.327*	-2.327	-2.119**	-2.243*	-2.243**
	(1.321)	(1.322)	(1.037)	(1.321)	(1.321)	(1.126)
Staten Island	-5.500***	-5.306***	-5.306***	-5.469***	-5.277***	-5.277***
	(1.697)	(1.703)	(1.076)	(1.696)	(1.703)	(1.122)
Constant						
Manhattan	12.405***					
	(1.107)					
Bronx	9.391***					
	(1.248)					
Brooklyn	6.385***					
	(0.635)					
Queens	9.114***					
	(1.285)					
Staten Island	10.748***					
	(1.601)					
P-value of H0 : ui = 0						
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0140	0.0208	0.0004
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis. Sample restricted to 2011.

Table 38: Summons Issued, Overall and by Borough, by Gender

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.867**	-0.847**	-0.847*	-0.669*	-0.648*	-0.648
	(0.371)	(0.370)	(0.484)	(0.370)	(0.369)	(0.503)
Constant	6.679***					
	(0.358)					
P-value of H0 : ui = 0				0.0000	0.0000	0.0000
Male						
Manhattan	-0.224	-0.197	-0.197	0.087	0.094	0.094
	(0.942)	(0.941)	(0.867)	(0.925)	(0.924)	(0.905)
Bronx	-2.492**	-2.539**	-2.539	-2.277*	-2.329*	-2.329
	(1.213)	(1.214)	(1.828)	(1.214)	(1.214)	(1.800)
Brooklyn	-2.528***	-2.475***	-2.475**	-2.304***	-2.254***	-2.254**
	(0.759)	(0.759)	(1.047)	(0.756)	(0.756)	(1.069)
Queens	0.338	0.338	0.338	0.119	0.121	0.121
	(1.115)	(1.115)	(0.915)	(1.117)	(1.117)	(0.976)
Staten Island	-2.310	-2.388	-2.388	-2.270	-2.344	-2.344
	(1.884)	(1.888)	(1.879)	(1.882)	(1.887)	(1.840)
Constant						
Manhattan	5.900***					
	(0.901)					
Bronx	8.122***					
	(1.176)					
Brooklyn	8.451***					
	(0.733)					
Queens	4.557***					
	(1.084)					
Staten Island	9.346***					
	(1.777)					
P-value of H0 : ui = 0						
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0044	0.0044	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0022	0.0019	0.0000
Staten Island				0.0098	0.0135	0.0000
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis. Sample restricted to 2011.

Appendix 9 – Discrimination against Black Males

Table 39: Arrest Made, Overall and by Borough

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black Male	-0.323	-0.333	-0.333	-0.484	-0.481	-0.481
	(1.176)	(1.176)	(1.400)	(1.172)	(1.172)	(1.352)
Constant	9.318***					
	(0.987)					
P-value of H0 : ui = 0						
Black Male						
Manhattan	1.841	1.856	1.856	1.485	1.528	1.528
	2.773	2.774	3.440	2.762	2.764	3.311
Bronx	-0.248	-0.584	-0.584	0.467	0.123	0.123
	4.068	4.060	3.203	4.063	4.056	3.102
Brooklyn	-0.843	-0.843	-0.843	-1.204	-1.197	-1.197
	1.820	1.820	2.555	1.824	1.823	2.636
Queen's	-4.329	-4.410	-4.410	-3.911	-3.923	-3.923
	3.000	3.004	2.838	2.993	2.996	2.981
Staten Island	-2.385	-2.535	-2.535*	-2.529	-2.661	-2.661*
	3.419	3.427	1.422	3.421	3.430	1.400
Constant						
Manhattan	14.103***					
	2.280					
Bronx	8.333**					
	3.577					
Brooklyn	7.101***					
	1.594					
Queens	6.542***					
	2.468					
Staten Island	9.091***					
	2.129					
P-value of H0 : ui = 0						
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0477	0.0677	0.0028
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero. Sample restricted to 2011. Individual indicator variables for black and male are also included in the estimations.

Table 40: Summons Issued, Overall and by Borough

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black Male	-1.116	-1.061	-1.061	-1.056	-1.003	-1.003
	(1.135)	(1.135)	(1.256)	(1.130)	(1.130)	(1.254)
Constant	6.988***					
	(0.953)					
P-value of H0 : ui = 0						
Black Male						
Manhattan	-0.124	-0.017	-0.017	0.627	0.745	0.745
	2.257	2.254	2.562	2.214	2.211	2.603
Bronx	-1.865	-1.877	-1.877	-2.038	-2.007	-2.007
	3.831	3.834	5.114	3.836	3.839	5.308
Brooklyn	-2.987	-2.866	-2.866	-3.506*	-3.412	-3.412
	2.101	2.100	2.564	2.099	2.098	2.484
Queen's	0.185	0.154	0.154	-0.096	-0.152	-0.152
	2.531	2.534	2.075	2.530	2.533	2.089
Staten Island	2.371	2.475	2.475	2.157	2.269	2.269
	3.798	3.802	2.494	3.795	3.800	2.054
Constant						
Manhattan	5.128***					
	1.856					
Bronx	8.333**					
	3.368					
Brooklyn	7.101***					
	1.839					
Queens	5.607***					
	2.082					
Staten Island	9.917***					
	2.364					
P-value of H0 : ui = 0						
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0068	0.0071	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0011	0.0009	0.0000
Staten Island				0.0054	0.0068	0.0061
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero. Sample restricted to 2011. Individual indicator variables for black and male are also included in the estimations.

Appendix 10 – Oaxaca Decompositions

From the results of the decomposition by race, shown in Table 41, whites are 0.76% more likely to be arrested when stopped. This may not constitute evidence of discrimination given that only the explained part of the gap is significant. It predicts that, if whites had the same predictors as blacks, they would be 1.21% more likely to be arrested. Most of the difference is driven by unemployment, age and gender. Around 10% of whites in the sample are women but only 6% for blacks. The average age of white suspects is 30.08 while it is 27.79 for blacks. This is equivalent to saying that, if whites were less likely to be employed, younger and more predominantly male, they would be more likely to be arrested.

For gender, one striking element is the vast difference in the number of men and women stopped by the NYPD. Approximately 93% of all stops targeted male suspects in 2011. Another striking difference, shown in Table 43, is the discrepancy in the probability of being arrested conditional on being stopped, with female suspects being over three percent more likely to be arrested. This difference could come from a starker contrast between female criminals and non-criminals. In other words, an officer who usually focuses on men - given their higher rate of criminal participation – may need strong evidence to divert his attention to women. If this is true, then women who are stopped would be more likely to be arrested. The decomposition of the difference in arrests seems to point towards this possibility, given that it is mostly unexplained by predictors. Specifically, a male suspect with the same covariates as the average female suspect would be 0.1% more likely to be arrested. On the other hand, if a male suspect with the same characteristics were instead a female, he would be 2.85% more likely to be arrested. The reasoning also holds for summons, though crimes which lead to being issued a court summons are usually less grave. The “discrimination” element is only significant at the 10% level and there are no large differences between IV and OLS for the “Overall” category.

Given the importance of precinct FE, as shown by the hit rates test, we also present the decompositions with both precinct FE and clustering at precinct level. The results are presented in Table 44. The difference between blacks and whites is no longer significant for arrests, while it remains significant for summons. This is in exact concordance with the hit rates test results obtained in the larger sample. Also in accordance with past results, the gap between men and women does remain significant.

Table 41: Oaxaca Decomposition, by Race

	Arrests		Summons	
	OLS	IV	OLS	IV
Overall				
White (N = 5671)	6.912***	6.912***	6.419***	6.419***
	(0.338)	(0.338)	(0.326)	(0.326)
Black (N = 33305)	6.149***	6.149***	5.726***	5.726***
	(0.132)	(0.132)	(0.127)	(0.127)
Difference	0.763**	0.763**	0.693***	0.693***
	(0.202)	(0.362)	(0.350)	(0.350)
Endowments	1.212***	1.212***	-0.362***	-0.362***
	(0.199)	(0.199)	(0.195)	(0.195)
Coefficients	-0.048	-0.035*	1.646***	1.676***
	(0.229)	(0.552)	(0.644)	(0.643)
Interaction	-0.402	-0.413	-0.591	-0.62
	(0.49)	(0.488)	(0.549)	(0.549)
Endowments				
Volunteering	-0.033**	-0.056***	-0.024**	-0.047***
	(0.015)	(0.022)	(0.009)	(0.017)
Age	0.15***	0.151***	0.329***	0.33
	(0.029)	(0.351)	(0.038)	(0.038)
Revenue	-0.139	-0.031	-3.05***	-2.94***
	(0.346)	(0.093)	(0.323)	(0.322)
Education	0.363	0.306	1.68***	1.627
	(0.227)	(0.228)	(0.209)	(0.208)
Age Distribution	0.354	0.260	0.569***	0.476***
	(0.227)	(0.235)	(0.202)	(0.206)
Housing Density	0.034	0.028	0.301***	0.296
	(0.046)	(0.046)	(0.052)	(0.052)
Unemployment	0.307**	0.276*	-0.295**	-0.326
	(0.155)	(0.155)	(0.149)	(0.15)
Single Mothers	0.018	0.127	0.051	0.16
	(0.386)	(0.388)	(0.354)	(0.358)
Gender	0.117***	0.117***	0.066***	0.065**
	(0.029)	(0.029)	(0.025)	(0.025)
Coefficients				
Volunteering	-0.004	0.015	0.061	-0.42
	(0.067)	(0.1)	(0.062)	(0.66)
Age	-1.009	-1.02	-3.913***	-3.915***
	(0.845)	(0.844)	(0.807)	(0.805)
Revenue	-6.431	-6.55	-8.825**	-8.179*
	(4.51)	(4.53)	(4.454)	(4.504)
Education	1.444	1.35	4.548	3.593
	(2.993)	(2.99)	(2.004)	(3.069)
Age Distribution	1.723	2.183	-2.053	-0.14
	(9.199)	(9.34)	(8.222)	(8.164)
Housing Density	1.303	1.341	3.187***	3.35***
	(1.22)	(1.221)	(1.245)	(1.242)
Unemployment	0.477	0.575	-3.801*	-2.384
	(2.413)	(2.422)	(2.296)	(2.3)
Single Mothers	-3.299	-3.458	6.321***	5.645***
	(2.077)	(2.122)	(2.012)	(2.028)
Constant	4.864	4.625	8.231	5.279
	(12.72)	(12.909)	(12.091)	(12.033)
Gender	0.071	0.079	1.06*	1.09
	(1.305)	(1.301)	(1.179)	(1.177)

Notes: The dependent variable Arrests is the probability of being arrested and the dependent variable Summons is the probability of being issued a summons, both conditional on being stopped. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Revenue includes income and poverty, Education includes high school completion rates and proportion with less than high school, Age Distribution regroups median age and proportion aged 15 to 24. Borough and month dummies are included in all estimations. Detailed coefficients for interaction are omitted since they provide no relevant information. Sample restricted to 2011.

Table 42: Oaxaca Decomposition, by Race, with Hispanics

	Arrests		Summons	
	OLS	IV	OLS	IV
Overall				
White and White-Hispanic (N = 22524)	6.118***	6.118***	6.22***	6.22***
	(0.16)	(0.16)	(0.161)	(0.161)
Black and Black-Hispanic (N = 37873)	6.155***	6.155***	5.695***	5.695***
	(0.124)	(0.124)	(0.119)	(0.119)
Difference	-0.037	-0.037	0.525***	0.525***
	(0.202)	(0.202)	(0.2)	(0.2)
Endowments	0.626***	0.629***	-0.653***	-0.65***
	(0.109)	(0.109)	(0.106)	(0.106)
Coefficients	-0.431*	-0.434*	1.103***	1.098***
	(0.229)	(0.229)	(0.257)	(0.256)
Interaction	-0.233	-0.232	0.075	0.076
	(0.162)	(0.162)	(0.18)	(0.18)
Endowments				
Volunteering	0.004	0.007	0.003	0.007
	(0.003)	(0.005)	(0.003)	(0.005)
Age	0.035***	0.035***	0.079***	0.079***
	(0.009)	(0.009)	(0.015)	(0.015)
Revenue	-0.12	-0.107	-0.7***	-0.685***
	(0.093)	(0.093)	(0.093)	(0.093)
Education	0.082	0.073	-0.157***	-0.167***
	(0.054)	(0.054)	(0.059)	(0.059)
Age Distribution	0.111	0.072	0.218***	0.176***
	(0.074)	(0.077)	(0.063)	(0.065)
Housing Density	0.002	0.001	0.009	0.009
	(0.003)	(0.002)	(0.013)	(0.013)
Unemployment	0.129**	0.115*	-0.048	-0.062
	(0.061)	(0.061)	(0.058)	(0.058)
Single Mothers	0.073	0.121	-0.09	-0.037
	(0.152)	(0.153)	(0.139)	(0.141)
Gender	0.037***	0.037***	0.016**	0.016**
	(0.01)	(0.01)	(0.007)	(0.007)
Coefficients				
Volunteering	0.041	-0.372	0.037	-4.319
	(0.052)	(8.238)	(0.037)	(7.514)
Age	-0.946*	-0.943*	-2.068***	-2.064***
	(0.512)	(0.512)	(0.521)	(0.521)
Revenue	-0.529	-0.452	-6.345**	-6.161**
	(2.732)	(2.75)	(2.524)	(2.562)
Education	-0.349	-0.461	4.548**	4.46**
	(2.012)	(2.014)	(2.004)	(2.008)
Age Distribution	1.029	1.529	3.32	4.023
	(5.701)	(5.744)	(5.144)	(5.145)
Housing Density	-1.119	-1.071	3.162***	3.229***
	(0.709)	(0.71)	(0.707)	(0.706)
Unemployment	0.588	0.601	-2.56**	-2.527**
	(1.038)	(1.039)	(1.031)	(1.034)
Single Mothers	-1.745	-1.916	4.861***	4.622***
	(1.167)	(1.182)	(1.148)	(1.161)
Constant	0.025	-0.004	-3.607	-0.026
	(8.174)	(0.07)	(7.488)	(0.049)
Gender	0.186	0.189	1.282*	1.286*
	(0.886)	(0.886)	(0.776)	(0.776)

Notes: The dependent variable Arrests is the probability of being arrested conditional on being frisked and the dependent variable Summons is the probability of being issued a summons conditional on being frisked. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Revenue includes income and poverty in the zip area, Education includes high school completion rates and proportion with less than high school, Age Distribution regroups median age and proportion aged 15 to 24, Borough regroups the effects of the five boroughs of New York City. Month regroups the effect of the 12 different months.

Table 43: Oaxaca Decomposition, by Gender

	Arrests		Summons	
	OLS	IV	OLS	IV
Overall				
Women (N = 4249)	8.99***	8.99***	6.707***	6.707***
	(0.44)	(0.439)	(0.385)	(0.384)
Men (N = 60032)	5.882***	5.882***	5.832***	5.832***
	(0.096)	(0.096)	(0.096)	(0.096)
Difference	3.108***	3.108***	0.876**	0.876**
	(0.451)	(0.449)	(0.397)	(0.396)
Endowments	0.1**	0.107***	0.119***	0.129***
	(0.04)	(0.041)	(0.04)	(0.041)
Coefficients	2.853***	2.851***	0.766*	0.763*
	(0.451)	(0.449)	(0.401)	(0.399)
Interaction	0.155	0.151	-0.009	-0.017
	(0.124)	(0.124)	(0.1)	(0.1)
Endowments				
Volunteering	-0.007	-0.017*	-0.008*	0.029
	(0.005)	(0.009)	(0.005)	(0.023)
Age	0.011	0.011	0.029	-0.196***
	(0.009)	(0.009)	(0.022)	(0.048)
Revenue	-0.021	0.007	-0.241***	0.237***
	(0.042)	(0.043)	(0.051)	(0.047)
Education	0.096**	0.086**	0.253***	0.022*
	(0.039)	(0.038)	(0.048)	(0.012)
Age Distribution	0.019	0.014	0.03**	-0.017
	(0.012)	(0.012)	(0.014)	(0.012)
Housing Density	-0.012	-0.011	-0.019	0.014
	(0.009)	(0.009)	(0.013)	(0.01)
Unemployment	0.014	0.012	0.018*	-0.037**
	(0.01)	(0.01)	(0.01)	(0.018)
Single Mothers	0.019	0.024	-0.045**	0.078***
	(0.014)	(0.015)	(0.021)	(0.024)
Borough	0.002	0.001	0.08***	0.023*
	(0.032)	(0.032)	(0.024)	(0.013)
Month	-0.021	-0.021	0.023*	-0.024**
	(0.013)	(0.013)	(0.013)	(0.012)
Coefficients				
Volunteering	-0.133	1.506	-0.02	0.04
	(0.111)	(1.084)	(0.043)	(0.985)
Age	1.498	-2.359	0.032	1.607
	(1.088)	(5.382)	(0.988)	(4.581)
Revenue	-2.071	-2.535	2.149	2.111
	(5.345)	(4.321)	(4.592)	(3.588)
Education	-2.433	-10.299	2.321	-4.934
	(4.3)	(11.937)	(3.607)	(9.463)
Age Distribution	-10.113	1.78	-4.508	1.011
	(11.951)	(1.443)	(9.516)	(1.352)
Housing Density	1.782	-4.476**	1.031	-0.365
	(1.441)	(2.148)	(1.357)	(1.918)
Unemployment	-4.423**	1.404	-0.266	0.071
	(2.152)	(2.371)	(1.925)	(2.084)
Single Mothers	1.376	-2.409	-0.017	-1.268
	(2.358)	(2.347)	(2.078)	(2.157)
Borough	-2.407	2.437*	-1.249	-0.651
	(2.352)	(1.348)	(2.162)	(1.321)
Month	2.434*	-0.134	-0.656	-0.008
	(1.351)	(0.151)	(1.326)	(0.055)
Constant	17.341	17.937	1.948	3.15
	(16.441)	(16.501)	(14.004)	(13.899)

Notes: The dependent variable Arrests is the probability of being arrested conditional on being frisked and the dependent variable Summons is the probability of being issued a summons conditional on being frisked. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Revenue includes income and poverty in the zip area, Education includes high school completion rates and proportion with less than high school, Age Distribution regroups median age and proportion aged 15 to 24. Borough and month dummies were included in all estimations. Detailed coefficients for interaction are omitted since they provide no relevant information. Sample restricted to 2011.

Table 44: Overall Oaxaca Decompositions, Precinct FE and Clustering

	Arrests		Summons	
	OLS	IV	OLS	IV
Black				
White	6.912***	6.912***	6.419***	6.419***
	(0.601)	(0.686)	(0.447)	(0.550)
Black	6.149***	6.149***	5.726***	5.726***
	(0.411)	(0.432)	(0.511)	(0.526)
Difference	0.763	0.763	0.693	0.693
	(0.625)	(0.810)	(0.555)	(0.761)
Endowments	0.55	0.551	-1.004*	-1.005
	(0.626)	(0.871)	(0.544)	(0.772)
Coefficients	-0.95	-0.952	2.664***	2.663**
	(0.949)	(1.275)	(0.713)	(1.098)
Interaction	1.164	1.165	-0.967	-0.965
	0.913	1.411	0.673	1.186
Black with Hispanic				
White and White-Hispanic	6.118***	6.118***	6.22***	6.22***
	(0.353)	(0.387)	(0.357)	(0.391)
Black and Black-Hispanic	6.155***	6.155***	5.695***	5.695***
	(0.386)	(0.405)	(0.465)	(0.479)
Difference	-0.037	-0.037	0.525	0.525
	(0.379)	(0.560)	(0.448)	(0.619)
Endowments	0.325	0.326	-0.569	-0.57
	(0.349)	(0.596)	(0.402)	(0.581)
Coefficients	-0.499	-0.504	1.7***	1.698***
	(0.352)	(0.465)	(0.386)	(0.491)
Interaction	0.138	0.141	-0.606*	-0.604
	(0.304)	(0.550)	(0.316)	(0.548)
Gender				
Female	8.99***	8.99***	6.708***	6.708***
	(0.663)	(0.789)	(0.587)	(0.698)
Male	5.882***	5.882***	5.832***	5.832***
	(0.310)	(0.324)	(0.356)	(0.369)
Difference	3.109***	3.109***	0.876*	0.876
	(0.494)	(0.853)	(0.494)	(0.789)
Endowments	0.175*	0.179	0.273**	0.273
	(0.095)	(0.462)	(0.134)	(0.572)
Coefficients	2.627***	2.623***	0.939*	0.936
	(0.485)	(0.671)	(0.515)	(0.658)
Interaction	0.307*	0.307	-0.336*	-0.333
	(0.172)	(0.700)	(0.172)	(0.733)

Notes: The dependent variable Arrests is the probability of being arrested conditional on being frisked and the dependent variable Summons is the probability of being issued a summons conditional on being frisked. Robust standard errors are in parentheses. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. Borough and month dummies were included in all estimations. Sample restricted to 2011.

Appendix 11 – Discrimination for Drug Crimes Only (Weapon Crimes Omitted)

Table 45: Arrest Made, Overall and by Borough, Drug Crimes

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-1.465***	-1.293***	-1.293	-1.7***	-1.66***	-1.66***
	(0.161)	(0.161)	(1.029)	(0.198)	(0.197)	(0.558)
Constant	11.72***					
	(0.148)					
P-value of H0 : ui = 0						
Black						
Manhattan	-4.051***	-3.990***	-3.990	-3.370***	-3.225***	-3.225**
	0.344	0.343	2.610	0.377	0.375	1.405
Bronx	0.436	0.343	0.343	-1.582**	-1.762**	-1.762
	0.665	0.662	1.607	0.724	0.720	1.608
Brooklyn	-1.636***	-1.443***	-1.443*	0.105	0.090	0.090
	0.258	0.258	0.783	0.314	0.313	0.566
Queens	1.454***	1.307***	1.307	-1.407**	-1.469**	-1.469*
	0.453	0.454	2.132	0.576	0.575	0.843
Staten Island	1.641***	1.897***	1.897	-2.350***	-2.178***	-2.178**
	0.416	0.419	1.671	0.529	0.528	0.825
Constant						
Manhattan	15.059***					
	0.317					
Bronx	11.496***					
	0.635					
Brooklyn	9.639***					
	0.244					
Queens	14.181***					
	0.398					
Staten Island	8.789***					
	0.305					
P-value of H0 : ui = 0				0.0000	0.0000	0.0000
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0000	0.0000	0.003
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero.

Table 46: Arrest Made, by Year, Drug Crimes

Model	OLS	OLS	FE	FE
	(1)	(2)	(3)	(4)
Black				
2004	-1.332	-1.332	-1.212	-1.212
	1.871	2.319	1.921	1.610
2005	-2.413**	-2.413	-0.904	-0.904
	0.921	2.017	1.073	1.607
2006	-2.229***	-2.229*	-1.190***	-1.190
	0.344	1.205	0.411	1.122
2007	-2.291***	-2.291*	-1.530***	-1.530
	0.394	1.172	0.485	0.960
2008	0.583	0.583	0.048	0.048
	0.426	1.140	0.543	0.857
2009	-1.233***	-1.233	-1.537***	-1.537*
	0.443	1.206	0.551	0.898
2010	-0.473	-0.473	-1.626***	-1.626
	0.476	1.693	0.582	1.003
2011	-1.834***	-1.834	-2.288***	-2.288***
	0.461	1.721	0.550	0.856
2012	-1.411***	-1.411	-2.169***	-2.169*
	0.489	2.011	0.586	1.149
P-value of H0 : ui = 0				
2004			0.0000	0.4673
2005			0.0000	0.0000
2006			0.0000	0.0000
2007			0.0000	0.0000
2008			0.0000	0.0000
2009			0.0000	0.0000
2010			0.0000	0.5354
2011			0.0000	0.0000
2012			0.0000	0.3137
Clustered SE	no	yes	no	yes
Time FE	no	no	no	no
Precinct FE	no	no	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero.

Table 47: Summons Issued, Overall and by Borough, Drug Crimes

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-0.605***	-0.593***	-0.594	-1.401***	-1.406***	-1.406**
	(0.141)	(0.141)	(0.775)	(0.174)	(0.174)	(0.563)
Constant	8.279***					
	(0.129)					
P-value of H0 : ui = 0						
Black						
Manhattan	-0.058	-0.085	-0.085	-1.891***	-1.871***	-1.871**
	0.290	0.289	1.071	0.320	0.319	0.902
Bronx	-0.427	-0.453	-0.453	0.476	0.440	0.440
	0.590	0.590	0.757	0.644	0.644	0.885
Brooklyn	-0.347	-0.369	-0.369	-1.487***	-1.514***	-1.514
	0.251	0.251	1.182	0.305	0.304	1.186
Queens	-1.026***	-1.023***	-1.023	-1.075**	-1.000**	-1.000
	0.327	0.328	1.398	0.421	0.421	0.716
Staten Island	-3.687***	-3.341***	-3.341	-1.668***	-1.533***	-1.533
	0.356	0.359	2.597	0.453	0.452	2.611
Constant						
Manhattan	7.949***					
	0.267					
Bronx	9.458***					
	0.563					
Brooklyn	7.983***					
	0.237					
Queens	8.056***					
	0.287					
Staten Island	8.907***					
	0.261					
P-value of H0 : ui = 0				0.0000	0.0000	0.0000
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0000	0.0000	0.0004
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero.

Table 48: Summons Issued, by Year, Drug Crimes

Model	OLS	OLS	FE	FE
	(1)	(2)	(3)	(4)
Black				
2004	-1.667	-1.667	-1.697	-1.697
	2.423	4.802	2.480	4.977
2005	0.347	0.347	-1.924	-1.924
	1.313	1.993	1.532	2.045
2006	-1.737***	-1.737	-0.787**	-0.787
	0.332	1.634	0.401	0.690
2007	-1.892***	-1.892	-1.059**	-1.059
	0.375	2.193	0.463	0.962
2008	0.478	0.478	-1.499***	-1.499*
	0.385	1.106	0.491	0.871
2009	0.344	0.344	0.512**	-1.450**
	0.409	0.821	0.000	0.621
2010	-1.362***	-1.362*	-3.301***	-3.301***
	0.408	0.793	0.504	1.102
2011	-0.582*	-0.582	-1.161***	-1.161
	0.351	0.648	0.426	0.774
2012	0.814**	0.814	-0.363	-0.363
	0.366	0.549	0.451	0.642
P-value of H0 : ui = 0			0.0000	0.0000
2004			0.0000	0.7395
2005			0.0000	0.3916
2006			0.0000	0.0000
2007			0.0000	0.0000
2008			0.0000	0.0000
2009			0.0000	0.0268
2010			0.0000	0.8851
2011			0.0000	0.0004
2012			0.0000	0.0000
Clustered SE	no	yes	no	yes
Time FE	no	no	no	no
Precinct FE	no	no	yes	yes

The dependent variable is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero.

Appendix 12 – Discrimination, War on Drugs, Year and Borough

Table 49: Arrest Made, by Year and Borough, War on Drugs Related Crimes

Model			OLS					OLS					FE					FE		
			(1)					(2)					(3)					(4)		
	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.
Black																				
2006	-3.814***	0.144	-1.423***	0.202	1.479*	-3.814**	0.144	-1.423*	0.202	1.479***	-2.388***	-0.248	0.470	-1.896***	0.273	-2.388	-0.248	0.470	-1.896	0.273
	(0.441)	(0.640)	(0.292)	(0.564)	(0.826)	(1.721)	(0.846)	(0.776)	(1.934)	(0.578)	(0.466)	(0.671)	(0.339)	(0.668)	(1.005)	(1.733)	(0.575)	(0.716)	(1.817)	(1.368)
2007	-5.102***	-0.609	-1.762***	-0.761	2.591**	-5.102**	-0.609	-1.762***	-0.761	***	-4.059***	-0.916	-0.474	-1.396**	-0.321	-4.059**	-0.916	-0.474	-1.396	-0.321*
	(0.520)	(0.673)	(0.351)	(0.578)	(0.998)	(2.126)	(1.229)	(0.594)	(1.336)	(0.457)	(0.556)	(0.722)	(0.422)	(0.677)	(1.205)	(1.952)	(0.886)	(0.514)	(1.187)	(0.185)
2008	-2.006***	-0.280	-0.441	-0.219	1.302*	-2.006	-0.280	-0.441	-0.219	1.302***	-1.709***	-0.392	0.255	-0.950	-0.986	-1.709	-0.392	0.255	-0.950	-0.986
	(0.538)	(0.651)	(0.389)	(0.573)	(0.726)	(1.753)	(1.265)	(1.061)	(1.163)	(0.410)	(0.599)	(0.688)	(0.462)	(0.694)	(0.910)	(1.423)	(1.035)	(0.724)	(1.214)	(0.632)
2009	-1.999***	-1.194**	-3.385***	-1.802***	-2.084***	-1.999	-1.194	-3.385***	-1.802	-2.084*	-1.173*	-1.197**	-0.951**	-4.556***	-1.977***	-1.173	-1.197	-0.951	-4.556**	-1.977***
	(0.552)	(0.563)	(0.382)	(0.562)	(0.598)	(1.867)	(1.109)	(1.311)	(1.868)	(1.267)	(0.615)	(0.604)	(0.443)	(0.710)	(0.724)	(1.813)	(1.206)	(0.866)	(2.297)	(0.727)
2010	-1.537***	-1.508**	-3.346***	-2.456***	-0.673	-1.537	-1.508	-3.346***	-2.456	-0.673	-2.327***	-2.260***	-1.468***	-2.248***	-3.104***	-2.327*	-2.260*	-1.468*	-2.248**	-3.104*
	(0.542)	(0.690)	(0.409)	(0.604)	(0.670)	(1.866)	(1.661)	(0.990)	(1.773)	(2.299)	(0.609)	(0.722)	(0.474)	(0.755)	(0.811)	(1.314)	(1.333)	(0.804)	(1.056)	(1.767)
2011	-4.268***	-0.996*	-2.395***	-3.041***	-1.731**	-4.268	-0.996	-2.395***	-3.041**	-1.731	-2.916***	-1.368**	-1.749***	-2.145***	-3.720***	-2.916*	-1.368**	-1.749***	-2.145**	-3.720***
	(0.463)	(0.576)	(0.377)	(0.603)	(0.747)	(2.599)	(0.725)	(0.650)	(1.381)	(1.337)	(0.522)	(0.597)	(0.439)	(0.705)	(0.862)	(1.637)	(0.663)	(0.521)	(0.864)	(0.304)
2012	-3.960***	-1.741***	-3.432***	-3.023***	0.496	-3.960	-1.741	-3.432***	-3.023	0.496	-3.949***	-1.967***	-3.320***	-2.031**	-2.449**	-3.949	-1.967	-3.320***	-2.031*	-2.449***
	(0.538)	(0.623)	(0.386)	(0.736)	(0.839)	(3.474)	(1.093)	(0.968)	(1.866)	(1.051)	(0.605)	(0.657)	(0.458)	(0.860)	(1.013)	(2.491)	(1.279)	(1.155)	(1.072)	(0.253)
H0 : ui = 0																				
2006											0.0000	0.0000	0.0000	0.0000	0.0961	0.0000	0.0000	0.0000	0.0000	0.1075
2007											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0642
2008											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0145
2009											0.0000	0.0000	0.0000	0.0000	0.0006	0.0000	0.0000	0.0000	0.0000	0.0140
2010											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0011
2011											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0121
2012											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0011
Clustered SE	no	no	no	no	no	yes	yes	yes	yes	yes	no	no	no	no	no	yes	yes	yes	yes	yes
Time FE	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Precinct FE	no	no	no	no	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero. Indicator variables for each different crime type are included.

Table 50: Summons Issued, by Year and Borough, War on Drugs Related Crimes

Model			OLS					OLS					FE					FE		
			(1)					(2)					(3)					(4)		
	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.	Ma.	Bronx	Brook.	Queens	St. Is.
Black																				
2006	1.619***	-1.524	-2.701***	2.185***	-4.299***	1.619	-1.524	-2.701	2.185	-4.299	-0.523	-3.231***	-1.535***	2.666***	-1.941**	-0.523	-3.231*	-1.535**	2.666	-1.941
	0.605	0.976	0.432	0.650	0.723	1.338	2.229	1.992	2.313	3.255	0.638	1.016	0.501	0.746	0.870	1.004	1.691	0.677	3.448	3.151
2007	1.452**	0.217	-2.211***	-0.752	-7.960***	1.452	0.217	-2.211	-0.752	-7.960	-2.121***	-2.053*	-2.540***	1.511**	-4.725***	-2.121**	-2.053	-2.540**	1.511	-4.725
	0.601	1.024	0.509	0.656	0.907	2.255	1.845	1.704	1.331	5.949	0.629	1.085	0.612	0.754	1.069	0.861	1.413	1.076	2.431	5.001
2008	0.936	-0.756	-2.783***	-3.259***	-4.396***	0.936	-0.756	-2.783	-3.259**	-4.396**	-2.052***	-0.743	-4.164***	-3.449***	-3.310***	-2.052**	-0.743	-4.164**	-3.449***	-3.310
	0.590	0.921	0.537	0.541	0.583	1.240	1.571	1.945	1.585	1.751	0.657	0.969	0.635	0.654	0.731	0.989	1.644	2.037	1.131	2.058
2009	-0.728	-3.608***	-3.147***	-0.392	-3.719***	-0.728	-3.608**	-3.147	-0.392	-3.719***	-2.503***	-3.037***	-6.511***	-1.340**	-3.206***	-2.503***	-3.037**	-6.511*	-1.340	-3.206***
	0.574	0.750	0.528	0.528	0.564	1.145	1.512	2.979	2.205	0.686	0.638	0.803	0.608	0.669	0.683	0.721	1.269	3.406	1.164	0.361
2010	-0.441	-6.181	-3.790	0.737	-2.944***	-0.441	-6.181**	-3.790**	0.737	-2.944***	-5.054***	-5.728***	-5.886***	-1.158	-2.246***	-5.054***	-5.728**	-5.886	-1.158	-2.246***
	0.601	0.798	0.511	0.611	0.607	1.933	2.652	1.782	1.935	0.740	0.670	0.835	0.589	0.764	0.736	1.674	2.019	2.098	0.965	0.552
2011	0.448	-4.865***	-3.480***	-2.021***	2.495***	0.448	-4.865***	-2.031**	-2.021***	2.495**	-2.641***	-3.480***	-2.871***	-1.917***	1.603*	-2.641***	-3.480***	-2.871***	-1.917	1.603***
	0.505	0.646	0.814	0.534	0.711	2.201	1.369	0.841	0.669	1.091	0.559	0.669	0.486	0.632	0.821	0.801	0.814	0.899	0.736	0.464
2012	0.501	-1.586**	-1.455***	-0.664	-1.188	0.501	-1.586	-1.455*	-0.664	-1.188	-1.919***	-1.715**	-2.350***	-1.037	-3.676***	-1.919**	-1.715	-2.350***	-1.037	-3.676***
	0.530	0.716	0.417	0.598	0.887	1.626	1.198	0.810	0.665	2.261	0.598	0.756	0.494	0.709	1.072	0.845	1.340	0.851	0.792	1.190
H0 : ui = 0																				
2006											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002
2007											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
2008											0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0272
2009											0.0000	0.0000	0.0000	0.0000	0.0885	0.0000	0.0000	0.0000	0.0000	0.0875
2010											0.0000	0.0000	0.0000	0.0000	0.0699	0.0000	0.0000	0.0000	0.0000	0.0191
2011											0.0000	0.0000	0.0000	0.0000	0.0556	0.0000	0.0000	0.0000	0.0000	0.0000
2012											0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0009
Clustered SE	no	no	no	no	no	yes	yes	yes	yes	yes	no	no	no	no	no	yes	yes	yes	yes	yes
Time FE	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Precinct FE	no	no	no	no	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

The dependent variable is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precincts fixed effects equal to zero. Indicator variables for each different crime type are included.

Appendix 13– Discrimination, Other Crimes

Table 51: Arrest Made, Overall and by Borough, Selected Other Crimes

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-0.393***	-0.477***	-0.477	0.536***	0.483***	0.483*
	(0.076)	(0.076)	(0.505)	(0.091)	(0.091)	(0.267)
Constant	5.628***					
	(0.730)					
P-value of H0 : ui = 0						
Black						
Manhattan	-0.212	-0.197	-0.197	1.114***	1.108***	1.108**
	(0.184)	(0.184)	(0.665)	(0.191)	(0.190)	(0.479)
Bronx	0.440	0.294	0.294	0.401	0.178	0.178
	(0.358)	(0.358)	(0.867)	(0.377)	(0.377)	(0.688)
Brooklyn	-1.638***	-1.801***	-1.801***	0.486***	0.411***	0.411
	(0.110)	(0.110)	(0.668)	(0.143)	(0.143)	(0.400)
Queens	0.398**	0.331**	0.331	0.180	0.080	0.080
	(0.167)	(0.167)	(0.648)	(0.197)	(0.197)	(0.524)
Staten Island	1.605***	1.643***	1.643***	0.904***	0.909***	0.909*
	(0.202)	(0.203)	(0.410)	(0.235)	(0.235)	(0.506)
Constant						
Manhattan	6.469***					
	(1.801)					
Bronx	9.785***					
	(2.285)					
Brooklyn	4.572***					
	(0.979)					
Queens	10.547***					
	(2.029)					
Staten Island	32.722***					
	(4.433)					
P-value of H0 : ui = 0						
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0000	0.0000	0.0118
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis. All estimations include dummy variables for each crime type in the sample.

Table 52: Arrest Made, by Year, Selected Other Crimes

Model	OLS	OLS	FE	FE
	(1)	(2)	(3)	(4)
Black				
2004	-3.813***	-3.813*	-1.723	-1.723
	(1.075)	(2.124)	(1.172)	(2.349)
2005	-0.236	-0.236	1.366**	1.366
	(0.467)	(0.625)	(0.622)	(0.958)
2006	-1.142***	-1.142**	0.137	0.137
	(0.175)	(0.464)	(0.210)	(0.376)
2007	-0.925***	-0.925	0.035	0.035
	(0.211)	(0.593)	(0.252)	(0.439)
2008	0.011	0.011	0.29	0.29
	(0.205)	(0.702)	(0.247)	(0.360)
2009	-0.231	-0.231	0.394	0.394
	(0.210)	(0.665)	(0.250)	(0.488)
2010	0.286	0.286	0.660***	0.660*
	(0.212)	(0.625)	(0.252)	(0.386)
2011	-0.476***	-0.476	0.499**	0.499
	(0.200)	(0.554)	(0.235)	(0.315)
2012	-0.544***	-0.544	1.206***	1.206***
	(0.221)	(0.634)	(0.259)	(0.432)
P-value of H0 : ui = 0				
2004			0.0000	0.0000
2005			0.0000	0.0000
2006			0.0000	0.0000
2007			0.0000	0.0000
2008			0.0000	0.0000
2009			0.0000	0.0000
2010			0.0000	0.0000
2011			0.0000	0.0000
2012			0.0000	0.0000
Clustered SE	no	yes	no	yes
Time FE	no	no	no	no
Precinct FE	no	no	yes	yes

The dependent variable is the probability of being arrested conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0 : ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis. All estimations include dummy variables for each crime type in the sample.

Table 53: Summons Issued, Overall and by Borough, Selected Other Crimes

Model	OLS	OLS	OLS	FE	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-0.463***	-0.455***	-0.455***	-0.557***	-0.552***	-0.552***
	-(0.047)	-(0.047)	-(0.098)	-(0.057)	-(0.057)	-(0.106)
Constant	3.015***					
	-(0.451)					
P-value of H0 : ui = 0						
Black						
Manhattan	-0.655***	-0.649***	-0.649***	-0.665***	-0.665***	-0.665***
	(0.089)	(0.089)	(0.208)	(0.092)	(0.092)	(0.135)
Bronx	-0.664***	-0.614***	-0.614	-0.490**	-0.425*	-0.425*
	(0.237)	(0.237)	(0.399)	(0.250)	(0.250)	(0.243)
Brooklyn	-0.457***	-0.458***	-0.458**	-0.537***	-0.562***	-0.562***
	(0.090)	(0.090)	(0.218)	(0.117)	(0.117)	(0.193)
Queens	-0.401***	-0.400***	-0.400**	-0.290***	-0.287***	-0.287
	(0.083)	(0.083)	(0.165)	(0.099)	(0.099)	(0.232)
Staten Island	-0.740***	-0.767***	-0.767***	-0.868***	-0.895***	-0.895***
	(0.125)	(0.126)	(0.111)	(0.146)	(0.146)	(0.088)
Constant						
Manhattan	1.778**					
	(0.868)					
Bronx	1.887					
	(1.515)					
Brooklyn	3.255***					
	(0.798)					
Queens	2.789****					
	(1.015)					
Staten Island	0.282					
	(2.745)					
P-value of H0 : ui = 0				0.0000	0.0000	0.0000
Manhattan				0.0000	0.0000	0.0000
Bronx				0.0000	0.0000	0.0000
Brooklyn				0.0000	0.0000	0.0000
Queens				0.0000	0.0000	0.0000
Staten Island				0.0000	0.0751	0.0789
Clustered SE	no	no	yes	no	no	yes
Time FE	no	yes	yes	no	yes	yes
Precinct FE	no	no	no	yes	yes	yes

The dependent variable is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0: ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis. All estimations include dummy variables for each crime type in the sample.

Table 54: Summons Issued, by Year, Selected Other Crimes

Model	OLS	OLS	FE	FE
	(1)	(2)	(3)	(4)
Black				
2004	-0.734	-0.734	-1.164	-1.164
	(1.051)	(1.049)	(1.147)	(1.117)
2005	0.349	0.349	-0.138	-0.138
	(0.477)	(0.398)	(0.641)	(0.494)
2006	-0.462***	-0.462**	-0.201	-0.201
	(0.131)	(0.211)	(0.158)	(0.225)
2007	-0.441***	-0.441**	-0.284	-0.284
	(0.148)	(0.180)	(0.178)	(0.194)
2008	-0.454***	-0.454**	-0.767***	-0.767***
	(0.131)	(0.179)	(0.159)	(0.185)
2009	-0.431***	-0.431***	-0.746***	-0.746***
	(0.126)	(0.147)	(0.152)	(0.231)
2010	-0.473***	-0.473***	-0.762***	-0.762***
	(0.119)	(0.142)	(0.143)	(0.167)
2011	-0.597***	-0.597***	-0.693***	-0.693***
	(0.106)	(0.189)	(0.126)	(0.185)
2012	-0.405***	-0.405**	-0.305**	-0.305
	(0.118)	(0.184)	(0.140)	(0.225)
P-value of H0 : ui = 0				
2004			0.0000	0.0000
2005			0.0000	0.0000
2006			0.0000	0.0000
2007			0.0000	0.0000
2008			0.0000	0.0000
2009			0.0000	0.0000
2010			0.0000	0.0000
2011			0.0000	0.0000
2012			0.0000	0.0000
Clustered SE	no	yes	no	yes
Time FE	no	no	no	no
Precinct FE	no	no	yes	yes

The dependent variable is the probability of being issued a summons conditional on being stopped. * denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. P-value of H0: ui = 0 is the p-value for the joint test of all precinct fixed effects equal to zero. Standard errors are presented in parenthesis. All estimations include dummy variables for each crime type in the sample.