The Effects of Stop, Question and Frisk on Young Adolescents' Educational Outcomes

Bachelor's Thesis

Presented to the
Department of Economics at the
Rheinische Friedrich-Wilhelms-Universität Bonn

In Partial Fulfillment of the Requirements for the Degree of
Bachelor of Science (B.Sc.)

Supervisor: JProf. Julia Mink

Submitted in August 2023 by

Kiana Kavyar

Matriculation Number: 3228917

Contents

List of Figures

List of Tables

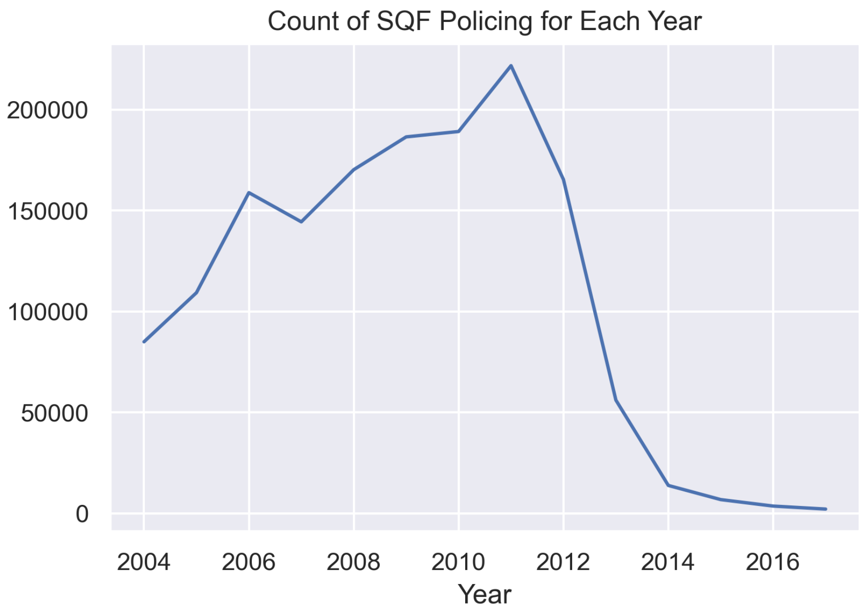


Figure 1. This plot shows the amount of stops for each year

Figure 2. This map shows the 32 school districts in New York City

Figure 3. This map shows the 123 precincts in New York City

Figure 4. This is the joined map when considering school districts and precincts that overlap with each other

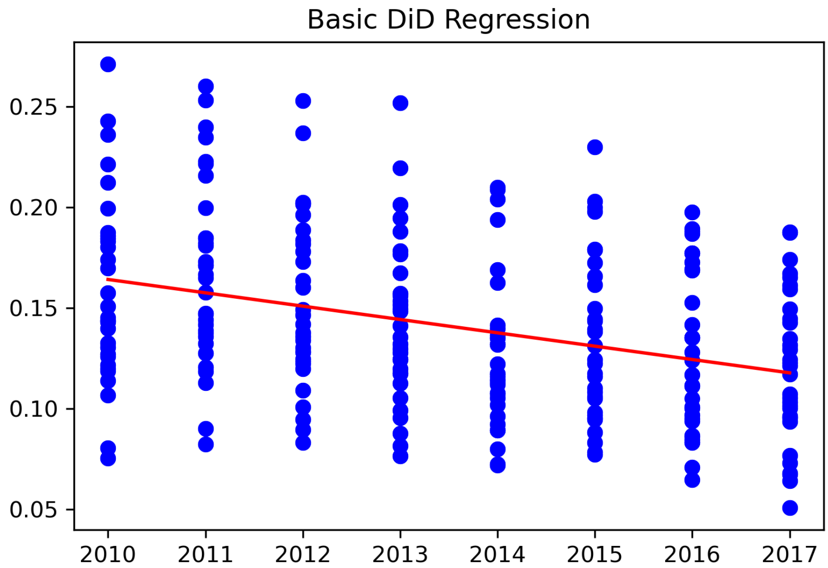


Figure 5. Basic DiD Regression with one Dummy Variable

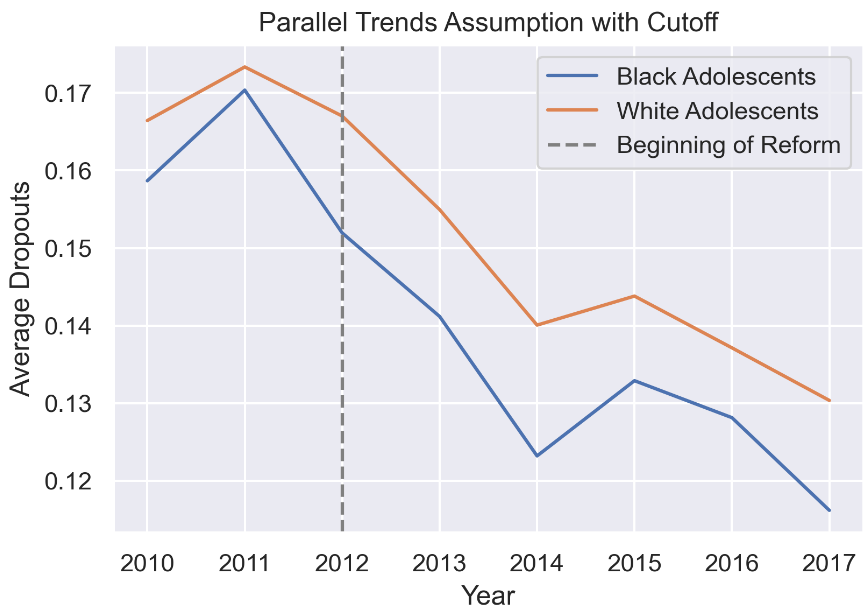


Figure 6. Parallel Trends Assumption

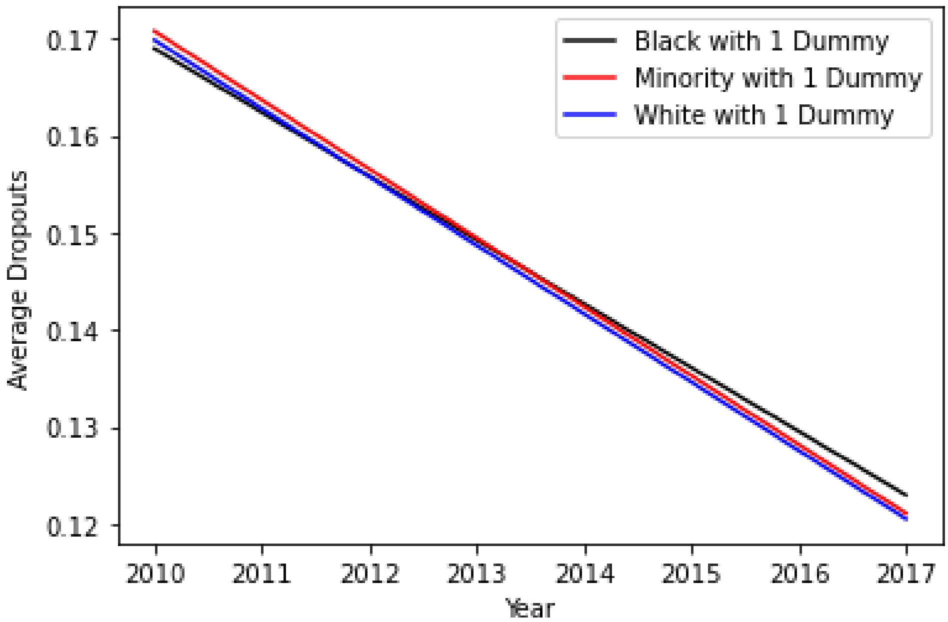


Figure 7. Differences within Race

Table 1. Regression Results

	Dependent variable:		
	(1)	(2)	(3)
Intercept	0.170***	0.171***	0.157***
	(0.007)	(0.005)	(0.005)
black	-0.005	-0.005	-0.005
	(0.009)	(0.006)	(0.006)
black_neighborhood			0.035***
			(0.004)
male		-0.002	-0.001
		(0.003)	(0.003)
post_12	-0.024***	-0.023***	-0.023***
	(0.008)	(0.006)	(0.006)
post_12:black	-0.008	-0.009	-0.009
	(0.011)	(0.008)	(0.007)
Observations	510	965	965
R^2	0.066	0.065	0.165
Adjusted R^2	0.061	0.061	0.160
Residual Std. Error	0.051(df = 506)	0.052(df = 960)	0.049(df = 959)
F Statistic	13.808*** (df = 3.0; 506.0)	19.361*** (df = 4.0; 960.0)	36.643*** (df = 5.0; 959.0)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2. Regression Results 2

	Dependent variable:		
	(1)	(2)	(3)
Intercept	0.170***	0.170***	0.151***
	(0.007)	(0.005)	(0.005)
male		-0.002	-0.001
		(0.003)	(0.003)
minority	-0.003	-0.002	-0.002
	(0.008)	(0.006)	(0.005)
minority_neighborhood			0.046***
			(0.003)
post_12	-0.031***	-0.031***	-0.026***
	(0.008)	(0.006)	(0.005)
post_12:minority	0.003	0.004	-0.001
	(0.010)	(0.007)	(0.006)
Observations	731	1,326	1,326
R^2	0.055	0.054	0.224
Adjusted R^2	0.051	0.052	0.221
Residual Std. Error	0.053(df = 727)	0.053(df = 1321)	0.048(df = 1320)
F Statistic	14.180*** (df = 3.0; 727.0)	19.033*** (df = 4.0; 1321.0)	76.095*** (df = 5.0; 1320.0)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3. District Level Fixed Effects Regression Results

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
post_12	0.1456	0.0050	29.057	0.0000	0.1358	0.1554
black	0.1645	0.0041	39.686	0.0000	0.1564	0.1727
post_12_black_interaction	-0.1779	0.0043	-41.588	0.0000	-0.1863	-0.1695

1 Introduction

In recent years policing strategies and practices in the U.S. have been heavily criticized throughout the media, leading to uproar among citizens. With regard to this matter the following thesis aims to analyze the impact of Stop, Question and Frisk (SQF) policing in New York City on educational outcomes for young adolescents aged 16 and on wards, extending the work of Legewie and Fagan in their paper “Aggressive Policing and the Educational Performance of Minority Youth". With the use of a Difference-in-Differences (DiD) estimator, the impact of SQF policing will be analyzed using official SQF reports [1] and Graduation Rate data in New York City [2] to then assess whether SQF policing led to more arrests thus dropouts or if the outcome is the exact opposite. Moreover the likelihood of spillover-effects regarding neighborhood exposure on parts of people’s life (here: educational outcome) will be assessed.

The outline for this thesis is as follows. Section 2 gives in the first subsection a historical background on SQF policing and its advances over the years. Among these are: CompStat, PredPol and ShotSpotter who advanced the police’s approach for future stops. Furthermore the second subsection elaborates in more detail on the history of SQF policing by discussing the Terry v. Ohio case, which is the main contributor behind the SQF policing. The last subsection will then talk about the possible dangers of SQF policing which can be in the form of increased dropouts or other. In Section 3 I will discuss related papers and their outlooks on SQF policing in regards to educational outcomes. These include Legewie and Fagans paper that concentrates on kids aged 9 to 15 years old and resulted in a higher effect for educational outcomes for African American males from 12 on wards. Tebes and Fagan on the other hand discuss the effects of Operation impact for students older than 15 years old. Section 4 will then allow for a deep dive into the data that has been used, the challenges faced using the given data and most importantly the DiD Regressions that have been built to allow for inferences. These include basic DiD Regression and also a fixed effects regression for the school district level. Lastly in Section 5 I will conclude the findings of Section 4 in a comprehensive manner as well as add a policy recommendation for the goal of decreasing high school dropouts.

1. which can be found on the official NYPD site.
2. which can be found on the NYSED Data Site.

2 The History Behind Stop, Question and Frisk

2.1 What is Stop, Question and Frisk?

“Stop, Question and Frisk" is a policing technique that aims to act as a crime reduction strategy by allowing police “officers to stop, question, and search individuals under reasonable suspicion", (Das and Bruckner, 2023). The idea behind it is quite straightforward. By stopping enough people, you will prevent many petty crimes, if not also the more major crimes (O’Neil, 2017, p. 92). Mathematician Catherine Helen O’Neil criticized in her book “Weapons of Math Destruction", the amount of harm algorithms can do to humans. In one of those Chapters O’Neil decided to take into account SQF policing and its advances using technology, which ended up harming people more than help them. Not just O’Neil herself criticized SQF policing but also the New York Civil Liberties Union went as far to sue the the Bloomberg administration, claiming SQF policing to be racist. Additionally they were able to back up their claims, with the use of available data, stating that African American men were six times more likely to be incarcerated than White men and twenty times more likely to be killed by the police (O’Neil, 2017, p. 93).

The origins of SQF policing dates back to the “Terry v. Ohio" case in 1968, where the Supreme Court ruled that these stops only remain permissible if there is no racial bias, but actual suspicion of a crime (Meares, 2014, p. 335). Due to this court case, SQF is also often times referred to as “Terry Stops". Throughout the years, technology has advanced and so has the police’s approach to reduce crime using the SQF methodology, essentially pairing the two. This is where “CompStat" surfaced in New York City. The software was founded back in the 1990’s and allowed officers at the New York Police Department (NYPD) to graph crime patterns, using analytical tools with the usage of already obtained crime data (*Technology - NYPD* (05.03.2023)). This way the police could see which areas to look out for more due to this pattern recognition.

This advance was soon followed by “PredPol" (now: “Geolitica"), which other states made use of. This is a program based on seismic activity, which is able to predict the occurrence of the next crime by accumulating old crimes into a historical pattern. This way the police would be able to strategically choose the most "profitable" precincts regarding SQF policing. Nonetheless there has been critics claiming that the use of PredPol does nothing but focus their attention on the poor, by stopping them more and sending a number of those to prison (O’Neil, 2017, p. 91). Not only did the police make use of crime predicting software, but also software that will alert officials of already occurred crimes (here: shootings). Known as “ShotSpotter", this technology launched in March of 2015 and it detects the sound of gunfire with real-time locations without the need of a 911 call. This technology aims to enable a faster response to incidents for the police (*Technology - NYPD* (05.03.2023)).

2.2 What is the harm of Stop, Question and Frisk?

Now that there is a clearer picture of the police's advances for crime prediction one would think that these software are great tools in addition to help SQF policing. However the case "Floyd v. City of New York" would beg to differ. Floyd, an African-American man living in the Bronx, New York was a victim of SQF policing. On that day, Floyd helped the basement tenant, who locked himself out, get back into his apartment with the multiple keys he had for the building. That's the moment when three NYPD officers stopped both pedestrians, asking them what they were doing just to proceed to frisk them. Judge Shira A. Scheindlin ruled that the SQF practice of New York violates both the Fourth and Fourteenth Amendment[3]. This case led to a new reform limiting the use of pedestrian stops (Community, 24.05.2023). With the introduction of such reform I predicted that this policy could possibly affect educational outcomes in a positive manner, reducing dropouts. The idea of why that should hold is, if there are less stops occurring then people (here: young adolescents) will feel safer in their local environment. This effect then would spillover to educational outcomes, which is here in the form of less dropouts. So by taking advantage of this reform, I want to assess the question whether SQF policing impacted young adolescent's educational outcomes in a positive manner by using a Difference-in-Differences estimator.

3 Related Papers

Other papers tried to make a similar analysis regarding policing and educational outcomes. In Legewie and Fagan (2019), the authors used a Difference-in-Differences approach for New York based students aged 9 to 15 years old, comparing their educational outcomes and putting these findings in relation to the policing technique they chose to focus on. For the educational outcome, the authors chose the English Language Arts (ELA) and Mathematics exam results. The ELA exam is a standardized test which is designed to assess grade 3-8 students in their proficiency of the English language, (Eliza Nimmich, COO of Learnt (2023)). As for the chosen policing technique, Legewie and Fagan decided to take "Operation Impact" into consideration.

3. The Fourth Amendment protects people from unreasonable search. The Fourteenth Amendment grants equal rights to African Americans

Operation Impact is the NYPD's crime reduction strategy that started back in January of 2003 by making use of data to efficiently deploy officers to high-crime areas, which they refer to as “Impact Zones". These Impact Zones were continuously being updated between 2004 and 2012 by removing, adding or even expanding them, (NYPD (2008)). The authors main focus of the paper was to determine the social costs of policing by taking into account two main findings of previous research, which suggested that either policing might have a positive impact due to its reduction in neighborhood crime, allowing for increased school performance or that aggressive broken-window policing[4] might have a negative effect on trust in authorities leading to withdrawal and system avoidance. Brayne (2014) explains system avoidance as the avoidance of institutions that keep formal records such as educational institutions (i.e. schools).

Fagan and Legewie however suspect that high rates of direct or indirect contact with the police might also create stress and other health and emotional responses that weaken cognitive performance. The idea for that suspicion goes as follows: police encounters are often harsh, entail racial/ethnic degradation. This can trigger persistent stress, anxiety, etc, which then can reduce cognitive and educational performance. Due to the lack of convincing causal evidence about the effects of proactive policing on minority youths' educational performance, both authors decided to close this gap with their paper. In regards to this matter, they also exploit the limitations of already existing papers which are: (1) Big focus on parental incarceration than on youth's (2) potential consequences of policing for youth have largely been ignored and (3) only focus on direct contact with criminal justice system (e.g. incarceration) but not the effect that extends to entire communities.

The authors findings suggested that Operation Impact did reduce crime, however it also did lower school attendance. Not only that, but it also affected minority youth's educational performance. In particular, African American boys from age 12 on wards.

Another paper related to my analysis is by Tebes and Fagan (2022), where the authors tried to answer whether the use of pedestrian stops as a crime deterrence tool was effective. In order to answer the given question the authors exploit the 2012 reform which ruled for the use of pedestrian stops to be limited. Just as in the first paper the authors also make use of a Difference-in-Differences (DiD) approach. To make sure the parallel trends assumption holds, Tebes and Fagan start off by comparing neighborhoods with similar crime rates, but substantially different stop rates prior to the reform. Not only crime is the main outcome measured in this analysis but also high school dropouts. This additional measure of high school dropouts has

4. The Broken Windows Theory by Wilson & Kelling suggests that visible signs of crime such as a broken windows will encourage additional crimes due to a lack of prosecution

the purpose of finding out if neighborhood stop exposure will also spill over to other parts of peoples' lives (here: education).

Moreover, the authors use two educational outcomes. One is an indicator for whether a student was discharged by a non department of education (DOE) agency directive, which gives a better insight to determine school dismissals that are linked to criminal justice. The latter is an indicator on whether a student drops out or is discharged by an institutional directive, allowing to neglect a link to criminal justice-related school dismissals. Now with these outcomes, the authors needed a mapping for the areas that the police stops occurred. In order to do so, Tebes and Fagan ended up grouping schools into four quartiles based on the level of stops in the students' home areas. The fourth and third quartile schools would be the "treatment" schools and the first and second quartile schools would be seen as "control" schools due to their low stop exposure. This allowed them to examine the impact of SQF policies on students' experiences and outcomes, taking into account both the school environment and the broader areas where students spend time.

By exploiting such a methodology the authors found out that fourth quartile schools displayed a reduction in the likelihood of being discharged by an institutional directive relative to students attending control schools by 54%. Furthermore the authors results suggest that counter to the crime deterrence hypothesis, which suggests that reductions in violent crimes occurred through increased policing, that treatment neighborhoods experienced reductions in **non-violent** crimes. Lastly the paper contributes to understand crime deterrence, adding evidence that SQF policing imposes negative externalities on local communities of color and also shedding light on mechanisms underlying the impact of neighborhoods on social mobility (i.e. change in a person's socioeconomic situation).

By having the prior knowledge of the discussed paper's outcomes, I will take these results to my advantage by answering the question whether SQF led to more high school dropouts. Because Legewie and Fagan's paper ends at the age of 15, I decided to extend their work and consider students above 15 until 21. The reasoning is quite simple: A child will be put into school at 7 years old latest, maybe has to re-do one school year and will then graduate at the age of 21 years. Thus, similar to Tebes and Fagan, I consider dropouts for high school students. Moreover their results show that stops are 21 times more likely during high school than in middle school. However, unlike Tebes and Fagan, I will not differentiate how a student got discharged nor group schools into multiple categories but group them in "race" and use that to consider treatment and control groups. I specifically chose to group by race due to Legewie and Fagan's results which suggest for African American boys, from 12 on wards, to experience a bigger deterioration in educational outcome than other kids.

Simply put, via a DiD regression, I will analyze whether the presence of SQF policing had an impact on high school dropouts due to incarceration or system avoidance.

3.1 What is Difference-in-Differences and its use?

Differences-in-Differences (DiD) is a statistical method that uses panel data, which involves observations on multiple entities (e.g., individuals, firms, regions) over multiple time periods, to identify effects of a certain policy. Unlike simple before and after comparisons, that can only be used for two different time periods — DiD more reliably, by comparing the changes in the outcome variable for a treatment group (exposed to the policy) and a control group (not exposed to the policy), aims to estimate the causal effect of the policy.

When wanting to make a causal inference using a DiD design, it's important to take a look at the DiD estimator $\hat{\beta}^{DiD}$ which can be computed the following way:

$$\begin{aligned} \hat{\beta}^{DiD} &= (\bar{Y}_{T,After} - \bar{Y}_{C,After}) - (\bar{Y}_{T,Before} - \bar{Y}_{C,Before}) \\ &= \Delta\bar{Y}_{Treatment} - \Delta\bar{Y}_{Control} \end{aligned} \tag{1}$$

where **T** stands for the treatment **C** for the control group, *Before* tells us the observation is from before the policy and *After* is for an observation after the policy. $\bar{Y}_{T,After}$ is thus the sample average for the treatment group after the treatment (here: 2012 reform) has been introduced. By focusing on the change in Y over the course of the experiment, the DiD estimator removes the influence of initial values of Y that vary between the treatment and control groups, such as state fixed effects. The DiD estimator is then the difference between the groups before and after the policy i.e., the difference of the two differences (i.e. $\Delta\bar{Y}_{Treatment} - \Delta\bar{Y}_{Control}$).

However there are a few assumptions that need to hold to make use of DiD estimation. The first thing that would come in mind are the four OLS assumptions, these need to be adjusted for the inclusion of a time dimension, since the individual, i for which $i \in \{1,...,n\}$, is observed over time t for which $t \in \{1,...,T\}$(Stock and Watson, 2019, p. 375). The adjusted OLS assumptions for these panel fixed effects regression, has the following assumptions:

(1) Our error term, u_{it}, has a conditional mean 0 i.e., $E(u_{it}|X_{i1}, ..., X_{iT}, \alpha_i) = 0$

(2) $(X_{i1}, ..., X_{iT}, u_{i1}, ..., u_{iT})$,for $i = 1, ..., n$ are i.i.d. draws from their joint distribution

(3) Large outliers are unlikely i.e., (X_{it}, u_{it}) have nonzero finite fourth moments

(4) There is no perfect multi-collinearity

Aside from these technical assumptions, the DiD estimator for a causal interpretation must satisfy the parallel trends assumption, to ensure internal validity of the DiD estimator. The parallel trends assumption states that in the absence of treatment (here: the 2012 reform), both control and treatment would follow the same trends but just in a parallel manner. Therefore one can conclude that omitted variables will affect both the treatment and the control similarly. When checking if the parallel trends assumption holds, one can make the control and treatment group as similar as possible.

For example, when wanting to check the effect of minimum wage increase for fast food restaurants as in Card and Krueger (1994) did for New Jersey and Pennsylvania(Angrist, 2014, p. 228), one could just check graphically, if control and treatment group would have been similar in trends pre-treatment (i.e. so in the absence of treatment, here: rise in minimum wage). If that is not the case then the parallel trends assumption is violated and both the control and treatment group need to be reconsidered. However there are also solutions to account for the violation of the assumption. For example, there could be confounds (i.e., variables that influence both the dependent and independent variable) during the time of observation for treatment and control group, for which if controlled for, would satisfy the parallel trends assumption.

Why do people make use of DiD estimation to begin with? Because it is a simple yet very effective tool allowing one to eliminate omitted variables that are time invariant yet have an effect on $\hat{\beta}^{DiD}$ ($\rightarrow$ the causal effect of interest), by having multiple observations on treatment and control groups. When taking a look at the DiD estimator, $\hat{\beta}^{DiD}$, one can see that it is very simple to compute, hence the argument of it being a simple tool. Of course there are also other tools to make use of when looking for causal inference e.g., propensity score matching and regression discontinuity design, but they are not as simple to implement and the marginal gain is unclear.

DiD also has its weaknesses such as the presence of autocorrelation (also referred to as serial correlation). Autocorrelation means that individual i correlates with itself over different time periods t. Autocorrelation confounds the coefficient of individual i for given time t as it is partly explained by the same individual i for time $t - 1$. This makes interpretation unnecessarily difficult, but is easily fixed using clustered standard errors. Clustering means that observations are being taken and put into different subgroups (i.e. clusters) such that the groups are independent of each other. On top, by using clustered *robust* standard errors (i.e. HAR standard errors), inference would be robust to heteroskedasticity and to autocorrelation. These standard errors allow the regression errors to have an arbitrary correlation within a cluster but assume that the regression errors are uncorrelated across clusters. Making clustered standard errors to be a quite flexible strategy due to them being able to correlate within clusters.

4 Working with the Data

4.1 Cleaning the SQF data

This section will give insights to the data work that has been done and go more in depth regarding the identification strategy. As already mentioned in the introduction, I will begin my analysis by taking data from the SQF reports from 2004 until 2017, as well as data from graduation rates for New York City.

(1) Before doing so, I made sure to check my environment and change it accordingly using the `os` package. When importing the SQF data set, I decided to only upload the data set with the columns of interest due to the size of the data. Those columns were: `Date of Stop`, `Precinct`, `Inside or Outside Stop`, `Suspected Crime`, `Frisked`, `Searched`, `Sex`, `Race`, `Age`.

(2) Because the SQF data is not available as one large file for the given years, I had to separately upload them using a `for` loop. Because not all years had the same column name, I had to proceed cautiously and rename accordingly.

(3) Additionally, I added a month column by creating a function which extracts the first two elements of the `Date of Stop` column values due to its format being: `MM/YYYY`.

(4) Once that was finished, it was time to concatenate all separate data frames into one containing all years. I then used the `unique` command as a hack to check for empty values in each column so that these could be deleted.

(5) One example for that would be the value of `999` in the precinct column which made no sense due to NYC Precincts ranging from 1 to 123.

(6) In addition to these steps, I made sure to change column values of `Year` and `Precinct` from string to numerical (to define categories), so that a regression analysis can be done using these column values.

(7) As already mentioned in Section 2, I also subsetted to observations that have `Age` to be between 16 and 21 so that I can make sure to only consider high school students.

(8) Now that the SQF data has been cleaned, I saved it as a `csv` file so that I could upload the cleaned up version for the analysis (without having to re-run everything).

4.2 Cleaning the graduation rates data

For cleaning the graduation rates data, I use similar steps for cleaning the SQF data.

(1) Just like with the SQF data, the dropout data is too large to upload by itself so I chose a handful of columns so that I could upload the dropout data for each year, one by one.

(2) The considered columns here were: `Year, School District, Race, Graduate%, IEP Diploma%, Still Enrolled%, GED%, Dropped Out%`. For clarification, the IEP Diploma is awarded to students with a disability at the end of the school year they turned 21 or after a student has attended school for at least 12 years (*Special Education* (17.05.2023)).

(3) Then, as with the SQF data, I changed the column names so they are all alike. Unlike the SQF data, I had to filter out school districts that weren't in NYC. On top of that I created a `Year` column by using a for loop.

(4) When looking more closely at the columns, I realized that the years 2004-2009 had the AHSEP[5] and for the years 2010-2017 the GED was considered instead of the AHSEP. Since those are two different Diplomas and can't be mapped onto each other the decision was to start with the year of 2010 until 2017 and leave out the years 2004-2009.

(5) Once this issue was fixed all data frames have been concatenated into one big data frame. When looking at the column values which gave out percentages for graduates, drop outs etc. it was obvious that these needed to be adjusted. For example the column that had the values for percentage of graduates was filled with a mix of string and integer variables such as `20%, 0.04,...` The integer values weren't a problem here but the string values. So I created a function that would only consider string values with a percentage sign as their last element.

(6) Once those values were found the function proceeded to delete their last element (here: the %), change the string value to numeric and then divide it by 100. So that a string value such as `20%` gets turned to `0.2`. I save the data frame as a `csv` file.

(7) Now comes the tricky part. Because I wanted to work with both the SQF data and the dropout data, I needed to merge them both in a way that makes sense. Both data frames have information on their location, Precinct for the SQF data and `School District` for the dropout data.

(8) Now one way to connect these two locations was to map the precincts to school districts by creating a dictionary. At first the idea seems quite simple and straightforward, but unfortunately it was anything but simple. The main issue was that for each school district there are multiple precincts and at times, one precinct within multiple school district.

5. Stands for Alternative High School Equivalency Preparation Program (AHSEP)

(9) With the help of the geopandas package and the .geojson files for both NYC school districts and precincts[6] I was able to also visually show the given problem. As you can see in figures 2 and 3 I did not have a simple point to polygon[7] mapping, but a polygon to polygon mapping. So what I ended up doing was manually finding out which one of the precincts was the "main" precinct for a given school district using the NYC School site [8] and the Precinct finder site from the NYPD [9]. Resulting in each school district mapping to exactly one precinct and not multiple ones. Visually this can be seen in figure 4. Sure this doesn't make interpretation easier but makes the regression feasible. To apply the mapping of precincts to districts I used the `.map` method.

(10) Since I needed information from both data sets, I performed an inner merge on three different columns: `district, year, race`. This way when I use the merged data frame my inference can actually be linked to SQF policing. Now that both data sets have been merged into one I was able to create the dummy variables needed for my DiD Regression shown in Section 2, which are `black, hispanic, minority, white, asian, post_12, black neighborhood, minority neighborhood, white neighborhood`. When creating the race dummies I made sure to take a quick look at the available values for it. Those were: `Hispanic, Black,White, Asian/Pacific Islander, American Indian/Alaska Native`. Leaving me with 5 values for race I knew that I needed to compute 4 dummies to avoid the dummy variable trap.

4.3 Data descriptives

From the cleaned version of the data, I plotted the raw count of stops for the years 2004 until 2017 using the SQF data frame — to observe the development of stops for these years 1. When looking closely at the graph, despite not controlling for population growth or other confounds, we observe a sudden crash in 2011-2012. This observation aligns with the knowledge of the 2012 reform that ruled for a limitation of pedestrian stops. From 2011 on wards, the amount of stops kept declining, reaching a new low in 2017 with 2118 stops. This is almost one percent of the 221697 stops in 2011. Looking at these numbers, the conclusion could be drawn that the 2012 reform was successful.[10]

6. Found on GitHub
7. a flat two-dimensional closed shape bounded with straight sides
8. NYC Schools
9. NYPD Precinct Finder
10. There are accusations that the NYPD may be underreporting.

4.4 Identification Strategy

With the now obtained knowledge of DiD in subsection 3.2, I built a simple DiD regression with one race dummy, which has the following specfication:

$$Average\_Dropout_{it} = \beta_1 + \beta_2 * (Black_i) + \beta_3 * (Post\_12_t) + \rho * (Black_i * Post\_12_t) + \epsilon_{it}$$

where $Black_i$ is the race dummy having a value of 1 if individual i is African American and 0 otherwise. "$Post\_12_t$" is the time dummy, being 1 for years 2012 and on wards where the reform took place and else 0. Then we have the interaction effect (i.e. $\hat{\beta}^{DiD}$) that is "$(Black_i * Post\_12_t)$", which captures the "effect" of the reform on African Americans. Lastly there is the error term ϵ_{it}, which captures the unexplained variance in the dependent variable. Before directly plotting my above model I decided to add an extra step and train (i.e. fit) my independent and dependent variable. This way the parameters of the model will get adjusted, resulting in a more precise regression model that has more accurate predictions (Pramoditha (11.04.2022)). This concept of training the independent and dependent variable might be a hard concept to grasp so let me explain this through a baking analogy, inspired by Dr. Meike Zehlike's cooking recipe analogy for explaining Artificial Intelligence at a talk held in support of the Heinrich-Böll-Stiftung (29.07.2023). In order to bake a cake (here: our independent variable) a lot of ingredients (here: our dependent variables) will be used. However as you might know from experience it takes practice baking the "perfect" cake. So by baking the same cake multiple times, eventually I will get the hang of the perfect amount of sugar, flour, milk and eggs (our dependent variables) resulting in the best possible cake (our independent variable). Applying this logic to my regression analysis now will make it clearer why I decided to train my dependent and independent variable first before plotting the model straight ahead. Once it was time to plot the model I made sure to use robust standard errors to control for heteroskedasticity. Ideally as mentioned in the subsection 3.1 it would be best to make use of HAR standard errors when building the regression. However when choosing the HAR standard error in Python, the number of lags have to be known and defined. Because I did not fully know how to choose the right amount of lags, I decided to take the robust standard errors into account. As known through Stock and Watson whether we have a homoskedasticity or heteroskedasticity, it is always best to choose robust standard errors. That is because the robust standard errors will "produce valid statistical inferences whether the errors are heteroskedastic or homoskedastic"(Stock and Watson, 2019, p. 191). Once the construction of a linear regression model with robust standard errors was done it was time to plot the table. By using the stargazer package, leaving me with the necessary LateX code to add my regression results in a table, I added the regression table accordingly. Now that the procedures around the construction of the linear model have been thoroughly discussed, I will continue by discussing the regression results as can be found in tables 1 and 2. With the now known results for the β and the ρ coefficients, the regression considering only "$Black_i$" as a dummy variable in table 1 will have the following shape:

$$\begin{aligned} \text{Average Dropout}_{it} &= 0.170 - 0.005 \times \text{Black}_i - 0.024 \times Post\_12_t \\ &\quad - 0.008 \times (\text{Black}_i \times Post\_12_t) + \epsilon_{it} \end{aligned} \tag{2}$$

These results can be interpreted as the following: The average dropout rate for White people ($black_i = 0$), our control group, before the 2012 reform ($post\_12 = 0$) equals to 17%, which not only seems quite high but is also significant at the 1% level. Under the same circumstances, as stated above, the average dropout rate for African Americans is slightly lower with an average dropout rate of 16,5%. This is due to the coefficient of "black" having a value of -0.005. Now this might seem odd because there has been the suspicion that being African American would lead to higher dropouts and not lower ones. However because the "black" coefficient is not significant at any level one can suspect that this negative coefficient is there by chance (i.e. unrelated to the SQF policing). Now when looking at the results for the control group **after** the 2012 reform we can see that there is a significant decrease of 2,4% leaving the average dropout rate at 14,6%. For African Americans the result is again slightly lower being at 13,3%. This is due to the interaction effect having a value of -0,8%. This leads to believe that the 2012 reform did have a positive impact, leading to less dropouts. This is definitely true due to the significance of the "post_12" coefficient. On the other hand when taking a look at the coefficient results it is quite clear that being African American doesn't have any effect on the dropout rate . This can be seen in the interaction term, or in other words the "effect" of the reform on African Americans. Because the interaction coefficient is insignificant, the effect of the reform on African Americans is therefore none. Additionally this would lead to believe that SQF policing is blind to race. This claim can be supported when looking at the coefficient result for "black" which is insignificant, meaning there's a high likelihood that race does not have much of an impact on the average dropout rate as the 2012 reform, for example, has. Continuing with the analysis I decided to also compute the DiD estimator (i.e. interaction effect) manually, as seen in subsection 3.1. After implementation I was left with:

$$-0.008 = (0.1322 - 0.1456) - (0.1645 - 0.1699)$$

Leaving me with a $\hat{\beta}^{DiD}$ of -0.008 which is the same value as the interaction effect on the first column of the Regression table1 when considering that those coefficient values are all rounded up. Now being naive this result would lead us to believe that the effect of the 2012 reform when being African American lead to a slight decrease in dropouts by 0,8%. But as already discussed it is clear that due to the insignificance of the interaction term this naive assumption does not hold. Nonetheless the significance of the "post_12" coefficient gives evidence to a decrease in dropouts since beginning of the reform. This can also be visually assessed when taking a look at the plotted regression in figure 5. When looking at the plot there's a noticeable downward trend in dropouts over the years. Not only that but also the presence of heteroskedasticity is observed when taking a closer look at the scatter plot. Before continuing with the regression analysis I compared dropouts rates in figure 6 between White people and African Americans pre reform to see if the parallel trends assumption holds. The cutoff represented for the year of 2012 is there to visually represent the cutoff between average dropout rates before and after the introduction of reform. Now pre reform (i.e. in absence of treatment) both White people and African Americans do follow the same trend in a parallel manner. Therefore as known in subsection 3.1, the parallel trends assumption holds, resulting to a given internal validity of the DiD estimator. Surprisingly after implementation of the reform in 2012, both treatment and control group still follow the same trend, whereas White people are worse off. In most cases when comparing control and treatment group post policy implementation it is rather common to see both not following equal trends. Now one of the reasons on why both groups still follow the same trend after reform implementation could be due to spillover effects such as neighborhood stop exposure, as Tebes and Fagan (2022) suggested. Note that even though the plot suggests that White people have a higher dropout rate than African American, it is important to keep in mind that there are less African Americans represented in the data set than there are White people. Resulting in having a highly unbalanced data set which won't make inference any easier. Because the parallel trends assumption holds, which is needed here to make causal inference, I can continue with my regression analysis. Moving forward to the second regression being:

$$\begin{aligned}\text{Average Dropout}_{it} &= \beta_1 + \beta_2 \times \text{Black}_i + \beta_3 \times \text{Male}_i + \beta_4 \times Post\_12_t \\ &\rho \times (\text{Black}_i \times Post\_12_t) + \epsilon_{it}\end{aligned} \tag{3}$$

it is noticeable that all standard errors decreased indicating that the added gender dummy leads to a more precise regression estimation. The control group has now a slightly higher average dropout rate of 17,1%, for females prior to the 2012 reform. When being African American the average dropout rate again decreases by 0,5% leaving them with 16,6%. When being male the average dropout rate decreases barely with an insignificant value for the gender dummy. This finding violates the assumptions of Legewie and Fagan that especially African American boy's educational outcome is affected. After the introduction of the 2012 reform the average dropout rate drops by 2,3%, leaving White people at 14,8%. Whereas African Americans are left with an average dropout rate of 13,4% due to the interaction effect. Lastly the third regression, which can be seen here:

$$\text{Average Dropout}_{it} = \beta_1 + \beta_2 \times \text{Black}_i + \beta_3 \times \text{Male}_i + \beta_4 \times Black\_Neighborhood_i + \beta_5 \times Post\_12_t \\ \rho \times (\text{Black}_i \times Post\_12_t) + \epsilon_{it} \tag{4}$$

adds a neighborhood dummy. The idea behind it was to replicate the idea of Tebes and Fagan, seeing if neighborhood stop exposure also spills over to people's education. For the female control group the average dropout is at 15,7% pre reform. As suspected the treatment group's average dropout is 0,5% less than for the control group, leaving them with an average dropout rate of 15,2%. Out of all three regressions this is by far the lowest dropout rate pre reform. Considering the neighborhood dummy, which is significant at the 1% level, we can see that location really does matter. Leaving people with a 3,5% higher chance of dropping out just for living in the "wrong" neighborhood. This result aligns with the assumptions Tebes and Fagan made stating that neighborhood stops do spill over to people's education (here:leading to further dropouts). It would also explain the similar trends in figure 6 for Control and Treatment group post reform. Because the coefficient of black was not just very low but also statistically insignificant I wanted to see if the same holds for minority groups. Where the minority group consists of African Americans and Hispanic people. I did the three exact regressions as for African Americans but for minorities. This was done by changing the race dummy from "black" to "minority". The first regression in table 2 has the exact same value for the intercept, regarding the control group, as the 1. regression table does. Surprisingly the 2012 reform resulted in a bigger decrease of dropouts with 3,1% than the other regression analysis. Showing that Minorities and White people ended up having the same amount of average dropouts post reform (13,9%). This was due to the "minority" coefficient and the interaction term's coefficient being exact opposites and therefore cancelling each other out. This would show that race doesn't play a role post reform. However due to the insignificance of both the minority coefficient and the interaction effect there isn't a strong relationship between dropouts and being part of a Minority. To also get a visual representation of the average dropout rate for different races, the plot in figure 7 gives a good comparison when having only race as a dummy. All three races start off at almost the same average dropout rate when looking at the first year's development. Whereas for all races being part of a Minority lead one to be worse off before reform. Soon after reform the positions switched, leaving Minority and White people with less dropouts overall and African Americans with the most dropouts. Just as the analysis before, when continuing to look at the second regression it is clear that also here all standard errors decreased hinting to a more precise estimation. Again the male dummy suggests a 0,2% decrease in dropouts and the intercept's results are the same as for the first regression. The only noticeable difference is the additional 0,1% increase for the interaction effect. This leaves female minorities at a dropout rate of 14,1% ($17\% - 0,2\% - 3,1\% + 0,4\%$) post reform. For female non-minorities the dropout rate is at 13,9% suggesting that this group is better off. However since neither the interaction effect nor the minority coefficient are statistically significant this lower dropout rate can be purely be by chance and not due to being female or part of a minority. Lastly when taking a look at the third regression it is again noticeable that the addition of a neighborhood dummy shows significance at the 1% level. Compared to the other regression analysis, considering only African Americans, the dropout rate pre reform for female non-minorities who don't live in a minority neighborhood is at an all time low with 15,1%. For minorities under the same assumptions the average dropout rate is at 14,9% due to the minority coefficient having a value of -0,2%. The male coefficient here decreased to -0.1% compared to the second regression. Interestingly it is shown that living in a minority neighborhood increases chances of a dropout by 4,6%, which is 1,1% more than for living in an African American dominated neighborhood. Lastly I decided to replicate the first regression, considering only "black" as a race dummy, as a fixed effects regression. Additionally I was able to cluster for my standard errors. This way there won't be any occurrence of autocorrelation. This fixed effects regression in table 3 is fixed on the district level and leaves us with lower standard errors than seen for the first regression in table 1. This shows that the fixed effects regression makes up for a more rigorous estimation. When interpreting the results it is shown that the reform lead to $14,56\%$ dropouts in average, however being African American lead to $16,45\%$. Meaning that being African American goes hand in hand with a $1,89\%$ increase of dropouts. This is a lot more than is shown in the first regression, in figure1. Lastly the effect of introducing the reform when being African American is associated with a decrease of $17,79\%$ dropouts on average. This concludes that the effect of the reform when being African American only decreases the average dropouts by $1,34\%$. This value aligns when computing the sum for the "black" and interaction effect coefficient ($-0,5\% - 0,8\% = -1,3\%$) found for the first regression in table 1. With all these information at hand it is now time to assess whether these results all align with the conclusions the authors had in their papers? Well, partly. The literature (Legewie and Fagan) suggested that young African Americans get frisked more however that can't be taken away from the data due to the insignificant results for both the "black" and "minority" coefficient. Nonetheless as already mentioned the data does show that neighborhood stop exposure spills over people's education, just as Tebes and Fagan suggested in their paper "Stopped by the Police: The End of "Stop-and-Frisk" on Neighborhood Crime and High School Dropout Rates".

5 Conclusion

With all the above information at hand, a conclusion can be made. Just as Legewie and Fagan's and also Tebes and Fagan's paper the use of DiD estimation was done.

The results suggested that there has been a high dropout rate prior the introduction of the 2012 reform. Not just that but also comparing dropouts for African Americans and White people show that race doesn't influence the dropouts much for the effect being insignificant and at -0,5%. This result already contradicted the findings of Legewie and Fagan, which stated that especially young male African Americans aged 12 and older experienced the highest impact on educational outcome. However this contradiction could be explained by the considered age range of individuals. Unlike Legewie and Fagan, who had an age range of 9 to 15 years old, I decided to extend that analysis by looking at an age range of 16 to 21 years old. Therefore it could be assumed that the effect of race is higher when considering younger individuals. Because the age range differs quite a lot, the educational outcome used for comparison will also differ. The authors chose the ELA and mathematical exam while I decided to focus more on the amount of dropouts (in percent). On top of this the authors focus on SQF policing more in depth by mainly taking into account "Operation Impact". As already mentioned race did not affect the average dropout rate, but the introduction of reform definitely did. With an up to $3,1\%$ decrease for Minorities in average dropouts. But not just the reform had an impact on the average dropout rate but also different minority neighborhoods. This approach considering neighborhoods comes closer to what Tebes and Fagan did. Their idea was to not only consider the school environment but also the one at home cause those two might differ substantially within each other. For me adding a neighborhood dummy helped to discover that location does have an impact on educational outcomes with an increase of up to $4,6\%$ in average dropouts for Minorities. So as suspected by Tebes and Fagan there is a clear spill-over effect regarding neighborhood exposure on other parts of people's life (here: average dropout rate). When using a fixed effects regression on the district level I was able to not just cluster my standard errors but also get smaller standard errors, indicating my regression to be precise. When comparing the coefficient values for the fixed effects regression, it is soon clear that those do slightly differ to the ones of the linear regression, which only uses robust standard errors. This could be explained by having a regression for the district level (here: the Fixed Effects regression). In total the results suggest that it is not race that punishes individuals with higher dropout rates but the neighborhood one lives in. This would lead to believe that it is best to let go of "Operation Impact" due to neighborhood effects diminishing almost all positive impact of the 2012 reform. Such ruling could have a positive impact on individuals through less stops and therefore a lower likelihood of incarceration and dropouts.

References

Angrist, Joshua. 2014. *Metrics - How Economists Learn about Cause and Effect: The path from cause to effect*. Princeton, NJ, and Oxford: Princeton University Press. [7]

Brayne, Sarah. 2014. "Surveillance and System Avoidance: Criminal Justice Contact and Institutional Attachment." *American Sociological Review* 79 (05): 367–91. DOI: 10.1177/0003122414530398. [4]

Community. 24.05.2023. "Floyd v. City of N.Y. | Case Brief for Law School | LexisNexis." URL: https://www.lexisnexis.com/community/casebrief/p/casebrief-floyd-v-city-of-n-y. [3]

Das, Abhery, and Tim A. Bruckner. 2023. "New York City's Stop, Question, and Frisk Policy and Psychiatric Emergencies among Black Americans." *Journal of Urban Health* 100 (2): 255–68. DOI: 10.1007/s11524-022-00710-x. [2]

Eliza Nimmich, COO of Learnt. 2023. "What is ELA? and How to Study - Tips and Strategies | Learnt." *Learnt Blog*, URL: https://learnt.io/blog/what-is-ela/. [3]

Heinrich-Böll-Stiftung. 29.07.2023. "ChatGPT: KI verstehen - Der richtige Weg? | Heinrich Böll Stiftung Baden-Württemberg." URL: https://www.boell-bw.de/de/2023/06/12/chatgpt-ki-verstehen-der-richtige-weg. [11]

Legewie, Joscha, and Jeffrey Fagan. 2019. "Aggressive Policing and the Educational Performance of Minority Youth." *American Sociological Review* 84 (2): 220–47. DOI: 10.1177/0003122419826020. [3]

Meares, Tracey L. 2014. "The Law and Social Science of Stop and Frisk." *Annual Review of Law and Social Science* 10 (1): 335–52. DOI: 10.1146/annurev-lawsocsci-102612-134043. [2]

NYPD. 2008. "NYC Safety and Security Operation Impact." Edited by NYC Government. [4]

O'Neil, Cathy. 2017. *Weapons of math destruction: How big data increases inequality and threatens democracy*. First paperback edition. New York: Broadway Books. [2]

Pramoditha, Rukshan. 11.04.2022. "Why Do We Need a Validation Set in Addition to Training and Test Sets?" *Towards Data Science*, URL: https://towardsdatascience.com/why-do-we-need-a-validation-set-in-addition-to-training-and-test-sets-5cf4a65550e0. [11]

"Special Education." 17.05.2023. "Special Education." URL: https://www.p12.nysed.gov/specialed/publications/iepdiploma.htm. [9]

Stock, James, and Mark Watson. 2019. *Introduction to Econometrics Global Edition*. Pearson Deutschland, 800. URL: https://elibrary.pearson.de/book/99.150005/9781292264523. [6, 11]

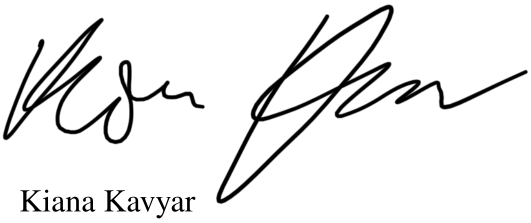
Tebes, Jonathan, and Jeffrey Fagan. 2022. "Stopped by the Police: The End of "Stop-and-Frisk" on Neighborhood Crime and High School Dropout Rates." [4, 13]

"Technology - NYPD." 5.03.2023. "Technology - NYPD." URL: https://www.nyc.gov/site/nypd/about/about-nypd/equipment-tech/technology.page. [2, 3]

Statement of Authorship

I hereby confirm that the work presented has been performed and interpreted solely by myself except for where I explicitly identified the contrary. I assure that this work has not been presented in any other form for the fulfillment of any other degree or qualification. Ideas taken from other works in letter and in spirit are identified in every single case.

6. August 2023



Kiana Kavyar