Evaluating the Dynamics of Racial Bias in Arrest Patterns: A Comparative Analysis Pre- and Post-George Floyd Protests in Chicago

Nafis Ahmed Munim

Introduction

The quest for an equitable criminal justice system in the United States remains at the forefront of national discourse, particularly concerning the systemic interactions between race and law enforcement. This research proposal addresses the question: "Is there evidence of racial bias in arrests for criminal offenses, and how has this potential bias evolved before and after the protests following the death of George Floyd?" It seeks to uncover the underpinnings of arrest patterns in Chicago, a city of America's urban diversity, and its complex relationship with policing.

The death of George Floyd on May 25, 2020, emerged as a catalyst for a global reckoning with racial injustices, casting a renewed spotlight on the prevalence of racial bias within police practices. Profound public demonstrations marked Chicago's reaction to this event and call for police reform, placing it at the heart of a movement that demands scrutiny and change. By leveraging rich datasets from the Chicago Data Portal and community-level insights from the Chicago Metropolitan Agency for Planning (CMAP), this research proposes a rigorous examination of arrest data to understand the correlation between racial demographics and the impact of the George Floyd protests.

The analytical journey is charted via a two-phase regression analysis: the baseline regression, which seeks to delineate the correlation between community-level demographics and arrests before the influence of the Floyd protests; and a Binomial Logistic Regression model that evaluates the causal impact of these protests on arrest rates. The focal point of this inquiry is to establish whether there is a causal relationship between racial demographics and arrests and to measure the extent to which the Floyd protests have served as an inflection point in this context.

The literature review accompanying this proposal will navigate through the scholarly discourse on racial biases in policing, framing the investigation within the wider academic and socio-political landscape. It aims to draw upon diverse sources to provide a comprehensive theoretical backdrop for the empirical analyses that will follow.

This research stands not only as an academic endeavor but as a timely societal investigation, poised to inform policy and potentially guide the trajectory toward a more just system of law enforcement. The following pages detail the proposed research design, methodology, and the potential impact of its findings, aiming to elucidate the mechanisms of racial bias and their evolution in a post-George Floyd era. Through answering the central research question, this study aspires to contribute a substantive, data-driven narrative to the ongoing conversation about racial equity in American policing.

Literature Review

The exploration of racial bias within the American criminal justice system has become increasingly significant, particularly following the widespread protests in response to George Floyd's death. This defining moment has catalyzed a national and international dialogue about the intricate relationship between race and law enforcement, underscoring the importance of conducting a thorough analysis of arrest patterns in major urban centers of the United States, such as Chicago. This research proposal gets into the multifaceted nature of racial bias in arrest processes, situating its inquiry within a broad spectrum of academic literature that has meticulously explored various aspects of discrimination and systemic bias within law enforcement.

Roland G. Fryer Jr.'s Study on Police Use of Force:

A cornerstone of this exploration is Roland G. Fryer Jr.'s seminal work, "An Empirical Analysis of Racial Differences in Police Use of Force." This study stands out for its comprehensive examination of both non-lethal and lethal uses of force within policing and their relation to racial dynamics. Fryer’s analysis uncovers that black and Hispanic individuals are significantly more likely to encounter non-lethal force in interactions with the police. This pattern of racial disparity holds even after accounting for a variety of influencing factors, suggesting a deep-seated racial bias in these types of police encounters. Surprisingly, Fryer's findings diverge when it comes to lethal force, particularly in officer-involved shootings. His research reveals no racial discrepancies in these incidents, a finding that remains consistent even when contextual variables are considered, thus challenging prevalent narratives about racial bias in the most extreme forms of the police force.

The methodological robustness of Fryer's study is evident in its use of data from a range of sources, including the New York City Stop, Question, and Frisk program, the Police-Public Contact Survey, and extensive records of officer-involved shootings from several cities across the United States. This comprehensive approach enables a detailed examination of the variables involved, such as civilian behavior and the characteristics of police encounters, offering a holistic view of these interactions. However, the study is not without its limitations. It predominantly relies on reported incidents, potentially missing unreported or misreported cases of police use of force. Additionally, the focus on documented cases of lethal force, namely officer-involved shootings, may exclude other forms of fatal encounters with law enforcement. These limitations notwithstanding, Fryer's research offers crucial insights into the complex racial dynamics in police use of force, highlighting the need for nuanced understandings of racial bias in law enforcement. This study is particularly relevant to the proposed research on racial bias in arrest patterns, as it underscores the importance of considering a broad spectrum of police behaviors and the potential disparities inherent in them.

"The George Floyd Effect" by Cassella et al.:

In the aftermath of George Floyd's death and the subsequent Black Lives Matter protests, the study titled "The George Floyd Effect: How Protests and Public Scrutiny Changed Police Behavior" by Cassella et al. undertakes a critical examination of contemporary law enforcement. This research digs into the impact of these widespread protests on police behavior and public safety, probing into the contentious notion of the "Ferguson Effect"—a hypothesized retreat of police from regular duties potentially leading to an escalation in crime rates.

The study presents key findings, notably a significant, sustained decrease in officer contact with civilians across cities like Seattle, Austin, Philadelphia, and Los Angeles. This observed reduction in policing activity, termed 'de-policing,' suggests a substantial shift in police operations in the wake of the protests. The study, however, reveals that this de-policing does not align neatly with either pro-social or anti-social policing behaviors. It points to a complex pattern where the decrease in policing could either be interpreted as a structured response to demands for reform or as an unstructured outcome stemming from individual-level stress and criticism. Importantly, the study contradicts fears that reduced policing would lead to an increase in violent crime, finding no clear evidence to support a link between de-policing and changes in violent crime rates. This challenges the narrative that proactive policing is indispensable for deterring serious crime and suggests a more nuanced understanding of the relationship between policing intensity and public safety.

In terms of methodology, Cassella et al. employed a sophisticated Regression Discontinuity-in-Time (RDiT) design to isolate the immediate effects of the BLM protests from other influencing factors. This approach is particularly noteworthy for its focus on the immediate changes in police behavior at the onset of the protests, thus minimizing potential biases from long-term trends. The study also accounted for the concurrent COVID-19 pandemic, ensuring that the observed changes in police behavior were attributable to the protests and not confounded by pandemic-related factors. Nevertheless, the study acknowledges its limitation in analyzing the 2020 BLM protests as a 'bundled treatment.' The multi-dimensional nature of these protests makes it challenging to isolate specific causal factors behind the observed changes in police behavior. The research also identified a significant decrease in discretionary policing activities across all cities involved, indicating a widespread response to the protests. However, the long-term implications of these changes for policing strategies and public safety remain areas requiring further exploration.

"The George Floyd Effect" study provides invaluable insights into how large-scale social movements can influence police behavior and practices. While it demonstrates changes in policing in response to the protests, it also highlights the complexity of these changes, the varied nature of police responses, and the necessity for further research to fully understand their implications. This study is particularly pertinent to the proposed research on racial bias in arrest patterns, as it provides evidence that societal events like the George Floyd protests can have a significant influence on policing behaviors and strategies. Such insights are crucial for understanding the dynamics of law enforcement in the context of social upheaval and public scrutiny.

Legewie and Fagan (2019) on Aggressive Policing:

Legewie and Fagan's 2019 study offers an insightful analysis of the intersection between aggressive policing and its repercussions on the educational performance of minority youth. This research is vital in understanding how policing extends beyond immediate street-level interactions, affecting the socio-economic fabric of minority communities and perpetuating cycles of disadvantage. The study utilizes regression analysis to explore the relationship between aggressive policing and educational outcomes, taking into account various socio-economic factors. This methodological approach allows for a detailed exploration of how police interactions influence the long-term educational trajectories of minority youth.

However, the study acknowledges the challenge of isolating the direct impact of aggressive policing from other socioeconomic factors that could also affect educational performance. The complex interplay of these variables makes it difficult to determine the exclusive effect of aggressive policing. Despite this limitation, the findings of the study are significant in understanding the extended consequences of policing on minority youth. The research suggests that systemic patterns of discrimination in law enforcement extend beyond immediate physical or psychological impacts, permeating deeper into the fabric of community life and affecting future opportunities for young individuals. This aspect of the research is particularly relevant to the proposed exploration of systemic racial bias in arrest patterns. Understanding the extended impact of policing on education adds a crucial dimension to the conversation about systemic discrimination within law enforcement.

Edwards, Lee, and Esposito (2019) on Mortality Risk:

The research conducted by Edwards, Lee, and Esposito in 2019 sheds light on a grim aspect of law enforcement: the mortality risk due to police use of force. Their findings indicate that African American, Indigenous, and Latino individuals face significantly higher lifetime risks of being killed by police compared to their white counterparts. This study positions police violence as one of the leading causes of death among young men, particularly young men of color. To estimate these mortality risks, the researchers employed Bayesian simulation and multilevel models, creating detailed risk profiles that consider age, race, and sex. This methodological approach provides a comprehensive picture of the risks associated with police encounters across different demographic groups.

However, the study's reliance on risk profiles observed during a specific period might not capture potential changes in police practices over time. The relevance of this study to the proposed research is profound. By highlighting the extreme consequences of racial bias in law enforcement, it underscores the severity of the issue. The mortality risks associated with police encounters offer a stark reminder of the ultimate costs of systemic bias. This perspective is crucial for the proposed examination of arrest patterns, as it situates the research within the broader context of life-and-death implications of racial bias in policing. Understanding these risks provides a more holistic view of the systemic racial disparities in law enforcement, reinforcing the importance of addressing these issues in policy and practice.

The synthesis of these critical studies forms a comprehensive backdrop for the proposed research. It supports a hypothesis that suggests the existence of systemic racial bias in arrest patterns, which manifests in various forms of discrimination, including disparities in educational outcomes, mortality risks due to police use of force, and the psychological impact of policing on minority communities. The research seeks to test this hypothesis by examining arrest patterns before and after the George Floyd protests, positing that such societal events have the potential to mitigate discriminatory practices within law enforcement. In conclusion, the extended literature review shapes a nuanced hypothesis that underpins the proposed research. It anticipates that, if systemic racial bias is indeed present in arrest patterns, it would be evidenced by a disproportionate impact on minority communities across various dimensions of life. Furthermore, it hypothesizes that the protests following George Floyd's death might have influenced a shift in these patterns, reflecting changes in policing behavior. This research will rigorously analyze data to determine if there has been any significant alteration in arrest trends post-protests, aiming to contribute valuable insights into the dynamics of racial bias in law enforcement and its implications for policy and social justice initiatives. Ultimately, the study aims to deepen the understanding of systemic racial bias in policing.

Data Overview

The research utilizes the Chicago Crime Dataset from the Chicago Data Portal as its primary data source. This extensive dataset, covering 2018 to 2022, provides a comprehensive record of reported crime incidents in Chicago, including details like offense type, location, date, time, and arrest outcomes. This period is crucial, encompassing key socio-political events like the George Floyd protests, and offering insights into crime and law enforcement patterns in the city. The data is sourced from the Chicago Police Department's CLEAR system, ensuring reliability and relevance.

As a secondary source, the research incorporates demographic data from the Chicago Metropolitan Agency for Planning (CMAP). Updated to 2022, CMAP's datasets offer in-depth demographic information for Chicago's diverse communities, including racial composition, median income, and unemployment rates. This integration allows the research to examine crime incidents within the broader context of socio-economic and demographic factors influencing arrest patterns.

The unit of observation in this study is each reported crime incident, allowing for a detailed analysis of arrest patterns and their potential biases. The combination of crime data with demographic information provides a robust foundation for evaluating arrest patterns in Chicago, contributing to informed, data-driven policy recommendations for a more equitable criminal justice system.

Unit of Observation and Key Variables

The dataset employed for this research comprises 931,168 observations, each representing a reported crime in Chicago's 77 community areas from 2018 to 2022. This extensive collection offers a comprehensive resource for analyzing crime and arrest patterns across a significant time frame.

At the core of the analysis are individual crime incidents, each serving as a unit of observation. The primary outcome of interest is the arrest record, denoted by a binary variable that captures whether an arrest was made (1) or not (0) following the reported incident. The dataset encompasses eight distinct categories under 'Primary Type,' [11] representing the main types of crimes reported. These categories were selected based on their frequency within the data, ensuring that the analysis focuses on the most common and impactful crime types within the city. Additionally, the data is segmented into 77 'GEOG' categories, corresponding to the 77 officially recognized community areas in Chicago. This categorization is vital as it allows for the examination of crime and arrest patterns across the diverse tapestry of neighborhoods, each with its unique socio-economic and demographic makeup.

Primary.Type	Arrest_Percentage
ASSAULT	13.225610
BATTERY	17.697270
CRIMINAL DAMAGE	4.788014
DECEPTIVE PRACTICE	3.092904
NARCOTICS	99.309927
ROBBERY	7.262478
THEFT	7.110287
WEAPONS VIOLATION	64.305837

[1] ASSAULT: An unlawful attack or attempt to inflict physical harm on another person, often involving threats or use of force without the consent of the victim.
BATTERY: The intentional and unlawful physical contact or use of force against another person, resulting in injury or harm.
CRIMINAL DAMAGE: Involves the destruction or damage of property without the consent of the owner, including acts like vandalism, graffiti, and defacement.
DECEPTIVE PRACTICE: Encompasses a range of fraudulent activities and crimes of deception, such as identity theft, forgery, fraud, and embezzlement.
NARCOTICS: Relates to offenses involving the illegal possession, use, sale, or distribution of controlled substances or illegal drugs.
ROBBERY: The act of taking property from another person through force or threat of force, often characterized by elements of physical confrontation or assault.
THEFT: Involves the unauthorized taking or stealing of property belonging to another person, ranging from petty theft to grand larceny.
WEAPONS VIOLATION: Crimes related to the illegal possession, carrying, use, or sale of firearms or other weapons, including carrying without a permit or possession by prohibited persons.

The numeric variables in the dataset primarily consist of demographic statistics for the community areas. These variables provide a quantitative backdrop for the analysis, encompassing the percentage of Black, Hispanic, and White populations, among other racial demographics. The summary statistics of these variables are crucial for identifying potential patterns or disparities in crime and arrests across different racial compositions.

From the histograms of these numeric variables, it is observed that some, such as the percentages of Black, Hispanic, and White populations within community areas, exhibit right-skewness. This skewness is indicative of a concentration of certain racial demographics within specific community areas, which could be reflective of the city's residential patterns and may inform arrest patterns.

Data summary

Name	AfterFloydCutoffMay25.202…
Number of rows	3075
Number of columns	21
_______________________	
Column type frequency:	
character	2
numeric	19
________________________	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Primary.Type	0	1	5	18	0	8	0
Predominant	0	1	4	5	0	4	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Year	0	1	2020.00	1.41	2018.00	2019.00	2020.00	2021.00	2022.00	▇▇▇▇▇
Community.Area	0	1	39.05	22.22	1.00	20.00	39.00	58.00	77.00	▇▇▇▇▇
Arrest_Percentage	0	1	25.83	33.16	0.00	3.97	10.29	25.00	100.00	▇▁▁▁▂
Total_Observations	0	1	302.82	444.11	1.00	60.00	161.00	369.50	6146.00	▇▁▁▁▁
Total_Arrests	0	1	50.01	120.97	0.00	5.00	16.00	45.00	2187.00	▇▁▁▁▁
TOT_POP	0	1	35086.47	23037.40	2180.00	18523.00	29429.00	46698.00	101428.00	▇▇▃▂▁
WHITE	0	1	27.82	26.14	0.83	4.40	14.71	48.75	82.72	▇▁▃▂▂
HISP	0	1	26.43	27.35	0.00	5.37	13.18	46.01	90.97	▇▂▁▁▂
BLACK	0	1	36.99	38.35	0.42	2.98	13.26	82.66	96.46	▇▂▁▁▅
ASIAN	0	1	6.46	10.75	0.00	0.34	2.20	8.59	70.66	▇▁▁▁▁
UNEMP	0	1	10.55	6.63	0.67	5.14	8.47	16.10	29.86	▇▅▅▂▁
MEDINC	0	1	59293.56	27727.56	17216.79	37484.19	54757.83	73616.00	133537.46	▅▇▃▂▂
OTHER	0	1	2.30	1.52	0.00	1.14	2.06	3.37	6.86	▇▅▅▂▁
afterfloyd	0	1	0.60	0.49	0.00	0.00	1.00	1.00	1.00	▅▁▁▁▇
PredominantBLACK	0	1	0.38	0.48	0.00	0.00	0.00	1.00	1.00	▇▁▁▁▅
PredominantWHITE	0	1	0.34	0.47	0.00	0.00	0.00	1.00	1.00	▇▁▁▁▅
PredominantHISP	0	1	0.26	0.44	0.00	0.00	0.00	1.00	1.00	▇▁▁▁▃
PredominantASIAN	0	1	0.03	0.16	0.00	0.00	0.00	0.00	1.00	▇▁▁▁▁

Data summary

Name	AfterFloydExcluding2020
Number of rows	2460
Number of columns	21
________________________	
Column type frequency:	
character	2
numeric	19
________________________	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Primary.Type	0	1	5	18	0	8	0
Predominant	0	1	4	5	0	4	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Year	0	1	2020.00	1.58	2018.00	2018.75	2020.00	2022.00	2022.00	
Community.Area	0	1	39.05	22.22	1.00	20.00	39.00	58.00	77.00	
Arrest_Percentage	0	1	25.97	33.11	0.00	3.98	10.14	26.29	100.00	
Total_Observations	0	1	310.81	465.11	1.00	60.00	163.50	377.25	6146.00	
Total_Arrests	0	1	51.49	127.45	0.00	5.00	16.00	48.00	2187.00	
TOT_POP	0	1	35089.38	23035.86	2180.00	18523.00	29429.00	46698.00	101428.00	
WHITE	0	1	27.82	26.13	0.83	4.40	14.71	48.75	82.72	
HISP	0	1	26.43	27.35	0.00	5.37	13.18	46.01	90.97	
BLACK	0	1	36.99	38.35	0.42	2.98	13.26	82.66	96.46	
ASIAN	0	1	6.46	10.75	0.00	0.34	2.20	8.59	70.66	
UNEMP	0	1	10.55	6.63	0.67	5.14	8.47	16.10	29.86	
MEDINC	0	1	59298.61	27740.08	17216.79	37484.19	54757.83	73616.00	133537.46	
OTHER	0	1	2.30	1.52	0.00	1.14	2.06	3.37	6.86	
afterfloyd	0	1	0.50	0.50	0.00	0.00	0.50	1.00	1.00	
PredominantBLACK	0	1	0.38	0.48	0.00	0.00	0.00	1.00	1.00	
PredominantWHITE	0	1	0.34	0.47	0.00	0.00	0.00	1.00	1.00	
PredominantHISP	0	1	0.26	0.44	0.00	0.00	0.00	1.00	1.00	
PredominantASIAN	0	1	0.03	0.16	0.00	0.00	0.00	0.00	1.00	

Demographic and Socio-Economic Indicators

Enriching the crime data is the demographic and socio-economic context gleaned from CMAP's community-level data. Key variables included:

BLACK, ASIAN, HISP, WHITE: Percentage of the total population of the race of an individual community area.

Unemployment (UNEMP): The unemployment rate within each community

Median Income (MEDINC): This variable reflects the median income level in thousands of dollars within the community areas

After Floyd (AfterFloyd): A binary indicator crafted to observe the potential influence of the George Floyd protests on policing and arrest practices. There are two different data frames created to handle the after-floyd variable. One of the data frames is where it differentiates incidents that occurred before (0) and after (1) May 25th, 2020. The other data frame completely excludes the year 2020 from the dataset and flags the reported crimes before 2020 as before (0) and after(1).

Community Area: The specific geographic location within Chicago, with data representing all 77 designated community areas.

Total Population: Reflecting the number of inhabitants in each area, providing a scale for the community demographics.

PreDominantBLACK: A binary indicator of the maximum population percentage in a community area. 1 if the maximum population percentage is BLACK, 0 otherwise

PreDominantWHITE: A binary indicator of the maximum population percentage in a community area. 1 if the maximum population percentage is WHITE , 0 otherwise

PreDominantHISP: A binary indicator of the maximum population percentage in a community area. 1 if the maximum population percentage is HISP, 0 otherwise

PreDominantASIAN: A binary indicator of the maximum population percentage in a community area. 1 if the maximum population percentage is ASIAN, 0 otherwise

Methodology

To explore the research questions effectively, this study adopts a two-fold methodological approach, incorporating both a Baseline Fixed Effects Model and a Binomial Regression Model. This approach is specifically designed to examine the influence of racial bias on arrest patterns, particularly in the context of the aftermath of the George Floyd protests.

The first part of the methodology employs a Baseline Fixed Effects Model, incorporating an 'AfterFloyd' variable. This model is intended to establish a foundational understanding of arrest patterns to community demographics, while specifically accounting for the period following the George Floyd protests. Including the 'AfterFloyd' variable allows for the analysis of changes in arrest rates that may be attributed to the protests, thereby providing a nuanced understanding of the baseline relationships between demographic factors and arrest probabilities.

In addition to the Baseline Fixed Effects Model, the study further utilizes a Binomial Regression Model to assess the likelihood of a reported crime resulting in an arrest. This model is particularly apt for this analysis as the outcome of interest – an arrest – is inherently a binary (yes/no) variable. Moreover, given that arrest data is a count with an upper bound set by the total number of reported crimes for a specific primary crime type in a specific community area in a given year, the Binomial Regression Model is well-suited for analyzing these bounded count outcomes. This model is instrumental in calculating odds ratios, offering a detailed perspective on the probability of arrest outcomes in different demographic contexts. By employing this model, the study aims to quantify the extent to which certain demographic factors influence the likelihood of an arrest following a reported crime.

Together, these two models provide a comprehensive framework for investigating the central research questions. The Baseline Fixed Effects Model offers insights into the general patterns of arrest probabilities across different communities and how these patterns may have shifted post-George Floyd protests. In contrast, the Binomial Regression Model provides a more focused analysis of the specific odds of a reported crime leading to an arrest, taking into account various demographic variables, and specifically accommodating the binary nature and bounded count of arrest data. This dual-model approach is designed to yield a thorough and nuanced understanding of the potential racial biases influencing arrest patterns.

Estimation Strategy for Baseline Regression Analysis

The estimation strategy for the baseline regression analysis in this study employs a fixed effects model that adeptly navigates the intricacies of arrest patterns to crime types, community demographics, and the post-George Floyd era. This model particularly shines in controlling for the distinct, unchanging characteristics inherent to each crime category, allowing for a more accurate understanding of how factors like community demographics influence arrest percentages.

Key to this model are variables like 'Predominant', which categorizes community areas based on their primary racial demographics, and 'afterfloyd', capturing the period after the George Floyd protests. These are complemented by socioeconomic indicators like unemployment rates ('UNEMP') and median income ('MEDINC'). The focus here is on the arrest percentage for each crime in specific community areas across different years, offering a clear picture of how often crimes lead to arrests in these areas.

By concentrating on variations within these groups, the model is uniquely positioned to dissect how changes in community racial composition and significant events like the George Floyd protests correlate with arrest percentages while factoring out the static nature of different crime types.

This study also includes an exploration of another model that incorporates the interaction between the 'Predominant' race variable and the 'afterfloyd' variable. This interaction analysis is pivotal, as it digs into how the relationship between community racial demographics and arrest rates might have altered in the aftermath of the George Floyd protests. By examining this interaction, the study aims to uncover deeper insights into the dynamics of racial biases in arrest patterns, particularly how these biases may have evolved or been influenced by significant societal events.

This approach is crucial for setting the stage for further analysis using a binomial logistic regression. It ensures that the observed relationships between demographic variables and arrest percentages are reliably attributed to these factors, rather than the intrinsic nature of the crimes. In essence, this model lays a foundational and unbiased groundwork for a subsequent, more focused examination of the odds of an arrest in different scenarios, enhancing our understanding of potential racial biases in arrest patterns.

Regression Equation1(Without Interaction):

$$Y_i = \alpha + \beta_1 PredominantBLACK_i + \beta_2 PredominantHISP_i + \beta_3 PredominantASIAN_i + \beta_4 afterfloyd_i + \beta_5 UNEMP_i + \beta_6 MEDINC_i + \gamma_i + \varepsilon_i$$

Regression Equation1(With Interaction):

$$Y_i = \alpha + \beta_1 PredominantBLACK_i + \beta_2 PredominantHISP_i + \beta_3 PredominantASIAN_i + \beta_4 afterfloyd_i + \beta_5 UNEMP_i + \beta_6 MEDINC_i + \beta_7 PredominantBLACK_i * afterfloyd_i + \beta_8 PredominantHISP_i * afterfloyd_i + \beta_9 PredominantASIAN_i * afterfloyd_i + \gamma_i + \varepsilon_i$$

Predictions and Hypotheses

In this research, we're looking into whether there's a bias in how the police make arrests, focusing on race and money matters. We're guessing that in places with more Black residents, there might be more arrests, hinting at possible police bias. Also, we think the time around the George Floyd protests might have changed things. Before the protests (2018-2019), arrests might have been more common, but after the protests started (2020-2022), maybe less so, as police practices faced more public scrutiny and calls for change.

We're also curious about how unemployment and income in an area affect arrests. You might think places with more unemployment or higher income would see more arrests, but it's not that straightforward. We're considering that maybe in poorer areas, the police don't respond to crimes as much, leading to fewer arrests. This part of our guesswork looks at another possible bias where police might not be as active in lower-income areas. We're aiming to get a fuller picture of how different factors like race, income, and public events like protests might play into police arrest patterns.

Results For Equation 1:

With the model utilizing the data frame that excludes the year 2020, the results demonstrate notable trends in arrest percentages with respect to various factors. According to the model output, the variable 'afterfloyd' is associated with a significant decrease in the arrest percentage, as indicated by its negative coefficient and a highly significant p-value ($p < 0.001$). This suggests that the period following the George Floyd protests saw a substantial decline in arrest rates, which could reflect changes in policing behavior in response to heightened public scrutiny and calls for reform.

In terms of community demographics, both 'PredominantHISP' (predominantly Hispanic communities) and 'PredominantBLACK' (predominantly Black communities) have negative coefficients, with

'PredominantHISP' showing a highly significant effect ($p < 0.001$) and 'PredominantBLACK' also being significant ($p < 0.05$). This indicates that, compared to the 'PredominantWHITE' (Predominantly White Group) these community areas saw lower arrest percentages. The variable 'PredominantASIAN' (predominantly Asian communities), however, was not found to be statistically significant.

Socioeconomic variables appear to have mixed effects. Median income ('MEDINC') showed no significant impact on arrest percentages, as its p-value did not indicate statistical significance. On the other hand, unemployment rates ('UNEMP') showed a positive coefficient, but this was not statistically significant either, indicating that within the scope of this model, there was no clear evidence to suggest that unemployment rates in a community had a significant effect on arrest percentages.

Overall, the model accounts for approximately 10.48% of the variability in arrest percentages (R-squared = 0.10483), which, while modest, is not unexpected given the complex nature of crime and arrest dynamics. The model's F-statistic is significant ($p < 0.001$), indicating that the variables, as a set, do contribute to the model's ability to predict arrest percentages.

These results provide an empirical foundation for discussions on the influence of racial demographics and the aftermath of the George Floyd protests on policing practices. They also raise questions about the role of socioeconomic factors in arrest patterns, suggesting that factors beyond median income and unemployment may be at play.

```
Oneway (individual) effect Within Model

Call:
plm(formula = Arrest_Percentage ~ afterfloyd + MEDINC + UNEMP + 
    PredominantHISP + PredominantBLACK + PredominantASIAN, data = joined_df2, 
    model = "within", index = c("Primary.Type"), efffect = "individual")

Unbalanced Panel: n = 8, T = 305-308, N = 2460

Residuals:
    Min.  1st Qu.   Median  3rd Qu.     Max. 
-62.9238  -2.4260  -0.1471   2.3270  88.8016 

Coefficients:
                    Estimate  Std. Error  t-value  Pr(>|t|)    
afterfloyd       -5.1188e+00  3.2283e-01 -15.8560 < 2.2e-16 ***
MEDINC           -2.6835e-06  1.0352e-05  -0.2592   0.79549    
UNEMP             5.5949e-02  5.0189e-02   1.1148   0.26507    
PredominantHISP  -2.7457e+00  5.2659e-01  -5.2142 2.001e-07 ***
PredominantBLACK -1.7511e+00  7.0953e-01  -2.4680   0.01366 *  
PredominantASIAN -2.1293e+00  1.1022e+00  -1.9318   0.05350 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    175130
Residual Sum of Squares: 156770
R-Squared:      0.10483
Adj. R-Squared: 0.10007
F-statistic: 47.7407 on 6 and 2446 DF, p-value: < 2.22e-16
```

Results For Equation 2:

With the inclusion of interaction terms in the model and utilizing a dataset that omits the year 2020, the results suggest a more intricate relationship between arrest percentages and the examined variables. The 'afterfloyd' coefficient is significantly negative, indicating a decrease in arrest percentages in the period following the George Floyd protests.

Regarding racial demographics, 'PredominantHISP' shows a negative coefficient, suggesting lower arrest percentages in predominantly Hispanic communities, with the result being statistically significant. 'PredominantBLACK' communities have a negative coefficient as well, although it is not statistically significant in this model.

Socioeconomic variables, median income ('MEDINC'), and unemployment rates ('UNEMP') do not exhibit a significant effect on arrest percentages, suggesting that these factors may not have a strong standalone impact on arrest rates within the parameters of this model.

The interaction terms add depth to the analysis. The 'afterfloyd:PredominantHISP' interaction term is not statistically significant, which implies that the post-protest changes in arrest percentages are not distinctly different in predominantly Hispanic communities relative to other communities. However, the 'afterfloyd:PredominantBLACK' interaction term is significant, pointing to a more pronounced decrease in arrest percentages in predominantly Black communities after the George Floyd protests. This could reflect a specific change in policing behavior in these communities during the post-protest period.

The lack of significance in the 'afterfloyd:PredominantASIAN' term suggests that the interaction between the post-protest period and being in a predominantly Asian community does not have a statistically discernible effect on arrest percentages.

The F-statistic remains highly significant, confirming that the overall model, with the interaction terms included, provides valuable predictive insights into arrest percentages. These results underscore the complexity of the relationships between race, socio-economic factors, and significant societal events, and how these relationships may have shifted in the time since the George Floyd protests.

```
Oneway (individual) effect Within Model

Call:
plm(formula = Arrest_Percentage ~ afterfloyd + MEDINC + UNEMP +
    PredominantHISP + PredominantBLACK + PredominantASIAN + afterfloyd *
    PredominantHISP + afterfloyd * PredominantBLACK + afterfloyd *
    PredominantASIAN, data = joined_df2, model = "within", index = c("Primary.Type"),
    efffect = "individual")

Unbalanced Panel: n = 8, T = 305-308, N = 2460

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-62.39988  -2.42278  -0.20557   2.43202  89.32378

Coefficients:
                              Estimate  Std. Error t-value  Pr(>|t|)
afterfloyd                 -4.0744e+00  5.5605e-01 -7.3274 3.172e-13 ***
MEDINC                     -2.6836e-06  1.0345e-05 -0.2594  0.795340
UNEMP                       5.5949e-02  5.0153e-02  1.1155  0.264725
PredominantHISP            -2.0144e+00  6.7394e-01 -2.9890  0.002827 **
PredominantBLACK           -8.5678e-01  8.0559e-01 -1.0635  0.287641
PredominantASIAN           -2.3392e+00  1.5134e+00 -1.5456  0.122325
afterfloyd:PredominantHISP  -1.4627e+00  8.4214e-01 -1.7369  0.082532 .
afterfloyd:PredominantBLACK -1.7886e+00  7.6489e-01 -2.3384  0.019445 *
afterfloyd:PredominantASIAN  4.1987e-01  2.0759e+00  0.2023  0.839731
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    175130
Residual Sum of Squares: 156360
R-Squared:      0.10721
Adj. R-Squared: 0.10136
F-statistic: 32.596 on 9 and 2443 DF, p-value: < 2.22e-16
```

Estimation Strategy for Binomial Logistic Regression:

For the binomial logistic regression analysis in this study, the estimation strategy is designed to evaluate the likelihood of a reported crime resulting in an arrest. The dependent variable in this model is a binary outcome: whether a reported crime did (1) or did not (0) lead to an arrest. This model choice is suitable because it addresses the yes/no nature of arrest outcomes and accounts for the fact that arrest counts are bounded by the total number of reported crimes for a specific primary crime type in a specific community area in a given year.

The logistic regression model will utilize predictors similar to those in the fixed effects model, including variables that reflect community demographics (such as 'PredominantRACE' variables), socio-economic indicators (like 'UNEMP' and 'MEDINC'), and a variable ('afterfloyd') to capture the period after the George Floyd protests.

In addition to these predictors, the binomial logistic regression can also include interaction terms to explore how the relationship between demographic factors and the likelihood of arrest might change in the period after the protests. This model will allow us to calculate odds ratios, which express the likelihood of arrest as a function of the independent variables. An odds ratio greater than 1 indicates a higher likelihood of arrest associated with the predictor, while an odds ratio less than 1 indicates a lower likelihood.

This estimation strategy is particularly powerful for examining the nuances of how reported crimes translate into arrests and for exploring potential systemic biases. By focusing on the binary outcome of arrest, this strategy can directly assess the influence of race and socioeconomic factors on policing practices, providing a clearer understanding of whether certain communities are disproportionately targeted or neglected in the arrest process following reported crimes.

Binomial Logistic Regression Equation3(Without Interaction):

$$\text{logit}(pi)=\log(pi/1-pi) = \alpha + \beta_1\text{PredominantBLACK}_i + \beta_2\text{PredominantHISP}_i + \beta_3\text{PredominantASIAN}_i + \beta_4\text{afterfloyd}_i + \beta_5\text{UNEMP}_i + \beta_6\text{MEDINC}_i + \gamma_i + \varepsilon_i$$

Binomial Regression Equation4(With Interaction):

$$\text{logit}(pi)=\log(pi/1-pi) = \alpha + \beta_1\text{PredominantBLACK}_i + \beta_2\text{PredominantHISP}_i + \beta_3\text{PredominantASIAN}_i + \beta_4\text{afterfloyd}_i + \beta_5\text{UNEMP}_i + \beta_6\text{MEDINC}_i + \beta_7\text{PredominantBLACK}_i * \text{afterfloyd}_i + \beta_8\text{PredominantHISP}_i * \text{afterfloyd}_i + \beta_9\text{PredominantASIAN}_i * \text{afterfloyd}_i + \gamma_i + \varepsilon_i$$

Results For Binomial Logistic Regression:

	Dependent variable:			
	cbind(Total_Arrests, Total_Observations - Total_Arrests)			
	(BinReg-ex2020)	(BinReg-ex2020int)	(BinReg-in2020)	(BinReg-in2020int)
afterfloyd	-0.628***	-0.538***	-0.522***	-0.449***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
PredominantHISP	-0.117***	-0.073***	-0.104***	-0.064***
	p = 0.000	p = 0.0001	p = 0.000	p = 0.0003
PredominantBLACK	-0.046**	0.009	-0.047***	0.011
	p = 0.015	p = 0.635	p = 0.006	p = 0.542
PredominantASIAN	-0.149***	-0.077	-0.153***	-0.062
	p = 0.0005	p = 0.151	p = 0.0001	p = 0.249
MEDINC	0.00000***	0.00000***	0.00000***	0.00000***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
UNEMP	0.0002	0.0003	0.002*	0.002*
	p = 0.889	p = 0.825	p = 0.075	p = 0.059
factor(Primary.Type)BATTERY	0.326***	0.325***	0.328***	0.328***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
factor(Primary.Type)CRIMINAL DAMAGE	-1.133***	-1.133***	-1.121***	-1.121***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
factor(Primary.Type)DECEPTIVE PRACTICE	-1.572***	-1.569***	-1.621***	-1.620***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
factor(Primary.Type)NARCOTICS	6.626***	6.626***	6.825***	6.824***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
factor(Primary.Type)ROBBERY	-0.693***	-0.694***	-0.682***	-0.683***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
factor(Primary.Type)THEFT	-0.782***	-0.780***	-0.777***	-0.775***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
factor(Primary.Type)WEAPONS VIOLATION	2.594***	2.602***	2.577***	2.583***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
afterfloyd:PredominantHISP		-0.116***		-0.080***
		p = 0.00000		p = 0.0001
afterfloyd:PredominantBLACK		-0.144***		-0.118***
		p = 0.000		p = 0.000
afterfloyd:PredominantASIAN		-0.184**		-0.183**
		p = 0.031		p = 0.016
Constant	-1.630***	-1.670***	-1.681***	-1.722***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	2,460	2,460	3,075	3,075
Note:				**p<0.1;**p<0.05;***p<0.01*

The results from the binomial logistic regression analyses present a multifaceted view of the factors influencing the likelihood of a reported crime resulting in an arrest. The models, which include variations with and without interaction terms, and with certain models excluding the year 2020 or using a cutoff at May 25th, 2020, show consistent patterns across different specifications.

In every model configuration, the 'afterfloyd' variable is associated with a significant decrease in the likelihood of arrest, indicating that following the George Floyd protests, the odds of a reported crime leading to an arrest have declined. This outcome is robust across the different model structures, suggesting a substantial change in arrest probabilities in the aftermath of these events.

The demographic variables of 'PredominantHISP', 'PredominantBLACK', and 'PredominantASIAN' communities show varying degrees of significance across models. In particular, communities predominantly composed of Hispanic and Black residents are consistently associated with a lower likelihood of arrest across most models, with the 'PredominantBLACK' variable showing significance in all model variations. This could indicate a potential shift in policing strategy or behavior in these communities. The 'PredominantASIAN' variable, however, does not show a consistent significant relationship with arrest likelihood.

Socioeconomic factors such as median income ('MEDINC') and unemployment ('UNEMP') generally do not exhibit a significant association with the likelihood of arrest in the models without interaction terms. However, when interaction terms are introduced, these relationships may shift, highlighting the complex nature of these socioeconomic influences on arrest patterns.

The primary type of crime, represented by various dummy variables (e.g., 'factor(Primary.Type)NARCOTICS', 'factor(Primary.Type)THEFT'), shows strong significance in influencing arrest likelihood. Certain crime types, like narcotics offenses, are highly predictive of arrest, whereas others, like deceptive practices and theft, are associated with a lower likelihood of arrest.

In models that include interaction terms, the post-protest period's influence on arrest likelihood varies by the predominant racial demographic of the community area. The interaction terms between 'afterfloyd' and the demographic variables ('afterfloyd:PredominantHISP', 'afterfloyd:PredominantBLACK', 'afterfloyd:PredominantASIAN') are significant in most models, indicating that the effect of the George Floyd protests on arrest likelihood is differentially experienced across communities with different racial compositions.

The consistency of these results across models that exclude 2020 entirely, as well as those that consider data up to May 25th, 2020, reinforces the robustness of the findings. Overall, these results paint a picture of a policing landscape that has been notably affected by the events following the George Floyd protests, with arrest probabilities shifting in a way that varies according to racial demographics and crime types.

Weaknesses of the Approach

One primary limitation is the temporal dimension. The 'afterfloyd' variable is treated as a binary indicator, potentially oversimplifying the complex temporal shifts that may have occurred before and after the George Floyd protests. Such an approach may not capture the nuances of how these shifts unfolded over time and may misrepresent the sustained or changing impacts of the protests.

Another challenge arises from potential biases in crime reporting. The models rely on reported crimes, and any systemic biases in reporting or recording these crimes could affect the accuracy of the models' outcomes. For example, if certain types of crimes are underreported in particular communities, the models may underestimate the true likelihood of arrests in those areas.

There's also the concern of unobserved heterogeneity. Despite the inclusion of several control variables, there may still be unmeasured factors specific to individual communities that influence arrest rates, which the models do not account for. This unobserved heterogeneity can confound the results, making it difficult to draw precise conclusions about the factors driving arrest likelihood. The ability of the models to establish causation is also limited. Observational studies inherently face challenges in distinguishing correlation from causation, and while the models identify patterns and associations, they cannot confirm that one factor causes another.

The models' inclusion of interaction terms, while enriching the analysis, also introduces complexity in interpreting the results. Disentangling the individual and combined effects of demographics and the post-protest period can be difficult, and misunderstandings could arise when policymakers or practitioners attempt to apply these findings. Moreover, logistic regression models come with certain statistical assumptions. If the assumptions about the independence of observations and the linearity of relationships in the log odds are not met, the models may yield biased estimates, leading to inaccurate interpretations.

Results derived from the dataset and context used in this research may not necessarily apply to different locations, crime types, or times. Additionally, the exclusion of data from the year 2020 in some of the model variations could mean important information about that year's arrest patterns is not considered, potentially skewing the results. Furthermore, changes in policing policies or practices that might have occurred alongside the protests are not explicitly accounted for in the models. Such changes could independently affect arrest probabilities and are difficult to isolate and measure.

Finally, the socioeconomic variables used, such as median income and unemployment rates, are broad indicators. These measures might not capture all the relevant economic factors that influence policing, such as wealth distribution, economic mobility, or informal employment, which could lead to an incomplete understanding of the socioeconomic factors at play. Recognizing these limitations is crucial for a balanced understanding of the study's implications. Future research could improve upon these areas by using more sophisticated statistical methods, incorporating richer datasets, or exploring broader contexts.

Limitations of the Methodology:

In addition to the previously discussed strengths and limitations, another important consideration in this study's methodology is the potential reporting bias in the dataset, given that the unit of observation is crime reported. This bias can manifest in two significant ways, both of which could impact the analysis and its findings.

Firstly, there is the issue of variability in crime reporting across different crime types. Certain types of crimes might be underreported compared to others, potentially skewing the data and the analysis. For example, some crimes like theft or minor property offenses might not always be reported to the authorities, while more severe crimes such as narcotics offenses or violent crimes are more likely to be officially recorded. This discrepancy in reporting rates can lead to an overrepresentation of certain types of crimes in the dataset, potentially influencing the analysis of arrest probabilities and the perceived correlation with demographic variables.

Secondly, the propensity to report crimes can vary significantly across different community areas, influenced by factors such as trust in law enforcement, community norms, or the perceived seriousness of the crime. Some communities might be more inclined to report certain types of crimes, while others might underreport due to various socio-cultural reasons. For instance, communities with a strained relationship with law enforcement might be less likely to report crimes, or there might be underreporting of domestic incidents in certain areas due to cultural or social factors. This variation in reporting practices can introduce another layer of bias in the dataset, affecting the accuracy of the analysis when correlating crime rates with demographic variables.

These aspects of reporting bias - both in terms of the type of crime and the community-specific propensity to report - add another dimension to the limitations of the study. They underscore the need for caution in interpreting the results, as the data may not fully capture the actual incidence of different types of crimes across various communities. Recognizing and accounting for these potential biases is crucial for a robust analysis of arrest probabilities and for concluding the systemic patterns of bias in policing practices in Chicago.

Conclusion

To sum it all up, we took a close look at whether arrests in Chicago might be leaning unfairly against certain races or depending on whether folks are rolling in dough or barely scraping by. We were especially curious about whether the huge wave of protests after George Floyd's death changed how often people from different backgrounds ended up arrested.

Our investigation was a bit like detective work. We used two methods to hunt for clues: one looked at arrest rates across different neighborhoods, and the other focused on the chances of someone getting arrested after a crime was reported. The findings were pretty eye-opening. It looks like, after the protests, there were generally fewer arrests in neighborhoods with mostly Black and Hispanic residents. And this pattern stuck even when we considered other things like income levels and unemployment.

But we’ve got to admit, our methods weren’t perfect. We had to simplify some complex situations, which might’ve missed some of the finer details. Plus, we can't be sure if every crime was reported or if they were reported the same way in every neighborhood, which could tilt our results a bit. Also, our findings aren't one-size-fits-all. They’re about Chicago during a very specific and intense time. What we found here might not hold true in another place or at a different time.

What we can say is that there seems to be a change blowing through the city—maybe a little less quick to arrest in some communities. It’s a glimpse at how big events like the George Floyd protests can send ripples through a city and touch the way policing is done. It shows us that looking at how fair policing is isn't just about digging into the numbers. It's about real people and the neighborhoods they call home. We're hoping this research gets people talking and might even help steer us toward a justice system that's more even-handed for everyone.

Works Cited

Fryer Jr, R. G. (2019). An empirical analysis of racial differences in police use of force. *Journal of Political Economy*, *127*(3), 1210-1261.

Legewie, J., & Fagan, J. (2019). Aggressive policing and the educational performance of minority youth. *American Sociological Review*, *84*(2), 220-247.

Edwards, F., Lee, H., & Esposito, M. (2019). Risk of being killed by police use of force in the United States by age, race–ethnicity, and sex. *Proceedings of the national academy of sciences*, *116*(34), 16793-16798.

Cassella, C., Epp, D., Fredriksson, K., Roman, M., & Walker, H. (2022). The George Floyd effect: How protests and public scrutiny change police behavior in seattle.

Data

1. Data.gov The Home of the U.S. Government's Open Data
   https://catalog.data.gov/dataset/crimes-2001-to-present
2. Community Data Snapshots (Chicago Metropolitan Agency For Planning)
   https://datahub.cmap.illinois.gov/maps/CMAPGIS::community-data-snapshots-raw-data-2014-2022/about

Use of ChatGPT:

ChatGPT was used for grammatical errors and for paraphrasing some of the paragraphs to make the writing more formal. No Idea was generated from ChatGPT.