THE UNIVERSITY OF CHICAGO

ESSAYS IN LABOR ECONOMICS AND POLICE ACCOUNTABILITY

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE IRVING B. HARRIS

GRADUATE SCHOOL OF PUBLIC POLICY STUDIES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

BOCAR ABDOULAYE BA

CHICAGO, ILLINOIS

JUNE 2018

To Robert J. Lalonde, for his kindness, devotion, and endless support.

Pour Aimé Césaire

“J’ai l’impression d’être ridicule
dans leurs salons
dans leurs manières
dans leurs courbettes
dans leur multiple besoin de singeries

J’ai l’impression d’être ridicule
avec tout ce qu’ils racontent
jusqu’à ce qu’ils vous servent l’après-midi
un peu d’eau chaude
et des gâteaux enrhumés

J’ai l’impression d’être ridicule
avec les théories qu’ils assaisonnent
au goût de leurs besoins de leurs passions
de leurs instincts ouverts la nuit
en forme de paillasson

J’ai l’impression d’être ridicule
parmi eux complice parmi eux souteneur
parmi eux égorgeur
les mains effroyablement rouges
du sang de leur ci-vi-li-sa-tion”

Damas, Léon-Gontran, Solde , Pigments, Présence Africaine Éditions, 1972

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

I thank my family in Senegal and France for their love, support, trust, and patience during this long period away from home. I pray that I will repay the sacrifices you made for me. Cynthia Cook-Conley, Robert J. Lalonde, and Amy Claessens provided much needed emotional support throughout my time at the University of Chicago. I am grateful to my uncle Oumar Kane, who mentored and coached me during this long journey. I am extremely grateful to Roman Rivera, Dan Kowalski, Julio Rivera, Jo Mamou, Gerardo Manzo, Leah Calvo, Erin Patrick, and Rajab Ghazzaoui for always being there for me.

I am greatly indebted to my dissertation committee, Professors Robert Lalonde, Steve Levitt, Christian Hansen, Jens Ludwig, and Scott Ashworth for their comments, critiques, encouragement, and support on this project. A special mention goes to Jeffrey Grogger, for our insightful and detailed discussions about policing and crime. I am grateful to Dan Black, Chris Blattman, Tom Coleman, and Austin Wright for guidance, support, and mentorship. I will forever be thankful to Damon Jones, Kerwin Charles, and my Chairs for being patient with me and being a source of inspiration. I also owe a substantial debt to Pablo Montagnes, John Ham, Philip Eisenhauer, and John MacDonald who introduced me to tools and topics covered in this dissertation between 2011 and 2017. Many thanks to my Professors in Montreal and Vancouver for instilling within me a love of learning: Phil Merrigan, Pierre Lefebvre, Steve Ambler, Pierre-Carl Michaud, and Vadim Marmer. I would like to thank my peers at the Harris School who provided wonderful intellectual support. Special thanks to Jocelyn Hare and Cynthia Cook-Conley provided useful insights to have access to the data.

Special thanks go to Craig Futterman, Chaclyn Hunt, Jamie Kalven, Rajiv Sinclair, Sam Stecklow, Andrew Fan, Maira Khwaja, Kahari Black, Trina Reynolds Tyler, Harry Backlund, Henri Dumas, Emma Herman, Jason Schumer, the Invisible Institute, and City Bureau for providing the data, trusting me, and answering numerous questions.

In my seven years living in Chicago, my experience here was made richer because of a

group of young boys and girls from the South Side who met me at Fiske Elementary, Gary Comer Youth Center, Youth Police Project, and Blackstone Bicycle Works. Whether I was tutoring them in mathematics and French, or just hanging out, I learned valuable life lessons from these resilient and wise young people. I count our time together as some of my most important moments in Chicago.

During these long years away from home, I have been blessed to receive the support of Brice Garcia-Mochales, Abdoulaye Magassa, la famille Ghariani, Redouane Nafi, Jean-Sébastien Chevrier, Princess Allen, Huda Oweis, Kareem Haggag, Abdulkareem Agunbiade, Noumaan Shamsi, Vicky Stavropoulos, Sebastian Gallegos, Maria Rosales-Rueda, Ruthie Coffman, Galen Pierce-Gardner, Lamine Danfadkha, Sawdatou Diallo, Henri Lauzière, Carlos Gotay, Liz Curnutte, Milvia Rodriguez-Gomez, Maggie Newman, Charles Bouis, Ricardo Avillez, Mouna Mahdi, Moussa Sy, Nico Petty, and Mourad Aloulou.

The first two chapters of this dissertation benefitted from valuable insights provided by officers from the "Brigade anti-criminalité" of the French National Police, the University of Chicago Police Department, and the Chicago Police Department. My often uncomfortable interactions with these agencies, both as a young boy and an adult, provided useful human capital and motivation for the development of this thesis. This research helps me understand how difficult the profession of a police officer is. This work was also, sadly, motivated by my observation that there is little understanding and relevant research on police accountability. There is a lack of research about the death of civilians via police interactions in countries such as the U.S. and France (one of my home countries). I have tried my best to bring my modest contribution to this discussion. I send my thoughts and prayers to the families of Laquan McDonald, Zyed Benna, Bouna Traoré, Adama Traoré, my friend Michael Cohen, and many others.

The Harris School and NORC provided generous financial support. All errors are mine.

ABSTRACT

There is a growing debate over police accountability in the United States, and, in particular, a growing interest from the public in police use of force and incidents of officer misconduct. Surprisingly, there is limited empirical evidence on whether or not law enforcement misconduct affects the level of crime, citizens' perceptions of the police, and citizens' criminal records (and subsequently their education and labor market outcomes).

The first two chapters of this dissertation focus on issues related to police accountability while the final chapter delves into understanding the methodologies behind the identification of treatment effects using duration models.

In Chapter 1, "Going the Extra Mile: the Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago," I study civilian willingness to pay to file a complaint against a police officer. As officers fulfill their duties to protect and serve the community, they often receive complaints from civilians with whom they have interacted. This setting makes policing fraught with agency problems. I use new, detailed administrative data to study the costs and benefits associated with filing a complaint against the police in Chicago. I exploit the fact that allegations without affidavits signed by complainants are considered null. I use variation in people's distance to the oversight agency where they sign affidavits as a way to study the effect of civilian oversight on policing. An administrative change of location of the reporting center provides a quasi-experimental setup for the identification strategy. A difference-in-difference analysis suggests that a one-standard deviation increase in traveling distance to the reporting center decreases the likelihood of a signed complaint by 6.2 percent for allegations of constitutional violations and 16.3 percent for complaints alleging the police's failure to provide service. In non-white residential areas, higher injury rates due to use of force and a higher level of force used per arrest were observed as distance from the reporting center increased. Individuals who benefit most from oversight are those with lowest valuation of complaining. I simulate counterfactual scenarios under a policy that would reduce the cost

of signing the complaint. This policy would largely increase the number of investigations and the sustained rates for failure to provide service complaints in the most violent police districts. On the other hand, for allegations of constitutional violations, this policy would reduce sustained rates overall and marginally increase the number of investigations. This research sheds light on the tradeoffs that arise when increasing the cost of reporting police misconduct.

While this paper documents the role of civilians' oversight on police performance and their willingness to complain based on their incentives, it is also important to understand police officers' behaviors. As society delegates to police the authority to enforce laws, including the right to use force when needed, the question raised is whether or not police use force more than necessary.

Chapter 2 addresses this question in joint work with J. Grogger, "The Introduction of Tasers and Police Use of Force: Evidence from the Chicago Police Department," where we study police use of force. In March 2010, the Chicago Police Department changed its Taser policy, issuing the weapons to patrol officers instead of largely restricting their use to sergeants. We used that policy change to obtain difference-in-difference estimates of how the availability of Tasers affected the types of force employed by police, the total number of use-of-force incidents, injury rates per incident, the total number of injuries, and the race distribution of civilians involved in use-of-force incidents. The policy change initially led to a large increase in the use of Tasers, with limited substitution from other types of force. After a period of retraining, substitution between Tasers and other types of force, both greater and lesser, increased. Police injuries fell, but neither injury rates nor the number of injuries to civilians were affected. Contrary to popular arguments in support of Tasers, we find there is no evidence that Tasers led to a reduction in police use of firearms.

Chapter 3 focuses on understanding the methodologies behind the identification of treatment effects using duration models. In joint work with J. Ham, R. Lalonde, and X. Li, “On the Identification, Estimation, and Use of Dynamic Treatment Effects,” we estimate the effect of endogenous training participation on transitions in and out of employment for disadvantaged women in the Job Training Partnership Act (JTPA) study. Decomposing the effect of training on employment into its effects on transitions in and out of employment has the potential to develop more effective programs. We also consider a potentially serious identification problem that arises when individuals do not undertake training immediately, and we propose a test to shed light on this problem. We find that this problem is not important in our context. JTPA classroom training substantially reduced unemployment durations, and thus it complements programs that increase employment durations.

CHAPTER 1

GOING THE EXTRA MILE: THE COST OF COMPLAINT FILING, ACCOUNTABILITY, AND LAW ENFORCEMENT OUTCOMES IN CHICAGO

Bocar A. Ba

1.1 Introduction

Law enforcement's mission in the United States is to both protect society and preserve community trust (Lum and Nagin [2017]). By the very nature of their work, police officers interact with a broad cross-Section of society— bystander civilians, victims of a crime, and offenders. On occasion, civilians may seek to file a complaint against an officer. Both the police and civilians have private information about their interaction with each other, and mixed incentives to reveal it truthfully: this makes policing fraught with agency problems. However, understanding the causes and consequences of alleged officer misconduct is crucial because society delegates to the police the authority to enforce its laws, including the right to use force when needed.

Although there is a growing public concern regarding police behavior, there is not yet research that credibly attempts to quantify the willingness of citizens to hold officers accountable by filing complaints against the police. This paper studies the costs and benefits of civilian oversight of the police.

Information surrounding allegations of misconduct against the police is often transmitted via a complaint process, as is common in other bureaucratic organizations (Prendergast [2003]). Yet, complainants and others often wonder whether their actions matter. Members of the public can expose an officer's unlawful behavior (excessive use of force, false arrest, or verbal abuse) or mistakes (failure to provide service) through the complaint process. Beyond the issue of effective management of officer misconduct, however, it is also important to understand the ways in which the exercise of accountability (through the civilian complaint process) affects the quality of policing. While on one hand, civilian complaints can deter police from engaging in misconduct in the first place, officers may also react to complaints by cutting down on their policing to avoid complaints (Prendergast [2001, 2002], Shi [2009]).

Police officers are required to act within the confines of the US Constitution and other relevant law. If a police officer violates the Constitution or other applicable law during an encounter with a civilian, there are limited ways the violation can be detected. An oversight agency or the department may have a system to routinely audit police behavior without provocation, though the effectiveness of these audits has not been empirically analyzed (Walker [2007]; Hickman and Poore [2016]). A civilian who believes their Constitutional rights have been violated has two options: to sue law enforcement agencies or to report the violation to an oversight agency. Civil lawsuits are very costly for local governments which often bear the financial burdens of liability (Schwartz [2014, 2016])[1]. Complaints are generally easier to file than a lawsuit, and are the most readily available form of civilian feedback a police department can access (Walker and Macdonald [2008]). Mandated reform agreements have included provisions for "improved citizen complaint systems." To give a sense of scope, the Department of Justice[2] has required sweeping overhauls of forty law enforcement agencies since 1997[3]. Finally, yet importantly, when citizens do not trust their police to act within the law, they are less likely to contribute to public safety by reporting crimes and

1. The payouts in misconduct cases for the ten largest local police departments was $1.02 billion from 2010 to 2014 (Elinson and Frosch [2015])

2. The Violent Crime Control and Law Enforcement Act of 1994, 42 U.S.C. § 14141 allows the U.S. Justice Department (DOJ) to review the practices of law enforcement agencies that may be violating people's federal rights.

3. See details here

testifying against suspected criminals (Tyler and Fagan [2008], Desmond et al. [2016], Lum and Nagin [2017]).

A primary obstacle to the empirical study of police misconduct has been the lack of readily available public data that links allegations of misconduct to demographic information about the complainant and administrative information about the officers (Bureau of Justice Statistics). Another challenge is to find a measure of cost of reporting misconduct that varies exogenously in order to make causal arguments. There is very little relevant academic research that directly tries to assess the influence of civilian oversight on police misconduct, arrests, crime, and other measures of police performance. Last but not least, there is also little empirical work on the tradeoff between enforcing the law and building community trust (Manski and Nagin [2017]).

I address these questions by using novel datasets from the Chicago Police Department on allegations of officer misconduct, use of force, and crime data in the city from January 2011 to July 2014. I take advantage of an administrative change in Chicago that changed the cost of registering an allegation of misconduct against a police officer. Chicago is unusual in that complainants may initiate complaints remotely, but, in order for the complaint to be investigated, the complainant must sign an affidavit in person at a single location within the city. At the end of 2011, the city changed the location. This altered the cost of signing the complaint in a manner that varied geographically. For residents who live close to the original location, the cost of signing the affidavit rose; for residents who live close to the new location, it fell. I use these changes to study the ways in which the varied costs of filing a complaint affects several aspects of policing and civilian willingness to complete a complaint.

The analysis considers two types of civilian complaints: those alleging a failure to provide service and those alleging a violation of an individual's constitutional rights during an attempt by an officer to enforce the law (search, verbal abuse, and excessive use of force). This division in the classification of complaints enables me to distinguish civilians who filed a complaint because they desired help from the police (e.g.: potential victim of a crime) from civilians who are treated as potential suspects by the police and feels that their constitutional rights are violated. One can reasonably believe that these two types of complainants have different incentives when alleging an officer misbehaved.

This paper makes the following contributions. First, I provide the first empirical estimates of civilian willingness to complain against the police. I find that increased distance to the reporting center deters civilians from completing their allegation of police misconduct. Second, I use distance to the oversight agency as a proxy for cost of monitoring the police, since areas that are farther away from the reporting center are less likely to be investigated because of the higher likelihood of missing affidavits (see Figure 1.5). This enables me to study the effect of civilian oversight on complaints, use of force incidents, and crime outcomes. I find that non-white residential areas that are harder to monitor because of their increased distance from the reporting center are subject to more aggressive policing, and that Black residential areas report lower rates of index crimes. Finally, I estimate a simple model of a complainant's decision to complete the affidavit and investigator 's decision to sustain the complaint provided that the affidavit requirement is met. This model allows me to compute the civilian's valuation of their complaint and to perform counterfactual scenarios where I evaluate the effect of removing the affidavit requirement on the number of officer that are going to be held accountable of their actions and on the rates at which complaints are sustained. Individuals who benefit most from oversight are those with the lowest valuation of their complaint.

I start the analysis by using a difference-in-difference design based on the administrative change of location for the oversight agency. I show that civilians who live farther from the oversight agency are much less likely to sign the affidavit, which is necessary for an investigation of the misconduct and full feedback to the police department. In fact, each additional 3.6 miles an incident (i.e a one standard deviation) occurs from the reporting center relates to a decrease in the likelihood of a signed affidavit by 8.6%. Civilians are less sensitive to distance when the allegation is related to a serious offense from the officer: a standard deviation in traveling distance to the reporting center relates to a decrease in the likelihood of a signed affidavit by 6.2% for serious complaints and 16.3% for allegations related to failure to provide service.

In this context, using distance[4] as a proxy for monitoring price derives from the idea that interactions between officers and civilians that occur farther away from the reporting center are less likely to be investigated because of the civilian's higher opportunity cost of traveling to complete the complaint by signing the affidavit. I employ a reduced form approach to study the effect of distance on complaint outcomes, crimes, arrests, and use of force patterns. This approach sheds the light on the effect of monitoring price on police performance. This paper is not the first to study the effect of oversight on police performance. Prendergast [2001, 2002], Shi [2009], and Heaton [2010] study the impact of police scandals on police performance. A primary difference between my design and those studies is that I study oversight in an environment free of major scandals involving the police. A scandal could potentially alter both the behavior of civilians and the police. An advantage of my research design is that the administrative change of the oversight agency is unknown to the majority of the officers and the civilians, at least in the short run, and thus not a catalyst for major behavioral changes.

Beats[5] that are farther away from the oversight agency are more difficult to monitor because civilians are less likely both to complain and to complete their complaint by signing the affidavit. Moreover, for incidents that occurred in Hispanic residential areas, distance does not seem to have a statistically significant effect on the probability of a signed affidavit.

4. Or other traveling cost such as traveling time by car or public transit

5. A police beat is a tract of land designated for police patrol. Currently, the Chicago Police Department has divided the city into 22 geographical Police districts. Each of the 22 police districts currently has between 9 and 16 beats.

The effect of distance is small compared to white and Black residential areas, and non-significant). Data aggregated at the beat level also suggests that complaints about incidents that occurred farther away from the oversight agency have a higher likelihood of being sustained. The results are, however, only significant for Black residential areas.

The number of use of force incidents increases with distance in non-white residential areas. In particular, civilians in Hispanic neighborhoods are more likely to be subject to a high level of force the farther away the incident occurs from the oversight agency. Incidents in Black residential areas have a higher number of civilian injuries if they occurred farther away from the oversight agency. I also find that, in the short run (12 months after the location change of the oversight agency), residential areas that are more difficult to monitor had a significantly higher number of use of force reports and a higher reported level of force per arrest. The results for the complaint and use of force outcomes suggest that non-white areas that are more difficult to monitor are more likely to be subject to more aggressive policing.

For the complaints and use of force outcomes, I use a placebo test to supplement the regression evidence on robustness of prior location change and differential trends. The placebo test replaces the actual location change that occurred in December 2011 with a location change in May 2011 and then reruns the difference-in-difference analysis from January to November 2011 to show that nothing mechanical drives the results. I find that the results of the falsification tests are relatively small and statistically non-significant; therefore, I conclude that the estimates for the complaints and use of force outcomes are not related to events that would have potentially occurred prior to the true location change.

One justification of aggressive or proactive policing is that it can help reducing crime (Manski and Nagin [2017]). Prendergast [2001, 2002] suggest that more police oversight yields to worse crime outcomes, due in part to less aggressive policing. I find that more oversight yields to lower reported crimes for the whole city, but that the results seem to be driven by Black residential areas. The effect of reporting cost is non-significant for non-Black residential areas. The results for arrests go in opposite directions for index and non-index crimes: areas that are harder to monitor have a higher number of arrests for index crimes, but a lower number of arrests for non-index crimes. Unlike with the use of force and complaint results and because placebo tests fail for some crime outcomes of interest, it is difficult to conclude, given my design, that civilian oversight causally impacts the number of crimes, arrests, and clearance rates.

To perform a counterfactual scenario and compute civilians' valuation of their complaint, I estimate a model of civilian willingness to complete their complaint, accounting for the investigator's decision to sustain the complaint. This model provides insight into the effect of reducing the cost of completing a complaint and makes out-of-sample predictions. Using maximum likelihood, the model estimates both the probability that the civilian signs the affidavit, and the probability that the investigator sustains the complaint, assuming that the affidavit is signed. The model assumes that the unobserved heterogeneity of the investigator and the beat affect both the civilian and investigator decision. To ease the interpretation of the parameter estimates, I compute civilian willingness to pay to complete the complaint using Capps et al. [2003]. Closed form solution of the willingness to pay can be derived using the fact that the error term from the civilian's utility function are type 1 extreme value (Small and Rosen [1981]).

For the counterfactual scenario, I evaluate the effects of removing the traveling requirement to the oversight agency to complete the complaint, while holding everything else constant. Under the alternative scenario, the share of investigated complaints would increase by 5.11% and 37.58%, respectively, for serious and failure to provide service allegations. This alternative policy would significantly increase the number of investigated complaints that affect potential victim of a crime (failure to provide service), especially in violent neighborhoods. Overall, this alternative policy would, on average, raise the share of sustained complaints for failure to provide service by 8.1%, but lower the share of sustained complaints about serious allegations by 9.77%.

Prendergast [2003] and Shi [2009] provide theoretical models of officer behavior when police departments want to minimize crime, minimize errors (misconduct), and minimize wage expenditures. One of the key elements of the officer utility function is oversight (when an officer receives a complaint), which is a function of the probability of investigation, the probability of a sustained complaint, and the penalty for the officer if found guilty. By estimating civilians' willingness to complete their complaint, the proposed model provides the first empirical estimates of civilian oversight. This research shows that although minority civilians have a high valuation of their complaint, there is little chance that their allegation will be sustained.

Although the public debate on police misconduct mainly focuses on police use of force (Fryer [2016]), this paper documents that increasing the cost of reporting officers' alleged wrongdoing mainly hurts civilians requesting the help from the police in the city's most violent neighborhoods. However, there is an inverse relationship between addressing complaints that allege a constitutional violation and complaints that allege a failure to provide service. A higher level of contact between officers and civilians—i.e., more aggressive policing—should intuitively lead to lower crime rates. A side effect of more aggressive policing, however, is an increase in the likelihood of officer use of force. A greater rate of incidents of force though leads to an increase in the other type of complaint: those alleging violations of rights. Meanwhile, minimal levels of policing will only exacerbate the number of complaints alleging a failure to provide service. Thus, a single policy change cannot effectively minimize both type of complaints simultaneously.

Non-white civilians are more likely to live farther away from the oversight agency and are less likely to complete a complaint. The mechanisms of accountability—namely the complaint process—begin to fail as distance from the oversight agency increases. Officers assigned to these far-flung areas, who are statistically less likely to be disciplined as a result of a complaint than their colleagues, can operate with a greater degree of impunity. Moreover, the levels of violence in these distant neighborhoods seems to demand an aggressive response from the police—a response that leads to higher levels of use-of-force incidents. As a result of these dynamics, the cost of police violence in Chicago falls disproportionately on the shoulders of the city's most vulnerable residents: non-white civilians living in poor and violent neighborhoods with the least access to/ability to engage in the processes of oversight [mechanisms of accountability]. These dynamics may suggest why police can afford to be more aggressive with non-white civilians.

The paper is organized as follows. Section 1.2 briefly reviews related literature. Section 1.3 provides background information on the complaint process in Chicago. Section 1.4 describes the administrative data used for the analysis. Section 1.5 presents the empirical methods and the results. In Section 1.6, I develop a model of civilian willingness to complete her complaint accounting for the investigator decision to sustain the complaint if the affidavit is signed. Section 1.7 and 1.8 estimate civilians' valuation of their complaint and presents counterfactual scenarios. Section 1.9 concludes.

1.2 Motivation and relation to the literature

This paper aims to shed some light on the issue of officer misconduct by quantifying civilians' willingness to complain against the police and by analyzing the effect of oversight on other measures of police performance, such as use of force and misconduct outcomes. To ground this work, I reference three main areas of research on policing: that which investigates the relationship between race and the practice of policing, that which looks at the factors governing public perceptions of the police, and that which examines the impact of increased civic oversight on policing.

First, the economic literature on crime and policing that originated with the seminal work of Becker [1968] mostly focuses on the objective of protecting society through the deterrence of crime (Glaeser [1999]). Chalfin and McCrary [2017] provides a detailed overview on the effect of police, punishments, and work on crime from the last two decades. This paper analyzes the effect of oversight on traditional measures of police performance taken from the crime literature, such as crime rates, arrest rates, and clearance rates.

Although crime outcomes are important measures of police performance, the economic literature gives little importance to other aspects of policing such as its impact on community trust, use of force, and protection of individual rights. Lum and Nagin [2017] and Manski and Nagin [2017] argue that crime prevention and community relations should be the predominant feature of policing objective functions. Lum and Nagin [2017] suggest some steps to recalibrate organizational incentives within agencies, including taking into account citizen's complaints. However, Walker (2007) argues that there is no credible social science research that studies police misconduct and tries to propose solutions to reduce civilian complaints or police use of force. This paper aims to shed some lights on the issue by quantifying citizens' willingness to complain against the police and by analyzing the effect of oversight on other measures of police performance such as use of force and misconduct outcomes.

This article relates to a number of articles that demonstrate that police reduces enforcement efforts when faced with increased oversight, a dynamic that eventually yields higher levels of crime. Prendergast [2001, 2002] examines time series data from the Los Angeles Police Department from 1994 to 2001; he finds that as a result of increased oversight, officers were less likely to pursue "aggressive" policing in factors such as use-of-force, officer-involved shootings, assaults on officers, and arrest rates. Shi [2009] and Heaton [2010]] exploited quasi-experimental events, such as public scandals that generated increased media attention and judicial scrutiny of police agencies, to examine the effect of increased oversight on officer conduct. Shi [2009] studied the effect of the heightened scrutiny as a result of the 2001 Cincinnati riots and found that communities with higher percentage of Black people experienced even greater reductions in arrests. Heaton [2010] notes a similar trend with data on motor vehicle theft in New Jersey, suggesting that a public racial-profiling scandal contributed to the significant decline of minority arrests for motor vehicle theft. Because the identification strategy of these papers rely on an increase in oversight generated by an increased in media attention and judicial scrutiny (DOJ) after a scandal[6], it is difficult to: (1) assess the role of civilians in police oversight independently of the other monitors (DOJ and other institutions), and (2) understand the causes and consequences of the misconduct of individual officers independently of the stigmatization of the whole force. This paper investigates the impact of civilian oversight on police performance in an environment that only focuses on allegedly "poorly" behaved officers, rather than an environment that villainized and scrutinized the whole force. Moreover, in response to a scandal, civilians may file more complaints, officers might reduce their efforts to police, or both. The potential of altered behavior from both civilians and the police makes it difficult to identify the effect of oversight on policing (e.g: Desmond et al. [2016]).

A different strain of the literature documents public perceptions of the police. Studies have found public opinion of police is influenced by factors like race, neighborhood conditions, the media, public scandal, and personal experience with law enforcement personnel (Van Craen and Skogan [2015]). Within neighborhoods, Weitzer [1999] finds that class is an important factor in explaining citizens' perception of police misconduct by drawing on survey data from Washington, DC. Kane [2002] finds that structural disadvantage,[7] population mobility, and racial tension yielded higher rates of officer misconduct and argues that social disorganization makes it more difficult for citizens to organize against abusive/poor policing. With regards to race, minorities and liberal citizens tend to have a more negative

6. In those three studies, the DOJ was running an investigation on those police Departments at some point during the sampling period of those studies.

7. The author defines structural disadvantage as a function of percentages of persons in poverty, households receiving public assistance, male unemployment, adult unemployment, low educational attainment, youth, and female-headed households with children.

perception of the police compared to their white and conservative counterparts (Weitzer and Tuch [2004]). This literature provides a starting point to describe citizen sentiment on police misconduct and understand the factors that are correlated with those opinions. However, those studies do not provide any information on: (1) citizen willingness to report misbehavior through complaints and their perception about law enforcement using exogenous source of variations, and (2) how complaint outcomes might vary depending on citizen characteristics and incentives. This paper aims to address these two points.

Finally, this paper contributes to the literature on race and policing in America. Researchers have tried to assess the level of racial disparities in police interactions with the public. Numerous studies document the presence of discrimination in policing for traffic stops (Knowles et al. [2001], Anwar and Fang [2006], Grogger and Ridgeway [2007], Ritter and Bael [2009]), stop-and-frisk (Gelman et al. [2007], Ridgeway and MacDonald [2009], Goel et al. [2016]), traffic tickets (Anbarci and Lee [2014], West [2015], Goncalves and Mello [2017]), and uses of force (Fryer [2016]). However, some of these studies assume that the behavior of the civilian is exogenous. One can argue that civilians change their behavior in response to their belief that police is biased (e.g: Kalinowski et al. [2017]); if civilians do alter their behavior, it is difficult to identify the effect of race on policing. This study documents the racial disparity that exists in civilians' decision to report police misconduct. I document that there are racial disparities in the cost of complaining for civilians. This cost is a function the civilian's willingness to pay to complete a complaint and the probability that the complaint is sustained. If one believes that the cost of policing is a function of detecting officer error, as in the theoretical models of Prendergast [2001] and Shi [2009], these results have important implications for the contemporary practice of policing.

1.3 Background

1.3.1 Institutional context

Since 2007, Chicago's oversight agency, the Independent Police Review Authority (IPRA), has collected all allegations of misconduct against Chicago Police Department (CPD) members. Allegations originate from the public or from other officers in the Department. Complaints are classified according to one of the twenty main categories[8]. This research is primarily interested in categories of complaint that involve civilians: failure to provide service, use of force, verbal abuse, arrest or locked up procedures, and search. The process of filing a complaint has two main stages:

1. The complainant initiates his/her complaint by phone, in person at the oversight agency's location, by mail, with any CPD supervisor at any district station, or over the internet.

2. The complainant must then sign a sworn affidavit to certify that the allegation is true and correct. Since the end of 2010, the oversight agency has required that the complainant must physically appear at the oversight agency office to sign the affidavit[9]. At this point, the complaint is filed and investigated. The State of Illinois requires that any person making an allegation of misconduct against a Chicago police officer complete this step. In the event the complainant does not sign the affidavit, the investigation is terminated and the allegations are classified as "not sustained".

The first stage to file a complaint is straightforward and easy for the complainant because of the number of alternatives available and the option of not being physically present to file a complaint. However, the second stage requires the complainant to be physically available to

8. See here for details about allegation categories

9. See Rapid Pilot Program, page 19: The oversight agency changed its intake procedure in order to make the investigative process more efficient.

sign the affidavit at the oversight agency location. This requirement may be difficult to meet for individuals who live far away from the oversight agency or for those who are working during weekdays. Hence, individuals might fail to meet the affidavit requirement because they have a high opportunity cost of signing the affidavit (i.e. commuting time, commuting fees, and forgone wages if working).

Once a complaint is received and the affidavit is signed, an investigator is assigned to conduct a comprehensive investigation of each complaint. When the investigation is completed the allegations are classified as “sustained,” “not sustained,” “exonerated,” or “unfounded.” For the remainder of the paper, I classify “exonerated” and “unfounded” as “not sustained.” Additionally, if an accused officer cannot be identified, I classify the complaint as “not sustained.” The officer does not need to answer the investigators until the affidavit requirement is met. In the event that no affidavit is received, the investigation is terminated, classified as “not sustained,” and no record of the complaint remains on the officer’s disciplinary history. A complaint with a signed affidavit remains on the officer’s disciplinary history for five years after the complaint is issued. At the end of the investigation, the oversight agency sends a letter to both the complainant and the CPD members reflecting the findings and recommended discipline, if any. The police department is in charge of enforcing the final discipline determined for each allegation. It is important to stress that the complainant does not receive any compensation at the end of the process. If the complainant wants to sue the member of CPD, it has to be done independently from the complaint process through the oversight agency.

This research focuses on the period between January 2011 and July 2014 for two reasons. First, complaints filed prior to 2011 are excluded from the analysis because the complainant was not required to be physically present at the oversight agency and the investigator could travel to the complainant. For those cases, it is difficult to isolate civilian willingness to travel to complete the complaint. Secondly, since late March 2014, lists of past complaints against Chicago police officers are available to the public for review. These complaints were made public as the result of a lawsuit, Kalven vs. the City of Chicago & the Chicago Police Department, under the Freedom of Information Act (FOIA). The data was made available to a nonprofit organization, Invisible Institute, in July 2014 and the Chicago police union, the Fraternal Order of Police (FOP), notified its members in August 2014[10].

1.3.2 Oversight agency location

On December 19, 2011, the oversight agency moved from a rented space on the South Side of Chicago to the Near West Side of the City. The new site of the oversight agency is located in a building owned by the City of Chicago.

The former South Side location was near the Chicago police headquarters and accessible by two subway lines and an expressway. The new location is accessible by bus but not by subway. Hence, the change of location increased the travel time to file a complaint for someone residing on the South Side, and decreased the travel time for someone living on the Near West Side. The two location are about 6.3 miles apart from each other using the Manhattan Distance metric and 4.9 miles by euclidian distance.

Figures 1.1-1.3 depict the oversight agency before and after its location change with respect to the distribution of race, income, and complaints within the city. Figure 1.1 indicates that the oversight agency moved closer to neighborhoods with a high proportion of Hispanic and white residents, but farther away from the black population of the South Side. Figure 1.2 suggests that the North Side of Chicago tends to be wealthier than the South and West Sides. Thus, the oversight agency moved closer to more affluent neighborhoods. Finally, Figure 1.3 shows that the oversight agency moved away from neighborhoods with a high concentration of allegations of police misconduct.

10. See Kalven Court Decision

1.4 Data

1.4.1 Data sources

This Section describes the datasets that are used for the empirical analysis. My primary analysis focuses on complaints, crime data, and civilian injuries data spanning January 2011 to December 2014. I supplement those datasets with distance and travel times from each beat centroid to the oversight agency locations using data from Google Distance API. I merge complaints and crime data with demographics and socioeconomic indicators by census block obtained from the U.S. Census Bureau 2010-2014 American Community Survey's five-estimates.

Complaints Data The complaint data contains all recorded allegations of misconduct filed against an officer from 2001 to 2016. The allegations can come from another officer (an internal complaint) or from a civilian (an external complaint). Each complaint contains information on involved police officers, complainant demographics, and incident location. The data includes both internal and external complaints filed against officers and does not account for appeals or subsequent hearings. I do not have information on the residence of the complainant; however, the location of the incident might provide some useful information, since it can serve as a loose proxy indicator of the complainant's residence.

The dataset provides information about the final finding of the investigation. The outcome variable has the following classification: disciplined, not-sustained, open- investigation, sustained, and unknown. The finding variable provides an explanation of the outcome and has the following descriptions: exonerated, missing affidavit, not sustained, sustained, unfounded, unknown, sustained. I also supplement the finding variable by identifying the complaints with missing officer information. For the purpose of this study, I restrict attention to general conclusions from the investigation rather than their recommendations. I restrict my attention on civilian complaints with identified officers and final outcomes (i.e I do not consider unknown outcomes or unknown officers). Because I am interested in whether or not the complaint was investigated, I use another classification for the purpose of my research question: no affidavit, not sustained (not sustained, exonerated, or unfounded), and sustained (sustained or disciplined). Additional information about the classification is provided in Appendix 1.10.2

Use of force and civilian injuries Information on civilian injuries comes from the 2004-2016 Tactical Response Reports (TRRs). The TRR is a form that officers are required to complete for incidents involving the use of force. The form provides detailed information about the incident, type of force, subject demographics, and involved officer information (demographics, unit assignment, injury, etc). Moreover, a TRR is also necessary if the subject is injured or allegedly injured.

Chicago crime data The crime data reflects reported incidents that occurred in Chicago from 2001 to present. Each incident contains information about the crime, such as location, date, type, and whether or not an arrest was made. For each beat, I compute the monthly number of total incidents and arrests for index crimes (violent and property crimes), and non-index crimes (less serious offenses). I do not include beats that are located outside of Chicago. Beats that do not have any residents are also removed from the sample. The resulting sample contains a total of 265 beats with 43 months of observational data.

Traveling cost I use the proximity of the incident location to the oversight agency as a proxy for the opportunity cost of filing a complaint. I consider spherical distance, driving time, and travel time by public transportation. For each beat, I compute the travel times (car and public transit) from beat centroid to the oversight agency locations using data from Google Distance API in December 2016[11].

11. I use the "sphdist" routine from Stata to compute the spherical distance.

American Community Survey (ACS) American Community Survey (ACS): Using the 2010-2014 ACS data, I compute block level aggregates to characterize demographics and neighborhood demographics. The numbers are adjusted to 2014 dollars by Social Explorer. I average the block median income at the beat level. I also compute the share of Black, Hispanic, white, and other race at the beat level to identify whether or not the beat has a majority of Black civilians.

1.4.2 Sample construction

In order to study the effect of allegations of misconduct on police performance, I construct two datasets: complainant's data and beat level data.

Complainants data The analysis focuses on incidents that occurred between January 2011 and July 2014. As discussed in Section 1.3.1, this period is appropriate for studying civilian willingness to complain against the police. Because misconduct records were not accessible to the public and there was no major scandal, the behavior change of police officers or civilians is less of a concern. I exclude from the sample incidents with missing location or if they occurred outside of Chicago, including the suburbs. The resulting sample contains a total of 15,039 complaints for the analysis of the Raw data provided in the Appendix 1.11. The Raw complainants data is helpful in order to describe the type of misconduct that officers are accused of. Among those complaints, 6,763 come from civilians (e.g. failure to provide service, use of force, verbal abuse, arrest, locked up procedures, and search). Table 1.16 provides details about the sample construction.

Beat Level Data The beat level data helps us understand the effect of signed affidavits and traveling cost on different policing outcomes. I restrict the analysis on allegations of misconduct that are coming from civilians. For each beat, I merge the monthly crime data to the use of force and injuries data, ACS data, and complaint data.

1.4.3 Descriptive analysis

1.4.3.1 Civilian complaints data

Table 1.1 reports the frequency distribution of civilian complaint categories by complainant race for incidents that occurred between January 2011 and July 2014. That sample only consider com- plaints with known final outcome (i.e., sustained, not sustained, and those with a missing affidavit). About 70.5 percent of the complaints are related to use of force, verbal abuse, arrest, locked up procedures, and searches. This suggests the presence of a large degree of heterogeneity among complainants. Hispanic and Black civilians are more likely to complain about events that have greater consequences for their criminal records, but are also seen as a positive indicator for officer performance, whereas white civilians are more likely to complain about issues related to quality of the service provided by the police (e.g, response time for a 911 call or a failure to arrest a suspect). Accounting for 29.5 percent of the total complaints, failure to provide service (FPS) complaints represents significant portion of the complaints. It is important to note that FPS complaints are most likely filed by complainants who are potentially victims of a crime or who seek police help, whereas serious allegations are filed by complainants who are potentially viewed as criminal by police officers.

Table 1.2 presents some summary statistics of civilian complaints by distance bins from the over- sight agency location. About 71 percent of the complaints are serious and 29 percent are complaints related to failure to provide service. Incidents that occurred closed to the oversight agency, i.e., less than three miles away, have a higher share of failure provide service (33 percent). On the other hand, incidents that occurred three to six miles from the oversight agency, have a higher share of serious complaints (73 percent). The average travel time to sign the affidavit is about 21 minutes by car and 50 minutes by public transportation. Incidents that occurred less than three miles from the oversight agency have a more affluent population. The average hourly wage within the three-mile radius of the new oversight agency is about $26, versus the $18-$20 of the rest of the city. Blacks, Hispanics, whites, and individuals of unknown race respectively represent 68 percent, 12 percent, 15 percent, and 5 percent of the complainants. The share of Black complainants in- creases as the distance from the oversight agency increases, whereas whites have the highest share of complainants living less than three miles from the oversight agency. The highest share of Hispanic complainants live less than nine miles from the oversight agency. Looking at age characteristics, we find that approximately 14 percent of the complainants are less than 30 years old and 49 percent are between 30 and 49 years old. The median age of the accused officers is slightly higher for incidents that occurred less than 3 miles from the oversight agency: 42 years old, versus 40-41 years old for incidents that occurred more than three miles from the agency. Between 31 and 33 percent of the incidents involved a non-regular police officers, i.e. higher ranked or specialized officers. The racial distribution of the accused officers varies geographically. Incidents that occurred farther away from the oversight agency have a higher share of accused Black officers. The share of accused Hispanic and white officers is higher for incidents closed to the oversight agency. The summary statistics on the racial distribution of civilians, and on where officers get their complaints (weakly) suggest that there is some (voluntary or not) geographical allocation of officers based on their race and ethnicity. As a confirmation to Figure 1.1, which depicts racial segregation in Chicago, Hispanic and white civilians are more likely to live closer to the oversight agency than Black civilians. Table 1.2 suggests the need to condition for complainant demographics and incident characteristics, as distance to the oversight agency is clearly correlated with other complainant characteristics that may themselves affect an individual's decision to sign the affidavit.

To explore the effects of distance and location change on the probability of a completed complaint, Figure 1.4 plots the probability of a signed affidavit by distance to the oversight agency at both the old and new reporting centers. According to the graphic, the share of signed affidavit was higher at the old location (50.7%) than at the new location (43.7%). Moreover, at the old location, complaints resulting from incidents that occurred twelve or more miles or more from the reporting center were less likely to be completed (i.e., to contain a signed affidavit) than those resulting from incidents that occurred less than twelve miles away. For the new location, complaints are on average less likely to have a signed affidavit the farther away the incident occurred from the oversight agency. This graph suggests that the location change had an impact on the likelihood of signed affidavit. Because the oversight agency moved farther away from the areas with a high concentration of complaints (see figure 1.3), one can infer that the increased distance and/or traveling cost had a deterring effect on the willingness of individual civilians to complete their complaint by signing the affidavit.

In order to provide additional evidence of the negative relationship between traveling cost and the number of completed complaints, I then analyze how travel distance to the oversight agency affects an individual's willingness to file a complaint and to sign the affidavit. Figure 1.5 plots the probability that a complaint will include a signed affidavit by distance to the oversight agency. Figure 1.5 plots the probability of a signed affidavit and includes a histogram of the number of complaints by traveling distance. This graphic displays the expected result: incidents that occur farther away from the oversight agency are less likely to have a signed affidavit. The number of complaints declined with distance, for incidents that occurred more than three miles away from the oversight agency. Figure 1.6 plots the demeaned signed affidavits rate as well as two residualized versions. The first residualized signed affidavits rate using complainants characteristics. The second accounts for accused officers and incidents characteristics. Consistent with what I will present in the regression specifications below, Figure 1.6 suggests that the overall relationship between the probability of signed affidavits and travel distance is not very sensitive to the inclusion of these covariates.

1.4.3.2 Beat level data

In order to understand the context in which complaints are filed, it is important to account for the environment in which police officers are working. The aggregation of the data at the beat level helps us to understand how the cost of traveling to the reporting center affect misconduct, use of force, and policing.

Table 1.3 presents some summary statistics by beats from January 2011 to July 2014. I present the data for the whole city, beats with a majority[12] of Black residents, beats with a majority of Hispanic residents, and beats with a majority of white residents. According to that table, compared to non-Black residential areas, Black residential areas are, on average, farther away from the reporting center (about 2-2.20 miles farther away). White residential areas have a lower traveling time by car and transit compared to member of minority groups. Income disparity and racial segregation that are depicted in Figures 1.2 and 1.1 are also confirmed by Table 1.3. White residential areas are more affluent than those of other racial and ethnic groups. For instance, the average income in white beats is more than twice the average income in Black beats. Finally, the average number of police officers in the district is higher in non-White areas.

1.5 Empirical analysis

This Section analyses the effects of traveling costs on the likelihood of having a full investigation, provided that the officer is identified by the individual when they filed the complaint. I use distance and traveling times by car and public transit from the incident to oversight agency location as a proxy for the cost of filing a complaint against a police officer. The change of location of oversight agency after December 2011 provides us a quasi-experimental design to study the opportunity cost of filing a complaint, and more specifically, to sign the

12. Majority: share of a racial or ethnic group (Black, Hispanic, or White) larger than 50 percent in a beat

affidavit. I exploit the fact that the cost of signing the affidavit varies by police beat, and that this cost exogenously changes when the location of the oversight agency change.

The economic literature often uses distance or traveling time as a proxy for opportunity cost or a measure of the individual's willingness to travel to get education (Card [1995], Duflo [2001], Dahl [2002]), receive healthcare treatment (Einav et al. [2016], Gowrisankaran et al. [2014]), migrate (Black et al. [2015]), consume (Davis et al. [2016]), or apply to disability programs (Deshpande and Li [2017]). Provided that the accused officers are identified, a complaint is fully investigated if the affidavit is signed. Thus, I restrict the analysis to closed allegations with known final outcomes and consider the following final findings: sustained, non-sustained, and missing affidavit.

1.5.1 Complainants data

1.5.1.1 Setup

I start my analysis by examining the relationship between distance traveling costs (distance or traveling time) and the probability to sign the affidavit, I estimate the following linear probability model for individual i in beat b and in month t:

$$Sign_{ibt} = C_{bt}\beta + X_i^{'}\delta + \alpha + \gamma_t + \varepsilon_{ibt} \tag{1.1}$$

The variable $Sign_{ibt}$ is equal to 1 if the outcome is sustained or not sustained and to 0 otherwise. The vector X_i contains a set of controls for complainant age, race, gender, and information about the incident. The time fixed effects and the constant are given by γ_t and α. The error term ε_{ibt} is assumed to have non-constant variance and standard errors are clustered at the police district and community area level. The vector of traveling costs is captured by C_{bt} which is specified as a linear function. The coefficient of interest, β, captures the effect of traveling cost on probability to sign the affidavit. The estimated coefficients from equation 1.1 might suffer from an omitted variable bias due to unobserved factors that are correlated with traveling cost. The bias can come from permanent differences between police beats, as well as biases from different trends over time. In order to address the bias due to unobserved heterogeneity, I estimate the following difference-in-difference model:

$$Sign_{ibt} = C_{bt}^{'}\beta + X_i^{'}\delta + \alpha_b + \gamma_t + \varepsilon_{ibt} \tag{1.2}$$

such that equation 1.2 augments model from equation 1.1 by including police beat fixed effects that are given α_b. The vector of traveling costs is captured by $C_{bt}^{'}$ which is specified as either linear or step functions. I am interested in coefficients β which capture the effect of distance on the probability of signing the affidavit. The fact that the cost is time varying at the beat level and because of the exogenous location change of the oversight agency over time, enables me to identify the causal effect of β. The identification strategy also relies on the notion that the site change of the oversight agency locations and traveling cost should not be correlated with complainant and incident characteristics. This would be consistent with the idea that traveling cost varies quasi-randomly from incident to incident given the change of location of the oversight agency.

1.5.1.2 Results

Table 1.4 presents the results from equation 1.4. Panel A displays the estimates for all complainants. Panel B, C, and D present the results for Black, Hispanic, and white complainants respectively. Separating the results by race enables me to document the heterogeneity of the results by complainant's race. Columns 1-3 present the results for all type of civilian complaints, columns 4-6 only accounts for serious civilian complaints, and columns 7-9 present the results for complaints related to failure to provide service. Columns 1-3 present the results for all type of civilian complaints, columns 4-6 only accounts for serious civilian complaints, and columns 7-9 present the results for complaints related to failure to provide service. For Panel A, the average probability that an individual will sign the affidavit is 46 percent, 52 percent, and 31 percent for the three categories. A standard deviation (3.6 miles) increase in distance reduces the probability of signed affidavit by 2.3 percent -3.5 percent depending on the type of complaint[13]. Overall, Table 1.4 suggests that distance does not have a statistically significant impact on the likelihood to sign the affidavit. The effect is marginally significant when pooling failure to provide service and serious complaints together (columns 1 and 3). According to Panel B C, and D suggest that distance does not have a statistically significant effect on the probability to sign the affidavit for most of the specifications.

Panel A of Table 1.5 displays the estimates from the equation 1.2 which accounts for beat's unobserved heterogeneity of the beats. Incidents that occurred farther away from the oversight agency are less likely to have a signed affidavit. A standard deviation increase in distance reduces the probability of signed affidavit by 8.6 percent , 6.2 percent , and 16.3 percent for the whole sample, serious complaints, and failure to provide service complaints respectively. The results are statistically significant at the 5 percent level when controlling for officers and complainants' characteristics. Table 1.22 finds that the results for equation 1.2 hold when using a nonlinear specification of distance (step functions). The coefficients from Table 1.5 are two to three times larger relative to the results from Table 1.4. This suggests that not accounting for beat fixed effect biases down, in absolute value, the effect of distance on the probability to sign the affidavit.

Panel B, C, and D of Table 1.5 display the estimates from the equation 1.2 by complainant's race. On average, Black complainants are less likely to sign the affidavit compared to white and Hispanics. For instance, for all types of complaints, the average probability to sign the affidavit is 44 percent, 51 percent, and 55 percent for the all the Blacks, Hispanics, and white complainants respectively. The effect of distance is negative for most of the

13. The interpretation is given by $\frac{\partial y}{\partial C} \cdot \frac{SD(C)}{\bar{y}} = \beta \cdot \frac{SD(C)}{\bar{y}}$, based of coefficient β, standard deviation of the traveling cost ($SD(C)$), the dependent variable (y), and mean of the dependent variable ($\overline{y}$)

specifications, but it is statistically significant only for Black complainants for all type of complaints and for failure to provide service. For all types of complaints, a standard deviation increase in distance reduces the probability of signed affidavit by 10.6 percent and 5.6 percent for the Black and Hispanic complainants respectively. For white complainants, the effect of distance on the probability to sign the affidavit is positive but small and not statistically significant. For complaints related to failure to provide service, Black complainants are almost two times less likely to sign the affidavit compared to whites (25 percent for Blacks vs 48 percent for whites). Moreover, a standard deviation increase in distance reduces the probability of signed affidavit by 40.3 percent for allegations related to failure to provide service from a Black complainant. For incidents related to serious misconduct, Blacks are also more sensitive to distance compared to the other ethnic or racial groups. A standard deviation increase in distance reduces the probability of signed affidavit by 6.2 percent, 1.9 percent, and 1.2 percent for the Black, Hispanic, and White complainants respectively. The results for serious complaints are statistically significant when pooling the racial and ethnic groups (Panel A), but not significant when one separates the results by the complainant's racial or ethnic group. Once again, not accounting for beat fixed effects potentially attenuates the effect of distance on the probability to sign the affidavit.

In order to understand the short term effect of the policy, Table 1.6 presents the effect of distance on the probability of signed affidavit during a symmetrical period of time around the the location change of the oversight agency in December 2011. I look at the effect of distance in the six, nine, and twelve-month periods around the move of the reporting center. In order to have sufficient degrees of freedom for the estimation and to compute the standard errors, I estimate equation 1.2 with district fixed effects, rather than beat fixed effects[14]. This approach is necessary when conditioning on complainant race (Panel B, C, and D). Panel A of Table 1.6 suggests that distance has a statistically and economically significant impact for

14. In other context of the paper, I also adopt this approach when the specification does not have enough observation to perform inference.

all types of complaints and for complaints related to failure to provide service. For instance, from January 2011 to December 2012, a standard deviation increase in distance reduces the probability of signed affidavit by 7.5 percent and 20.9 percent for all type of allegations and for complaints related to failure to provide service, respectively. The results are smaller and not statistically significant for allegations involving serious misconduct: during the twelve-month period around the location change, a standard deviation increase in distance reduces the probability of signed affidavit by 4.6 percent . Panel B, C, and D suggest that the results are driven mostly by Black complainants who are less likely to complete a complaint that is related to failure to provide service.

1.5.1.3 Threats to validity and robustness

This Section briefly discusses supplemental evidence that supports the assumptions underlying the results and robustness tests. Appendix 1.13 explores the robustness of the results across different specifications. I briefly summarize the main conclusions.

I use a placebo test to supplement the regression evidence on robustness to prior location change and differential trends. The placebo test replaces the actual location change that occurred in December 2011 with a location change in May 2011 and then reruns the main specification from January to November 2011, equation 1.2, to show that nothing mechanical drives the results. Overall, according to Table 1.7, the results are relatively small and statistically non-significant; therefore, these falsification tests reveal that the estimates are not related to events that would have potentially occurred prior to the true location change.

Overall, for the effect of traveling costs on positive findings, the results remain reasonably robust across a variety of alternative specifications. I find that distance from the incident location to the oversight agency affects complainants' willingness to complete their complaint against a police officer. First, the coefficients of interest stay stable including and excluding different sets of controls, such as the demographics of the complainants and incident characteristics. Secondly, I show that the results are robust in response to a variety of travel cost measures. The first type is related to the opportunity cost of time that is captured by distance, travel time by car, and travel time by transit from the oversight agency to the incident location. The results using the alternative travel costs (time by car and transit) regressions can be found in Tables 1.26 and 1.27.

I find that the results for equation 1.2 hold when using alternative functional forms for: travel costs (step function), error term (logit), and district fixed effects rather than beat fixed effects. This indicates that the results are not sensitive to functional form. For nonlinear traveling distance, I display the coefficients in Table 1.22 and 1.23. Table 1.25 reports the results using a logit with district fixed effect specification, rather than a LPM with beat fixed effect. The results are similar to the main specification and confirm that a higher traveling cost reduces the likelihood of traveling to sign the affidavit.

Unless specified, standard errors are clustered at the district and community area level. As an additional robustness check, I also perform the analysis with standard errors clustered at the beat level and police district level. Those alternative specifications does not affect the inference of the analysis.

1.5.2 Beat level data

1.5.2.1 Setup

This Section perform reduced-form analysis to investigate the effect of the location of the oversight agency on various outcomes. Similarly to equation 1.2, I estimate a the following DID model at the beat b and month t level such that:

$$y_{bt} = C'_{bt}\beta + X'_{dt}\delta + \alpha_b + \gamma_t + \varepsilon_{bt} \tag{1.3}$$

The time and beat fixed effects are given by γ_t and α_b. The error term ε_{bt} is assumed to have non constant variance and standard errors are clustered at the police district and community area level. The vector of traveling costs is captured by C'_{bt} which is specified as either linear or step functions. The vector X_{dt} contains a set of controls for the police district characteristics over time such as average monthly salary and number of officers in the district.

The analysis at the beat level provides an opportunity to study the effect of monitoring (distance or traveling time) on police performance. Recall that police agencies are supposed to both prevent crime and preserve trust in the communities they are serving (Lum and Nagin [2017]), thus I do not limit my analysis to crime outcomes to measure police performance. Due to data limitation, I only consider the following set of outcomes, y_{bt}: complaints, use of force and civilian injuries, and crimes. I am interested in the effect of traveling cost on outcomes that measures both the extensive margin (per 1,000 capita) and the intensive margin. Because of the presence of zeros in the data (months without complaints, signed affidavits, crime, arrest, injuries, and use of force in a given beat), I chose to analyze per 1,000 capita levels, rather than percentage changes or logs.

For the complaint outcomes, I consider the number of complaints per 1,000 capita, the number of complaints with signed affidavit per 1,000 capita, the share of complaints with signed affidavit, and the share of complaints that are sustained. This set of outcomes helps our understanding of the relationship between civilian oversight and officers' likelihood to be investigated and penalized.

For the use of force and civilian injuries outcomes, I consider the number of TRRs per 1,000 capita (i.e. the number of use of force reports per 1,000 capita), the number of civilian injuries per 1,000 capita, whether or not there was incident with high level of force (interaction involving use of Taser or firearm), and the number of incidents involving reported force per arrest. This set of outcomes help us understand the relationship between civilian oversight and police use of force.

Finally, I analyze the relationship between civilian oversight and crime outcomes (index and non-index). For each type of crime, I consider the number of reported offenses per 1,000 capita, the number of reported offenses with an arrest per 1,000 capita, and the clearance rates which I define as the number of reported crimes with an arrest over the number of reported crimes.

To account for the fact that Blacks, Hispanics, and whites do not live in the same neighborhoods (because of the effects of segregation and income disparity), I perform the analysis separately for those groups. For each set of outcomes, I report the results for the whole city, beats with a majority of Black residents, beats with a majority of Hispanic residents, and beats with a majority of white residents. For the interpretation of the results, I rely on the standard deviations of traveling distance which are 3.5, 3.6, 2.4, and 3.4 the whole city, beats with a majority of Black residents, beats with a majority of Hispanic residents, and beats with a majority of white residents respectively.

1.5.2.2 Complaints

Table 1.8 presents the monthly effect of distance from the incident location to the oversight agency on complaint outcomes at the beat level: number of complaints per 1,000 capita, signed affidavit per 1,000 capita, share of signed affidavit, and share of sustained complaints.

This table suggests that beats that are farther away from the oversight agency are more difficult to monitor because civilians are: less likely to complain and less likely to complete their complaint. O A standard deviation in traveling distance to the reporting center causes a decrease in the number of complaints per 1,000 capita and signed affidavits per 1,000 capita by 11.6 percent and 26.1 percent . The share of signed affidavits drop by 10.8 percent for beats that are a standard deviation away from the oversight agency. The results are statistically and economically significant for beats that are majority Black: a standard deviation in traveling distance to the oversight agency causes a decrease in the number of complaints and signed affidavits by 7.2 percent and 17.9 percent, respectively. Beats that are predominantly composed of Hispanic and white residents tend to show similar patterns for the share of signed affidavits, but the effect is only statistically significant for shares of signed affidavits in white residential areas.

Moreover, columns (7) and (8) of Table 1.8 suggest that the rate at which complaints are sustained increases with distance, but the results are only marginally significant for beats that are predominantly composed of Black residents. In beats that are majority Black or Hispanic: a standard deviation in travel distance to the oversight agency relates to a 39.8 percent to 28.0 percent increase in sustained rate. A standard deviation in travel distance to the oversight agency relates to a 10.4 percent increase in sustained rate for white residential areas.

Table 1.31 finds qualitatively similar results when focusing on the short term effect of the policy nine and twelve months around the location change of the oversight agency.

This set of results strongly suggests that beats that are farther away from the oversight agency are more difficult to monitor. Civilians are on average less likely to report misconduct and to complete their complaint if they are farther away from the reporting center.

1.5.2.3 Use of force and civilian injuries

Table 1.9 presents the monthly effect of distance from the incident location to the oversight agency on use of force outcomes at the beat level: number of incidents reporting use of force (TRR), any use of high level of force (interaction involving use of Taser or firearm), number of incidents involving reported force per arrest, and number of civilian injuries. Overall, the results are not statistically significant when looking at the results for the whole city (Panel A) and when controlling for district characteristics.

Panel A from this Table 1.9 suggests that distance does not have a statistically significant impact on use of force outcomes. However, civilians in Hispanic residential areas are more likely to be subject to use of force if they are farther away from the oversight agency. For civilians in Hispanic residential areas, a single standard deviation increase in traveling distance to the reporting center relates to a 13.7 percent increase in the number of use of force, a 13.2 percent increase in the likelihood of high use of force, a 8.0 percent increase in the likelihood of being subject to use of force during an arrest, and a 12.0 percent increase in the injury rates due to use of force from an officer. Civilians in Black residential areas are more likely to be subject to force and injury during an arrest if they are farther away from the oversight agency. Distance does not have a statistically significant impact on use of force outcomes in white residential areas.

The short run results from Table 1.32 suggest that beats that are more difficult to monitor (in terms of distance from the oversight agency) had a significantly higher number of use of force reports and a higher level of force per arrest. Moreover, non-Black residential areas that are farther away from the reporting center had a higher likelihood of having an incident involving a higher level of force (firearm or Taser discharges). The results also show the number of injured civilians tends to increase in the specification that considered the nine-month period around the location change, but the effect seems to vanish when considering the twelve-month period around the policy change.

Overall the results suggest that non-white areas that are more distant from the reporting center exhibit a pattern of more aggressive policing. For Black residential areas, this type of policing increases the likelihood of injury and force per arrest. Hispanic residential areas exhibit a similar pattern and are also more likely to be subject to a higher level of force (Taser and firearm discharges) if they are farther away from the oversight agency.

1.5.2.4 Crime and arrest

Table 1.10 presents the monthly effect of distance from the incident location to the oversight agency on index and non-index crime outcomes at the beat level: number of offenses per

1,000 capita, arrest per 1,000 capita, and clearance per 1,000 capita. Results for the whole city suggest that beats that are farther away from the oversight agency have lower crime rates for both types of offenses, significantly higher clearance rates for index crimes, but lower arrest rates for non-index crimes. For readability, I only report the results that accounts for district's controls. The results are similar when not controlling for district's covariates.

Panel A of this table suggest that beats that more distant from the oversight agency have lower reported offenses, higher arrests (not statistically significant), and higher clearance rates for index crimes. For non-index crimes, beats that are closer to the oversight agency exhibit significantly have a higher number of arrests per 1,000 capita. Black residential areas, in Panel B, exhibit a similar pattern as the full sample. Beats with a majority of Hispanic residents tend to demonstrate worse index crime outcomes as they are farther away from the oversight agency. The results are not statistically significant. For non-Black residential areas, distance does not seem to have a statistically significant impact on the number of offenses and the number of arrests.

Results from non-Black residential areas suggest that it is difficult to conclude that more oversight yields to lower crime rates, higher number of arrests, and better clearance rates. Although this relationship is present in the whole sample and Black in Black residential areas, non-Black residential areas do not clearly exhibit that the higher cost of complaining yield to worst crime outcomes. The short run results from Table 1.33 suggest that beats that are farther away from the oversight agency had significantly lower crime rates and higher clearance rates for index crimes in non-White residential areas. Beats that more difficult to monitor have higher clearance rates for non-index crimes in Hispanic residential areas.

1.5.2.5 Threats to Validity and Robustness

As before, I use a placebo test to supplement the regression evidence on robustness to prior location change and differential trends. The placebo test replaces the actual location change that occurred in December 2011 with a location change in May 2011 and then reruns the main specification from January to November 2011, equation 1.3, to show that nothing mechanical drives the results.

Tables 1.28 and 1.29 report the placebo results for complaints and use of force outcomes respectively. This suggest that the results tend to be smaller, but are all statistically non-significant. The results from these falsification tests are encouraging and suggest that the estimates for the complaint and use of force outcomes are not related to event that would have potentially occurred prior to the true location change.

Table 1.30 reports the placebo results for crime outcomes. The results are mixed. None of the results are statistically significant for non-Black residential areas. For Black residential areas, results are non significant for index crimes; however, they are statistically significant for arrests and clearance rates for non-index crimes which makes it difficult to believe that the effect of distance on crime outcomes should be interpreted as causal. In Panel A, the placebo policy fails to reject the null for three out of six of the outcomes.

1.6 Model: Effect of the cost of signing the affidavit

The reduced-form analysis emphasizes how the cost of completing a complaint by signing the affidavit affects the number of investigated allegations of misconduct. I build on the key lessons from this experiment to shed light on what the impact would be on the number of investigated complaints and sustained rates if policies that influence the cost of investigating officers were to be implemented. To do so, I develop a model of civilian willingness to complete a complaint, accounting for the investigator decision to sustain the complaint if the affidavit is signed. I do not intend to model every feature of policing and I will later discuss how some of the simplifications I make might affect results. Rather, I show how a simple estimated model can provide insight into the effect of reducing the cost of completing a complaint and make out-of-sample predictions.

I do not model police officers' behavior because I do not have data on police-civilian encounters that did not result in a complaint. In other words, I do not have a risk set, or "benchmark" against which to compare officers that receive complaints. For instance, I do not observe officers' workloads and geographically assignments over time, and thus, it is difficult to clearly identify the impact of the oversight agency's location change on the behavior of individual officers.

1.6.1 Setup

The model has the following timing structure. First, after filing a complaint, the civilian decides whether or not to travel to oversight agency in order to sign the affidavit. If the affidavit is not signed, the complaint is dropped. Second, if the affidavit is signed, the investigator decides whether or not to sustain the complaint. Because the model requires information from the investigator, I drop complaints that have missing information about the investigator. Moreover, I keep complaints that have only one investigator assigned to the case. Table 1.17 provides details about the sample construction for this analysis.

I assume that the investigator is assigned when the complaint is filed. The model begins with the civilian complaining about the interaction with the police.

Complainant's preferences Given investigator j in beat b, and the severity of the allegation k, I assume that complainant i's utility to sign the affidavit is:

$$U^1_{k,idt}(jb) = \eta^1_{k,idt} + \mu^D_{k,jb} + \varepsilon^1_{k,it} \tag{1.4}$$

where

$$\eta^1_{k,idt} = -Cost_{bt}\beta_{k,c} + X'_{it}\beta_{k,x} + Z'_{ibt}\beta_{k,z} + \alpha_{k,d} + \gamma_{k,t}$$

where the disutility of signing the affidavit is function of traveling cost $Cost_{bt}$. The vector X_{it} contains a set of controls for complainant age, race, and gender. Vector Z_{ibt} is a set of controls related to the incident characteristics such as the number of officers, median age of the involved officers, and race of the officers. I assume that complainants have different utility depending on the severity of the allegation, k, so that k can be a serious complaint ($k = 1$) or a complaint related to failure to provide service ($k = 2$). The time and district fixed effects are given by $\gamma_{k,t}$ and $\alpha_{k,d}$. I assume that $\varepsilon^1_{k,it}$ is an i.i.d. error term that is distributed type 1 extreme value. For incidents that occurred in beat b, if the affidavit is signed, the unobserved characteristics of the investigator is given by $\mu^D_{k,jb}$. This parametrization of the unobserved heterogeneity allows the investigator unobserved characteristics to be correlated with the location of the incidents. The outside choice, denoted as choice 0, is not signing the affidavit. The utility from this option is given by $U^0_{k,idt}(jb) = \varepsilon^0_{k,it}$, where $\varepsilon^0_{k,it}$ is an i.i.d. error term that is distributed type 1 extreme value. I define $\eta^1_{k,idt} = Cost_{bt}\beta_{k,c} + X^{'}_{it}\beta_{k,x} + Z'_{ibt}\beta_{k,z} + \alpha_{k,d} + \gamma_{k,t}$ and $\eta^0_{idt} = 0$. These assumptions lend themselves to a logit regression:

$$\begin{aligned} \Pr(D_{k,idt} = 1|\mu^D_{k,jb}) &= \Pr(U^1_{k,idt}(jb) - U^0_{k,idt}(jb) > 0) \\ &= \Pr(\eta^1_{k,idt} + \mu^D_{k,jb} - \eta^0_{k,idt} > \varepsilon^0_{k,it} - \varepsilon^1_{k,it}) \\ &= q_{k,i}(\mu^D_{k,jb}) \end{aligned} \tag{1.5}$$

Where the variable $D_{k,idt}$ is equal to 1 if the complaint is signed and to 0 otherwise. To streamline notation in the last line of equation 1.5, I drop the d and t subscripts.

Investigator's preferences If the affidavit is signed, investigator j has the following utility to sustained the complaint of complainant i:

$$V^1_{k,it}(jb) = W'_{it}\beta_{k,W} + \tilde{\alpha}_{k,d} + \tilde{\gamma}_{k,t} + \mu^s_{k,jb} + \xi^1_{k,jbt} \tag{1.6}$$

Vector W_{it} is a set of controls related to the length of the investigation, the incident characteristics such as the number of officers, median age of the involved officers, and race of the officers. The time and district fixed effects are given by $\widetilde{\gamma}_{k,t}$ and $\tilde{\alpha}_{k,d}$. I assume that $\xi^1_{k,jbt}$ is an error term that is distributed type 1 extreme value. If the affidavit is signed, the unobserved characteristics of the investigator is given by $\mu^s_{k,jb}$. The outside choice, denoted as choice 0, is not sustaining the complaint. The utility from this option is given by $V^0_{k,it}(jb) = \xi^0_{k,jbt}$, where $\xi^0_{k,jbt}$ is an error term that is distributed type 1 extreme value. Conditional on signing the affidavit, these assumptions lend themselves to a logit regression, where the probability to sustained the complaint is:

$$\Pr(S_{jb,kit} = 1|D_{k,idt} = 1; \mu^s_{k,jb}) = \Pr(V^1_{k,it}(jb) - V^0_{k,it}(jb) > 0|D_{k,idt} = 1)$$

$$\Pr(S_{jb,kit} = 1|D_{k,idt} = 1; \mu^s_{k,jb}) = \Pr(W'_{it}\beta_{k,W} + \tilde{\alpha}_{k,d} + \tilde{\gamma}_{k,t} + \mu^s_{k,jb} > \xi^0_{k,jbt} - \xi^1_{k,jbt}|D_{k,idt} = 1)$$

$$\Pr(S_{jb,kit} = 1|D_{k,idt} = 1; \mu^s_{k,jb}) = p_{k,jb|D}(\mu^s_{k,jb}) \tag{1.7}$$

Where the variable $S_{jb,kdt}$ is equal to 1 if the complaint is sustained and to 0 otherwise. In other words, the investigator thinks that the benefit of having the complaint sustained is larger than the cost. To streamline notation in the last line of equation 1.7, I drop the i and t subscript. Recall that for a complaint to be sustained, the affidavit has to be signed i.e. $D_{k,idt}$ equals 1.

1.6.2 Estimation strategy

The model is estimated by maximum likelihood. The likelihood function of the full model will be derived. The parameter set of the full model consists of coefficients of covariates and parameters of unobserved heterogeneity. Unobserved heterogeneity of the investigator by beat enters into the model via the permanent components $\mu = (\mu^D_{1,jb}, \mu^s_{1,jb}, \mu^D_{2,jb}, \mu^s_{2,jb})$ which affect willingness to sign the affidavit and sustained rates in a similar manner to the covariates. This specification of the unobserved factor allows for the fact that investigator might affect the complainant decision to sign the affidavit which eventually has an impact on the sustained rates. For the unobserved heterogeneity, I use the McCall [1996] multivariate generalization of the Heckman and Singer [1984a] approach, where $\mu \sim G(\mu^D_{1,jb}, \mu^s_{1,jb}, \mu^D_{2,jb}, \mu^s_{2,jb})$ follows a discrete distribution with G points of support. In the model, unobserved heterogeneity of investigator by beat takes the form of discrete types and the error terms are i.i.d. when conditioned on type. The Panel structure of the data is sufficient to identify the parameters of unobserved heterogeneity and exclusion restrictions are included to facilitate estimation.

The following expression is the likelihood of signing the affidavit for individual i assigned to investigator j in beat b for allegation of type k:

$$\mathcal{L}^D_{k,i}(\mu) \quad = \quad q_{k,i}(\mu^D_{k,jb})^{D_{k,idt}}(1 - q_{k,i}(\mu^D_{k,jb}))^{1-D_{k,idt}}$$

Conditional on signing the affidavit for individual i, the likelihood contribution for sustaining the complaint for investigator j in beat b is:

$$\mathcal{L}^{S|D}_{k,jb}(\mu) \quad = \quad p_{k,jb|D}(\mu^s_{k,jb})^{S_{j,kdt}}(1 - p_{k,jb|D}(\mu^s_{k,jb}))^{1-S_{j,kdt}}$$

The likelihood function is

$$L \quad = \quad \prod_{jb} \sum_g \pi_g \prod_i \sum_k I_k \cdot \mathcal{L}^D_{k,i}(\mu_g) \left[\mathcal{L}^{S|D}_{k,jb}(\mu_g)\right]^{D_{k,idt}} \tag{1.8}$$

The log likelihood function is

$$logL \quad = \quad \sum_{jb} \log \sum_g \pi_g \prod_i \sum_k I_k \cdot \mathcal{L}^D_{k,i}(\mu_g) \left[\mathcal{L}^{S|D}_{k,jb}(\mu_g)\right]^{D_{k,idt}} \tag{1.9}$$

where I_k is equal one for complaints of type k, and zero otherwise. The probabilities for the points of support are given by $\pi_g = \exp(\kappa_g)/(1 + \exp(\kappa_1) + ... + \exp(\kappa_{G-1}))$. I use a likelihood ratio test to determine the number of support points (Heckman and Singer [1984a], Ham and LaLonde [1996], Eberwein et al. [1997]). As in Ba et al. [2017], I start by assuming no unobserved heterogeneity and then continue adding support points and keep the model with the fewest points of support that is not rejected by a standard likelihood ratio test.

1.6.3 Parameter estimates

I first focus on parameter estimates, from estimating equation 1.9, which are displayed in tables 1.11 and 1.12. I separate the results into two tables for readability. For the interpretation, I report the average marginal affect for the coefficients of the observable variables. Moreover, I also report the mean of the dependent variables for the whole sample and based on the race of the complainant table 1.13. The main takeaway from table 1.13 is that Black complainants have very low probability that their complaints will be sustained compared to non-Blacks. Provided that the affidavit is signed, the share of complaints that are sustained for serious allegations are 2.7 percent, 11.1 percent, and 30.4 percent for Blacks, Hispanics, and whites respectively. For failure to provide service complaints and provided that the affidavit is signed, the share of complaints that are sustained are 16.0 percent,36.9 percent, and 46.5 percent for Blacks, Hispanics, and whites, respectively.

Table 1.11 presents the estimates on the probability that the civilian will sign the affidavit by complaint types. Incidents that occurred farther away from the oversight agency are less likely to have a signed affidavit. The magnitude of the coefficients are similar to the DID estimation from Section 1.5. A standard deviation (3.6 miles) increase in distance reduces the probability of signed affidavit by 2.9 percentage points and 7.9 percentage points for serious complaints, and failure to provide service complaints respectively. The coefficients are not statistically significant for serious allegations. Table 1.11 suggests that males are significantly less likely to complete their complaint if the allegation is serious, the results are not significant for FPS allegations. For serious and FPS complaints, Blacks and Hispanics are less likely to sign the affidavit relative to white civilians. Older civilians are also more likely to complete their complaint. The race of the accused officers' race does not seem to significantly impact civilians' likelihood to complete the complaint.

Table 1.12 presents the estimates on the probability that the investigators will sustain the complaint by allegation type. Conditional on signing the affidavit, complainant and incident characteristics do not have a statistically significant impact on the likelihood that the investigator sustains the complaint. On the other hand, the race of the complainant has a significant impact for serious allegation of misconduct: non-white complainants are significantly less likely to have their complaint sustain. Incidents involving older officers or black officers are more likely to be sustained for serious allegations.

I find three points of support when estimating specification 1.9. The probability distributions of the three type of investigator-beat are 42 percent, 39 percent, and 19 percent. The points of support are statistically significant for serious allegations. The standard errors are relatively big for the point of supports of the non-serious allegations.

1.6.4 Model fit

Figures 1.7, 1.8, 1.9, and 1.10 present some results on the in-sample fit of the model. The predictions are based on the estimation of equation 1.9 and the results are reported in Tables 1.11 and 1.12.

To assess the fit, I generated 10,000 simulations for signed affidavits and sustained outcomes for each allegation based on the parameter estimates. I then compute the aggregate outcomes for the 22 police districts and report the predicted and actual frequency distributions of the outcomes variables for the police district cells. The police districts are ordered from most to least violent according to reported violent crime per 1,000 capita. Overall, these figures show that the model fits very closely the patterns observed in the data. Table 1.34 and figure 1.18 provide additional information about the districts and beats' characteristics from January 2011 to July 2014.

I also apply Chi-Squared goodness-of-fit tests (Heckman and Walker [1990]) to the estimated and actual frequency distributions. Recall that the predicted conditional distributions depends on estimated parameters from the model. I do not adjust the goodness-of-fit statistic to account for parameters estimation error because the adjustments are usually slight (Heckman and Walker [1990]). The Chi-Squared tests fail to reject the null hypothesis that the predicted values from the model are statistically different from the data, i.e. the model seems to fit the data relatively well.

1.7 Using valuation of the complaint to understand the parameter estimates

1.7.1 Willingness to pay

The parameters from the model can now be used to compute civilians' willingness to pay to complete their complaint. Here, the willingness to pay is the maximum amount of money a civilian is willing to sacrifice to complete her complaint by signing the affidavit. This quantity has a useful interpretation from an economic standpoint that provides, for example, some insights on the distribution of civilians' valuation of their allegation by race-age groups.

I now use the parameter estimates from Section 1.6 to show how to compute the willingness to pay to complete a complaint. Under the logit assumptions from equation 1.5, the "surplus" associated with a set of alternatives (signed or not) takes a closed form that is easy to calculate. I drop the time and location (beat and district) subscripts for expositional ease. To exposit expected utility, following Capps et al. [2003] the ex-ante expected utility of individual i related to the affidavit and complaint of type k is

$$E(CS_{k,i}|\mu_g) = \frac{1}{-\beta_{k,c}\theta_b} E\left[\max(\eta^1_{k,i} + \mu^D_{k,jb,g} + \varepsilon^1_{k,i}, \eta^0_{k,i} + \varepsilon^0_{k,i})\right] \tag{1.10}$$

The ex-ante expected utility has to account for the beat-investigator unobserved heterogeneity ($\mu^D_{k,jb,g}$) of type g. The division by $\beta_c\theta_b$[15] translates utility into dollars . The opportunity cost of time in each beat is captured by θ_b . As shown in Small and Rosen [1981], because the error terms are type 1 extreme value, and the utility is linear in traveling cost, the "complainant surplus" from equation 1.10 can be re-written as

$$E(CS_{k,i}|\mu_g) = \frac{1}{-\beta_c\theta_b} \log(\exp(\eta^1_{k,i} + \mu^D_{k,jb,g}) + \exp(\eta^0_{k,i})) \tag{1.11}$$

As in Capps et al. [2003], the expected utility gain of signing the affidavit or willingness to pay for signing the affidavit for individual i for complaint of type k is:

$$WTP_{k,i}(\mu_g) = \tfrac{1}{-\beta_{k,c}\theta_b}\left[\log(\exp(\eta^1_{k,i} + \mu^D_{k,jb,g}) + \exp(\eta^0_{k,i})) - \log(\exp(\eta^0_{k,i}))\right] \tag{1.12}$$

$$WTP_{k,i}\quad(\mu_g) = \tfrac{1}{-\beta_{k,c}\theta_b}\left[\tfrac{1}{1-\Pr(D_{k,i}=1|\mu^D_{k,jb,g})}\right] \tag{1.13}$$

I assume that θ_b is equal to the average hourly wage in each beat[16]. The main concern of this assumption is that the opportunity cost of time (by car) may be different across civilians, i.e. the opportunity cost of time of a civilian (potential suspect of crime or victims of a crime) interacting with the police may be higher or lower than working individuals that do not have any interaction with law enforcement. . Because I do not know the true

15. θ_b converts each unit of cost in dollar term.

16. I assume that the hourly average cost of time is captured by the average hourly wage =average annual income/(40 hours $\times$52 weeks)

wages of civilians, I have to assume there is no selection with respect to wage. However, the transformation is fairly straightforward that one could use the results to get estimates that rely on alternative values of θ_b .

In order to calculate the overall willingness to pay, one needs to integrate over the unobserved heterogeneity, μ_g, that follows a discrete distribution with G points of support such that

$$WTP_{k,i} = \sum_{g=1}^{G} \pi_g \cdot WTP_{k,i}(\mu_g) \tag{1.14}$$

Where the probabilities for the points of support are given by $\pi_g = \exp(\kappa_g)/(1 + \exp(\kappa_1) + ... + \exp(\kappa_{G-1}))$.

Conditional on the parameter estimates from the model, the estimated empirical willingness to pay is given by

$$W\hat{T}P_{k,i} = \sum_{g=1}^{G} \hat{\pi}_g \cdot \frac{1}{-\hat{\beta}_{k,c}\theta_b} \left[\frac{1}{1-\hat{\Pr}(D_{k,i}=1|\hat{\mu}^D_{k,jb,g})} \right] \tag{1.15}$$

Equation 1.15 can be calculated for each complainant, given her explanatory variables. Moreover, this quantity can be used to trace back the estimated empirical distribution of civilians' valuation of the complaint. Another quantity of interest is the average willingness to pay by race-age groups. Let Z denotes the group of interest. For each group Z, the sample average willingness to pay to complete a complaint of type k is

$$W\hat{T}P_{k,Z} = \frac{\sum_{i \in Z} W\hat{T}P_{k,i}}{n_Z} \tag{1.16}$$

where n_Z is the sample size of group Z. One can assume that Z is composed of each of the 22 police districts or different race-age groups of the complainants. To conduct inference on the quantity from equation 1.16, it is important to adjust for the sampling error and uncertainty from the parameter estimates. Additional information about the standard errors of the sample average of the willingness to pay is provided in Appendix 1.12.

1.7.2 Results

Figures 1.11,1.12,1.13, and 1.14 report civilians' willingness to pay in dollars to complete their complaint by racial-ethnic group. Conditional on the age and the race of the complainant, I report both the kernel density, using 1.15, and the average willingness to pay to sign the affidavit. To ease the interpretation of the results, I also interpret the results in terms of hours of work sacrificed to complete a complaint using the ratio between the willingness to pay and the hourly wages by complainant race reported in table 1.14. This alternative measure helps accounting for the fact that complainants who are willing to pay the same price for a similar complaint, might have to sacrifice a different number of hours of worked.

The distributions of the willingness to pay for both types of complaints are not symmetric (Figures 1.11 and 1.13). The results suggest that civilians' willingness to pay to complete their complaint for serious allegation ($68.1 or 3.5 hours of work on average) is higher than for FPS allegation ($19.9 or an hour of work on average). The median willingness to pay is lower than the average willingness to pay. The median amounts of money that complainants are willing to pay are $61.6 (3.1 hours of work) and $16.8 (55 minutes of work) for serious and FPS, respectively. For both types of complaint, Hispanics have a lower valuation of their complaint relative to their non- Hispanic counterparts. Blacks have the highest average valuation of their complaint for serious allegations, whereas whites have the highest average valuation of their complaint for FPS allegations. Finally, the kernel density plots suggest that the median valuations of complaints for Black civilians is far higher than for non-Black civilians. For instance, the median valuation for FPS is $18.4 (an hour of work), $15.5 (36 minutes of work), and $11.5 (half an hour of work) for Black, white, and Hispanic complainants. The median valuation for serious allegations is $68.6 (3.9 hours of work), $49.5 (two hours of work), and $45.31 (almost two hours of work) for Black, white, and

Hispanic complainants.

According to Figures 1.12 and 1.14, relative to non-Black civilians, Black civilians between the age of 18 and 49 years old have a significantly higher valuation of their complaint for serious allegation. For FPS, Black civilians between the age of 18 and 39 years old express a significantly higher valuation of their complaint relative to non-Blacks. There is a low share of white civilians below 30 years old who complain about failures to provide service. White civilians between the age of 40 and 74 years old have an economically large valuation of their complaint compare to other racial-ethnic groups.

Overall, the results suggest that people with the lowest valuation of complaining benefit the most of civilian oversight. Table 1.15 summarizes the cost and benefits of signing the affidavit by complainants' race for each type of complaint. Hispanics seem to exhibit a willingness to pay to complete a complaint that is similar to that of white civilians. Given that Blacks have high valuation and low returns on complaining, pooling minorities (Blacks and Hispanics) together or keeping those groups separate when studying discrimination might yield very different results.

1.8 Counterfactuals

This Section of the paper uses the parameter estimates from Section 1.6 to simulate the various impacts of a policy that lower the cost of completing a complaint. I consider a policy that requires a signed statement, but not to sign a sworn affidavit at the oversight agency. There is also the option of filling out a sworn affidavit, getting it notarized and then mailing it the oversight agency. For example, community organizations or local government agencies can be trained to assist with the filing of a complaint and notarizing the document[17]. Those two alternatives are not legally equivalent, but both policies set the traveling cost to the

17. For example, the Houston Police Department has adopted a similar system to file a complaint against Houston police officers.

oversight agency to zero (or close to).

I evaluate the effects of removing the traveling requirement to the oversight agency to complete the complaint, holding everything else constant. I assume that the number of complaints would have stayed constant under this alternative policy. This alternative policy would impact: (i) the share of complaints with signed affidavit, i.e. the share of investigations after a complaint is signed (share of officers held accountable of their action after a complaint), and (ii) the share of allegations of misconduct that yields to a sustained outcome.

As presented in the previous Section, let Z denotes the group of interest. Here, I consider that Z is composed of each of the 22 Chicago police districts. In a given group Z, the expected share of complaints of type k with signed affidavit is given by

$$E(D_{k,Z}) = \sum_{g=1}^{G} \pi_g \cdot \Pr(D_{k,idt} = 1 | \mu_g, Z) \tag{1.17}$$

Such that μ_g, that follows a discrete distribution with G points of support and the probabilities for the points of support are given by $\pi_g = \exp(\kappa_g)/(1 + \exp(\kappa_1) + ... + \exp(\kappa_{G-1}))$.

Conditional on the complaint being signed under the current policy, the expected share of complaints of type k that are sustained in group Z is given by

$$E(S_{k,Z} | D_{k,Z} = 1) = \sum_{g=1}^{G} \pi_g \cdot \Pr(S_{jb,kit} = 1 | D_{k,idt} = 1; \mu_g, Z) \tag{1.18}$$

There are two possible counterfactuals for the sustained rates. Let $D^A_{k,idt}$ equals one if the complainant signed the affidavit under the alternative policy, and zero otherwise. The first possible counterfactual is the sustained rates of complaint that are signed under the current policy and signed according to the counterfactual policy is given by

$$E(S_{k,Z} | D_{k,Z} = 1, D^A_{k,Z} = 1) = \sum_{g=1}^{G} \pi_g \cdot \Pr(S_{jb,kit} = 1 | D_{k,idt} = 1, D^A_{k,idt} = 1; \mu_g, Z) \tag{1.19}$$

The second possible counterfactual is sustained rates for complaint that are not signed under the current policy, but signed according to the counterfactual policy, which is given by

$$E(S_{k,Z}|D_{k,Z}=0,D^{A}_{k,Z}=1)=\sum_{g=1}^{G}\pi_{g}\cdot\Pr(S_{jb,kit}=1|D_{k,idt}=0,D^{A}_{k,idt}=1;\mu_{g},Z) \quad (1.20)$$

Signed Affidavits Figure 1.15 presents the effect of the alternative policy on the share of investigated complaints (i.e. with sign affidavit). Relative to the number of complaints that are predicted by the model, the share of complaint that should be investigated increases for both type of complaints under the alternative policy. In other words, I compute

$$\triangle_{k,Sign}=\frac{E(D^{A}_{k,Z})-E(D_{k,Z})}{E(D_{k,Z})} \quad (1.21)$$

such that $D^{A}_{k,Z}$ is the decision to sign the affidavit under the alternative policy and is $D_{k,Z}$ the decision to sign the affidavit under the current policy. Under the alternative scenario, the share of investigated complaints would increase by 5.11 percent and 37.58 percent respectively for serious and FPS allegations. This alternative policy would significantly increase the number of investigated com- plaints that affect potential victim of a crime (FPS). Complainants (with allegations of serious misconduct) who are potential criminal suspects would marginally respond to that policy. For FPS, the response to the policy is larger for districts with the highest rate of violent crime per 1,000 capita and which also have a majority of Black residents.

Sustained Complaints for Observed Signed Affidavits Figure 1.16 presents the effect of the alternative policy on the share of sustain complaints for complaints that were both: (i) signed under the current policy, and (ii) signed according to the counterfactual policy. In other words, I compute

$$\triangle_{Sust,D_{k,Z}=1} = \frac{E(S_{k,Z}|D_{k,Z}=1, D^A_{k,Z}=1) - E(S_{k,Z}|D_{k,Z}=1)}{E(S_{k,Z}|D_{k,Z}=1)} \tag{1.22}$$

such that $D^A_{k,Z}$ is the decision to sign the affidavit under the alternative policy and is $D_{k,Z}$ the decision to sign the affidavit under the current policy. Overall, this alternative policy would on average raise the share of sustained to complain for FPS by 8.1 percent, but lower the share of sustained complain about serious allegations 9.77 percent. For serious allegations, only the Austin and Morgan Park police districts, would see an increase in their sustained rates (of 6.31 percent and 1.47 percent respectively) under the alternative policy. For FPS allegations, five police districts (Albany Park, Morgan Park, Near West, South Chicago, and Town Hall) would have a decrease in their sustained rates under the alternative policy. Out of five police districts for which the alternative policy would have the largest increases in the sustained rates for FPS, three of them are the most violent police districts in the city: Englewood, Harrison, and Grand Crossing.

Sustained Complaints for Observed Not Signed Affidavits This Section attempts to assess the sustained rates for complaints that would have been signed under the alternative policy, but are not complete (no affidavit) under the current policy. Recall that the sustain rates under the current environment is zero, since those complaint are classified as non-sustained because of the lack of an affidavit. Conditional on a signed affidavit, Figure 1.17 presents the effect of the alternative policy on the number of sustained complaints. This figure restricts the sample to complaints that were both: (i) not signed under the current policy, and (ii) signed according to the counterfactual policy. In other words, I compute

$$\triangle_{Sust,D_{k,Z}=0} = E(S_{k,Z}|D_{k,Z}=0, D^A_{k,Z}=1) \tag{1.23}$$

such that $D^A_{k,Z}$ is the decision to sign the affidavit under the alternative policy and is $D_{k,Z}$ is the decision to sign the affidavit under the current policy. Out of the five police districts for which the alternative policy would have the largest the sustained rates for FPS, three of them are the least violent police districts in the city (Jefferson Park, Albany Park, and Lincoln) under the alternative policy. Harrison and Austin districts, which are two of the five most violent districts, are among the districts that would have experienced the highest sustained rates for FPS under the alternative policy. The sustained rates for serious allegations would be at most 5.0 percent for districts with a majority of Black residents (except for the Wentworth district, which would have a 13 percent sustained rate) and two out of three Hispanic districts. The remaining districts with sustained rates that would be higher than 10.0 percent for serious allegations either have no dominant racial-ethnic group or have predominantly white residents.

1.9 Conclusion

Placing unnecessary barriers on the civilian complaint process is expensive for society. It leads to public scandals, taxpayer payouts for civil rights lawsuits, and decreased police legitimacy and effectiveness. This paper uses detailed administrative data to study the cost and benefits of filing a complaint against the police in Chicago. As described earlier, I exploit the fact that complaints without affidavits are considered null and an administrative change of location of the reporting center to study the effect of civilian oversight on policing.

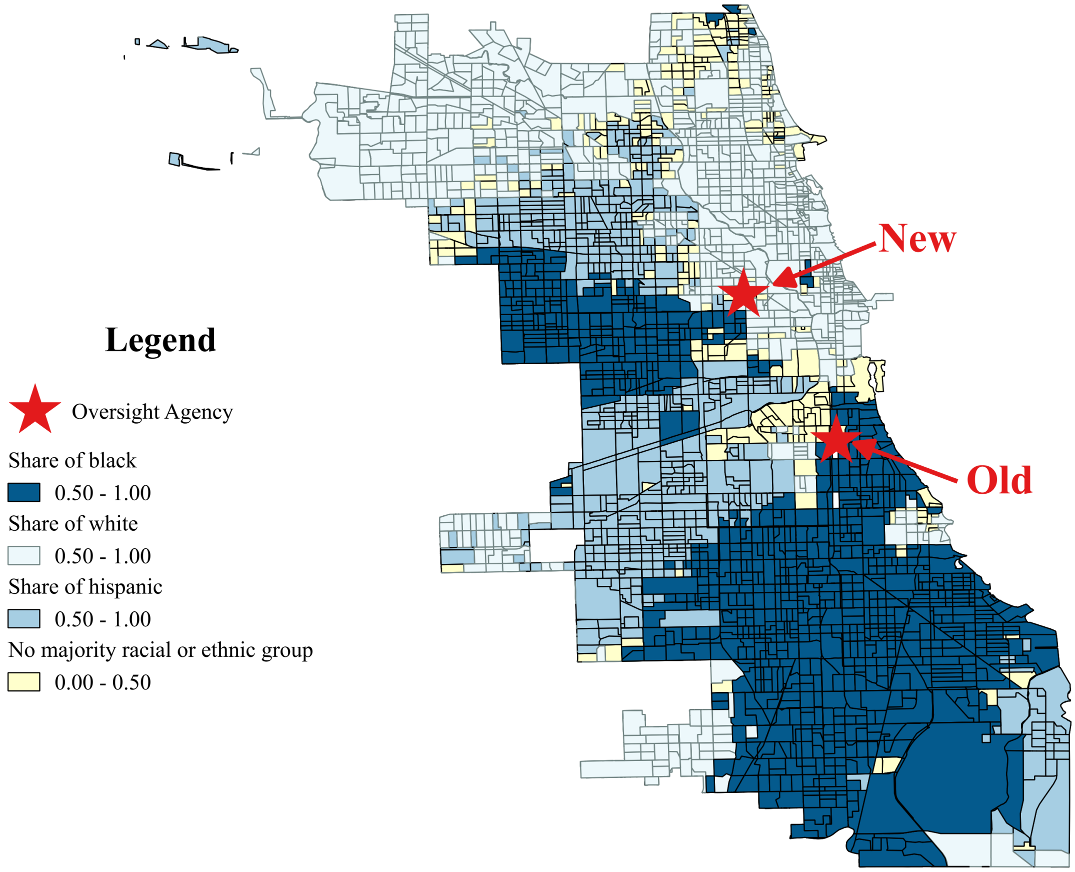
I present evidence that complaints filed as a result of interactions between civilians and officers that occurred farther away from the oversight agency are more likely to be lacking an affidavit for both serious allegations and allegations of a failure to provide service. I find that non-white residential areas that are harder to monitor due to their distance from the reporting location are subject to more aggressive policing. Black residential areas, meanwhile, report lower rates of index crimes. These results provide evidence that officers appear to respond to decreasing levels of oversight —defined by the lower likelihood that a complaint will be sustained and the officer investigated, and correlated with increased distance from the reporting center by engaging in more aggressive forms of policing. However, this aggressive policing only seems to exist in non-white residential areas and the effect of oversight on crime reduction only impacts index-crimes in Black residential areas.

To perform a counterfactual scenario, I estimate a model of civilian willingness to complete their complaint, accounting for the investigator decision to sustain the complaint. To ease the interpretation of the parameter estimates, I compute civilian willingness to pay to complete the complaint. I find that individuals who benefit the most from oversight are those with lowest valuation of complaining: non-Blacks have lower valuations of their complaints relative to Blacks, but that there is a higher likelihood complaints from non-Blacks will be sustained. Whites and Hispanics have similar valuation of their complaint, while a Black complainant is willing to sacrifice twice as much time to complete their complaint relative to his or her non-Black counterpart.

Finally, I use my model to simulate counterfactual scenarios under a policy that would remove the cost of signing the complaint. This policy would largely increase the number of investigations and sustained rates for failure to provide service in the city's most violent police districts. On the other hand, for allegations of constitutional violations, this policy would reduce sustained rates overall and only marginally increase the number of investigations.

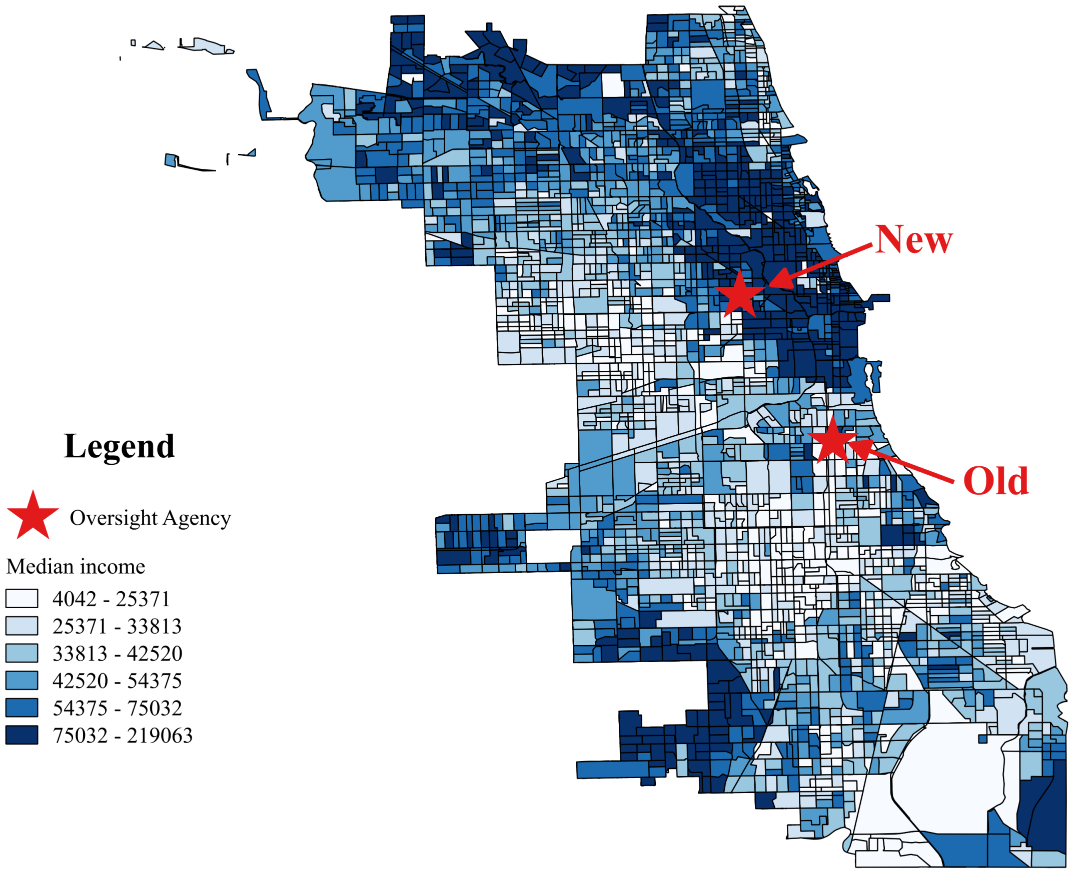
This research demonstrates a paradox inherent in efforts to use the complaint process as a primary mechanism in identifying police misconduct and ensuring accountability. To limit the number of complaints alleging excessive use of force or a violation of rights, an officer or a department may be tempted to engage in less-active policing. To address complaints alleging a failure to provide service, however, requires more proactive policing. A single policy change cannot resolve both issues. Rather, fair and effective police reform requires a nuanced understanding of the trade-offs involved in using the complaint process as a primary mechanism of civilian police oversight.

Figure 1.1: Racial distributions in Chicago



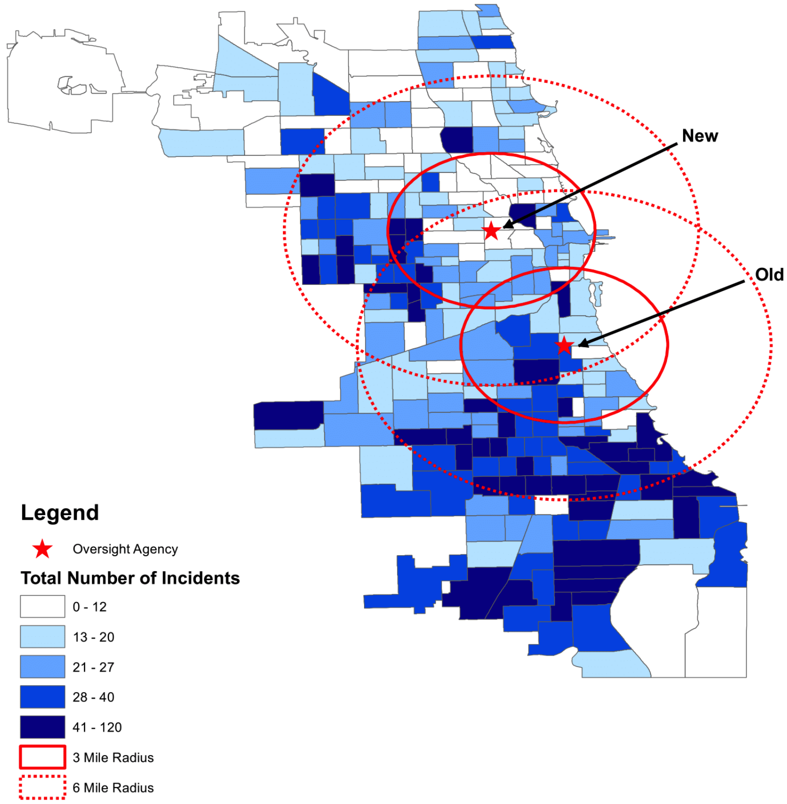
Notes: Figure 1.1 depicts the residential Chicago population in terms of four demographic categories that cover all the city population at the block level using the 2010-2014 ACS data. The fourth demographic category displays blocks where no racial or ethnic groups represents more than fifty percent of the block. The oversight agency locations (red star) moved from the South Side of Chicago to the Near West Side of the city on December 19, 2011.

Figure 1.2: Median income level by census blocks



Notes: Figure 1.2 presents the median income categories at the block level using the 2010-2014 American Community Survey (ACS) data. The oversight agency locations (red star) moved from the South Side of Chicago to the Near West Side of the city on December 19, 2011.

Figure 1.3: Allegations of misconduct from 2011 to 2014

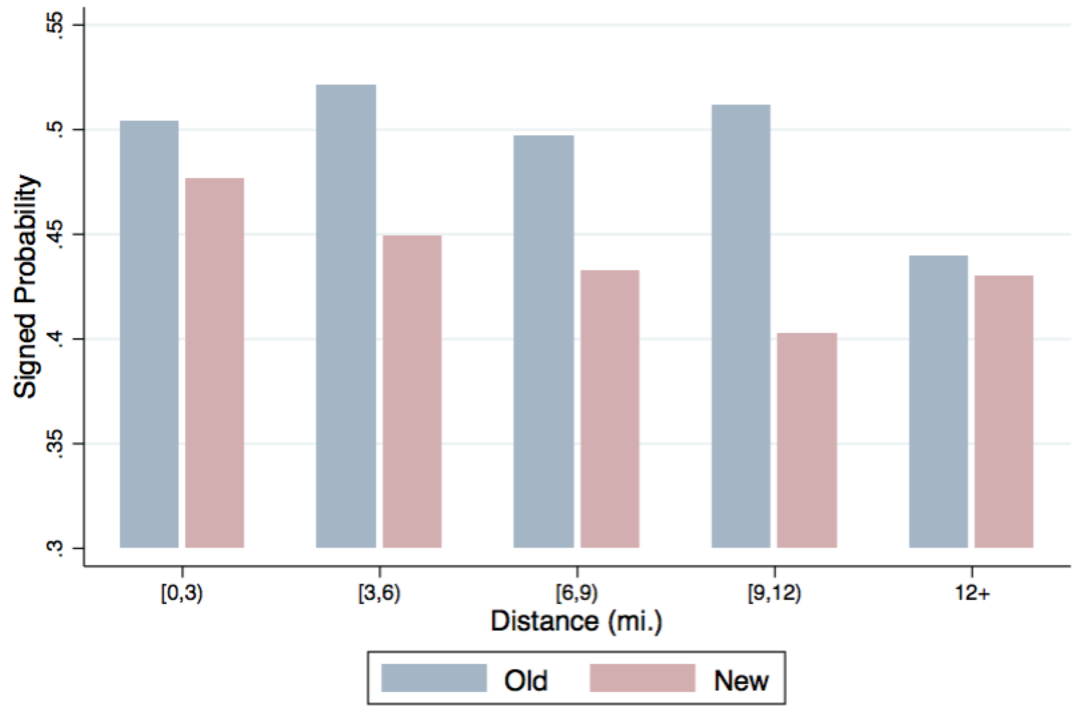


Notes: Figure 1.3 depicts the quintile distribution of civilian complaints filed against identified CPD officers from January 2011 to December 2014 at the police beat level. I consider that civilian complaints are allegations of misconduct which are classified as failure to provide service, use of force, verbal abuse, arrest or locked up procedures, and search. The oversight agency locations (red star) moved from the South Side of Chicago to the Near West Side of the city on December 19, 2011.

Table 1.1: Civilian complaint categories by complainant's race

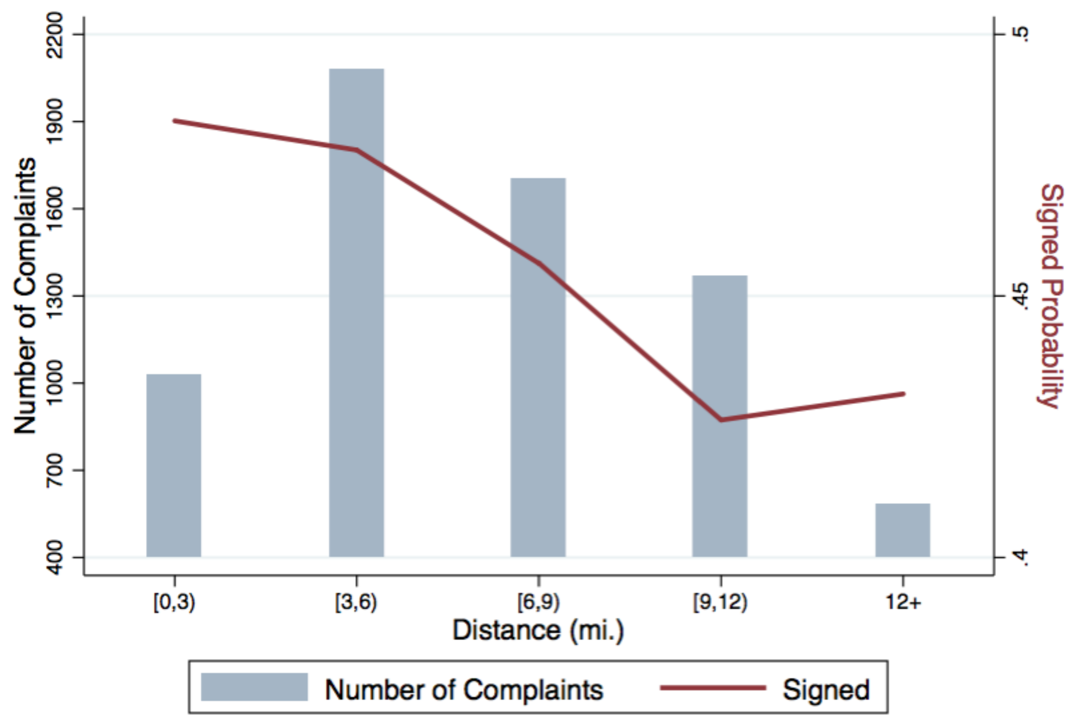
	Complainant's Race									
	Black		Hispanic		White		Unknown		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Use of Force/Verbal Abuse	1633	35.4	345	42.3	336	33.3	120	37.0	2434	36.0
Arrest/Locked up Procedure	334	7.2	59	7.2	82	8.1	14	4.3	489	7.2
Search	1415	30.7	166	20.4	146	14.5	120	37.0	1847	27.3
Failure to Provide Service	1234	26.7	245	30.1	444	44.0	70	21.6	1993	29.5
Total	4616		815		1008		324		6763	

Figure 1.4: Probability to sign the affidavit by oversight agency's location



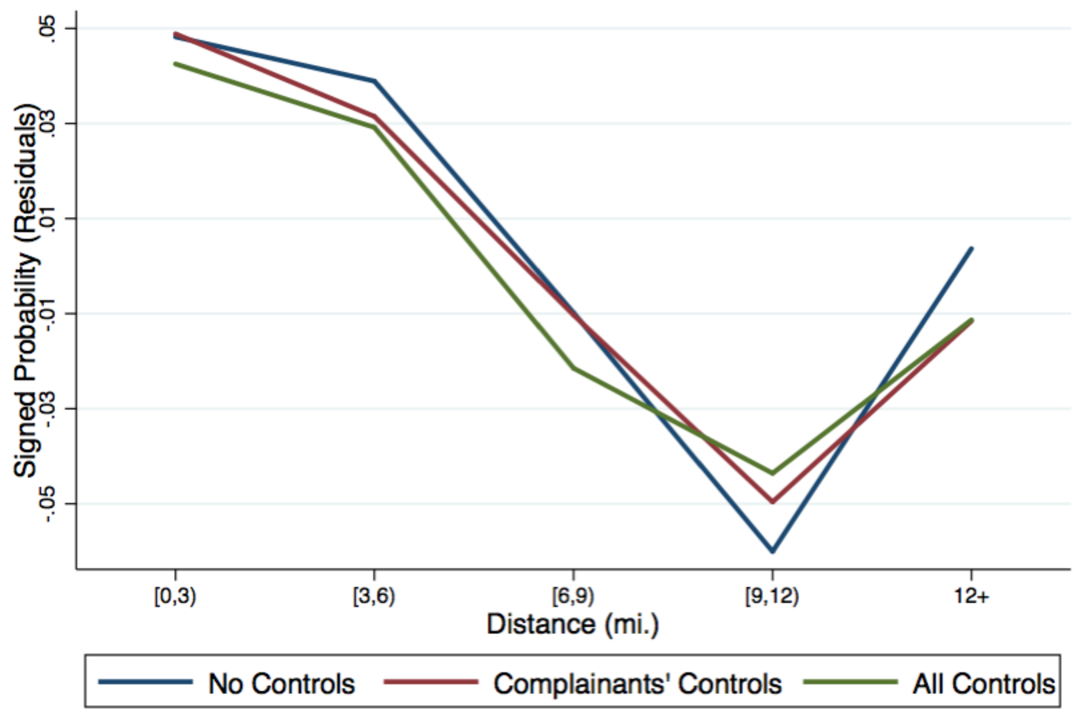
Notes: This Figure plots the probability of signed affidavits for civilian complaints by traveling distance to the oversight agency in miles from January 2011 to July 2014, at the old and new oversight agency's location.

Figure 1.5: Number of complaints and signed probability by distance



Notes: This Figure plots the probability of signed affidavits for civilian complaints by traveling distance to the oversight agency in miles from January 2011 to July 2014. Figure 1.5 plots the probability to sign the affidavit, and a histogram of the number of complaints by traveling distance.

Figure 1.6: Residualized signed probability by distance



Notes: This Figure plots the probability of signed affidavits for civilian complaints by traveling distance to the oversight agency in miles from January 2011 to July 2014. Figure 1.6 plots the demeaned signed affidavits rate as well as two residualized versions. The first residualized signed affidavits rate using complainants characteristics. The second accounts for accused officers and incidents characteristics. All covariates (except distance) are as described in Table 1.21.

Table 1.2: Summary statistics by distance bins for civilian complaints

	(1)	(2)	(3)	(4)	(5)	(6)
	All	[0,3)	[3,6)	[6,9)	[9,12)	12+
Serious	0.71	0.67	0.73	0.70	0.71	0.69
	(0.46)	(0.47)	(0.45)	(0.46)	(0.45)	(0.46)
Distance (mi.)	6.86	2.03	4.48	7.52	10.24	13.97
	(3.57)	(0.65)	(0.83)	(0.84)	(0.84)	(1.06)
Time by car (min.)	20.56	10.66	17.28	23.10	26.31	28.82
	(7.12)	(3.16)	(3.80)	(5.37)	(4.58)	(3.02)
Time by transit (min.)	49.28	23.91	40.17	51.98	64.44	83.05
	(18.91)	(7.91)	(8.19)	(9.63)	(10.63)	(11.55)
Hourly wage	19.74	26.26	18.24	18.78	18.18	20.04
	(9.28)	(13.15)	(8.56)	(8.18)	(6.35)	(7.12)
Male complainant	0.54	0.55	0.56	0.52	0.54	0.53
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
18-29yo	0.14	0.13	0.13	0.13	0.15	0.12
	(0.34)	(0.34)	(0.34)	(0.34)	(0.36)	(0.33)
30-39yo	0.26	0.25	0.27	0.27	0.27	0.24
	(0.44)	(0.43)	(0.44)	(0.44)	(0.45)	(0.43)
40-49yo	0.23	0.21	0.23	0.24	0.22	0.27
	(0.42)	(0.41)	(0.42)	(0.43)	(0.41)	(0.45)

Table 1.2, continued

	(1)	(2)	(3)	(4)	(5)	(6)
	All	[0,3)	[3,6)	[6,9)	[9,12)	12+
Black	0.68	0.59	0.66	0.68	0.77	0.72
	(0.47)	(0.49)	(0.47)	(0.47)	(0.42)	(0.45)
Hispanic/Other	0.12	0.16	0.14	0.12	0.07	0.08
	(0.33)	(0.37)	(0.35)	(0.32)	(0.26)	(0.27)
White	0.15	0.19	0.15	0.15	0.11	0.16
	(0.36)	(0.39)	(0.36)	(0.36)	(0.31)	(0.37)
Unknown race	0.05	0.05	0.05	0.05	0.05	0.05
	(0.21)	(0.23)	(0.21)	(0.21)	(0.21)	(0.21)
Median age of the PO	40.60	42.20	40.10	40.55	40.04	41.01
	(7.42)	(7.40)	(7.37)	(7.37)	(7.28)	(7.66)
Any non PO	0.32	0.32	0.33	0.32	0.32	0.31
	(0.47)	(0.46)	(0.47)	(0.47)	(0.47)	(0.46)
Any black PO	0.33	0.31	0.24	0.32	0.42	0.49
	(0.47)	(0.46)	(0.43)	(0.47)	(0.49)	(0.50)
Any hispanic PO	0.35	0.38	0.42	0.31	0.31	0.24
	(0.48)	(0.48)	(0.49)	(0.46)	(0.46)	(0.43)
Any white PO	0.62	0.60	0.65	0.64	0.59	0.55
	(0.49)	(0.49)	(0.48)	(0.48)	(0.49)	(0.50)
N	6763	1028	2080	1703	1370	582

Table 1.3: Summary statistics by beats

	(1)	(2)	(3)	(4)
	All	Black Beats	Hispanic Beats	White Beats
Distance (mi.)	6.34	7.46	5.49	5.20
	(3.48)	(3.58)	(2.42)	(3.37)
Time by car (min.)	20.23	20.75	21.66	18.51
	(7.30)	(6.55)	(6.60)	(7.67)
Time by transit (min.)	46.91	52.08	47.82	38.82
	(18.49)	(19.31)	(13.24)	(17.27)
Average income	46896.84	31684.26	40166.76	76133.45
	(21742.70)	(9204.22)	(7169.53)	(17931.79)
Number of PO in the District	351.06	387.84	344.98	310.41
	(75.06)	(60.71)	(54.46)	(78.11)
Average Monthly PO Salary in the District	6738.27	6669.45	6723.18	6849.69
	(189.08)	(158.87)	(171.55)	(192.98)
N	11395	5074	1978	2795

Table 1.4: Effect of distance on the probability of signed affidavit (Pooled)

		All			Serious			Failure to provide service	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A) All Comp.									
Distance (mi.)	-0.004*	-0.003	-0.003*	-0.003	-0.002	-0.003	-0.005	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mean Dependent Variable	0.46	0.46	0.46	0.52	0.52	0.52	0.31	0.31	0.31
Observations	6,763	6,763	6,763	4,770	4,770	4,770	1,993	1,993	1,993
R-squared	0.07	0.10	0.10	0.03	0.06	0.06	0.10	0.15	0.15
B) Black Comp.									
Distance (mi.)	-0.003	-0.003	-0.003	-0.002	-0.001	-0.002	-0.007	-0.007	-0.007*
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Mean Dependent Variable	0.44	0.44	0.44	0.52	0.52	0.52	0.25	0.25	0.25
Observations	4,616	4,616	4,616	3,382	3,382	3,382	1,234	1,234	1,234
R-squared	0.08	0.09	0.09	0.03	0.04	0.05	0.07	0.08	0.09
C) Hispanic Comp.									
Distance (mi.)	-0.003	-0.004	-0.003	-0.006	-0.006	-0.006	0.007	0.010	0.013
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)
Mean Dependent Variable	0.51	0.51	0.51	0.57	0.57	0.57	0.35	0.35	0.35
Observations	815	815	815	570	570	570	245	245	245
R-squared	0.12	0.14	0.15	0.08	0.11	0.13	0.29	0.33	0.35
D) White Comp.									
Distance (mi.)	0.003	0.001	0.001	-0.001	-0.005	-0.007	0.002	0.002	0.001
	(0.004)	(0.004)	(0.004)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Mean Dependent Variable	0.55	0.55	0.55	0.60	0.60	0.60	0.48	0.48	0.48
Observations	1,008	1,008	1,008	564	564	564	444	444	444
R-squared	0.14	0.18	0.19	0.09	0.15	0.16	0.28	0.32	0.33
Complainant's Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Officers' Controls	No	No	Yes	No	No	Yes	No	No	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	No	No	No	No	No	No	No	No	No

Notes: This Table presents the effect of distance on the probability of signed affidavit from January 2011 and July 2014. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search). The specification controls for complainants and incident characteristics, incident location, fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all the complainants, Black complainants, Hispanic, and White complainants. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.5: Effect of distance on the probability of signed affidavit

	All			Serious			Failure to provide service		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A) All Comp.									
Distance (mi.)	-0.011***	-0.011***	-0.011***	-0.009*	-0.009**	-0.009**	-0.013**	-0.015**	-0.014**
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Mean Dependent Variable	0.46	0.46	0.46	0.52	0.52	0.52	0.31	0.31	0.31
Observations	6,763	6,763	6,763	4,770	4,770	4,770	1,993	1,993	1,993
R-squared	0.12	0.14	0.14	0.10	0.12	0.13	0.24	0.28	0.28
B) Black Comp.									
Distance (mi.)	-0.014***	-0.013**	-0.013**	-0.010	-0.009	-0.009	-0.028***	-0.029***	-0.028***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.009)	(0.010)	(0.010)
Mean Dependent Variable	0.44	0.44	0.44	0.52	0.52	0.52	0.25	0.25	0.25
Observations	4,616	4,616	4,616	3,382	3,382	3,382	1,234	1,234	1,234
R-squared	0.15	0.16	0.16	0.12	0.13	0.13	0.27	0.28	0.29
C) Hispanic Comp.									
Distance (mi.)	-0.011	-0.010	-0.008	-0.003	-0.002	-0.003	-0.031	-0.031	-0.033
	(0.015)	(0.015)	(0.015)	(0.018)	(0.018)	(0.018)	(0.051)	(0.057)	(0.056)
Mean Dependent Variable	0.51	0.51	0.51	0.57	0.57	0.57	0.35	0.35	0.35
Observations	815	815	815	570	570	570	245	245	245
R-squared	0.35	0.36	0.36	0.41	0.44	0.45	0.73	0.76	0.78
D) White Comp.									
Distance (mi.)	0.004	0.006	0.005	-0.002	-0.003	-0.002	-0.016	-0.015	-0.016
	(0.012)	(0.011)	(0.011)	(0.017)	(0.017)	(0.018)	(0.023)	(0.022)	(0.022)
Mean Dependent Variable	0.55	0.55	0.55	0.60	0.60	0.60	0.48	0.48	0.48
Observations	1,008	1,008	1,008	564	564	564	444	444	444
R-squared	0.35	0.37	0.37	0.43	0.45	0.45	0.62	0.63	0.65
Complainant's Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Officers' Controls	No	No	Yes	No	No	Yes	No	No	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on the probability of signed affidavit from January 2011 and July 2014. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search). The specification controls for complainants and incident characteristics, incident location, fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all the complainants, Black complainants, Hispanic, and White complainants. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.6: Effect of distance on the probability of signed affidavit (Short Run)

	All			Serious			Failure to provide service		
	[-6;6]	[-9;9]	[-12;12]	[-6;6]	[-9;9]	[-12;12]	[-6;6]	[-9;9]	[-12;12]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A) All Comp.									
Distance (mi.)	-0.007	-0.007*	-0.010***	0.003	-0.002	-0.007	-0.035***	-0.021***	-0.018***
	(0.004)	(0.004)	(0.003)	(0.006)	(0.005)	(0.005)	-0.009	-0.006	-0.006
Mean Dependent Variable	0.49	0.48	0.48	0.55	0.55	0.55	0.33	0.31	0.31
Observations	2,293	3,396	4,098	1,666	2,432	2,920	627	964	1,178
R-squared	0.13	0.12	0.11	0.11	0.08	0.07	0.18	0.17	0.15
B) Black Comp.									
Distance (mi.)	-0.005	-0.007	-0.011**	0.006	-0.001	-0.006	-0.044***	-0.031***	-0.031***
	(0.006)	(0.005)	(0.005)	(0.008)	(0.007)	(0.006)	-0.013	-0.009	-0.009
Mean Dependent Variable	0.48	0.48	0.47	0.56	0.56	0.55	0.28	0.25	0.25
Observations	1,578	2,327	2,789	1,171	1,711	2,044	407	616	745
R-squared	0.12	0.12	0.11	0.09	0.06	0.05	0.15	0.15	0.14
C) Hispanic Comp.									
Distance (mi.)	-0.002	-0.007	-0.016	0.012	-0.012	-0.020	-0.026	0.014	0.010
	(0.019)	(0.015)	(0.012)	(0.021)	(0.019)	(0.015)	(0.043)	(0.033)	(0.029)
Mean Dependent Variable	0.53	0.54	0.52	0.61	0.63	0.61	0.32	0.32	0.30
Observations	256	383	476	184	269	341	72	114	135
R-squared	0.28	0.20	0.22	0.29	0.21	0.25	0.61	0.36	0.41
D) White Comp.									
Distance (mi.)	-0.007	0.001	-0.004	-0.009	0.001	-0.003	-0.016	-0.011	-0.01
	(0.013)	(0.011)	(0.009)	(0.019)	(0.016)	(0.015)	(0.025)	(0.017)	(0.015)
Mean Dependent Variable	0.56	0.56	0.56	0.62	0.60	0.62	0.48	0.49	0.48
Observations	331	506	625	201	308	369	130	198	256
R-squared	0.30	0.23	0.22	0.38	0.26	0.22	0.56	0.44	0.4
Complainant's Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Officers' Controls	No	No	Yes	No	No	Yes	No	No	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on the probability of signed affidavit from -/+ t months from the location change of the oversight agency in December 2011, such that $t = \{6, 9, 12\}$ months. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search). The specification controls for complainants and incident characteristics, incident location, beat fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all the complainants, Black complainants, Hispanic, and White complainants. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.7: Effect of distance on the probability of signed affidavit (Placebo)

	All			Serious			Failure to provide service		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A) All Comp.									
Distance (mi.)	-0.002	-0.003	-0.002	-0.008	-0.010	-0.009	0.010	0.012	0.013
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)	(0.011)
Mean Dependent Variable	0.51	0.51	0.51	0.59	0.59	0.59	0.32	0.32	0.32
Observations	1,998	1,998	1,998	1,410	1,410	1,410	588	588	588
R-squared	0.08	0.10	0.11	0.03	0.05	0.06	0.06	0.11	0.13
B) Black Comp.									
Distance (mi.)	-0.007	-0.008	-0.006	-0.010	-0.010	-0.009	0.002	0.001	0.002
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.012)	(0.013)	(0.014)
Mean Dependent Variable	0.49	0.49	0.49	0.57	0.57	0.57	0.27	0.27	0.27
Observations	1,411	1,411	1,411	1,021	1,021	1,021	390	390	390
R-squared	0.10	0.11	0.12	0.04	0.05	0.06	0.09	0.12	0.14
C) Hispanic Comp.									
Distance (mi.)	-0.004	-0.013	-0.017	-0.005	-0.012	-0.013	0.016	0.013	0.108
	(0.017)	(0.018)	(0.018)	(0.021)	(0.023)	(0.021)	(0.043)	(0.055)	(0.074)
Mean Dependent Variable	0.56	0.56	0.56	0.62	0.62	0.62	0.34	0.34	0.34
Observations	226	226	226	176	176	176	50	50	50
R-squared	0.15	0.20	0.23	0.12	0.20	0.26	0.51	0.70	0.82
D) White Comp.									
Distance (mi.)	0.008	0.007	0.009	0.003	0.001	0.002	-0.003	0.003	0.002
	(0.015)	(0.016)	(0.015)	(0.022)	(0.022)	(0.021)	(0.024)	(0.028)	(0.026)
Mean Dependent Variable	0.56	0.56	0.56	0.62	0.62	0.62	0.46	0.46	0.46
Observations	313	313	313	183	183	183	130	130	130
R-squared	0.21	0.25	0.26	0.18	0.24	0.28	0.38	0.41	0.45
Complainant's Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Officers' Controls	No	No	Yes	No	No	Yes	No	No	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of placebo-distance on the probability of signed affidavit from January 2011 and November 2011. The placebo policy occurred in June 2011. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search). The specification controls for complainants and incident characteristics, incident location, district fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all the complainants, Black complainants, Hispanic, and White complainants.Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value $<$ 0.10, **p-value $<$ 0.05, *** p-value $<$ 0.01.

Table 1.8: Effect of distance on complaint outcomes

	Number of complaints		Number of signed affidavits		Share of signed affidavits		Share of sustained	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) All Beats								
Distance (mi.)	-0.003***	-0.003***	-0.003***	-0.003***	-0.014***	-0.014***	0.004	0.004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.003)	(0.003)
Mean Dependent Variable	0.09	0.09	0.04	0.04	0.45	0.45	0.11	0.11
Observations	11,395	11,395	11,395	11,395	4,653	4,653	2,513	2,513
R-squared	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02
B) Black Beats								
Distance (mi.)	-0.003**	-0.003*	-0.003**	-0.003***	-0.022***	-0.023***	0.010*	0.010*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.007)	(0.007)	(0.006)	(0.005)
Mean Dependent Variable	0.15	0.15	0.06	0.06	0.44	0.44	0.09	0.09
Observations	5,074	5,074	5,074	5,074	2,644	2,644	1,427	1,427
R-squared	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03
C) Hispanic Beats								
Distance (mi.)	-0.001	-0.001	-0.001	-0.001	-0.003	-0.005	0.015	0.014
	(0.001)	(0.001)	0.000	(0.001)	(0.015)	(0.015)	(0.010)	(0.010)
Mean Dependent Variable	0.04	0.04	0.02	0.02	0.46	0.46	0.12	0.12
Observations	1,978	1,978	1,978	1,978	669	669	354	354
R-squared	0.01	0.01	0.01	0.02	0.04	0.04	0.03	0.03
D) White Beats								
Distance (mi.)	0.003	0.003	-0.001*	-0.001	-0.029*	-0.032**	0.003	0.004
	(0.002)	(0.002)	(0.001)	(0.001)	(0.014)	(0.015)	(0.011)	(0.013)
Mean Dependent Variable	0.05	0.05	0.03	0.03	0.49	0.49	0.13	0.13
Observations	2,795	2,795	2,795	2,795	800	800	444	444
R-squared	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.03
District's Controls	No	Yes	No	Yes	No	Yes	No	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on complaint outcomes from January 2011 and July 2014. The complaint outcomes are the number of complaints per 1,000 capita (columns 1-2), the number of complaints with signed affidavit per 1,000 capita (columns 3-4), the share of complaints with signed affidavit (columns 4-6), and the share of complaints that are sustained (columns 7-8). The specification controls for beat fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all beats, beats with a majority (>50 percent) of Black residents, beats with a majority Hispanic residents, and beats with a majority White residents. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.9: Effect of distance on use of force outcomes

	Number of TRR		Any use of high force		Force per arrest		Number of injuries	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) All Beats								
Distance (mi.)	0.002	0.002	0.002	0.002	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.003)	(0.003)	0.000	0.000	0.000	0.000
Mean Dependent Variable	0.17	0.17	0.34	0.34	0.04	0.04	0.04	0.04
Observations	11,395	11,395	11,395	11,395	11,390	11,390	11,395	11,395
R-squared	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01
B) Black Beats								
Distance (mi.)	0.004	0.005	0.004	0.006	0.001	0.001**	0.002**	0.002**
	(0.003)	(0.003)	(0.005)	(0.006)	(0.001)	(0.001)	(0.001)	(0.001)
Mean Dependent Variable	0.26	0.26	0.43	0.43	0.04	0.04	0.06	0.06
Observations	5,074	5,074	5,074	5,074	5,074	5,074	5,074	5,074
R-squared	0.03	0.03	0.01	0.01	0.01	0.02	0.01	0.01
C) Hispanic Beats								
Distance (mi.)	0.004***	0.004***	0.018***	0.016**	0.001**	0.001**	0.001*	0.001
	(0.001)	(0.001)	(0.006)	(0.006)	(0.001)	0.000	0.000	(0.001)
Mean Dependent Variable	0.07	0.07	0.29	0.29	0.03	0.03	0.02	0.02
Observations	1,978	1,978	1,978	1,978	1,978	1,978	1,978	1,978
R-squared	0.02	0.03	0.02	0.03	0.01	0.01	0.01	0.01
D) White Beats								
Distance (mi.)	0.007	0.008	0.009	0.010	0.002	0.002	0.002	0.002
	(0.005)	(0.005)	(0.006)	(0.006)	(0.001)	(0.001)	(0.002)	(0.002)
Mean Dependent Variable	0.11	0.11	0.26	0.26	0.05	0.05	0.03	0.03
Observations	2,795	2,795	2,795	2,795	2,790	2,790	2,795	2,795
R-squared	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
District's Controls	No	Yes	No	Yes	No	Yes	No	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on use of force outcomes from January 2011 and July 2014. The use of force outcomes are the the number of TRRs per 1,000 capita (columns 1-2), whether or not there was incident with high level of force (columns 3-4), i.e. involving use of Taser or firearm, the number of incidents involving reported force per arrest (columns 5-6), and the number of civilian injuries per 1,000 capita (columns 7-8) . The specification controls for beat fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all beats, beats with a majority (>50 percent) of Black residents, beats with a majority Hispanic residents, and beats with a majority White residents. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value $<$ 0.10, **p-value $<$ 0.05, *** p-value $<$ 0.01.

Table 1.10: Effect of distance on crime outcomes

	Index crimes			Non Index crimes		
	Offenses	Arrests	Clearance	Offenses	Arrests	Clearance
	(1)	(2)	(3)	(4)	(5)	(6)
A) All Beats						
Distance (mi.)	-0.071***	0.005	0.001***	-0.098***	-0.047**	0.001
	(0.013)	(0.005)	0.000	(0.027)	(0.023)	(0.001)
Mean Dependent Variable	5.90	0.73	0.11	8.32	3.59	0.37
Observations	11,395	11,395	11,395	11,395	11,395	11,395
R-squared	0.19	0.02	0.01	0.11	0.03	0.03
B) Black Beats						
Distance (mi.)	-0.046*	0.007	0.001**	-0.194**	-0.120	0.000
	(0.027)	(0.007)	(0.001)	(0.094)	(0.094)	(0.003)
Mean Dependent Variable	7.17	0.75	0.10	12.84	5.86	0.43
Observations	5,074	5,074	5,074	5,074	5,074	5,074
R-squared	0.37	0.04	0.02	0.18	0.05	0.05
C) Hispanic Beats						
Distance (mi.)	0.001	0.005	0.001	-0.027	-0.024	-0.001
	(0.013)	(0.005)	(0.002)	(0.027)	(0.024)	(0.002)
Mean Dependent Variable	3.13	0.40	0.12	4.41	1.74	0.37
Observations	1,978	1,978	1,978	1,978	1,978	1,978
R-squared	0.34	0.03	0.03	0.31	0.18	0.05
D) White Beats						
Distance (mi.)	-0.060	-0.075	-0.001	0.104	-0.010	-0.007**
	(0.049)	(0.062)	(0.002)	(0.102)	(0.017)	(0.003)
Mean Dependent Variable	6.76	1.09	0.11	5.01	1.86	0.30
Observations	2,795	2,795	2,795	2,795	2,795	2,795
R-squared	0.10	0.03	0.01	0.01	0.01	0.02
District's Controls	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on outcomes from January 2011 and July 2014. Offenses and arrests are expressed per 1,000 capita, and clearance rates are define as the number of reported crimes with an arrest over the number of reported crimes. The specification controls for beat fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all beats, beats with a majority (>50 percent) of Black residents, beats with a majority Hispanic residents, and beats with a majority White residents. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.11: Probability to sign the affidavit parameter estimates

	Serious Allegations		Failure to Provide Service	
Variables	Coeff (Std. Err)	AME	Coeff (Std. Err)	AME
Observables				
Distance (mi.)	-0.035 (0.036)	-0.008	-0.092 (0.030)***	-0.022
Male	-0.231 (0.140)*	-0.054	0.005 (0.130)	0.001
Black	-0.648 (0.231)***	-0.151	-1.015 (0.179)***	-0.237
Hispanic/Other	-0.219 (0.291)	-0.051	-0.510 (0.223)**	-0.119
Unknown race	-2.716 (0.391)***	-0.635	-1.061 (0.386)***	-0.248
30-39yo	0.252 (0.217)	0.059	0.585 (0.292)**	0.137
40-49yo	0.640 (0.232)***	0.149	0.973 (0.285)***	0.227
50-59yo	1.266 (0.252)***	0.296	1.191 (0.291)***	0.278
60-74yo	1.006 (0.329)***	0.235	1.376 (0.311)***	0.321
>74yo/missing	-0.101 (0.276)	-0.024	0.615 (0.309)**	0.144
Median age of the PO	0.026 (0.011)**	0.006	0.006 (0.009)	0.001
Any non PO	0.124 (0.159)	0.029	0.252 (0.137)*	0.059
Any black PO	0.207 (0.227)	0.048	0.249 (0.201)	0.058
Any hispanic PO	-0.112 (0.182)	-0.026	0.275 (0.179)	0.064
Any white PO	0.003 (0.214)	0.001	-0.011 (0.195)	-0.003
Number of PO	0.109 (0.049)**	0.025	-0.134 (0.091)	-0.031
Public Location	-0.531 (0.269)**	-0.124	-1.241 (0.195)***	-0.29

Notes: Notes: See next page.

Table 1.11, continued

	Serious Allegations		Failure to Provide Service	
Variables	Coeff (Std. Err)	AME	Coeff (Std. Err)	AME
Unobserved Heterogeneity				
κ_1	0.782 (0.341)**	---	0.782 (0.341)**	---
κ_2	0.694 (0.525)	---	0.694 (0.525)	---
Type 1: $\mu^D_{k,jb}$	2.620 (0.799)***	---	0.826 (0.706)	---
Type 2: $\mu^D_{k,jb}$	-2.421 (0.871)***	---	0.360 (0.720)	---
Type 3: $\mu^D_{k,jb}$	-1.021 (1.013)	---	-1.253 (0.796)	---
N	4,303		1,986	
llk		-11845.118		

Notes: This Table presents the set of estimates on the probability of signed affidavit. The sample considers complaint that were filed between January 2011 and July 2014, with non missing investigator, and only one investigator assigned. The parameter estimates are based on the specification depicted in equations 1.5 and 1.9 in the text. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search). The specification controls for district and quarter fixed effects, but not reported. The probabilities for the points of support are given by $\pi_g = \exp(\kappa_g)/(1 + \exp(\kappa_1) + \exp(\kappa_2))$ for $g = \{1, 2\}$. For interpretation of the coefficients, I report the average marginal effect (AME). Standard errors are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.12: Probability to sustain a complaint conditional on signing the affidavit parameter estimates

	Serious Allegations		Failure to Provide Service	
Variables	Coeff (Std. Err)	AME	Coeff (Std. Err)	AME
Observables				
Investigator Experience	-0.027 (0.019)	-0.006	0.066 (0.061)	0.015
log(duration of inv.)	2.846 (0.447)***	0.665	7.080 (2.154)***	1.655
log(duration of inv.)ˆ2/10	-3.976 (2.239)*	-0.929	-35.285 (22.887)	-8.247
log(duration of inv.)ˆ3/100	-42.804 (15.642)***	-10.004	123.184 (240.800)	28.79
CPD Investigator	2.263 (0.547)***	0.529	2.720 (2.646)	0.636
Other Type of Investigator	1.779 (0.391)***	0.416	2.991 (2.633)	0.699
Male	0.381 (0.264)	0.089	0.649 (0.859)	0.152
Black	-3.201 (0.382)***	-0.748	-2.148 (1.326)	-0.502
Hispanic/Other	-1.239 (0.315)***	-0.29	0.652 (1.277)	0.152
Unknown race	-1.769 (1.328)	-0.413	-4.430 (3.554)	-1.035
Median age of the PO	0.040 (0.019)**	0.009	-0.058 (0.060)	-0.014
Any non PO	-0.277 (0.319)	-0.065	-0.721 (0.934)	-0.168
Any black PO	0.947 (0.469)**	0.221	1.296 (1.680)	0.303
Any hispanic PO	0.622 (0.432)	0.145	0.332 (1.679)	0.078
Any white PO	0.248 (0.420)	0.058	0.015 (1.747)	0.004
Number of PO	-0.512 (0.180)***	-0.12	-0.465 (0.791)	-0.109

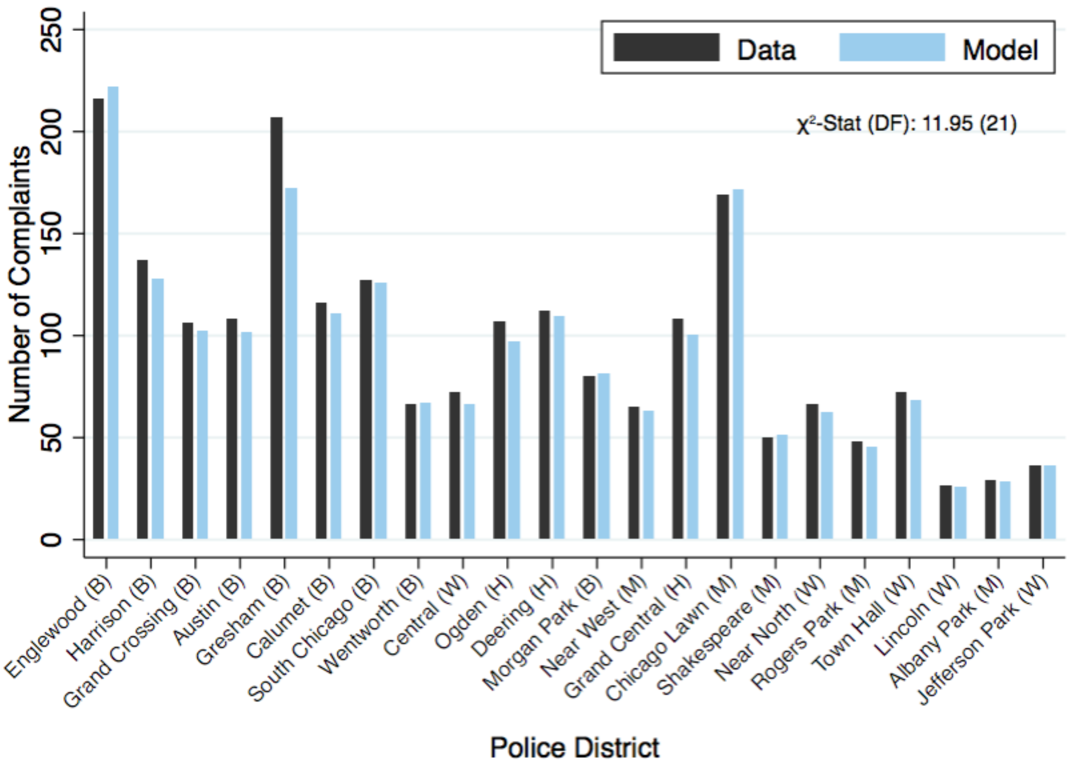
Notes: Notes: See next page.

Table 1.12, continued

	Serious Allegations		Failure to Provide Service	
Variables	Coeff (Std. Err)	AME	Coeff (Std. Err)	AME
Unobserved Heterogeneity				
κ_1	0.782 (0.341)**	---	0.782 (0.341)**	---
κ_2	0.694 (0.525)	---	0.694 (0.525)	---
Type 1: $\mu^s_{k,jb}$	-3.122 (1.142)***	---	-4.659 (4.600)	---
Type 2: $\mu^s_{k,jb}$	-1.988 (1.932)	---	-3.011 (5.099)	---
Type 3: $\mu^s_{k,jb}$	-2.639 (2.118)	---	7.480 (6.907)	---
N	2,123		618	
llk		-11845.118		

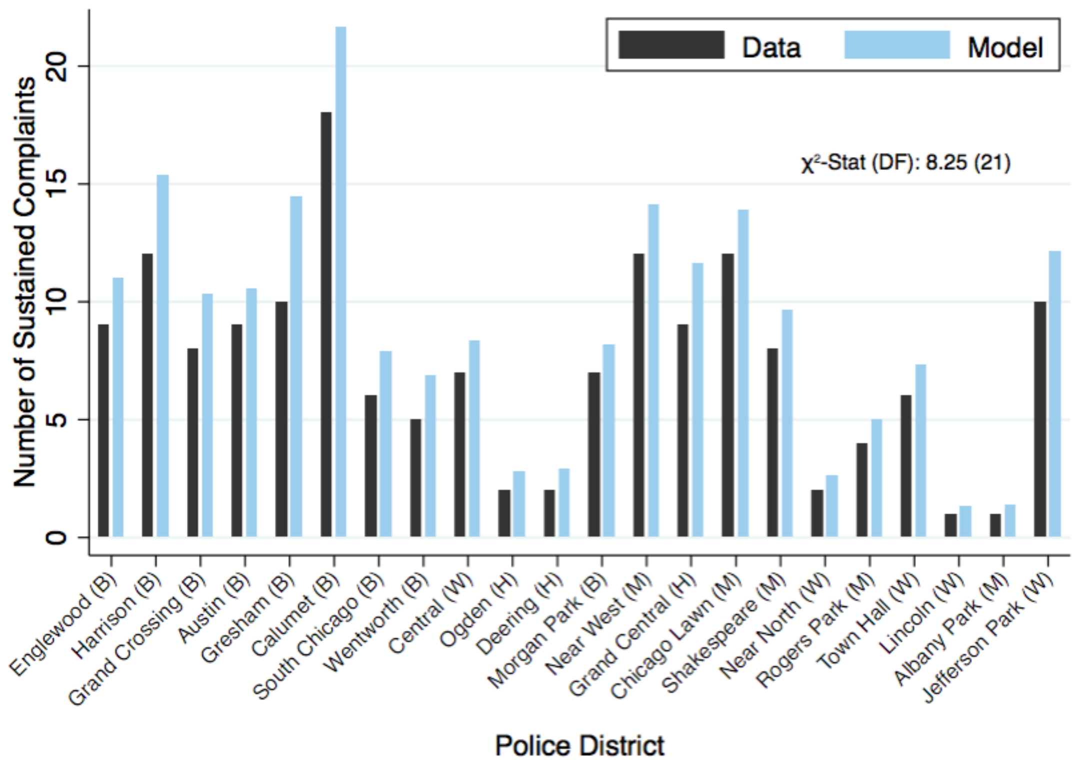
Notes: This Table presents the set of estimates on the probability that the investigator sustain the complaint Conditional on the complainant signed the affidavit. The sample considers complaint that were filed between January 2011 and July 2014, with non missing investigator, and only one investigator assigned. The parameter estimates are based on the specification depicted in equations 1.7 and 1.9 in the text. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search). The specification controls for district and quarter fixed effects, but not reported. There are three types of investigator: police officer, investigator from the oversight agency (reference category), and other type of investigator (City of Chicago employees, FBI, ...).The probabilities for the points of support are given by $\pi_g = \exp(\kappa_g)/(1 + \exp(\kappa_1) + \exp(\kappa_2))$ for $g = \{1, 2\}$. For interpretation of the coefficients, I report the average marginal effect (AME). Standard errors are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Figure 1.7: Model versus Data: Signed affidavit for serious allegations



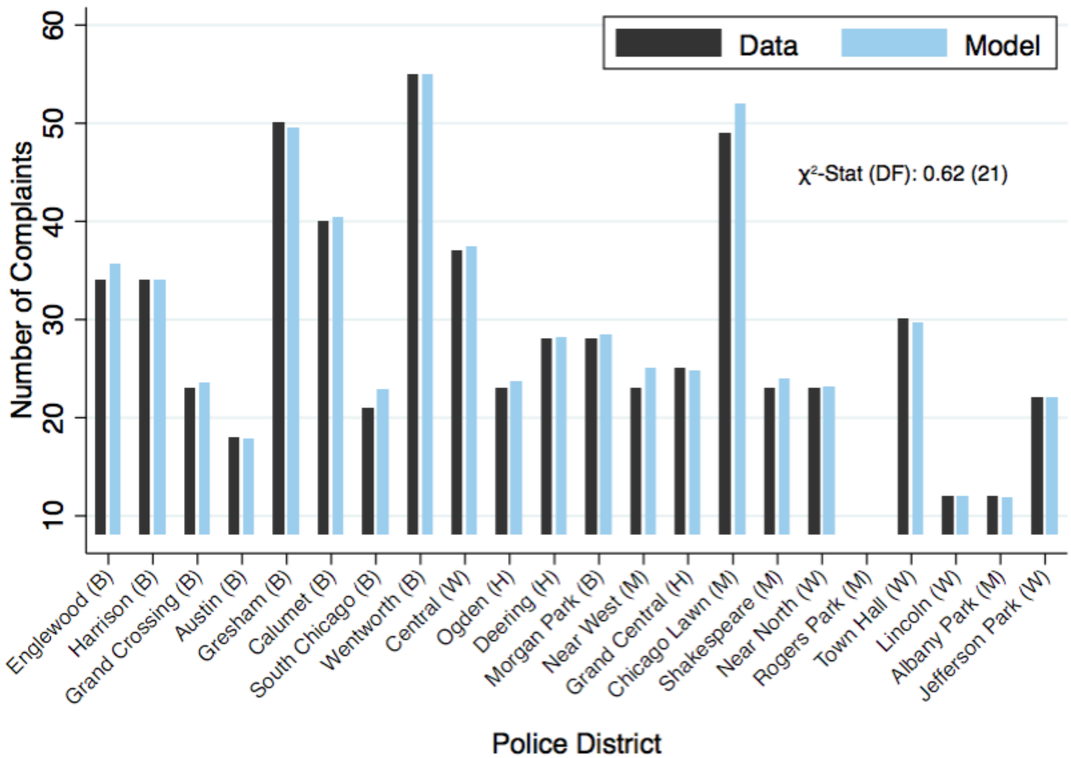
Notes: This figure presents the frequencies distribution of the signed complaints by police districts from both the model and the data. Predictions from the model are based on the results from tables1.11 and 1.12. The critical value from a Chi-Squared distribution with 21 degrees of freedom at the 10 percent level of confidence is 29.6. The police Districts (x-axis) are ordered from the most to the least violent regarding reported crime per 1,000 capita. The majority racial-ethnic group for each district is reported in parenthesis (Black (B),Hispanic (H),White (W), No majority (M)).

Figure 1.8: Model versus Data: Sustained complaint for serious allegations|Signed



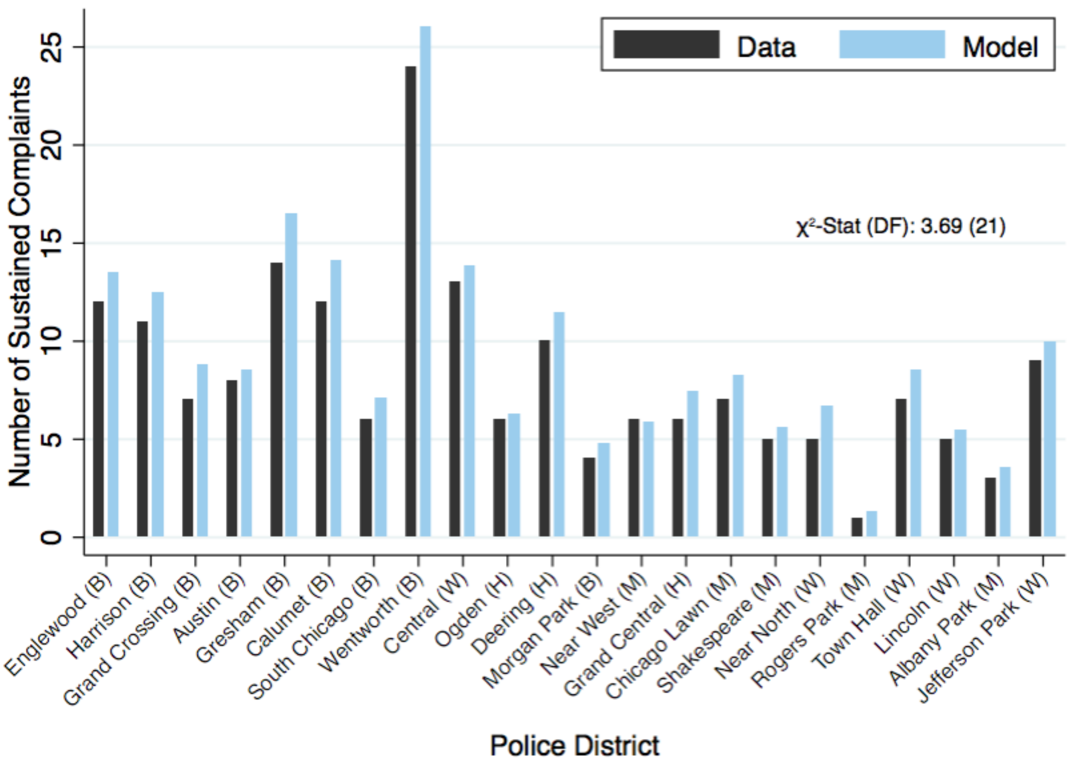
This figure presents the frequencies distribution of the sustained complaints (conditional on being signed) by police districts from both the model and the data. Predictions from the model are based on the results from tables1.11 and 1.12. The critical value from a Chi-Squared distribution with 21 degrees of freedom at the 10 percent level of confidence is 29.6. The police Districts (x-axis) are ordered from the most to the least violent regarding reported crime per 1,000 capita. The majority racial-ethnic group for each district is reported in parenthesis (Black (B),Hispanic (H),White (W), No majority (M)).

Figure 1.9: Model versus Data: Signed affidavit for FPS allegations



Notes: This figure presents the frequencies distribution of the signed complaints by police districts from both the model and the data. Predictions from the model are based on the results from tables1.11 and 1.12. The critical value from a Chi-Squared distribution with 21 degrees of freedom at the 10 percent level of confidence is 29.6. The police Districts (x-axis) are ordered from the most to the least violent regarding reported crime per 1,000 capita. The majority racial-ethnic group for each district is reported in parenthesis (Black (B),Hispanic (H),White (W), No majority (M)).

Figure 1.10: Model versus Data: Sustained complaint for FPS allegations|Signed



Notes: This figure presents the frequencies distribution of the sustained complaints (conditional on being signed) by police districts from both the model and the data. Predictions from the model are based on the results from tables1.11 and 1.12. The critical value from a Chi-Squared distribution with 21 degrees of freedom at the 10 percent level of confidence is 29.6. The police Districts (x-axis) are ordered from the most to the least violent regarding reported crime per 1,000 capita. The majority racial-ethnic group for each district is reported in parenthesis (Black (B),Hispanic (H),White (W), No majority (M)).

Table 1.13: Complaint outcomes by complainant's race

	Serious		Failure to Provide Service	
	$Pr(Signed)$	$Pr(Sustained \mid Signed)$	$Pr(Signed)$	$Pr(Sustained \mid Signed)$
All	49.3%	7.5%	31.1%	29.3%
N	4303	2123	1986	618
Black	48.8%	2.7%	24.9%	16.0%
N	3081	1505	1231	307
Hispanic	55.1%	11.1%	34.4%	36.9%
N	508	280	244	84
White	57.4%	30.4%	48.3%	46.5%
N	493	283	441	213
Unknown race	24.9%	5.5%	20.0%	14.3%
N	221	55	70	14

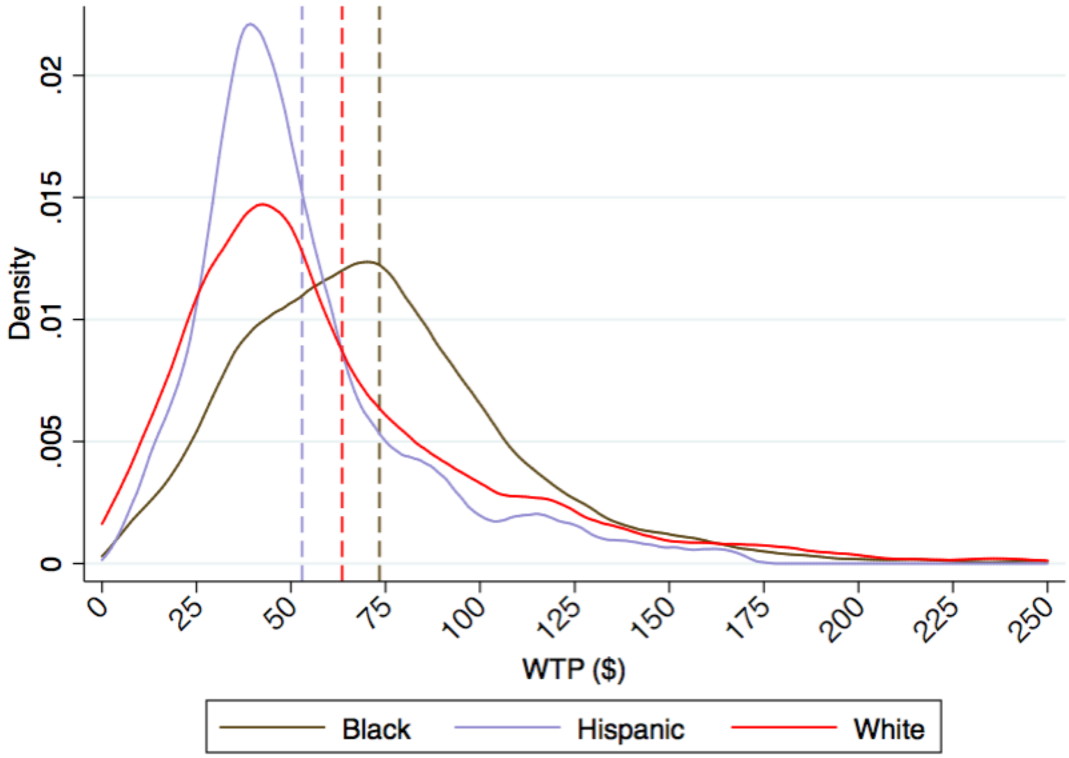
Notes: This table reports the probability to sign the affidavit and the probability that the complaint is sustained given that the affidavit is signed. The sample considers complaint that were filed between January 2011 and July 2014, with non missing investigator, and only one investigator assigned. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search).

Table 1.14: Wage per hour

	N	Mean ($)	Std. Dev
All	6,289	19.6	9.2
Black	4,312	17.7	8.1
Hispanic	752	22.4	8.4
White	934	25.6	10.6
Unknown race	291	21.4	10.2

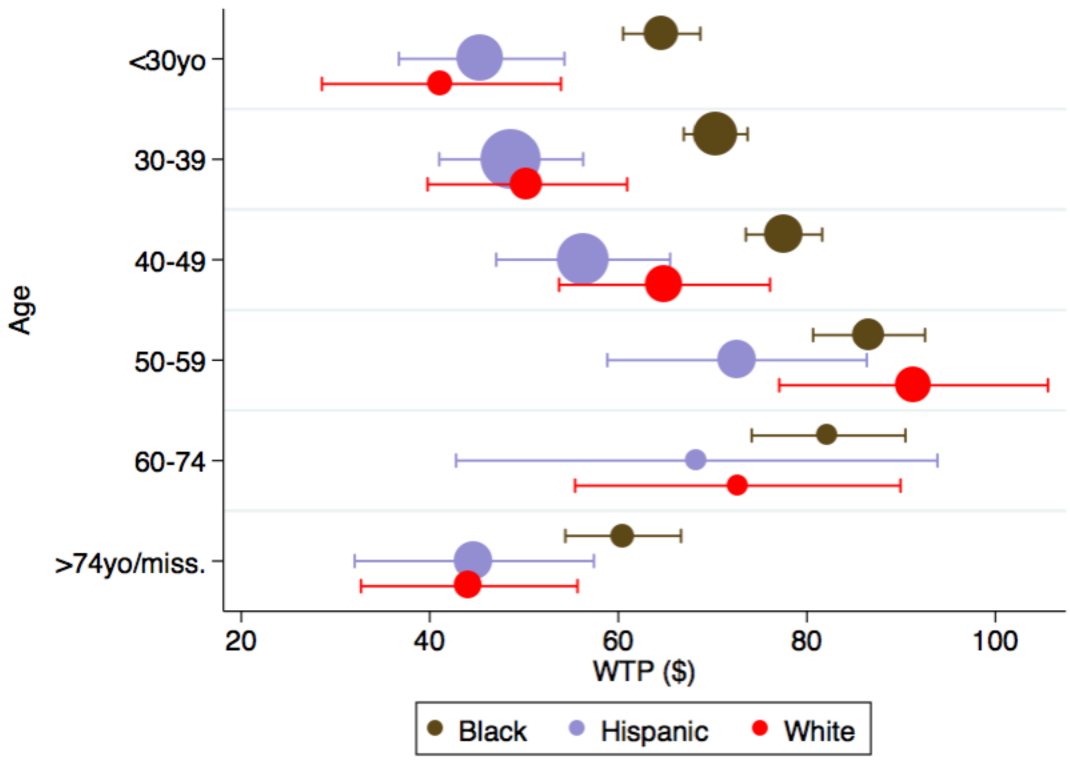
Notes: This table reports the average wage per hour (average annual income/(40 hours $\times$52 weeks)) in the beat where the incident occurred. The income is computed by using the 2010-2014 American Community Survey (ACS) data.

Figure 1.11: Kernel density for willingness to pay (WTP) serious allegations



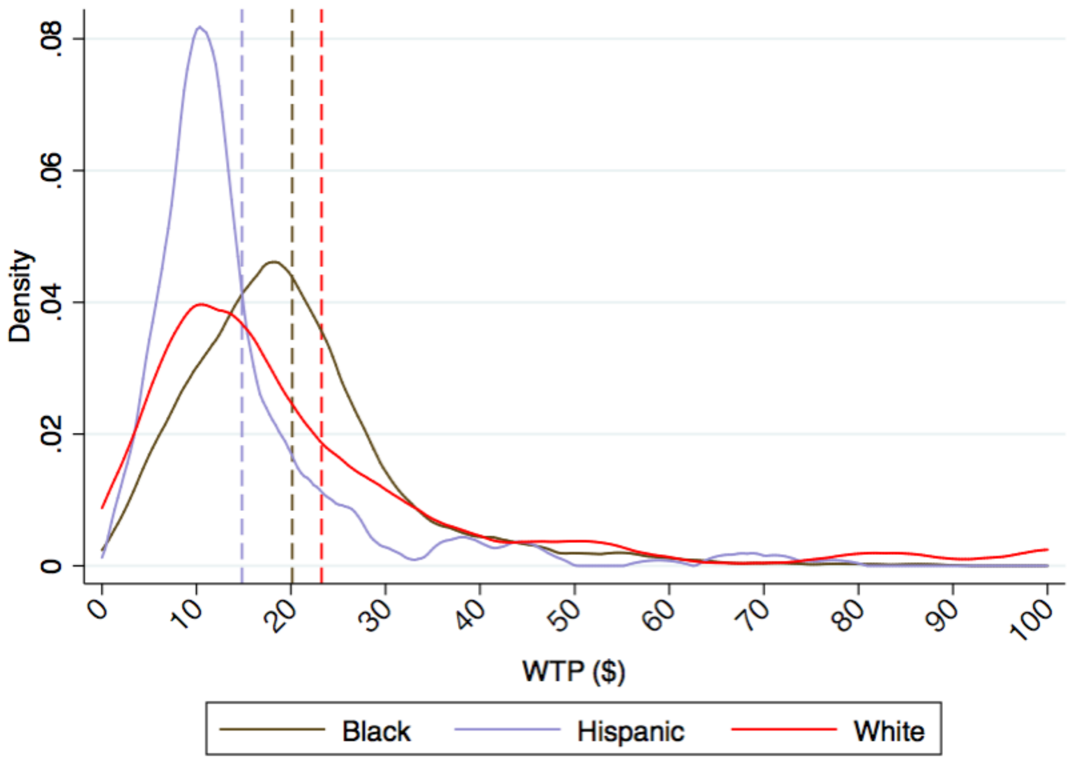
Notes: This figure presents the estimated distribution of the willingness to pay race of the complainants. Willingness to pay are computed using equation 1.13. Predictions from the model are based on the results from tables 1.11 and 1.12.The dashed lines in figures 1.11 and 1.13 represent the average willingness to pay by racial-ethnic group.

Figure 1.12: Average willingness to pay (WTP) by allegation type, race and age groups for serious allegations



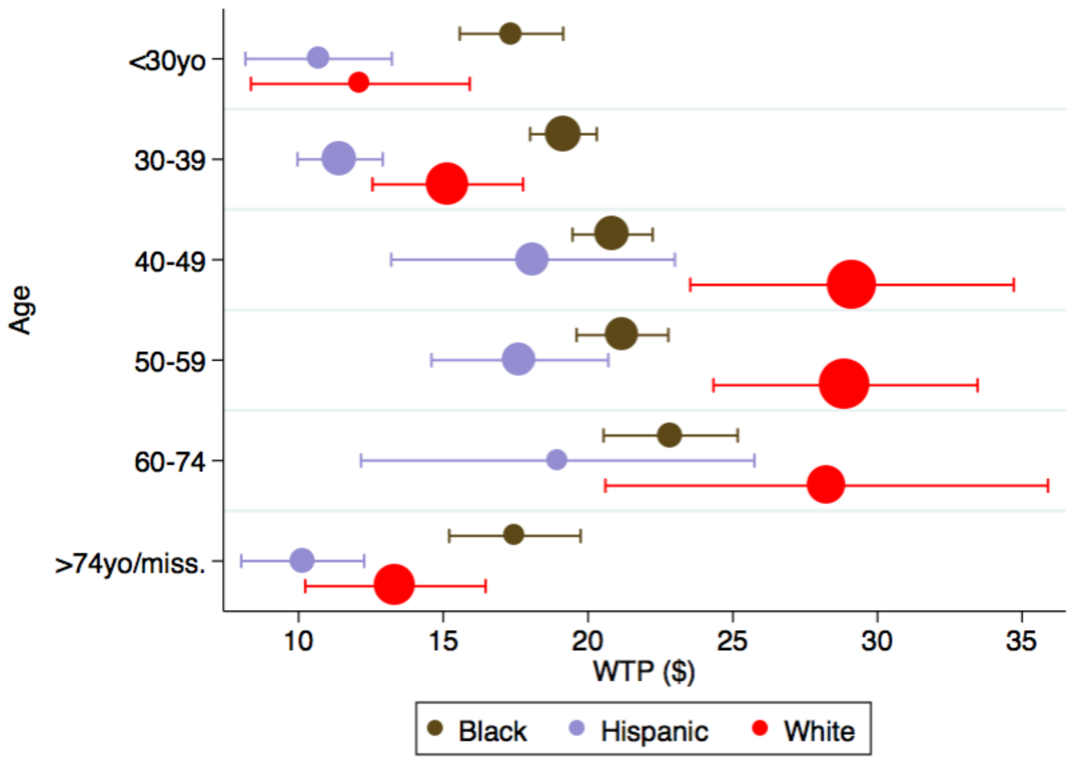
Notes: This figure presents the average willingness to pay by allegation type, race and age groups of the complainants. Willingness to pay are computed using equation 1.13. Predictions from the model are based on the results from tables 1.11 and 1.12. Conditional on race of the complainant, the area of each circle is proportional to the age group weights. The Confidence Intervals are computed at the 95 percent level and accounts for estimation uncertainties.

Figure 1.13: Kernel density for willingness to pay (WTP) FPS allegations



Notes: This figure presents the estimated distribution of the willingness to pay race of the complainants. Willingness to pay are computed using equation 1.13. Predictions from the model are based on the results from tables 1.11 and 1.12.The dashed lines in figures 1.11 and 1.13 represent the average willingness to pay by racial-ethnic group.

Figure 1.14: Average willingness to pay (WTP) by allegation type, race and age groups for FPS allegations



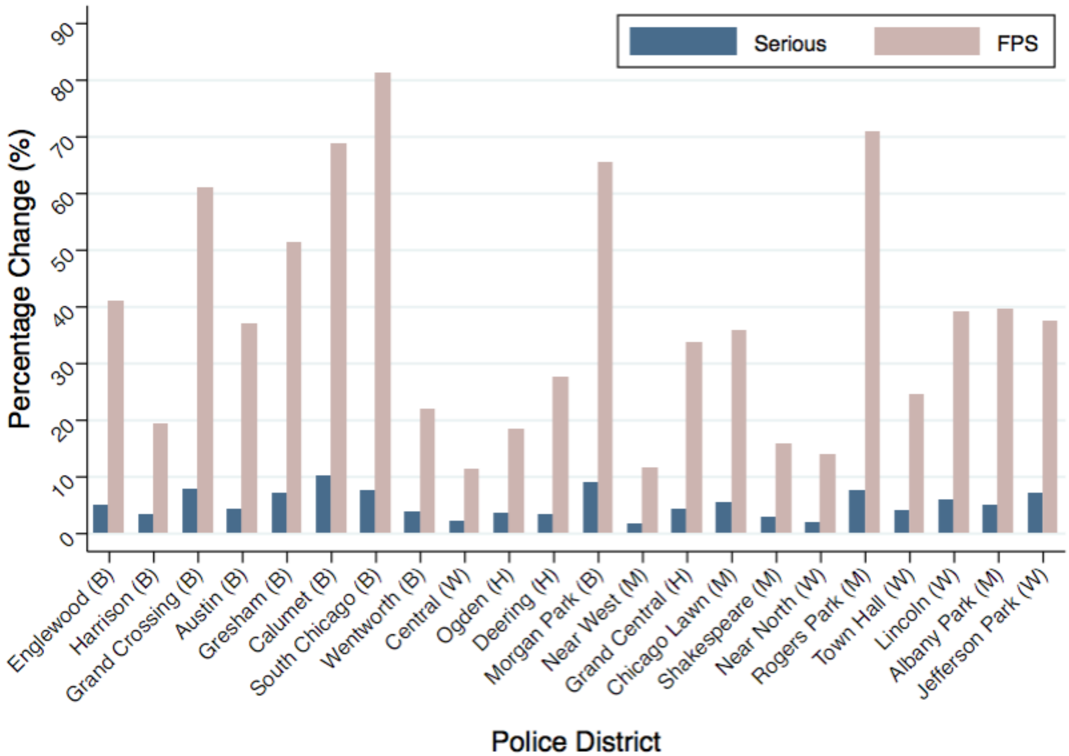
Notes: This figure presents the average willingness to pay by allegation type, race and age groups of the complainants. Willingness to pay are computed using equation 1.13. Predictions from the model are based on the results from tables 1.11 and 1.12. Conditional on race of the complainant, the area of each circle is proportional to the age group weights. The Confidence Intervals are computed at the 95 percent level and accounts for estimation uncertainties.

Table 1.15: Costs and benefits of signing the affidavit for the complainant

	Serious			Failure to Provide Service		
	WTP $	Time Hours	Sustained Rates	WTP $	Time Min.	Sustained Rates
Black	68.6	3.9	2.7 %	18.4	60	16.0%
Hispanic	45.3	2.0	11.1%	11.5	30	36.9%
White	49.5	2.0	30.4%	15.5	36	46.5%

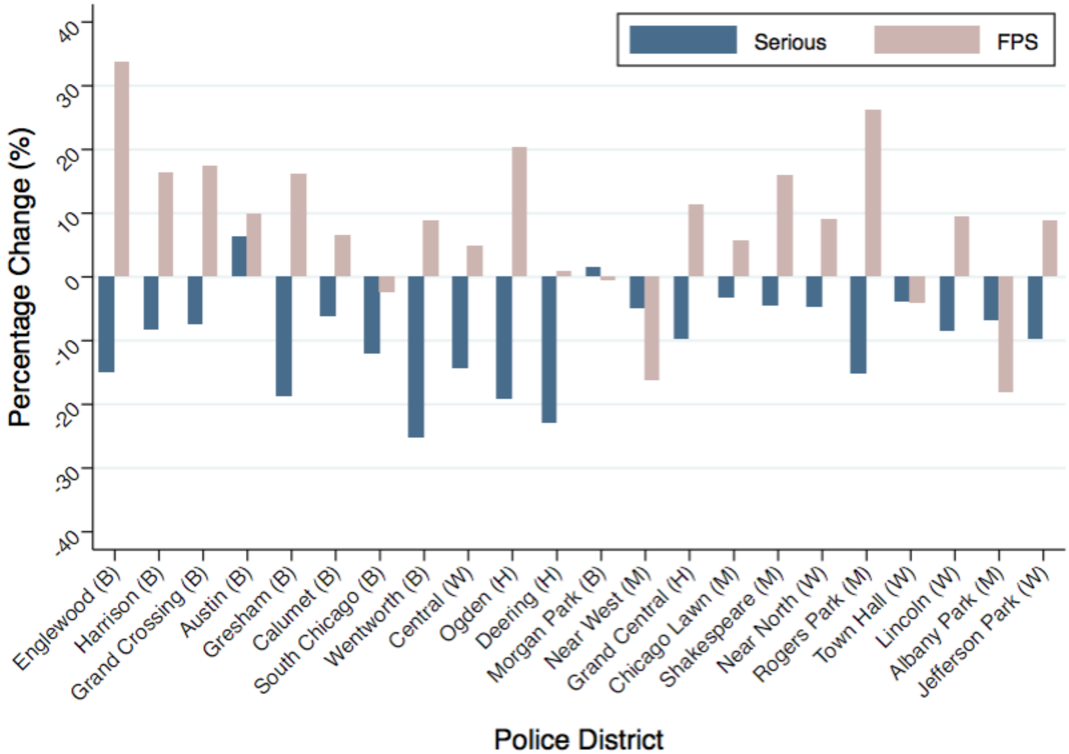
Notes: This table reports costs and benefits of signing the affidavit for the complainant: sustained rates, median willingness to pay (WTP) and time sacrificed to sign the affidavit by race of the complainant and type of complaints.

Figure 1.15: Counterfactuals for signed affidavits



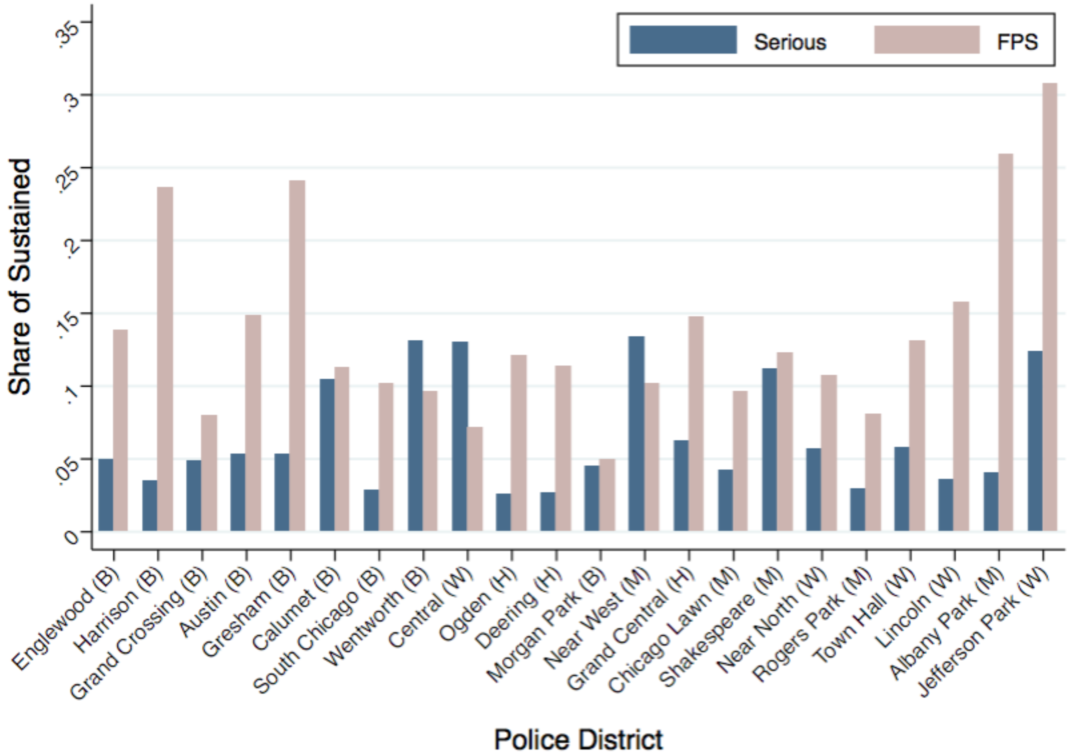
Notes: This figure presents the effect of a policy that does not require to travel to sign the affidavit at the oversight agency for both serious and failure to provide service (FPS) allegations. The y-axis presents the percentage change in sign affidavit when the alternative policy is implemented relative to the predictions from the model. Predictions from the model are based on the results from tables 1.11 and 1.12. The police Districts (x-axis) are ordered from the most to the least violent regarding reported crime per 1,000 capita. The majority racial-ethnic group for each district is reported in parenthesis (Black (B),Hispanic (H),White (W), No majority (M)).

Figure 1.16: Counterfactuals for sustained complaint|Signed in the data



Notes: This figure presents the effect of a policy that does not require to travel to sign the affidavit at the oversight agency for both serious and failure to provide service (FPS) allegations. The y-axis presents the percentage change in sustained rates, conditional on signing the affidavit (observed and predicted), when the alternative policy is implemented relative to the predictions from the model. This figure restricts the sample to complaints that were both: (i) signed in the data, and (ii) signed according to the counterfactual policy. Predictions from the model are based on the results from tables 1.11 and 1.12. The police Districts (x-axis) are ordered from the most to the least violent regarding reported crime per 1,000 capita. The majority racial-ethnic group for each district is reported in parenthesis (Black (B),Hispanic (H),White (W), No majority (M)).

Figure 1.17: Counterfactuals for sustained complaint|Not signed in the data



Notes: This figure presents the effect of a policy that does not require to travel to sign the affidavit at the oversight agency for both serious and failure to provide service (FPS) allegations. The y-axis presents the probability of sustaining a complaint, conditional on signing the affidavit (counterfactual), when the alternative policy is implemented relative to the predictions from the model. This figure considers complaints that were not signed in the data, but with signed counterfactual. Predictions from the model are based on the results from tables 1.11 and 1.12. The police Districts (x-axis) are ordered from the most to the least violent regarding reported crime per 1,000 capita. The majority racial-ethnic group for each district is reported in parenthesis (Black (B), Hispanic (H), White (W), No majority (M)).

1.10 Appendix A: Data

1.10.1 Sample Selection

Table 1.16: Sample construction for Section 1.5

Step	Description	Number of observations	Number of complainants	Number of complaints
1	Raw	48214	48214	47042
2	Keep if only one complainant	45956	45956	45956
3	Drop if missing location	43360	43360	43360
4	Keep if complaint/incident occurred after December 2010	18489	18489	18489
5	Keep if complaint/incident occurred before January 2015	16187	16187	16187
6	Keep if Investigated	9083	9083	9083
7	Keep if serious/FPS incident	7211	7211	7211
8	Sample if complaint/incident occurred before August 2014	6763	6763	6763

Table 1.17: Sample construction for Sections 1.6-1.8

Step	Description	Number of observations	Number of complainants	Number of complaints
1	Data from the empirical analysis	6763	6763	6763
2	Keep if only one investigator	6760	6760	6760
3	Drop if missing investigator	6639	6639	6639
4	Drop if missing tenure of investigator	6296	6296	6296
5	Drop if missing investigation duration	6289	6289	6289

1.10.2 Complaints classification

Table 1.18: Allegation categories

Classification	Allegation name
1.Use of Force/Verbal Abuse	01A-USE OF PROFANITY
(Civilian Complaints)	01B-RACIAL/ETHNIC, ETC.
	01C-MISCELLANEOUS
	03E-INJURY/DEATH (UNDER COLOR OF LAW)
	04H-PROPER CARE, INJURY/DEATH
	05A-ARRESTEE - DURING ARREST
	05B-ARRESTEE - AFTER ARREST, PRIOR TO LOCKUP
	05C-ARRESTEE - LOCKUP/DETENTION
	05D-NO ARREST
	05E-TRAFFIC
	05F-DOMESTIC
	05G-WEAPON, USE/DISPLAY OF
	05H-MISCELLANEOUS
	05J-""""U"""" CONVERTED TO C.R.
	(RECORDS KEEPING ONLY, INITIAL)
	05K-DOMESTIC ALTERCATION/INCIDENT - OFF DUTY
	05L-UNNECESSARY PHYSICAL CONTACT - ON DUTY
	05M-UNNECESSARY PHYSICAL CONTACT - OFF DUTY
	05N-WEAPON - UNNECESSARY DISPLAY OF
	05P-EXCESSIVE FORCE - OFF DUTY
	(INCLUDES NEIGHBOR, TRAFFIC, TAV)
	05Q-CIVIL SUIT - THIRD PARTY
	05T-EXCESSIVE FORCE - Taser - USE OF

Table 1.18, continued

Classification	Allegation name
2.Arrest/Locked up	04E-PRISONER'S PROPERTY - INVENTORY/RECEIPT
(Civilian Complaints)	04B-ARREST/IMPROPER
	04A-BONDING/BOOKING/PROCESSING
	04D-SEARCH, PERSON/PROPERTY
	04F-ESCAPE
	04J-MISCELLANEOUS
	04G-TELEPHONE -
	ATTORNEY/RELATIVE PRIVILEGES
	04C-EXCESSIVE DETENTION
3.Search	03A-FIRST AMENDMENT
(Civilian Complaints)	03B-SEARCH OF PERSON WITHOUT WARRANT
	03C-SEARCH OF PREMISE/VEHICLE WITHOUT WARRANT
	03D-ILLEGAL ARREST
	03F-FAILURE TO INSURE
	03G-MISCELLANEOUS
	03P-RACIAL PROFILING
	(ADVOCATE USE ON CLOSING ONLY)
4.Failure to Provide Service	10J-NEGLECT OF DUTY/CONDUCT UNBECOMING - ON DUTY
(Civilian Complaints)	10U-INADEQUATE/FAILURE TO PROVIDE SERVICE

Table 1.18, continued

Classification	Allegation name
5.Operation and Personnel Violations	07A-MISCONDUCT DURING ISSUANCE OF CITATION
	07B-IMPROPER PROCESSING/REPORTING/PROCEDURES
	07C-VIOLATION (OTHER THAN D.U.I.) - ON DUTY
	07D-PARKING COMPLAINTS
	07E-FAIL TO ENFORCE TRAFFIC REGULATIONS
	07F-MISCELLANEOUS
	07T-PREVENTable TRAFFIC ACCIDENT
	10A-ABSENT WITHOUT PERMISSION
	10B-MEDICAL ROLL
	10C-COMPENSATORY TIME
	10D-COMMUNICATION OPERATIONS PROCEDURES
	10E-SECONDARY/SPECIAL EMPLOYMENT
	10F-COURT IRREGULARITIES
	10G-UNFIT FOR DUTY
	10H-LEAVING ASSIGNMENT (DISTRICT, BEAT, SECTOR, COURT)
	10K-LATE - ROLL CALL/ASSIGNMENT/COURT
	10L-WEAPON/AMMUNITION/UNIFORM DEVIATION

Table 1.18, continued

Classification	Allegation name
5.Operation and	10M-INSUBORDINATION
Personnel Violations	10N-LUNCH/PERSONAL VIOLATIONS
	10P-MISUSE OF DEPARTMENT EQUIPMENT/SUPPLIES
	10Q-MISUSE DEPARTMENT RECORDS
	10R-RESIDENCY
	10S-SEXUAL HARASSMENT
	10T-REPORTS - FAILED TO SUBMIT/IMPROPER
	10V-INVENTORY PROCEDURES
	10W-VEHICLE LICENSING - CITY
	10X-VEHICLE LICENSING - STATE
	10Y-ACT TO CIRCUMVENT PROPER ADMINISTRATIVE ACTION
	10Z-MISCELLANEOUS
	12A-PROPER ACTION, INITIATE
	12B-PROPER DIRECTION - SUBORDINATE
	12C-PROPER ACTION REVIEW/INSPECT - SUBORDINATE
	12D-FAIL TO OBTAIN A COMPLAINT REGISTER NUMBER
	12E-IMPROPER/INADEQUATE INVESTIGATION
	12F-MISCELLANEOUS

Table 1.18, continued

Classification	Allegation name
6.Others	02A-INTOXICATED ON DUTY
	02B-INTOXICATED OFF DUTY
	02C-D.U.I. - ON DUTY
	02D-D.U.I. - OFF DUTY
	02E-POSSESSION/DRINKING ALCOHOL - ON DUTY
	02G-MISCELLANEOUS
	06A-SOLICIT/ACCEPT BRIBE (NON-TRAFFIC)
	06B-SOLICIT/ACCEPT BRIBE (TRAFFIC)
	06C-EXTORTION
	06D-BRIBE, FAILURE TO REPORT
	06E-GRATUITY
	06F-RECOMMEND PROFESSIONAL SERVICE
	06G-USE OFFICIAL POSITION
	06H-AN ACT TO CIRCUMVENT CRIMINAL PROSECUTION
	06J-MISCELLANEOUS

Table 1.18, continued

Classification	Allegation name
6.Others	08A-MURDER/MANSLAUGHTER, ETC.
	08B-ASSAULT/BATTERY, ETC.
	08C-RAPE/SEX OFFENSES
	08D-BURGLARY
	08E-AUTO THEFT
	08F-THEFT
	08G-SHOPLIFTING
	08H-ROBBERY
	08J-DRUGS/CONTR. SUB., POSSESSION OR SALE
	08K-DAMAGE/TRESPASSING PROPERTY
	08L-ARSON
	08M-OTHER FELONY
	08N-MISCELLANEOUS
	08P-POLICE IMPERSONATOR - ADV Section USE ON CLOSING ONLY

Table 1.18, continued

Classification	Allegation name
6.Others	09A-ALTERCATION/DISTURBANCE - DOMESTIC
	09B-ALTERCATION/DISTURBANCE - NEIGHBOR
	09C-ALTERCATION/DISTURBANCE - TRAFFIC
	09D-TRAFFIC VIOLATION (OTHER THAN D.U.I.)
	09E-MISDEMEANOR ARREST
	09F-SEXUAL MISCONDUCT
	09G-ABUSE OF AUTHORITY
	09H-JUDICIAL PROCESS/DIRECTIVE - CONTEMPT
	09J-MISCELLANEOUS
	09K-INDEBTEDNESS TO CITY
	09L-DRIVER'S LICENSE REVOKED/SUSPENDED
	11A-FORWARDED TO O.E.C.
	14A-STATE CIVIL SUIT
	14B-FEDERAL CIVIL SUIT
	15A-USE/ABUSE DRUGS/CONTR. SUBSTANCE - ON DUTY
	15B-USE/ABUSE DRUGS/CONTR. SUBSTANCE - OFF DUTY
	15C-D.U.I., DRUGS/ CONTR. SUB. - ON DUTY
	15D-D.U.I., DRUGS/ CONTR. SUB. - OFF DUTY
	15E-POSITIVE DRUG SCREEN - ORIGINATED FROM COMPLAINT
	15H-POSITIVE DRUG SCREEN - OTHER PHYSICAL EXAM
	15J-REFUSAL OF DIRECT ORDER TO PROVIDE DRUG SCREEN SPECIMEN
	15K-MISCELLANEOUS
	NA

1.11 Appendix B: Analysis of the Raw data

As complementary evidence to the analysis, this Section studies the Raw data to understand the overall patterns behind officers' internal (Police Department) and external (mostly civilians) allegations of misconduct.

Table 1.19 reports the annual frequency distribution of complaint outcomes for incidents that occurred between January 2011 and July 2014. Overall, about 5.8 percent of the complaints are sustained, and 5.7 percent of the complaints have an unknown outcome. About 67.4 percent of the complaints are not sustained because of administrative procedure (29.1 percent for missing affidavit and 38.3 percent for missing officer identifier). Hence, 21.6 percent of the complaints are not sustained after full investigation.

Table 1.20 reports the frequency distribution of complaint categories by complainant race for incidents that occurred between January 2011 and July 2014. About 34.1 percent of the complaint are related to use of force, verbal abuse, arrest, locked up procedures, and search. Failure to provide service (FPS) and operation-personnel violation (OPV) respectively represent 14.3 percent and 9.39 percent of the complaints. The remaining complaints (42.2 percent) have unknown or miscellaneous categories. When the complaint category is known, Table 1.20 suggests that Black and Hispanic civilians mainly complain about use of force and verbal abuse whereas whites mainly complain about failure to provide service.

I use a a multinomial model to describe the risk factors associated with each observed outcome presented in observed outcome presented in Table 1.19. This approach enables us to analyze the full sample where I consider both known outcomes (sustained, non-sustained, and missing affidavit) and unknown outcomes (unknown officer and unknown outcome). To conduct the analysis, I estimate a multinomial logit where I use the "not sustained" outcome as a reference category. For individual i in police district d during year t, the probability that outcome $y_{idt,j}$ occurs among alternative $j \in$ {sustained, non-sustained, no affidavit, unknown officer, unknown outcome} is

$$P(y_{idt,j}|X) = \frac{\exp(X'_{idt,j}\beta_j)}{\sum_h \exp(X'_{idt,h}\beta_h)} \tag{1.24}$$

where the vector $X_{idt,h}$ is a set of characteristics for individual i in police district d during year t, who experienced outcome j. Table 1.21 displays the relative risk ratios from equation 1.24. Overall, complaints attached to incidents that occur farther away from the oversight agency have a higher likelihood of missing affidavit or unknown officer. Male complainants are more likely to sign the affidavit (not statistically significant) and to have their complaint sustained. White and Hispanic complainants are about 7.3 and 4 times more likely, respectively, to have a sustained complaint compared to Black complainants. Overall, older complainants are more likely to sign the affidavit, know the officer's identifier, and have a sustained complaint. Beat with higher hourly wage are more likely to have a signed affidavit. Incidents that occurred in a police beat with a higher share of Black population are significantly more likely to have an unidentified officer.

1.12 Appendix C: Variance for the average willingness to pay

The parameter set of the full model consists of parameter ϕ. Without loss of generality and to streamline notation, I drop the k subscript.The average willingness to pay for group Z is given by

$$WTP_Z = E\left[g(z_i, \phi_0)\right] \tag{1.25}$$

where ϕ_0 is the true parameter vector, z_i contains the characteristics of complainant i in race-age group Z. The sample average willingness to pay to pay for group Z is given by

$$W\hat{T}P_Z = \frac{1}{n_Z} \overset{[}{i}= 1]n_Z \sum g(z_i, \hat{\phi}) \tag{1.26}$$

where $\hat{\phi}$ is the estimated parameter vector using equation 1.9. The variance of the willingness to pay is function of two components: sampling error and uncertainty from the parameter estimates. The asymptotic variance of the sample average of the willingness to pay is a function of

$$\sqrt{n_Z}\left(W\hat{T}P_Z - WTP_Z\right) = \frac{1}{\sqrt{n_Z}} [i = 1]n_Z \sum \left[g(z_i,\hat{\phi}) - E\left[g(z_i,\phi_0)\right]\right] \tag{1.27}$$

By adding and subtracting $g(z_i, \phi_0)$, this expression can be written as

$$\begin{aligned} \sqrt{n_Z}\left(W\hat{T}P_Z - WTP_Z\right) &= \frac{1}{\sqrt{n_Z}} [i = 1]n_Z \sum \left[A_i + B_i\right] \\ A_i &= g(z_i,\hat{\phi}) - g(z_i,\phi_0) \\ B_i &= g(z_i,\phi_0) - E\left[g(z_i,\phi_0)\right] \end{aligned} \tag{1.28}$$

Using the mean value theorem, A_i can be written as

$$\begin{aligned} g(z_i,\hat{\phi}) &= g(z_i,\phi_0) + \frac{\partial g(z_i,\overline{\phi})}{\partial \phi'}(\hat{\phi} - \phi_0) \\ A_i &= \frac{\partial g(z_i,\overline{\phi})}{\partial \phi'}(\hat{\phi} - \phi_0) \end{aligned} \tag{1.29}$$

Recall that $\hat{\phi}$ maximizes the likelihood function that is given by $L_n(\phi)$. Using the first order condition and the mean value theorem

$$\begin{aligned} \frac{\partial L_n(\hat{\phi})}{\partial \phi} &= \frac{\partial L_n(\phi_0)}{\partial \phi} + \frac{\partial^2 L_n(\overline{\phi})}{\partial \phi \partial \phi'}(\hat{\phi} - \phi_0) \\ 0 &= \frac{\partial L_n(\phi_0)}{\partial \phi} + \frac{\partial^2 L_n(\overline{\phi})}{\partial \phi \partial \phi'}(\hat{\phi} - \phi_0) \\ \hat{\phi} - \phi_0 &= -\left[\frac{\partial^2 L_n(\overline{\phi})}{\partial \phi \partial \phi'}\right]^{-1} \frac{\partial L_n(\phi_0)}{\partial \phi} \end{aligned} \tag{1.30}$$

Combining equations 1.29 and 1.30, A_i becomes

$$A_i = -\frac{\partial g(z_i,\overline{\phi})}{\partial \phi'}\left[\frac{\partial^2 L_n(\overline{\phi})}{\partial \phi \partial \phi'}\right]^{-1} \frac{\partial L_n(\phi_0)}{\partial \phi} \tag{1.31}$$

$$\sqrt{n_Z}\left(\hat{WTP}_Z - WTP_Z\right) = \underbrace{\frac{1}{\sqrt{n_Z}} [i = 1]n_Z \sum A_i}_{A} + \underbrace{\frac{1}{\sqrt{n_Z}} [i = 1]n_Z \sum B_i}_{B} \tag{1.32}$$

The asymptotic variance for the average willingness to pay that accounts for both sampling error and uncertainty from the parameter estimates is given by

$$Var(A + B) = Var(A) + Var(B) + 2Cov(A, B) \tag{1.33}$$

In practice, I use the sample analog of equation 1.33.

1.13 Appendix D: Additional Tables and Figures

Table 1.19: Complaint outcomes

	2011		2012		2013		2014		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Unknown Officer	1584	36.0	1774	41.0	1544	37.6	852	38.5	5754	38.3
Unknown Outcome	136	3.1	265	6.1	267	6.5	140	6.3	808	5.4
No Affidavit	1284	29.2	1186	27.4	1177	28.7	724	32.7	4371	29.1
Not Sustained	1164	26.5	896	20.7	810	19.7	379	17.1	3249	21.6
Sustained	228	5.2	205	4.7	308	7.5	116	5.2	857	5.7
Total	4396		4326		4106		2211		15039	29.5

Table 1.20: Complaint categories by complainant's race

	Complainant's Race									
	Black		Hispanic		White		Unknown		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Unknown	3767	39.2	731	39.1	1056	36.1	200	31.6	5754	38.3
Use of Force/Verbal Abuse	1707	17.8	364	19.5	380	13.0	127	20.1	2578	17.1
Arrest/Locked up Procedure	359	3.7	66	3.5	94	3.2	16	2.5	535	3.6
Search	1530	15.9	178	9.5	165	5.6	141	22.3	2014	13.4
Failure to Provide Service	1320	13.7	277	14.8	476	16.3	77	12.2	2150	14.3
Operation/Personnel Violations	697	7.2	190	10.2	478	16.4	47	7.4	1412	9.4
Others	234	2.4	63	3.4	274	9.4	25	3.9	596	4.0
Total	9614		1869		2923		633		15039	

Table 1.21: Risk factors associated with complaints' outcome from January 2011 and July 2014

	Unknown Officer	Unknown Outcome	No Affidavit	Sustained
Distance (mi.)	1.025*	1.015	1.041***	1.044
	(0.011)	(0.023)	(0.013)	(0.025)
Hourly wage	0.994	1.011	0.986*	0.995
	(0.005)	(0.009)	(0.006)	(0.012)
Male complainant	0.985	0.962	0.943	1.252*
	(0.037)	(0.091)	(0.042)	(0.131)
30-39yo	0.791***	0.724*	0.868	1.965*
	(0.055)	(0.109)	(0.067)	(0.646)
40-49yo	0.688***	0.792	0.756***	5.583***
	(0.059)	(0.113)	(0.055)	(1.649)
50-59yo	0.703***	0.853	0.627***	8.406***
	(0.053)	(0.121)	(0.051)	(2.592)
60-74yo	0.689***	0.753	0.729**	5.892***
	(0.058)	(0.125)	(0.077)	(1.863)
+74yo or missing	0.907	1.080	1.015	3.186***
	(0.079)	(0.159)	(0.086)	(1.092)

Notes: See next page.

Table 1.21, continued

	Unknown Officer	Unknown Outcome	No Affidavit	Sustained
	(0.079)	(0.159)	(0.086)	(1.092)
White	1.317***	2.174***	0.983	7.336***
	(0.097)	(0.278)	(0.094)	(1.539)
Hispanic/Other	1.091	1.540**	0.903	3.971***
	(0.079)	(0.219)	(0.077)	(0.866)
Unknown race	1.140	2.092***	1.935***	1.732
	(0.185)	(0.402)	(0.258)	(0.560)
Share of Black	1.863**	1.818*	1.037	1.325
	(0.363)	(0.547)	(0.225)	(0.549)
Share of Hispanic	1.556	1.274	1.215	0.766
	(0.368)	(0.383)	(0.326)	(0.342)
N	15039			
ll	-19519.3			

Notes: This Table presents the risk factors associated with complaints' outcome from January 2011 and July 2014. The Table presents the relative risk ratios by running a multinomial logistic regression. Non-sustained outcome is the reference category. The specification controls for incident location, district fixed effects, and quarter fixed effects, but not reported. Standard errors are clustered at the police district and community area level are reported in parentheses. *p-value $<$ 0.10, **p-value $<$ 0.05, *** p-value $<$ 0.01.

Table 1.22: Effect of distance on the probability of signed affidavit

	All			Serious			Failure to provide service		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance:<3	0.121***	0.123***	0.122***	0.114**	0.112**	0.113**	0.125**	0.129**	0.123**
	(0.031)	(0.032)	(0.032)	(0.044)	(0.046)	(0.047)	(0.056)	(0.058)	(0.058)
Distance:[3,6)	0.084***	0.081***	0.081***	0.073**	0.067**	0.071**	0.073*	0.080**	0.075*
	(0.025)	(0.024)	(0.025)	(0.031)	(0.030)	(0.031)	(0.038)	(0.039)	(0.040)
Distance:[6,9)	0.058**	0.057**	0.057**	0.051	0.045	0.045	0.035	0.039	0.041
	(0.024)	(0.023)	(0.023)	(0.030)	(0.028)	(0.028)	(0.040)	(0.041)	(0.041)
Serious	0.231***	0.264***	0.275***						
	(0.016)	(0.016)	(0.016)						
Observations	6,763	6,763	6,763	4,770	4,770	4,770	1,993	1,993	1,993
R-squared	0.116	0.142	0.145	0.096	0.124	0.128	0.243	0.279	0.283
Complainant's Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Officers' Controls	No	No	Yes	No	No	Yes	No	No	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.459	0.459	0.459	0.520	0.520	0.520	0.313	0.313	0.313

Notes: This Table presents the effect of distance on the probability of signed affidavit from January 2011 and July 2014. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search).The specification controls for complainants and incident characteristics, incident location, fixed effects, and time fixed effects, but not reported. Standard errors are clustered at the police district and community area level are reported in parentheses. *p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.23: Effect of distance on the probability of signed affidavit by race of the complainant

	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic	White
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance:<3	0.107**	0.158	-0.058	0.080	0.156	0.024	0.181*	0.262	0.106
	(0.050)	(0.118)	(0.108)	(0.066)	(0.172)	(0.181)	(0.098)	(0.472)	(0.235)
Distance:[3,6)	0.102***	0.078	-0.110	0.085**	0.055	-0.129	0.162**	0.363	0.090
	(0.033)	(0.103)	(0.085)	(0.041)	(0.103)	(0.137)	(0.071)	(0.303)	(0.221)
Distance:[6,9)	0.052	0.116	-0.076	0.049	0.143	-0.113	0.057	0.249	0.061
	(0.032)	(0.071)	(0.075)						
Serious	0.311***	0.249***	0.208***						
	(0.019)	(0.057)	(0.047)						
Observations	4,616	815	1,008	3,382	570	564	1,234	245	444
R-squared	0.162	0.365	0.377	0.134	0.451	0.459	0.285	0.779	0.646
Type of Complaint	All	All	All	Serious	Serious	Serious	FPS	FPS	FPS
Complainant's Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Officers' Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.445	0.506	0.550	0.515	0.574	0.601	0.251	0.347	0.484

Notes: This Table presents the effect of distance on the probability of signed affidavit from January 2011 and July 2014. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search).The specification controls for complainants and incident characteristics, incident location, fixed effects, and time fixed effects, but not reported. Standard errors are clustered at the police district and community area level are reported in parentheses. *p-value $<$ 0.10, **p-value $<$ 0.05, *** p-value $<$ 0.01.

Table 1.24: Effect of distance on the probability of signed affidavit (District)

		All			Serious			Failure to provide service	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A) All Comp.									
Distance (mi.)	-0.009***	-0.011***	-0.011***	-0.008*	-0.008*	-0.009**	-0.015***	-0.017***	-0.017***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Mean Dependent Variable	0.46	0.46	0.46	0.52	0.52	0.52	0.31	0.31	0.31
Observations	6,763	6,763	6,763	4,770	4,770	4,770	1,993	1,993	1,993
R-squared	0.08	0.10	0.11	0.04	0.07	0.07	0.13	0.17	0.17
B) Black Comp.									
Distance (mi.)	-0.011**	-0.010**	-0.010**	-0.007	-0.007	-0.007	-0.025***	-0.025***	-0.025***
	(0.004)	(0.004)	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
Mean Dependent Variable	0.44	0.44	0.44	0.52	0.52	0.52	0.25	0.25	0.25
Observations	4,616	4,616	4,616	3,382	3,382	3,382	1,234	1,234	1,234
R-squared	0.09	0.10	0.11	0.04	0.05	0.06	0.11	0.13	0.13
C) Hispanic Comp.									
Distance (mi.)	-0.012	-0.012	-0.011	-0.007	-0.005	-0.004	-0.022	-0.017	-0.013
	(0.009)	(0.009)	(0.009)	(0.012)	(0.011)	(0.011)	(0.019)	(0.019)	(0.019)
Mean Dependent Variable	0.51	0.51	0.51	0.57	0.57	0.57	0.35	0.35	0.35
Observations	815	815	815	570	570	570	245	245	245
R-squared	0.15	0.17	0.18	0.13	0.18	0.19	0.37	0.40	0.43
D) White Comp.									
Distance (mi.)	-0.002	-0.001	-0.001	-0.012	-0.011	-0.011	0.001	0.001	0.000
	(0.008)	(0.008)	(0.008)	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)
Mean Dependent Variable	0.55	0.55	0.55	0.60	0.60	0.60	0.48	0.48	0.48
Observations	1,008	1,008	1,008	564	564	564	444	444	444
R-squared	0.16	0.20	0.21	0.12	0.17	0.18	0.31	0.34	0.35
Complainant's Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Officers' Controls	No	No	Yes	No	No	Yes	No	No	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on the probability of signed affidavit from January 2011 and July 2014. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search). The specification controls for complainants and incident characteristics, incident location, district fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all the complainants, Black complainants, Hispanic, and White complainants. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.25: Effect of distance on the probability of signed affidavit using a logit specification

	All			Serious			Failure to provide service		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A) All Comp.									
Distance (mi.)	-0.010***	-0.012***	-0.012***	-0.008*	-0.009**	-0.009**	-0.017***	-0.019***	-0.019***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Mean Dependent Variable	0.46	0.46	0.46	0.52	0.52	0.52	0.31	0.31	0.31
Observations	6,763	6,763	6,763	4,770	4,770	4,770	1,993	1,993	1,993
B) Black Comp.									
Distance (mi.)	-0.011**	-0.011**	-0.011**	-0.007	-0.006	-0.007	-0.024***	-0.025***	-0.024***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
Mean Dependent Variable	0.44	0.44	0.44	0.52	0.52	0.52	0.25	0.25	0.25
Observations	4,616	4,616	4,616	3,382	3,382	3,382	1,234	1,234	1,234
C) Hispanic Comp.									
Distance (mi.)	-0.013	-0.013	-0.012	-0.006	-0.003	-0.003	-0.026	-0.026	-0.019
	(0.009)	(0.009)	(0.009)	(0.012)	(0.012)	(0.012)	(0.017)	(0.019)	(0.018)
Mean Dependent Variable	0.50	0.50	0.50	0.57	0.57	0.57	0.34	0.34	0.34
Observations	807	807	807	557	557	557	242	242	242
D) White Comp.									
Distance (mi.)	-0.005	-0.004	-0.004	-0.011	-0.010	-0.010	0.004	0.003	0.001
	(0.009)	(0.009)	(0.009)	(0.013)	(0.013)	(0.013)	(0.016)	(0.015)	(0.015)
Mean Dependent Variable	0.55	0.55	0.55	0.60	0.60	0.60	0.48	0.48	0.48
Observations	1,008	1,008	1,008	564	564	564	444	444	444
Complainant's Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Officers' Controls	No	No	Yes	No	No	Yes	No	No	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on the probability of signed affidavit from January 2011 and July 2014 using a logit specification for the error term. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search). The specification controls for complainants and incident characteristics, incident location, district fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all the complainants, Black complainants, Hispanic, and White complainants. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.26: Effect of traveling time by car on the probability of signed affidavit

	All			Serious			Failure to provide service		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A) All Comp.									
Distance (mi.)	-0.006***	-0.006***	-0.006***	-0.005*	-0.005*	-0.005**	-0.008**	-0.009***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Mean Dependent Variable	0.46	0.46	0.46	0.52	0.52	0.52	0.31	0.31	0.31
Observations	6,763	6,763	6,763	4,770	4,770	4,770	1,993	1,993	1,993
R-squared	0.12	0.14	0.14	0.10	0.12	0.13	0.24	0.28	0.28
B) Black Comp.									
Time by car (min.)	-0.008***	-0.008**	-0.008**	-0.005	-0.004	-0.004	-0.018***	-0.019***	-0.018***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)
Mean Dependent Variable	0.44	0.44	0.44	0.52	0.52	0.52	0.25	0.25	0.25
Observations	4,616	4,616	4,616	3,382	3,382	3,382	1,234	1,234	1,234
R-squared	0.15	0.16	0.16	0.12	0.13	0.13	0.27	0.28	0.29
C) Hispanic Comp.									
Time by car (min.)	-0.005	-0.004	-0.003	0.001	0.002	0.001	-0.013	-0.016	-0.015
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.021)	(0.022)	(0.022)
Mean Dependent Variable	0.51	0.51	0.51	0.57	0.57	0.57	0.35	0.35	0.35
Observations	815	815	815	570	570	570	245	245	245
R-squared	0.35	0.36	0.36	0.41	0.44	0.45	0.73	0.76	0.78
D) White Comp.									
Time by car (min.)	0.002	0.003	0.003	-0.002	-0.002	-0.002	-0.008	-0.007	-0.007
	(0.005)	(0.005)	(0.005)	(0.008)	(0.008)	(0.008)	(0.010)	(0.010)	(0.009)
Mean Dependent Variable	0.55	0.55	0.55	0.60	0.60	0.60	0.48	0.48	0.48
Observations	1,008	1,008	1,008	564	564	564	444	444	444
R-squared	0.35	0.37	0.37	0.43	0.45	0.45	0.62	0.63	0.65
Complainant's Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Officers' Controls	No	No	Yes	No	No	Yes	No	No	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of traveling time by car on the probability of signed affidavit from January 2011 and July 2014. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search). The specification controls for complainants and incident characteristics, incident location, fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all the complainants, Black complainants, Hispanic, and White complainants. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.27: Effect of traveling time by public transit on the probability of signed affidavit

	All			Serious			Failure to provide service		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A) All Comp.									
Time by transit (min.)	-0.001**	-0.002**	-0.002**	-0.001*	-0.001*	-0.002*	-0.002	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean Dependent Variable	0.46	0.46	0.46	0.52	0.52	0.52	0.31	0.31	0.31
Observations	6,763	6,763	6,763	4,770	4,770	4,770	1,993	1,993	1,993
R-squared	0.11	0.14	0.14	0.09	0.12	0.13	0.24	0.28	0.28
B) Black Comp.									
Time by transit (min.)	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.003	-0.003	-0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Mean Dependent Variable	0.44	0.44	0.44	0.52	0.52	0.52	0.25	0.25	0.25
Observations	4,616	4,616	4,616	3,382	3,382	3,382	1,234	1,234	1,234
R-squared	0.15	0.16	0.16	0.12	0.13	0.13	0.27	0.28	0.28
C) Hispanic Comp.									
Time by transit (min.)	-0.003	-0.003	-0.002	-0.003	-0.003	-0.003	-0.001	-0.002	-0.002
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.008)	(0.009)	(0.008)
Mean Dependent Variable	0.51	0.51	0.51	0.57	0.57	0.57	0.35	0.35	0.35
Observations	815	815	815	570	570	570	245	245	245
R-squared	0.35	0.36	0.36	0.41	0.44	0.45	0.73	0.76	0.77
D) White Comp.									
Time by transit (min.)	-0.001	0.000	0.000	-0.002	-0.003	-0.003	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mean Dependent Variable	0.55	0.55	0.55	0.60	0.60	0.60	0.48	0.48	0.48
Observations	1,008	1,008	1,008	564	564	564	444	444	444
R-squared	0.35	0.37	0.37	0.43	0.45	0.45	0.62	0.63	0.65
Complainant's Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Officers' Controls	No	No	Yes	No	No	Yes	No	No	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of traveling time by public transportation on the probability of signed affidavit from January 2011 and July 2014. Civilian complaints are either failure to provide service or serious (use of force, verbal abuse, arrest, locked up procedures, and search). The specification controls for complainants and incident characteristics, incident location, fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all the complainants, Black complainants, Hispanic, and White complainants. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.28: Effect of distance on complaint outcomes (Placebo)

	Number of complaints	Number of signed affidavits	Share of signed affidavits	Share of sustained
	(1)	(2)	(3)	(4)
A) All Beats				
Placebo Distance (mi.)	0.001	0.000	-0.001	-0.002
	(0.001)	(0.001)	(0.007)	(0.006)
Mean Dependent Variable	0.10	0.05	0.50	0.10
Observations	2,915	2,915	1,297	786
R-squared	0.01	0.00	0.00	0.02
B) Black Beats				
Placebo Distance (mi.)	0.000	-0.002	-0.014	0.000
	(0.002)	(0.002)	(0.010)	(0.007)
Mean Dependent Variable	0.17	0.08	0.50	0.08
Observations	1,298	1,298	734	457
R-squared	0.00	0.00	0.01	0.02
C) Hispanic Beats				
Placebo Distance (mi.)	0.002	0.001	-0.004	0.006
	(0.001)	(0.001)	(0.021)	(0.024)
Mean Dependent Variable	0.04	0.02	0.54	0.11
Observations	506	506	188	116
R-squared	0.01	0.00	0.01	0.04
D) White Beats				
Placebo Distance (mi.)	0.003	-0.001	-0.011	0.003
	(0.004)	(0.002)	(0.028)	(0.021)
Mean Dependent Variable	0.05	0.03	0.49	0.12
Observations	715	715	214	123
R-squared	0.01	0.00	0.02	0.02
District's Controls	No	No	No	No
Quarter FE	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on complaint outcomes from January 2011 to November 2011. The placebo policy occurred in June 2011.The complaint outcomes are the number of complaints per 1,000 capita (column 1), the number of complaints with signed affidavit per 1,000 capita (column 2), the share of complaints with signed affidavit (column 3), and the share of complaints that are sustained (column 4).The specification controls for beat fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all beats, beats with a majority (>50 percent) of Black residents, beats with a majority Hispanic residents, and beats with a majority White residents. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.29: Effect of distance on use of force outcomes (Placebo)

	Number of TRR	Any use of high force	Force per arrest	Number of injuries
	(1)	(2)	(3)	(4)
A) All Beats				
Placebo Distance (mi.)	0.002	0.004	0.000	0.001
	(0.002)	(0.004)	(0.001)	(0.001)
Mean Dependent Variable	0.18	0.35	0.04	0.04
Observations	2,915	2,915	2,915	2,915
R-squared	0.01	0.01	0.00	0.00
B) Black Beats				
Placebo Distance (mi.)	0.006	0.007	0.001	0.000
	(0.004)	(0.006)	(0.001)	(0.002)
Mean Dependent Variable	0.27	0.44	0.04	0.07
Observations	1,298	1,298	1,298	1,298
R-squared	0.03	0.01	0.01	0.01
C) Hispanic Beats				
Placebo Distance (mi.)	-0.002	-0.012	-0.001	0.000
	(0.002)	(0.012)	(0.001)	(0.001)
Mean Dependent Variable	0.07	0.30	0.03	0.02
Observations	506	506	506	506
R-squared	0.02	0.03	0.01	0.02
D) White Beats				
Placebo Distance (mi.)	0.004	-0.002	-0.002	0.000
	(0.009)	(0.012)	(0.002)	(0.003)
Mean Dependent Variable	0.12	0.26	0.05	0.03
Observations	715	715	715	715
R-squared	0.02	0.00	0.00	0.01
District's Controls	No	No	No	No
Quarter FE	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on use of force outcomes from January 2011 to November 2011. The placebo policy occurred in June 2011. The use of force outcomes are the the number of TRRs per 1,000 capita (column 1), whether or not there was incident with high level of force (column 2), i.e. involving use of Taser or firearm, the number of incidents involving reported force per arrest (columns 3), and the number of civilian injuries per 1,000 capita (columns 4) .The specification controls for beat fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all beats, beats with a majority (>50 percent) of Black residents, beats with a majority Hispanic residents, and beats with a majority White residents. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.30: Effect of distance on crime outcomes (Placebo)

	Index crimes			Non Index crimes		
	Offenses	Arrests	Clearance	Offenses	Arrests	Clearance
	(1)	(2)	(3)	(4)	(5)	(6)
A) All Beats						
Placebo Distance (mi.)	0.046**	0.005	0.000	0.044**	0.031*	0.001
	(0.023)	(0.005)	0.000	(0.019)	(0.016)	(0.001)
Mean Dependent Variable	6.59	0.79	0.10	8.87	3.81	0.38
Observations	2,915	2,915	2,915	2,915	2,915	2,915
R-squared	0.13	0.02	0.00	0.07	0.02	0.03
B) Black Beats						
Placebo Distance (mi.)	0.054	0.006	0.000	0.035	0.061*	0.003***
	(0.038)	(0.006)	(0.001)	(0.038)	(0.030)	(0.001)
Mean Dependent Variable	8.20	0.80	0.10	13.62	6.11	0.42
Observations	1,298	1,298	1,298	1,298	1,298	1,298
R-squared	0.27	0.07	0.00	0.11	0.03	0.06
C) Hispanic Beats						
Placebo Distance (mi.)	0.029	0.006	0.001	0.048	0.031	0.002
	(0.021)	(0.009)	(0.002)	(0.029)	(0.028)	(0.003)
Mean Dependent Variable	3.58	0.42	0.11	4.95	2.02	0.39
Observations	506	506	506	506	506	506
R-squared	0.12	0.05	0.01	0.13	0.06	0.03
D) White Beats						
Placebo Distance (mi.)	0.165	0.077	0.001	0.010	0.000	0.003
	(0.161)	(0.071)	(0.001)	(0.039)	(0.022)	(0.003)
Mean Dependent Variable	7.18	1.22	0.11	5.22	2.00	0.31
Observations	715	715	715	715	715	715
R-squared	0.08	0.03	0.03	0.02	0.01	0.02
District's Controls	No	No	No	No	No	No
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on crime outcomes from January 2011 to November 2011. The placebo policy occurred in June 2011. This Table presents the effect of distance on outcomes from January 2011 and July 2014. Offenses and arrests are expressed per 1,000 capita, and clearance rates are define as the number of reported crimes with an arrest over the number of reported crimes. The specification controls for beat fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all beats, beats with a majority (>50 percent) of Black residents, beats with a majority Hispanic residents, and beats with a majority White residents. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.31: Effect of distance on complaint outcomes (Short Run)

	Number of complaints		Number of signed affidavits		Share of signed affidavit		Share of sustained	
	[-9;9]	[-12;12]	[-9;9]	[-12;12]	[-9;9]	[-12;12]	[-9;9]	[-12;12]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) All Beats								
Distance (mi.)	-0.002**	-0.002***	-0.002***	-0.002***	-0.013**	-0.017***	-0.003	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.005)	(0.004)	(0.005)	(0.004)
Mean Dependent Variable	0.10	0.09	0.05	0.05	0.49	0.48	0.09	0.10
Observations	5,035	6,360	5,035	6,360	2,198	2,704	1,277	1,558
R-squared	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
B) Black Beats								
Distance (mi.)	-0.002	-0.002	-0.003**	-0.003**	-0.021**	-0.026***	-0.006	0.000
	(0.002)	(0.001)	(0.001)	(0.001)	(0.009)	(0.008)	(0.009)	(0.007)
Mean Dependent Variable	0.16	0.15	0.08	0.07	0.48	0.47	0.08	0.08
Observations	2,242	2,832	2,242	2,832	1,261	1,537	746	890
R-squared	0.01	0.01	0.01	0.01	0.02	0.02	0.01	0.01
C) Hispanic Beats								
Distance (mi.)	0.000	0.000	-0.001	-0.001**	-0.012	-0.017	-0.003	0.006
	(0.002)	(0.002)	(0.001)	0.000	(0.020)	(0.016)	(0.011)	(0.011)
Mean Dependent Variable	0.04	0.04	0.02	0.02	0.50	0.51	0.09	0.11
Observations	874	1,104	874	1,104	300	370	170	213
R-squared	0.01	0.01	0.02	0.01	0.02	0.02	0.07	0.05
D) White Beats								
Distance (mi.)	0.006**	0.003**	0.000	-0.001	-0.025	-0.024	0.008	-0.002
	(0.003)	(0.002)	(0.001)	(0.001)	(0.018)	(0.016)	(0.017)	(0.015)
Mean Dependent Variable	0.06	0.05	0.03	0.03	0.49	0.50	0.13	0.12
Observations	1,235	1,560	1,235	1,560	372	468	210	265
R-squared	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02
District's Controls	No	No	No	No	No	No	No	No
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on complaint outcomes from January 2011 to December 2012. The complaint outcomes are the number of complaints per 1,000 capita (columns 1-2), the number of complaints with signed affidavit per 1,000 capita (columns 3-4), the share of complaints with signed affidavit (columns 4-6), and the share of complaints that are sustained (columns 7-8). The specification controls for beat fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all beats, beats with a majority (>50 percent) of Black residents, beats with a majority Hispanic residents, and beats with a majority White residents. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.32: Effect of distance on use of force outcomes (Short Run)

	Number of TRR		Any use of high force		Force per arrest		Number of injuries	
	[-9;9]	[-12;12]	[-9;9]	[-12;12]	[-9;9]	[-12;12]	[-9;9]	[-12;12]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) All Beats								
Distance (mi.)	0.003	0.003*	0.004	0.003	0.000	0.000	0.001	0.001
	(0.002)	(0.002)	(0.004)	(0.004)	0.000	0.000	(0.001)	(0.001)
Mean Dependent Variable	0.18	0.17	0.36	0.34	0.04	0.04	0.04	0.04
Observations	5,035	6,360	5,035	6,360	5,034	6,359	5,035	6,360
R-squared	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.00
B) Black Beats								
Distance (mi.)	0.006	0.008**	0.012*	0.010	0.001	0.001**	0.001	0.002*
	(0.004)	(0.003)	(0.006)	(0.006)	(0.001)	0.000	(0.001)	(0.001)
Mean Dependent Variable	0.28	0.27	0.44	0.43	0.04	0.04	0.07	0.07
Observations	2,242	2,832	2,242	2,832	2,242	2,832	2,242	2,832
R-squared	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01
C) Hispanic Beats								
Distance (mi.)	0.006***	0.005**	0.026***	0.020***	0.002***	0.002***	0.002**	0.001
	(0.002)	(0.002)	(0.008)	(0.007)	0.000	(0.001)	(0.001)	(0.001)
Mean Dependent Variable	0.08	0.07	0.31	0.30	0.03	0.03	0.02	0.02
Observations	874	1,104	874	1,104	874	1,104	874	1,104
R-squared	0.02	0.02	0.03	0.02	0.02	0.01	0.01	0.01
D) White Beats								
Distance (mi.)	0.018	0.015	0.014**	0.011*	0.005***	0.003**	0.003	0.002
	(0.012)	(0.010)	(0.006)	(0.006)	(0.001)	(0.001)	(0.002)	(0.001)
Mean Dependent Variable	0.13	0.12	0.26	0.26	0.05	0.05	0.03	0.03
Observations	1,235	1,560	1,235	1,560	1,234	1,559	1,235	1,560
R-squared	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
District's Controls	No	No	No	No	No	No	No	No
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table presents the effect of distance on use of force outcomes from January 2011 to December 2012. The use of force outcomes are the the number of TRRs per 1,000 capita (columns 1-2), whether or not there was incident with high level of force (columns 3-4), i.e. involving use of Taser or firearm, the number of incidents involving reported force per arrest (columns 5-6), and the number of civilian injuries per 1,000 capita (columns 7-8) . The specification controls for beat fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all beats, beats with a majority (>50 percent) of Black residents, beats with a majority Hispanic residents, and beats with a majority White residents. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.33: Effect of distance on crime outcomes (Short Run)

	Index crimes			Non Index crimes		
	Offenses	Arrests	Clearance	Offenses	Arrests	Clearance
	(1)	(2)	(3)	(4)	(5)	(6)
A) All Beats						
Distance (mi.)	-0.070***	0.004	0.001***	-0.040*	-0.018	0.000
	(0.012)	(0.004)	0.000	(0.022)	(0.019)	(0.001)
Mean Dependent Variable	6.44	0.78	0.10	8.67	3.70	0.37
Observations	6,360	6,360	6,360	6,360	6,360	6,360
R-squared	0.10	0.01	0.00	0.06	0.03	0.04
B) Black Beats						
Distance (mi.)	-0.081***	0.003	0.001**	-0.080	-0.028	0.001
	(0.021)	(0.006)	(0.001)	(0.073)	(0.065)	(0.002)
Mean Dependent Variable	7.89	0.80	0.10	13.35	5.97	0.42
Observations	2,832	2,832	2,832	2,832	2,832	2,832
R-squared	0.22	0.03	0.01	0.11	0.04	0.05
C) Hispanic Beats						
Distance (mi.)	-0.022	0.003	0.002	-0.028	-0.025**	-0.003**
	(0.015)	(0.004)	(0.001)	(0.017)	(0.010)	(0.001)
Mean Dependent Variable	3.45	0.41	0.11	4.74	1.90	0.38
Observations	1,104	1,104	1,104	1,104	1,104	1,104
R-squared	0.15	0.03	0.01	0.17	0.10	0.07
D) White Beats						
Distance (mi.)	0.118	-0.039	-0.001	0.079	-0.027**	-0.006***
	(0.108)	(0.037)	(0.002)	(0.064)	(0.012)	(0.002)
Mean Dependent Variable	7.25	1.18	0.11	5.14	1.93	0.30
Observations	1,560	1,560	1,560	1,560	1,560	1,560
R-squared	0.07	0.02	0.01	0.01	0.01	0.03
District's Controls	No	No	No	No	No	No
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Beat FE	Yes	Yes	Yes	Yes	Yes	Yes

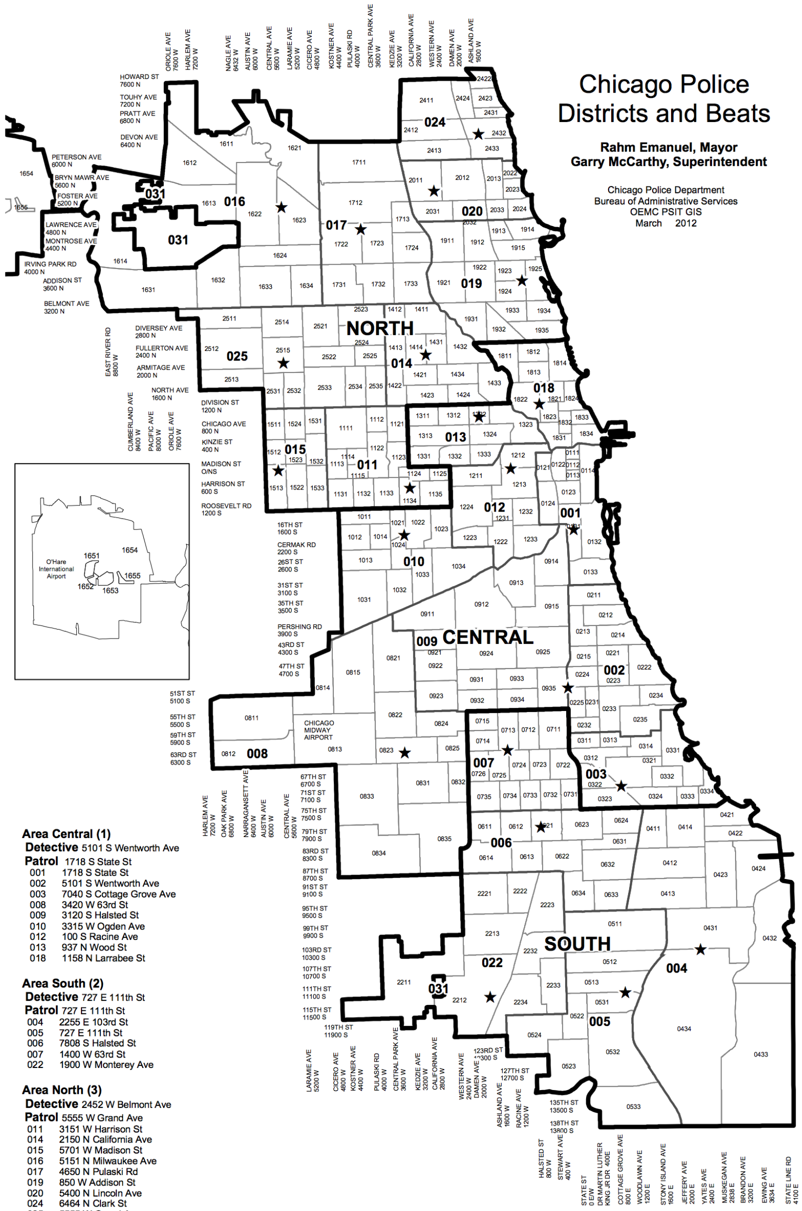
Notes: This Table presents the effect of distance on crime outcomes from January 2011 to December 2012. Offenses and arrests are expressed per 1,000 capita, and clearance rates are define as the number of reported crimes with an arrest over the number of reported crimes. The specification controls for beat fixed effects, and time fixed effects, but not reported. Panel A, B, C, and D respectively report the results for all beats, beats with a majority (>50 percent) of Black residents, beats with a majority Hispanic residents, and beats with a majority White residents. Standard errors are clustered at the police district and community area level are reported in parentheses.*p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 1.34: Summary statistics for police district

District #	Name	Distance (mi)	Time by Car (min.)	Time by Transit (min.)	Violent Crime Offense	Property Crime Offense	Non Index Crime Offense
7	Englewood	7.3	21.5	50.9	2.6	6.3	15.2
11	Harrison	3.7	13.7	35.1	2.5	5.1	19.0
3	Grand Crossing	8.0	21.9	52.5	2.2	5.6	12.5
15	Austin	5.7	19.2	41.1	1.9	4.1	14.6
6	Gresham	9.4	22.5	60.7	1.8	5.4	10.5
5	Calumet	13.6	26.5	77.4	1.8	4.7	11.5
4	South Chicago	10.9	26.4	67.6	1.7	5.1	9.9
2	Wentworth	5.4	17.4	41.7	1.7	4.9	10.2
1	Central	2.7	12.4	27.7	1.5	20.3	15.7
10	Ogden	4.1	17.8	44.9	1.4	3.3	8.7
9	Deering	4.8	19.5	43.1	1.0	3.2	6.7
22	Morgan Park	11.8	26.3	71.5	0.9	3.3	5.8
12	Near West	2.2	9.5	24.5	0.8	4.7	5.7
25	Grand Central	6.0	24.2	52.6	0.7	2.7	5.0
8	Chicago Lawn	7.9	29.4	61.6	0.7	2.8	4.9
14	Shakespeare	3.5	13.2	30.6	0.6	3.9	3.8
18	Near North	2.8	13.7	26.7	0.5	6.3	5.2
24	Rogers Park	9.0	30.9	55.5	0.5	2.0	3.8
19	Town Hall	4.9	20.9	39.5	0.4	3.1	3.2
20	Lincoln	7.1	26.1	52.2	0.3	2.1	3.1
17	Albany Park	6.7	18.6	47.1	0.3	2.3	2.7
16	Jefferson Park	9.0	22.0	52.5	0.2	1.5	2.2

Notes: This table reports the monthly average per police districts from January 2011 to July 2014. Offenses are calculated per 1,000 capita. The districts are sorted from the most to the least violent in terms of reported violent crime per 1,000 capita. The first three columns report the average distance, time by car, and time by transit to the oversight agency.

Figure 1.18: Districts and beats map



Notes: This figure depicts police beats and districts from Chicago Police Department. All the events in the analysis are geocoded according to that map in order to make events spatially com- parable with each other across time. I do not include beats that are located outside of Chicago. Beats that do not have any residents, according to the 2010-2014 ACS data, are removed from the sample.

CHAPTER 2

THE INTRODUCTION OF TASERS AND POLICE USE OF FORCE: EVIDENCE FROM THE CHICAGO POLICE DEPARTMENT

Bocar A. Ba and Jeffrey T. Grogger

2.1 Introduction

A key role of the police is to protect people and property. This often requires the use of force, since many would-be arrestees would rather take their chances with patrol officers than submit voluntarily to being arrested. Roughly 20 percent of arrests entail some use of force on the part of police (Eith and Durose [2011]).

At the same time, police use of force is controversial. One dimension of controversy involves the frequency of police use of force, with complaints that police use force even when conflicts could be resolved by other means. Another involves the level of force, with claims that the type of force deployed by police is disproportionate to the level of resistance offered by suspects.

Either way, excess force can undermine the legitimacy of the police (Tyler [2004], Ramsey and Robinson [2015], Lum and Nagin [2017], Manski and Nagin [2017]). It can also violate civil rights. In the US, courts have held that excess use of force violates the U.S. Constitution's protections against undue search and seizure (ACLU [nd], Neuscheler and Freidlin [2015]). Controversy over police use of force has led to public unrest, including recent cases of mass demonstrations in locales such as Baltimore, Chicago, and Ferguson, Missouri.

To alleviate the controversy, many police departments have adopted weapons involving less-than-lethal force. One example is pepper spray, which was widely deployed by US police agencies during the 1980s and 1990s (Smith et al. [2009]). A more recent example is the Taser, a conductive energy device (CED) that shoots wired barbs at its target, incapacitating him or her by means of an electrical charge. The idea behind such weapons is to provide patrol officers with a means of subduing resistant subjects that reduces the risk of injury or death relative to other means of force.

Indeed, injury risk has been the focus of most previous work on CED's. Most studies involve a cross-Sectional design, comparing incidents in which Tasers were used to those in which they were not, although a few studies have employed a before-and-after approach. The evidence is somewhat mixed, although most studies have found that CED's reduce the likelihood of injury, particularly to police, in the typical use-of-force incident (Department [2003], Jenkinson et al. [2006], Smith et al. [2007], MacDonald et al. [2009], Taylor et al. [2009], Lin and Jones [2010], Smith et al. [2009], Taylor and Woods [2010], Kaminski et al. [2015]).

In this paper we study injury risk, but we also analyze a range of important outcomes that have received less attention in the literature. We start by analyzing substitution between Tasers and other types of force. Subject to some data limitations, this lets us determine whether Tasers substitute for greater or lesser types of force. This includes firearms, which pose a particularly high risk for innocent bystanders as well as the target subject. The issue of substitution between Tasers and firearms is also important because it has rarely been tested, even as several large police departments have recently established or expanded their Taser arsenals on the grounds that they reduce police shootings (Bustamante [2017], Hinkel and Smith Richards [2017]).

We also ask how the availability of Tasers affects the total number of use-of-force incidents. If police become overly reliant on the devices, a phenomenon referred to by Alpert and Dunham [2010] as "lazy cop syndrome," then the availability of Tasers could lead to an increase in use of force. Relatedly, we ask how Tasers affect the total number of injuries to police and civilians. If the availability of Tasers increases police use of force, then injuries could potentially rise, even if injury rates fall.

Figure 2.1 helps motivate the analysis. There we plot use-of-force incidents involving Tasers for the period 2005-2015. We highlight three key features of the Figure here, postponing a detailed discussion of sources and definitions to Section 2.2 below. First, the use of Tasers rose sharply beginning in March 2010. Second, Taser use fell almost as sharply in late 2012. Finally, Taser use then stabilized at a rate roughly double that observed before March 2010.

The remainder of the paper proceeds as follows. The next Section provides background information. We provide more discussion of Tasers; weapons policy at CPD; prior work on the effect of intermediate-force weapons; and sketch a theoretical model that helps explain our findings. Section 2.3 discusses the data and Section 2.4 discusses our difference-in-difference approach in more detail. Section 2.5 presents results, followed by our conclusions in Section 2.6.

2.2 Background

2.2.1 What is a Taser and what does it do?

The Taser is a conducted energy device (CED) manufactured by Axon, formerly known as Taser International. Tasers are in widespread use, deployed by 15,000 police agencies (NIJ [2011]). They have been discharged over 2 million times (LAAW International [2017]).

When used in "probe mode" the hand-held weapon fires two probes, or darts, that are attached to the handset by means of thin wires that can be up to 30 feet long. It produces 50,000 volts of electricity at 0.004 amps, stunning and temporarily disabling its target by causing involuntary muscular contractions (Jenkinson et al. [2006], NIJ [2011]). It can also be used in "drive-stun" mode as a pain-compliance device, whereby the police officer presses the handset directly against the target and discharges a current.

Despite the apparent dangers, most individuals who are shot with Tasers experience no or only minor injuries (Bozeman et al. [2009]). At the same time, they cause involuntary muscular contractions, which can cause the subject to fall, leading to injuries (NIJ [2008]). More than 200 people have died after being shocked by a Taser (NIJ [2008, 2011]). The shock is said to be excruciating painful (Braidwood [2010]; Alpert and Dunham [2010]). Tasers have been controversial both because of the pain they inflict and because of the risks they pose.

2.2.2 CPD Taser policy

CPD first acquired Tasers in 2004. Initially they were issued only to sergeants and field training officers. Then as now, Taser use was governed by the CPD use-of-force model, which spells out the types of force that police officers are authorized to use in response to different levels of resistance offered by subjects. Since our data source is also linked to the CPD use-of-force model, we describe it here in some detail.

Table 2.1 shows subject resistance levels, police force options, our classification of those options, and reporting requirements. The first column shows that subjects are classified broadly as compliers, passive resisters, active resisters, or assailants. Types of force that are an option in response to a particular resistance level are also options against higher levels of resistance. In general, higher levels of force are authorized for higher levels of resistance. Tasers are an option for use against active resisters and assailants, but not against compliers or passive resisters (Department [2016a]).

In March 2010 CPD more than doubled the number of Tasers in their inventory, issuing them to patrol officers for the first time with the goal of equipping all squad cars with the weapon (Sweeney and Schorsch [2010]). This policy change, or treatment, provides the key explanatory variable for our analysis below. Officers must be certified in order to use Tasers;

Figure 2.1 shows that certifications spiked in March 2010. Use-of-force incidents involving Tasers rose sharply as well.

In late 2012, use-of-force incidents involving Tasers fell rather abruptly. Shortly prior to that time, CPD purchased a new model of Taser and required officers to become certified in its use (Hinkel and Smith Richards [2017]). Figure 2.1 shows a sharp increase in new certifications at that time. Between the 2010 and 2012 spikes in certification, Axon changed its Taser training materials. Materials introduced in May 2010 included health and safety issues that were not present in earlier materials, including discussions of cardiac, pulmonary, other physiological and metabolic risks involved with the use of Tasers, as well as higher-risk populations (International [2010]). Materials introduced in July 2011 added information on legal matters, including 4th Amendment issues and issues involving officers' personal legal liability for excessive use of force (International [2011]).

The new weapons and 2012 recertification involved all officers, both sergeants and patrol officers. As such, they do not constitute an intervention whose effects we can evaluate with our difference-in-difference approach. Nevertheless, since we thought they might change the treatment effect associated with the March 2010 introduction of Tasers to patrol officers, we allow our estimated treatment effects to vary before and after September 2012.

2.2.3 Theoretical framework

To see how the availability of a Taser might affect a police officer's behavior, consider a situation in which the officer seeks to arrest a subject. For simplicity, suppose the officer can use either no force, relying solely on verbal persuasion; a low level of force, such as physical coercion; or a high level of force, such as a firearm. Each option varies in its expected costs to the officer. The costs involve the time involved in resolving the incident; the probability and expected extent of injury to the officer; the probability and expected extent of injury to the subject, which the officer may at least partially internalize; and the probability and expected extent of administrative sanction. Under certain conditions, the police officer will choose her option on the basis of the level of resistance she perceives from the subject: for sufficiently low resistance, she will use no force; for somewhat greater resistance, she will use the low-force option; and for a high level of resistance, she will use the high level of force.

Now add an intermediate force option between the high and low levels, such as a CED. The important observation is that the new option may have lower costs than both the low- and high-force alternatives. Relative to the high-force option, the CED may have less-dire potential consequences for the suspect. The CED may also entail a lower likelihood of being labelled as an excessive use of force, and thus less likelihood of provoking an administrative sanction. Under these conditions, the officer would prefer the CED to the firearm, provided the suspect's resistance is not too high. This is how an intermediate-force option could reduce police use of firearms, often among the goals of adopting such weapons.

CED's may also entail less risk of physical injury to the officer than the low-force option, as discussed in Section 2.4 below. At intermediate levels of suspect resistance, the officer may choose the CED over physical coercion. If so, the availability of CED's may cause the officer to substitute away from both the low- and high-force options.

Beyond substitution, the availability of CED's could lead to an increase in the overall use of force. This may happen if the cost of using CED's is lower than that of using no force. For example, a CED may speed the resolution of an incident relative to verbal persuasion, which can be time-consuming. The notion that police could see CED's as an easy way to resolve even low-level incidents has been referred to as "lazy cop syndrome" and has led to concerns about police overuse of the weapon (NIJ [2011]).

2.2.4 Prior Research

The question of weapon substitution has received relatively little attention. Sousa et al. [2010] conducted a field-training experiment with 32 police officers each in the treatment and control groups. The control group was equipped with their usual set of weapons; to that the treatment group added Tasers. In response to low to moderate subject resistance, none of the officers deployed their firearms. In response to potentially lethal resistance, 17 members of the control group and 7 members of the treatment group deployed their firearms. Although the difference was statistically significant, the small samples and training-ground setting raise questions as to whether their results would generalize. Lin and Jones [2010] track weapon use among the Washington State Patrol from 2005 to 2007, corresponding to one year before and one year after the introduction of Tasers. Over the entire three-year period, they registered only 13 uses of firearms, which provide an insufficient basis for inference. Finally, Taylor et al. [2009]compared changes in firearm use among a set of law enforcement jurisdictions that adopted CEDs during their sample period with a set that did not. They found no significant effect of CED adoption on firearm use, although their sample seems to have included only about 100 incidents involving firearms.[1]

Most of the research on the effects of CED's has focused on the link between those weapons and injury risk to police and suspects. The studies differ in their details; some analyze the availability of CED's whereas other focus on their actual use. One feature common to the great majority of studies on injury risk is their use of individual use-of-force incidents as the unit of observation. A typical study relates the occurrence of an injury to characteristics of the incident as well as an indicator for whether a CED was used (Jenkinson et al. [2006], Smith et al. [2007], MacDonald et al. [2009], Lin and Jones [2010], Smith et al. [2009], Taylor and Woods [2010], Terrill and Paoline III [2012], Paoline III et al. [2012], Kaminski et al. [2015]). Although there are exceptions, most of these studies credit CED's for reducing injuries per incident among both subjects and police.

A limitation of most prior studies is that they speak to only one mechanism by which the availability of CED's may affect injuries. As discussed in the previous subsection, the

1. In their regression tables, the largest samples that the authors report include fewer than 10,000 incidents. They indicate that only about 1 percent of incidents involved firearms.

availability of CED's could affect the number of use-of-force incidents as well as injury rates per incident. The effect of CED's on the total number of subject or officer injuries involves their effects on both the injury rate per incident, an intensive margin, and the number of use-of-force incidents, an extensive margin. If CED's increase the use of force by a sufficient amount, total injuries could rise despite a reduction in injury rates.

Only two studies provide any evidence on the question of whether the availability of CED's increases police use of force. Both were longitudinal studies that focused primarily on other outcomes. However, both provided before-and-after tabulations on the number of use-of-force incidents. Lin and Jones [2010] reported that the number of use-of-force incidents involving the Washington State Patrol rose from 269 in 2005, the year before Tasers were adopted, to 469 in 2007, the year after. Smith et al. [2009] report a sharp increase in the number of use-of-force incidents per month in Orlando, Florida after the police department deployed CED's. They report a decrease in Austin, Texas, although they indicate that there is some ambiguity in their Austin data as to whether they are measuring unique incidents or the number of officers involved in the incidents.

The data from Washington State and Orlando are consistent with the notion that the availability of CED's could lead to an overall increase in use-of-force incidents. At the same time, they do not establish a causal link, and it is important to emphasize that the authors do not construe their results as such. Many factors that could contribute to an increase in use of force besides the introduction of CED's. One example would be an increase in crime; another could be a rise in levels of violence in the conduct of narcotics markets. In Section 2.4, we detail our method for distinguishing the effects of CED availability from trends in other factors that influence police use of force.

2.3 Data

Our main source of data comes from CPD Tactical Response Reports (TRR). CPD requires police officers to fill out a TRR after use-of-force incidents in a manner that is tied to the CPD use- of-force model. As shown in Table 2.1, a TRR must be filed after use-of-force incidents involving subjects classified as active resisters or assailants.[2] For subjects classified as cooperative or as passive resisters, police are required to fill out a TRR if the subject is injured or alleges an injury. A TRR must also be filed for "all incidents where a subject obstructs a police officer when the obstructing is a physical act between the subject and the officer" (Department [2016b]).

TRR's require a supervisor's approval. The supervisor is then supposed to notify the external oversight agency for incidents involving the use of deadly force, the discharge of a firearm, the discharge of a Taser, the discharge of pepper spray and other chemical weapon, or an allegation of excessive force. In principle, the oversight agency is required to investigate those incidents. However, because of limited resources, the agency does not investigate them all. Starting in 2009, for example, the oversight agency chose to investigate only those Taser incidents involving serious injury or death, a minor (17 and under) or a senior citizen (65 or older).

We obtained the TRR data by means of Freedom of Information Act requests for the period from 2004 through early 2016. We make use of data from 2005 to 2015 because reporting for 2004 and early 2016 seemed incomplete. The key variables in the dataset include the date of the incident, number of involved officers, injured officers, suspects' race and ethnicity, injured suspects, and the type of force used against the suspect. One limitation of our dataset is that it includes no records for incidents involving juvenile suspects.

We classify use-of-force incidents into the six categories shown in Table 2.1 according to

2. The exception to this rule involves subjects whose only active resistance involves fleeing and when the member's actions did not extend beyond verbal commands and/or control holds utilized in conjunction with handcuffing and searching techniques which do not result in injury or allegation of injury.

the highest type of force used in the incident, which is a common practice in the literature. Our type-of- force hierarchy comes from the CPD use-of-force model. Table 2.2 reports the distribution of use-of- force incidents over our sample period. Only 15 percent of incidents involved no or minor force. This probably reflects reporting requirements. No and minor force are the only types of force that are authorized for compliant or passively resistant subjects. As mentioned above, TRR's are not required for such incidents unless the subject is injured. We suspect that many incidents involving minor force or less that do not result in injuries are unreported.

Of the 36,112 use-of-force incidents reported during our sample period, 30,641 involved more than minor use of force. A majority involved the use of intermediate force; Tasers were involved in another 10 percent of the incidents. More than one in six of the incidents resulted in the use of major force. Seven hundred thirty-three of the incidents involved firearms, a much larger number than were available in previous studies.

The third and fourth columns of the Table report injury rates per incident to police and suspects, respectively. These injuries are reported by police; the TRR specifically asks whether the subject was injured by the police officer. The injury rates for incidents involving no and minor force are hard to understand. On the one hand, one might expect incidents involving low levels of force to result in low injury rates. On the other, most incidents involving low levels of force are only required to be reported if they result in injury. The observed injury rates may reflect some combination of reporting requirements and voluntary reporting.

Injury rates for incidents involving intermediate force are roughly similar for police and subjects. In contrast, injury rates for Taser-related incidents are quite different. Whereas 36.9 percent of Taser-involved incidents result in injuries to subjects, only 16.5 percent of such incidents result in injuries to police officers. The situation reverses for incidents involving major use of force, with injury rates of 53.6 percent for police and 41 percent for subjects.

For incidents involving firearms, 28.5 percent of police, and 58.1 percent of subjects, are injured. We suspect that many of the firearm incidents in which subjects were not injured involved shots that missed their mark.

It's worth highlighting that police injury rates are lower for Tasers than for both intermediate and major uses of force. In the context of the model above, this should provide police with an incentive to substitute toward Tasers from both lesser and greater types of force. Of course, this incentive could be mitigated to the extent that police officers internalize risks facing subjects when selecting among types of force, since injury rates for subjects rise with the level of force.

One limitation of our data is that we have no information about the severity of injuries.

Although this information is reported on the TRR, it was withheld from our FOIA request on the grounds that it referred to the subject's medical condition. One question we can speak to indirectly is whether probe punctures from Taser shots are included as injuries. The probes themselves are fairly small and the punctures they produce have generally not been classified as injuries by previous researchers (Kaminski et al. [2015]). Whether the probe punctures themselves are counted as injuries is important, because doing so leads to the conclusion that Tasers raise subject injury rates (Terrill and Paoline III [2012], Kaminski et al. [2015]). Kaminski et al. [2015] is the only study of which we are aware that calculates injury rates with and without probe punctures. They report that the subject injury rate from the use of Tasers is 81 percent when probe punctures are included, but only 32 percent when they are excluded. Table 2.2 shows a subject injury rate of 36.9 percent for use-of-force incidents involving Tasers, suggesting that probe punctures by themselves are generally not being counted as injuries in our data.

To calculate our difference-in-difference estimates, we aggregated these incident-level data into monthly time series. We constructed monthly data on the number of use-of-force incidents (by type of force and total); the number of incidents involving at least one injury, to police and subjects separately; and the number of incidents involving at least one black subject or one Hispanic subject. Incidents and injuries are normalized by the number of police officers on the CPD payroll each month, divided by 100,[3] in order to provide commensurable magnitudes for patrol officers and sergeants. We use the number of incidents and the number of injuries per 100 police to study the extensive margin: how the introduction of Tasers affected the number of use-of-force incidents and use-of-force-related injuries.

To study injury risk, we analyze monthly injury rates constructed as the number of incidents involving at least one injury divided by the number of incidents. Such injury rates have been the focus of most previous research on the effects of CED's. We use these variables to study the intensive margin: how the introduction of Tasers affected the probability of injury to police and subjects.

Finally, we construct two variables reflecting the race of subjects. The first is the number of incidents in which at least one subject was black divided by the number of incidents. The second is a comparable measure for Hispanics. We use these variables to analyze whether the introduction of Tasers changed the race distribution of subjects involved in use-of-force incidents.

Table 2.3 presents means of our dependent variables by time period and treatment group. We divide the sample period into the pre-intervention period, from January 2005 to February 2010; the first post-intervention period, from March 2010 to September 2012; and the second post-intervention period, following the change of Taser models and the re-training in late 2012, from September 2012 to December 2015. Our treatment observations are based on incidents that involved only patrol officers; our comparison observations are based on incidents that involved sergeants.

The top panel of the Table presents statistics for patrol officers. There we see that the number of incidents involving no force, minor force, or firearms was roughly constant over

3. CPD payroll data came from a separate FOIA request.

time. Intermediate-force incidents fell steadily. Taser-related incidents follow the pattern suggested by Figure 2.1, rising dramatically between the pre-period and first post-period before falling substantially during the second post-period. Incidents involving major force first rose and then fell, as did the total number of use-of-force incidents.

The monthly number of officer injuries per 100 officers fell steadily over the sample period. In contrast, the monthly number of subject injuries per 100 officers rose between the pre-period and the first post-period, then fell during the second post-period to a level slightly below where it started. Officer injury rates/incident fell sharply between the pre-period and the first post-period, then remained flat. Subject injury rates/incident were essentially constant. The shares of incidents involving black and Hispanic subjects likewise were roughly constant over the sample period.

The bottom panel of the Table presents statistics for sergeants. The monthly number of incidents per 100 sergeants involving either no, minor, intermediate, or major force rose between all three periods. In contrast, the number of incidents involving Tasers remained roughly constant during the pre-treatment and first post-treatment periods, then fell markedly after the re-training that occurred in late 2012. The use of firearms fluctuated little over the entire sample period.

2.4 Identification and Estimation

2.4.1 Identification

From the monthly data described above, we seek to estimate the average effect of treatment on the treatment group, that is, the average effect of making Tasers available to patrol officers. To analyze identification via our difference-in-difference estimator, we adopt potential outcomes notation that is adapted for our setting. Denote the treatment dummy by $D_t = 1$ if Tasers are available at time t and $D_t = 0$ otherwise, where time$t = 1, ..., T$ is measured in months, with $t = 1$ denoting January 2005 and $t = T$ denoting December 2015. Denote the treatment status indicator by $D = 1$ for the treatment group, i.e., patrol officers, and $D = 0$ for the comparison group, i.e., sergeants. For each month, we have one observation for each value of D, that is, one for the treatment group and one for the comparison group.

Let the potential outcome at time t, such as use-of-force incidents per 100 officers, be given by $Y_t(D_t)$. Thus $Y_t(1)$ is the potential treated outcome, when Tasers are available, and $Y_t(0)$ is the potential untreated outcome, when they are not. Let t' denote March 2010, so $t \geq t$' denotes the post- treatment period and $t < t$' denotes the pre-treatment period. The average effect of treatment on the treated (ATT) at some date $a > t$' is given by

$$ATT_a = E\left[Y_a(1)|D = 1\right] - E\left[Y_a(0)|D = 1\right]$$

This is the difference between the expected treated potential outcome and the expected untreated potential outcome among patrol officers, at some post-treatment period a. It should be thought of as an intent-to-treat effect, since if measures the effect of the policy of making Tasers available to patrol officers, not the effect of actually using them.

The problem for estimation is that $E\left[Y_a(0)|D = 1\right]$, the expectation of the untreated outcome among patrol officers after Tasers were made available, cannot be estimated from the data. It is a missing counterfactual. However, if we can invoke a parallel trends assumption, we can nevertheless identify the ATT via a difference-in-difference estimator.

To derive the parallel trends condition, let period $t = b$ denote some pre-treatment date. The population analogue of the difference-in-difference estimator can be written as

$$\triangle_a = \left\{E\left[Y_a(1)|D = 1\right] - E\left[Y_b(0)|D = 1\right]\right\} - \left\{E\left[Y_a(1)|D = 0\right] - E\left[Y_b(1)|D = 0\right]\right\}$$

The difference-in-difference estimator subtracts the before-after difference in the comparison group, $E[Y_a(1)|D=0] - E[Y_b(1)|D=0]$, from the before-after difference in the treatment group, $E[Y_a(1)|D=1] - E[Y_b(0)|D=1]$. It uses the comparison group to adjust for changes that would have taken place in the treatment group had Tasers not been issued to patrol officers. The before- after difference for the comparison group involves treated outcomes $Y_a(1)$ and $Y_b(1)$ because sergeants had access to Tasers during our entire sample period.

Adding and subtracting the missing counterfactual $E[Y_a(0)|D=1]$ and rearranging terms, we can rewrite the difference-in-difference estimator as

$$\triangle_a = ATT_a + \{E[Y_a(0)|D=1] - E[Y_b(0)|D=1]\} - \{E[Y_a(1)|D=0] - E[Y_b(1)|D=0]\}$$

In this expression, the first term in braces represents a counterfactual trend: it represents the trend in outcomes involving patrol officers that would have obtained had patrol officers never been issued Tasers. The second term in braces represents the trend in outcomes involving sergeants, who had access to Tasers the entire time. Our parallel trend assumption involves these two terms: if the counterfactual trend among outcomes involving patrol officers equals the trend in outcomes involving sergeants, then the difference-in-difference estimator identifies the ATT.

A sufficient condition for the parallel trends assumption to hold at all time periods can be written as for all t:

$$E[Y_t(0)|D=1] - E[Y_t(1)|D=0] = \lambda$$

For $t \geq t$', the first term on the left side of the above expression cannot be estimated because it represents a missing counterfactual. However, for $t < t$', the expressions can be estimated and used to test the validity of the assumption.

2.4.2 Estimation

For estimation, we can define the observed dependent variables, such as use-of-force incidents/100 police, in terms of the potential outcomes, the treatment dummy, and the treatment status dummy as

$$Y_{Dt} = (1 - D + DD_t)Y_t(1) + D(1 - D_t)Y_t(0)$$

which implies that $Y_{0t} = Y_t(1)$ for $t = 1, ..., T$; $Y_{1t} = Y_t(0)$, $t < t'$; and $Y_{1t} = Y_t(1)$, $t \geq t'$. In the development above, we allowed the ATT to vary by month. For estimation purposes, we impose more structure, allowing the ATT to vary between the first and second post-treatment periods, but assuming it to be constant within those periods. Letting t'' denote September 2012, we assume $ATT_t = \beta_1$for $t = t', ..., t'' - 1$, and $ATT_t = \beta_2$for $t = t'', ..., T$. Define two post-treatment dummy variables corresponding to these two time periods: $P_{1t} = 1(t' \leq t < t'')$ and $P_{2t} = 1(t \geq t'')$ where the expression $1(z) = 1$ if z is true and zero otherwise.

With these definitions, we could use ordinary least squares (OLS) to estimate β_1 and β_2 by means of the following regression:

$$Y_{Dt} = \beta_0 + \beta_1 DP_{1t} + \beta_2 DP_{2t} + \beta_3 P_{1t} + \beta_4 P_{2t} + \beta_5 D + \varepsilon_{Dt} \quad (1)$$

for $D = 1, 0$ and $t = 1, ..., T$. In (1), there are two observations for each time period, one for the patrol officers and one for the sergeants. The ATT's are estimated as the coefficients on interactions between variables defined in terms of treatment status and the post-treatment dummies. In addition, the regression includes the two post-treatment dummies and a treatment status variable D, which equals one for the patrol officers and equals zero for the sergeants. Finally, ε_{Dt} is an idiosyncratic disturbance term that varies by treatment status and over time.

An equivalent approach to estimating β_1 and β_2 can be obtained by evaluating (1) at $D = 0$ and $D = 1$, then subtracting the former from the latter. This yields:

$$Y_{1t} - Y_{0t} = \beta_1 P_{1t} + \beta_2 P_{2t} + \beta_5 + \varepsilon_{1t} - \varepsilon_{0t} \quad (2)$$

for $D = 1, 0$ and $t = 1, ..., T$. OLS applied to (2) yields estimates of β_1 and β_2 that are algebraically identical to OLS estimates obtained from (1). The advantage of estimating (2) is that it provides a straightforward means of dealing with potential temporal dependence in the composite disturbance term $\varepsilon_{1t} - \varepsilon_{0t}$. To test for dependence, we estimated AR(1) models that allowed for 1st-order autocorrelation in $\varepsilon_{1t} - \varepsilon_{0t}$. For those few outcomes for which the autocorrelation parameter was significant, we report maximum likelihood estimates. For the remaining outcomes, we report OLS estimates. In all cases, the two sets of coefficients were very similar.

We can augment this specification to test for parallel pre-treatment trends by adding pre- treatment period dummies to (2), which is equivalent to adding interactions between those period dummies and the treatment status dummy to (1). Since we have one monthly observation each for patrol officers and sergeants, we aggregate months into years in order to test for parallel trends. Define the year dummies as $A_{jt} = 1$ if month t falls within year $2000 + j$ and $A_{jt} = 0$ otherwise, for $j = 5, ..., 9$, and $A_{10t} = 1$ if month t is January or February 2010, and 0 otherwise. Choosing 2009 as the base year, we then generalize equation (2) as

$$Y_{1t} - Y_{0t} = \beta_1 P_{1t} + \beta_2 P_{2t} + \beta_5 + \sum_{j=6, j\neq 9}^{10} \gamma_j A_{jt} + \varepsilon_{1t} - \varepsilon_{0t} \quad (3)$$

The test for parallel trends is computed as a test for the joint hypothesis[4]

4. One might have concerns about the power of this test, considering the number of observations involved. Our concerns are at least partly allayed by the fact that the test often rejects for a non- linear alternative to the model above, as discussed in Section 2.5.

$$H_0 : \gamma_5 = ...\gamma_8 = \gamma_{10} = 0 \ (4)$$

2.4.3 Defining the comparison group

Up to this point, we have discussed the comparison group as consisting of sergeants, or more precisely, of observations constructed from incidents in which sergeants were involved. However, the specific definition of sergeant-involved incidents could have implications for our estimates. The reason has to do with the pre-treatment deployment of Tasers. Before they were assigned to patrol officers, Tasers were available only to sergeants. Thus, sergeants may have been called to incidents specifically to provide Taser support. This is a type of spillover effect and means that Taser use among sergeants may have been higher prior to treatment than it would have been otherwise. Presumably, sergeants' use of Tasers post-treatment would have fallen under this scenario, since there would have been no need to call in a sergeant to provide Taser support once the weapons were available to patrol officers.

To gain insight into this problem and provide a potential solution, we initially worked with two comparison groups. The first we refer to as sergeant-involved. These were constructed from all incidents at which sergeants were present, including those at which patrol officers were present. These would include incidents to which sergeants may have been called expressly to provide Taser support.

We refer to our second comparison group as sergeant-only. It was constructed from all incidents at which only sergeants were present, usually a single sergeant acting on his own. Presumably, these were incidents that the sergeants encountered while on patrol. It should exclude incidents to which they were called to provide Taser support, since Tasers were available to sergeants from the beginning of our sample period.

Figure 2.2 plots the number of use-of-force incidents involving Tasers per 100 officers for both of these potential comparison groups. Taser-related incidents were higher pre-treatment, and fell more post-treatment, for the sergeant-involved case than for the sergeant-only case. This is consistent with the spillover hypothesis.

However, what is important for estimation is not so much the effect of spillovers on sergeants' Taser use per se, but rather its effect on their use of other weapons. The reason is that we seek to estimate the effect of CPD's Taser policy on the use of other weapons (and injuries and other outcomes), and it is the before-after change in sergeants' use of those other weapons that affects our difference-in-difference estimates. If Tasers are substitutes for other weapons, then the spillover effect described above may have led sergeants to use other weapons less often pre- treatment, and more often post-treatment, than they would have in the absence of the spillover. This would impart a downward bias (in algebraic terms) to our estimates. Depending on the rate at which other weapons substitute for Tasers, it could also impart downward bias on our estimated effect of the introduction of Tasers on the total use of force.

To analyze the importance of spillovers, we estimated all regressions two ways, once with the sergeant-involved comparison group and once with the sergeant-only group. The estimates based on the sergeant-involved group were always algebraically smaller, consistent with the discussion above. For this reason, we present estimates based on the sergeant-only comparison group in the next Section.

2.5 Results

2.5.1 Main findings

Table 2.4 presents estimates of the coefficients of equation (2) above, where the dependent variables measure use-of-force incidents per 100 police, by type of force and for all types of force in total. The coefficient on P_1 is the estimate of β_1; the coefficient on P_2 is the estimate of β_2. Standard errors are reported in parentheses. The bottom row of the table presents the p-values for the tests of parallel trends from equation (4). They are based on the regression in equation (3). Except for the case of Tasers themselves, none of the tests rejects the null hypothesis of parallel trends. Figure 2.3 plots estimates of the γ_j coefficients from equation (3), together with 90 percent confidence intervals, for total use-of-force incidents. The joint test for parallel trends failed to reject, as seen in at the bottom of column (7). The Figure shows that none of the individual coefficients were significant individually, either.

The estimated treatment effects, reported in the top two rows of Table 2.4, differ between the two post-treatment time periods. We focus first on the first period, from March 2010 to August 2012, corresponding to P_1. The estimate in column (4) shows, not surprisingly, that the policy change greatly increased the use of Tasers among patrol officers. This is what Figure 2.1 would have led one to expect.

More interesting are the estimated effects of introducing Tasers on other uses of force. The effect on incidents in the "no force" category was negative, though small and insignificant. The availability of Tasers reduced the use of minor force by 0.064, which amounts to about 25 percent of the pre-treatment mean among patrol officers. This coefficient is significant at the 10 percent level. The use of intermediate and major force also fell, although only the coefficient for intermediate force is significant, and then only at the 10 percent level. The coefficient on firearm use is positive, contrary to expectation, though insignificant. Taken as a whole, there is marginally significant evidence that the introduction of Tasers led patrol officers to substitute away from lesser types of force, and no significant evidence that they substituted away from greater types of force.

The coefficient in the last column shows that total number of use-of-force incidents rose during the first period following the introduction of Tasers to patrol officers. The estimated increase of 0.277 incidents/100 officers/month amounts to just over 10 percent of the pre-treatment mean among patrol officers. It is significant at the 10 percent level. Total use-of-force incidents rose because Taser-involved incidents rose by more than enough to counteract the small decline in other types of force.

The second row of the Table reports estimates of β_2, the effect of Tasers during the second post-treatment period, following the re-training of late 2012. The coefficient in column (4), for uses of force involving Tasers, is essentially the same as its counterpart for the initial post-treatment period. This may seem surprising at first, considering the large drop in Taser incidents apparent in Figure 2.1 that took place in late 2012. The explanation has to do with the behavior of sergeants. Table 2.2 shows that Taser use fell among sergeants during this period. The estimate of β_2 reflects the fact that the change in Taser use between the second post-treatment period and the pre-treatment period is greater for patrol officers than for sergeants, even though both groups reduced their use of Tasers between the first and second post-treatment periods.

In the period after the re-training of late 2012, the availability of Tasers among patrol officers had different effects on the use of other weapons than it had during the earlier period. Estimates are reported in the second row of the Table. During this period, Tasers widely substituted for other weapons. The estimate in column (3) shows that they substituted for other intermediate uses of force, as one might expect. Column (5) shows that they also substituted for major force. This is consistent with the theoretical model in Section 2.2, considering that Table 2.2 shows that injury rates to police are lower for incidents involving Tasers than for incidents involving major force.

Columns (1) and (2) show that Tasers also substituted for lesser uses of force. Taken at face value, this is consistent with concerns raised by researchers and others regarding over-reliance on Tasers, or "lazy cop syndrome" (Alpert and Dunham [2010]). At the same time, CPD reporting requirements are such that it is difficult generalize from reported incidents involving no or low force to the population of such incidents. Column (6) shows that firearms are the exception to the rule. The estimated effect of Tasers during this time period is negative, but it is barely larger than its standard error. Thus, there is no significant evidence that Tasers substitute for firearms. Finally, the last column of the Table shows that the availability of Tasers led to a reduction of 0.368 use-of-force incidents/100 officers during the second post-treatment period. This amounts to 14 percent of the pre-treatment average among patrol officers.

Table 2.5 reports how the introduction of Tasers affected injury rates per incident and the number of injuries per 100 police officers. Results for injuries/use-of-force incident, the focus of most previous research, are reported in column (1) for police officers and column (2) for civilian subjects. The first column shows that the introduction of Tasers reduced the likelihood that officers were injured during an incident involving police use of force. This is true before and after the re- training in late 2012. However, the second column shows that the subjects involved in those incidents enjoyed no such benefit: the coefficients for both time periods are small and insignificant.

Although most studies of Tasers have found that they reduce injury rates, our findings to the contrary have several precedents. Smith et al. [2007] analysis of the Richland County, South Carolina Sheriff's Department, for example, showed no link between Taser use and police or civilian injuries (see also Smith et al. [2009]). Likewise,Lin and Jones [2010] found that Tasers reduced police injury rates, but reported mixed results for civilian subjects.

Columns (3) and (4) present results for the number of injuries per 100 police, the product of the number of incidents and the rate of injuries/incident. The estimates in column (3) show that police injuries fell after Tasers were introduced, although the effect was only significant after late 2012. Civilian injuries actually rose slightly, albeit insignificantly, in the first post-treatment period. They fell slightly during the second period, although that coefficient too is insignificant.

The results in Table 2.6 show how the introduction of Tasers affected the race distribution of civilians involved in police use-of-force incidents. Neither the share of blacks nor Hispanics changed significantly in either follow-up period. Put differently, Tasers did not exacerbate the disproportionate representation of minority civilians involved in use-of-force incidents, but they did not ameliorate it either.

2.5.2 Robustness

The model in equations (1) and (2) assumes additivity of the treatment effect; a natural alternative would be to assume proportionality. If our dependent variables were always positive, this would amount to merely specifying the dependent variable in logs. However, since our dependent variables sometimes equal zero, this option is not available to us. Instead, we estimated negative binomial regression models, which are a variant on the more familiar Poisson model that relaxes the mean-variance equality that is imposed by the Poisson.

For the negative binomial regressions, the dependent variable is specified to be an exponential, rather than linear, function of the treatment status dummy, period dummies, and interactions between the two, as appear on the right side of (1). The dependent variables for these models are the numbers of events, rather than their rate per 100 police or per incident. That is, they are the numerators of the dependent variables that appear in the linear regressions reported above. The log of the denominator is included on the right-hand side of the model with a unit coefficient, so that the other coefficients can be interpreted as proportionate changes in the rates.[5]

Estimates from these models were qualitatively similar to those from the linear models reported above. Almost all of the estimated treatment effects had the same signs, and most the estimated treatment effects that were significant at the 5 percent level or better remained significant at similar levels. An important difference involved the tests for parallel trends. Whereas only 1 of the 13 tests of the linear models reported in Tables 2.4 through 2.6 rejected the null of parallel trends, 7 of the 13 tests applied to the exponential models rejected. Based on this comparison, we prefer the additive to the proportional specification of the model.

5. We also estimated an unconstrained version of each model so as to test the null hypothesis that the coefficient on the denominator was equal to one. None of the tests rejected.

2.6 Conclusion

Tasers and other CED's have long been controversial. Civil libertarians complain that they represent excessive force and are often used when less aggressive tactics would suffice. At the same time, evidence suggests that they reduce injury rates per use-of-force incident. Several large police departments have sought to expand their Taser arsenals on the belief that doing so would reduce police shootings.

We took advantage of a change in policy at the Chicago Police Department to analyze how the availability of Tasers affects police behavior. Prior to the policy change, Tasers had been available only to sergeants and field training officers. Starting in March 2010, they were made available to patrol officers as well. We took advantage of that policy change to study how the availability of Tasers affected patrol officers' choice of tactics in use-of-force incidents, the total number of use-of-force incidents, injury rates per incidents, and total injuries. We also analyzed how Tasers affected the racial and ethnic makeup of civilians on the other side of those altercations.

Like many other studies, we found that Tasers reduced injury rates among police. Unlike most prior studies, we analyzed whether the availability of Tasers increased the overall use of force among patrol officers. During the first two years it did, although the effect was marginally significant. Total injuries to police fell, but injuries to civilians were unaffected. There is no evidence that Tasers affected police use of firearms.

Roughly two years after the policy change, the department purchased a new model of Taser and re-trained much of the force. After that point, Tasers became a substitute for other types of force, both greater and lesser. Total use of force fell, although subject injuries and police use of firearms remained unaffected.

One lesson is that training appears to matter. The producer of Tasers changed their training materials prior to the 2012 re-training at Chicago PD, adding materials on health risks and officer liability. It would be useful to analyze more systematically how different approaches to training affect police officers' choice of tactics in confrontational situations.

Figure 2.1: Taser-Related incidents for newly certified officers

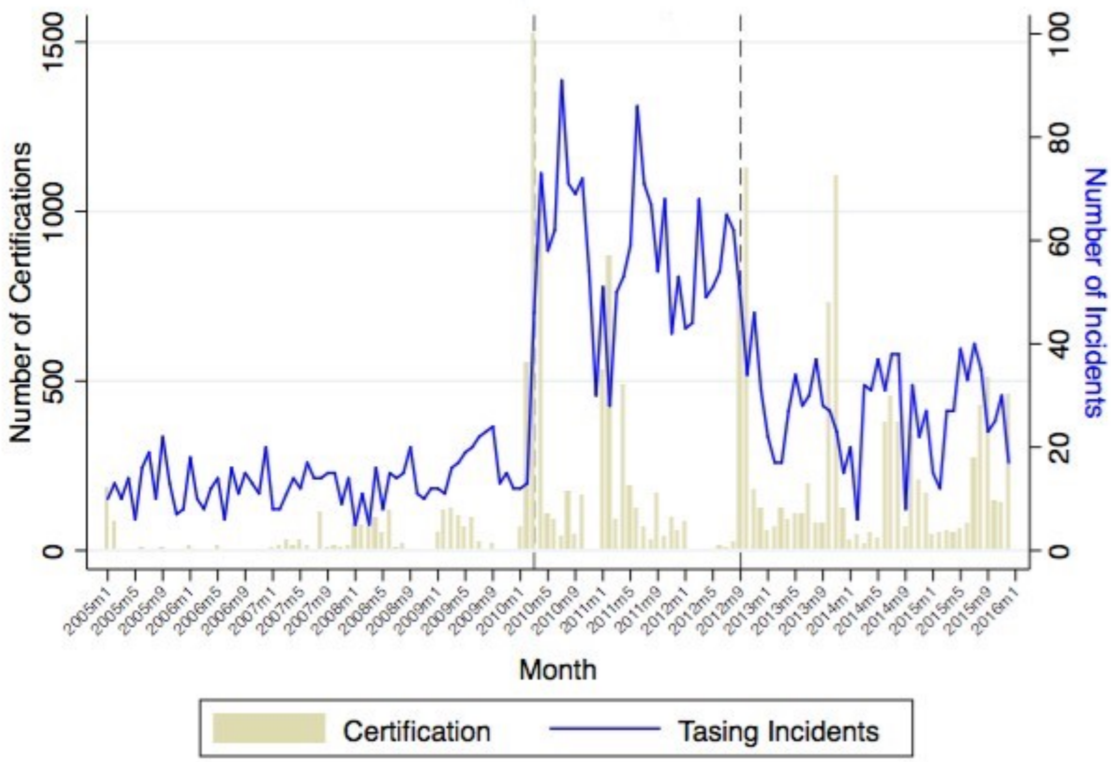


Figure 2.2: Taser-Related incidents per 100 sergeants

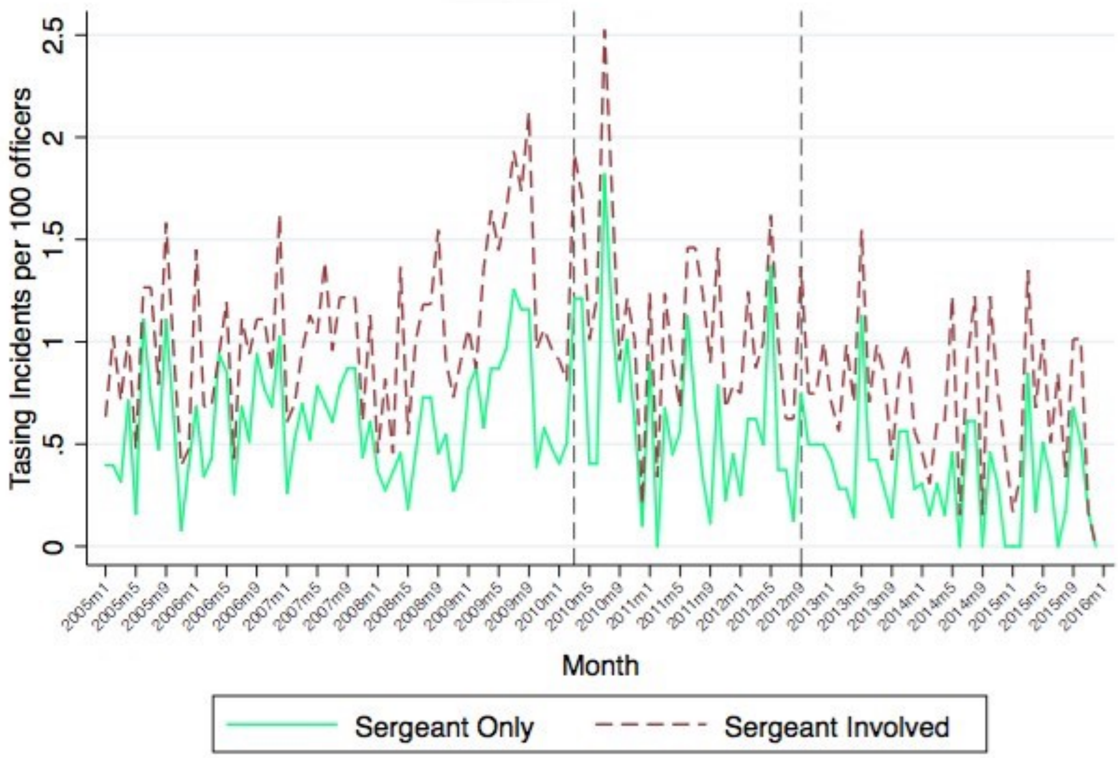


Figure 2.3: Use-of-Force incidents per 100 police

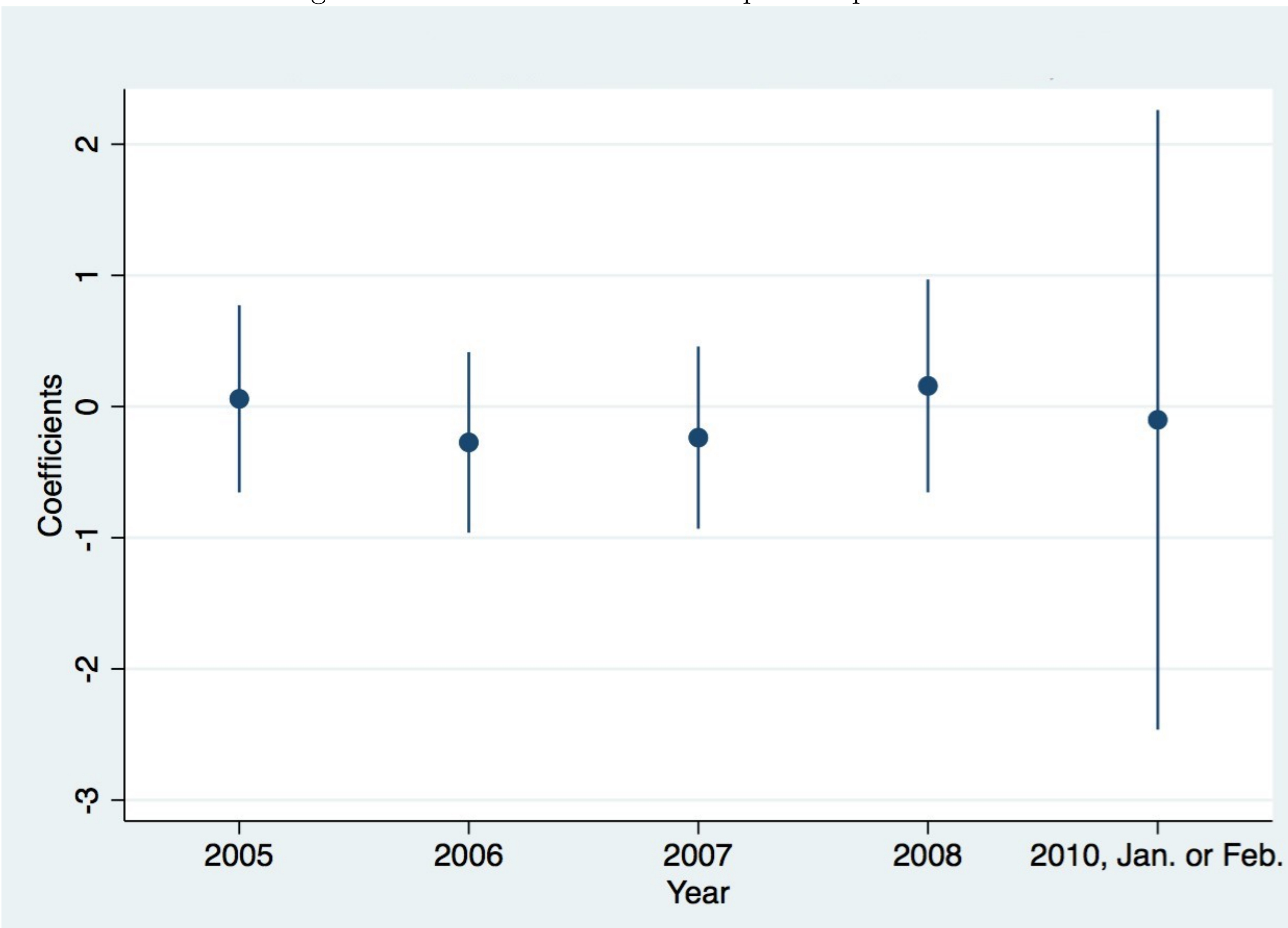


Table 2.1: Subject resistance, force options, and reporting requirements

Subject resistance	Force options	Our classification	TRR required?
Cooperative	Police presence, Verbal control	No force	If subject injured or alleges injury
Passive resistance			
	Holding techniques that do not involve pain-compliance, Pain-compliance techniques, Control instruments[a]	Minor force	If subject injured or alleges injury
Active resistance			
	Stunning; OC (pepper) spray, Capsaicin II powder, LRAD (acoustic device), Canines,	Intermediate force	Yes[b]
	Taser	Taser	Yes
Assailant			
Without weapons	Striking, Kicking, Powerful locks and pressure, Impact weapons, Impact munitions	Major force	Yes
Likely to cause injury	Same as above	Major force	Yes
Likely to cause serious Injury or death	Chokeholds	Major force	Yes
Likely to cause serious Injury or death	Firearms	Firearms	Yes

Notes: First and second columns adapted from Chicago Police Department (2012, 2016a). Third column gives our label for each group of force options. Fourth column adapted from Chicago Police Department (2016b). [a]OC (pepper) spray and LRAD (acoustic device) may also be used under certain circumstances with supervisor approval. [b]A TRR is not required when the subject's only action of resisting is fleeing, and the officer's actions did not extend beyond verbal commands and/or control holds utilized in conjunction with handcuffing and searching techniques which do not result in injury of allegation of injury (Chicago Police Department 2016b).

Table 2.2: Distribution of use-of-force incidents and injury rates to police and subjects by type of force

Type of force	Number	Percent	Injury rate/incident police (%)	Injury rate/incident subject (%)
1. No Force	2,214	6.1	19.2	9.8
2. Minor Force	3,257	9.0	16.8	11.9
3. Intermediate Force	20,128	55.7	26.0	22.3
4. Taser	3,678	10.2	16.5	36.9
5. Major Force	6,102	16.9	53.6	41.0
6. Firearm	733	2.0	26.5	58.1

Table 2.3: Mean monthly rates of use of force per 100 police, injuries/incident, injuries per incidents involving minority subjects, by time period and treatment group

	1/2005 to 2/2010	3/2010 to 8/2012	9/2012 to 12/2015
A. Patrol officers			
Use-of-force incidents/100 police			
1. No Force	0.17	0.20	0.17
2. Minor Force	0.26	0.25	0.25
3. Intermediate Force	1.65	1.56	1.48
4. Taser	0.02	0.55	0.26
5. Major Force	0.45	0.49	0.42
6. Firearm	0.05	0.07	0.05
Total	2.59	3.12	2.63
Officer injuries/incident	0.32	0.25	0.25
Subject injuries/incident	0.26	0.25	0.25
Officer injuries/100 police	0.83	0.77	0.67
Subject injuries/100 police	0.68	0.78	0.65
Share of UOF incidents involving at least one black subject	0.74	0.75	0.75
Share of UOF incidents involving at least one Hispanic subject	0.15	0.15	0.14
B. Sergeants			
Use-of-force incidents/100 police			
1. No Force	0.10	0.16	0.21
2. Minor Force	0.14	0.20	0.24
3. Intermediate Force	0.59	0.65	0.94
4. Taser	0.62	0.64	0.35
5. Major Force	0.14	0.20	0.24
6. Firearm	0.03	0.02	0.04
Total	1.62	1.87	2.03
Officer injuries/incident	0.18	0.17	0.16
Subject injuries/incident	0.24	0.23	0.21
Officer injuries/100 police	0.29	0.31	0.36
Subject injuries/ 100 police	0.39	0.43	0.45
Share of UOF incidents involving at least one black subject	0.69	0.72	0.69
Share of UOF incidents involving at least one Hispanic subject	0.16	0.15	0.16

Notes: UOF =use of force.

Table 2.4: Effect of Introducing Tasers on Use of Force Incidents by Type of Force

Variable	No Force (1)	Minor (2)	Intermed. (3)	Taser (4)	Major (5)	Firearm (6)	All (7)
P_1	-0.025	-0.064	-0.143	0.518	-0.023	0.021	0.277
	(0.026)	(0.036)	(0.086)	(0.067)	(0.042)	(0.015)	(0.168)
P_2	-0.108	-0.102	-0.516	0.511	-0.139	-0.016	-0.378
	(0.022)	(0.032)	(0.078)	(0.061)	(0.038)	-(.014)	(0.153)
Constant	0.064	0.114	1.055	-0.600	0.315	0.026	0.979
	(0.018)	(0.020)	(0.049)	(0.038)	(0.024)	(0.009)	(0.122)
AR(1)	-0.177						0.173
	(0.087)						(0.077)
N	132	132	132	132	132	132	132
P-value, test for parallel trends	0.611	0.583	0.567	0.031	0.537	0.525	0.894

Notes: Standard errors are in parentheses. Coefficients in columns (1) and (7) were estimated by maximum likelihood; those in columns (2)-(6) were estimated by OLS. AR(1) indicates the autoregressive parameter. Dependent variable is difference in number of UOF incidents/100 police between patrol officers and sergeants. UOF =use of force.

Table 2.5: Effect of Introducing Tasers on Injury Rates/Incident and Injuries

	Injuries/Incident		Injuries/100	Police
	Police	Subjects	Police	Subjects
Variable	(1)	(2)	(3)	(4)
P_1	-0.065	0.000	-0.080	0.072
	(0.025)	(0.025)	(0.055)	(0.057)
P_2	-0.045	0.020	-0.225	-0.085
	(0.023)	(0.023)	(0.050)	(0.052)
N	132	132	132	132
P-value, test for parallel trends	0.626	0.852	0.616	0.907

Notes: Standard errors are in parentheses. Dependent variable is difference between patrol officers and sergeants in injuries rates and injuries.

Table 2.6: Effect of introducing Tasers on race/ethnicity distribution of subjects

	Black	Hispanic
Variable	(1)	(2)
P_1	-0.016	0.006
	(0.029)	(0.022)
P_2	0.021	-0.006
	(0.027)	(0.020)
N	132	132
P-value, test for parallel trends	0.504	0.750

Notes: Standard errors are in parentheses. Dependent variable is difference between patrol officers and sergeants in share of incidents involving at least one Black or hispanic.

CHAPTER 3

ESTIMATING (EASILY INTERPRETED) DYNAMIC TRAINING EFFECTS FROM EXPERIMENTAL DATA

Bocar A. Ba, John C. Ham, Robert J. Lalonde, and Xianghong Li

3.1 Introduction

Analysis based on a randomized experiment is often considered the gold standard for evaluation. The simplest case is when we want to measure the effect of being assigned to the treatment group on an outcome of interest. A potentially more complicated but well known case occurs when the parameter of interest is the effect of participating in the program. For example, in the Oregon Health Experiment, individuals were randomly assigned either to a treatment group that received Medicaid eligibility, or to a control group that remained ineligible.[1] The effect of having Medicaid insurance on, e.g. mental health status, is not directly observable from the experimental data unless there is full take-up or compliance, i.e., unless all of those eligible for Medicaid actually obtain Medicaid coverage. With less than full compliance, having coverage is unlikely to be random, so one cannot simply regress mental health status on having Medicaid coverage. Of course, we can consistently estimate the effect of insurance coverage on mental health status using a two-stage least square regression with the random assignment dummy as the excluded instrument.

A second potential estimation problem with experimental data is less well known. Suppose we want to estimate the effect of random assignment on a subset of data, and in this

1. See Finkelstein et al. [2012] for a detailed description and analysis of the Oregon Health Experiment.

subset the treatments differ, on average, from the controls. For example, suppose we first look at the effect of random assignment on employment status one year after training ends. Since we conduct this comparison on the full data set, comparability between treatments and controls is guaranteed due to random assignment. Suppose that being in the treatment group increases employment, and that we are also interested in the effect of random assignment on wages (as a measure of increased productivity) for those who are employed. It is likely that treatment helped less-able members of the treatment group find employment and in this case, the average quality of the controls will be higher than the average quality of the treatments, i.e. random assignment does not hold among wage earners. Now to estimate the effect of random assignment on wages, we must use nonexperimental methods to model the selection into employment for the treatments and controls. Note that there is no further randomization, or simple solution, to this problem.

In this paper, using data from the National Job Training Partnership Act study (hereafter JTPA), we consider an estimation problem that involves both the endogeneity of training participation and the selection problems discussed above. In other words, we ask the question 'what is the effect of participating in the training provided by JTPA on current and future employment and unemployment spells?' In considering JTPA, it is important to focus on participation in training since the program had substantial non-compliance: only 73% of the treatments participated, while 45% of the controls managed to obtain equivalent training from other agencies. In the JTPA study, treatments could undertake JTPA training in any of the eighteen months after the baseline and the training was part-time, i.e. an individual could work at the same time that they received training. Also, the controls were able to receive equivalent training from other social agencies in the eighteen months after the baseline.

We estimate the effects of participating in JTPA-type training on employment and unemployment hazard functions to understand how JTPA worked, and to see whether there is any complementarity between JTPA and other training programs. For example, training provided by the National Supported Work Demonstration study (hereafter the NSW study) increased employment durations but did not affect unemployment duration of the participants – see Ham and LaLonde [1996] (hereafter HL96).[2] If JTPA decreased the unemployment duration of the participants, as we find below and as also was found in Eberwein et al. [1997, 2002] (hereafter EHL97 and EHL02 respectively), it would make sense to combine the JTPA and NSW programs. Once we have consistently estimated the parameters of the hazard functions, we transform them into counterfactual effects that are easily understood by policy makers and comparable across models; here we also provide some of the first consistent confidence intervals for these counterfactual effects.[3]

We also address an important identification issue that naturally arises in our dynamic context and implicitly arises in the more standard regression approach when treatment can occur at different times after the baseline. An intuitive way of viewing this problem is to consider the possibility that, before they enter training, the treatments may search less intensely than the controls because the treatments know that free easy-access training is available to them in future periods. If this is the case, employment status in each period will depend on treatment status, even conditional on training participation, and the standard identification assumption for estimating program participation will be violated. To offer more intuition on how future subsidized training can affect today's behaviour, we consider a perfect foresight life-cycle consumption model with two goods. We show that if we randomize the sample into a treatment group which receives a subsidy on one of the goods in the future, and a control group that does not receive this subsidy, the future subsidy will affect consumption of both goods in all periods for the treatments. Since we have two exclusion restrictions involving random assignment, we can shed light on this problem by testing the

2. HL96 actually looked at the effect of NSW on the treatments, but since the program had very high compliance, these effects can be expected to carry over to the participants. They also found that the selection effects were quite serious when looking at the effect of NSW training on future unemployment spells.

3. Ham et al. [2016] provided the first consistent estimates of confidence intervals for these effects.

overidentifying assumption in our model.

The outline of the paper is as follows. In Section 3.2 we review the JTPA program and provide thorough descriptive, non-causal, estimates of the effect of training participation on transitions between employment and nonemployment. In Section 3.3 we discuss estimating training participation effects with dynamic selection and endogeneity. In Section 3.4 we show how we consistently estimate our counterfactual effects and their confidence intervals. In Section 3.5 we discuss the above-mentioned identification problem and overidentifying test in more detail.

We present our estimates in Section 3.6 and confirm the results of EHL97, i.e. that participation in JTPA training reduced unemployment duration but did not affect employment duration, and hence would indeed complement the type of training considered in the NSW study. We next show that we do not reject the overidentifying restriction in our model. This suggests that, in our context, the standard approach of using treatment assignment as an instrument to estimating participation effects is still valid. Further, we find that our counterfactual estimates are informative and easy to interpret.

In Section 3.7 we compare the approach taken here to some of the recent work in the European context, which looks at the effect of training participation on labor market when one does not have access to randomized data. We also discuss the differences between the participation in conventional European labor market programs and in JTPA. On the basis of these comparisons, we argue that the approaches that we take and recent European studies take are converging in spite of differences in the respective training programs. We conclude in Section 3.8.

3.2 The JTPA Experimental Data and Empirical Estimates of Treatment and Participation Effects

3.2.1 The National JTPA Study, the JTPA Program, and Our Sample

The data that we use in this paper are from the adult women recommended for classroom training in the National JTPA study, the first experiment involving an ongoing social program in the United States. These are the same data used in EHL97 and EHL02. They also are a subset of the data used by Heckman and Smith [1999, 2000, 2004], Heckman et al. [1998], and Heckman et al. [1997].[4] The data from the JPTA study have been very influential in the development of new econometric methods for evaluating programs and policies.

The evaluation of this program is influenced by several institutional features of JTPA. First, although JTPA – like its successor program, the Workforce Investment Act (WIA) – was a federal program, the program was operated by local officials, who subcontracted training to many providers. These providers included the local community colleges, for-profit proprietary schools, and non-profit social service agencies. The decentralized design of JTPA means that the federal government could not prevent the controls from receiving similar services in the community, sometimes from the same provider. Indeed, about 45% of the control group in our sample acquired training similar to that of those in the treatment group.[5]

Second, JTPA provided a set of customized training and employment-related services. Program operators assigned each applicant a training plan according to her employment history. If an applicant had a relatively strong employment history, officials might decide to provide her with job search assistance first and occupational training only if she remained

4. We remain grateful for James Heckman's and Jeffrey Smith's help in our using these data here and in EHL97 and EHL02.

5. The phenomenon of controls obtaining training through other agencies is known as 'control group substitution' in the training literature.

unemployed. By contrast, an applicant with a very poor employment history might first receive adult basic education, followed by occupational training in a classroom setting, and finally job search assistance. Accordingly, in this study we evaluate the effect of JTPA plans that include classroom training (CT). Note that the design of JTPA's training plans implies that we cannot evaluate the effects of CT per se.

Third, JTPA, like the WIA, was not an entitlement program: being economically disadvantaged was not sufficient to receive JTPA services. The statute required program operators to admit people who would "benefit" from JTPA employment and training services. Prior to the baseline month, adult women, especially, were seeking to join the labor force (Heckman and Smith, 2004) and saw training as a potential job search tool. Accordingly, in the experimental data, we observe high rates of training entry near the baseline, including those in the control group. However, individuals could enter training at any time during the 18 months after the baseline. For the treatment group, 57% entered training within the first three months, 14% entered between month 4 and month 18, 2% after month 18, and 27% never entered training, while the corresponding distributions are 22% , 19% , 4% , and 55% respectively for the control group.

An important objective of those who designed the National JTPA Study was to preserve as much as possible how JTPA operated in the field (Bloom et al. [1993]). Two major steps were taken to serve this purpose. First, applicants were randomized not when eligibility was determined, but after their training plans had been explained to them and they had agreed to training under its terms. At that point, a woman assigned to the control group was prevented from receiving any JTPA services for 18 months but still possessed the information in the training plan and could obtain similar training within this 18-month period through other social agencies. Second, the size of the control group was kept to one-third of sample, not one-half of the sample. This was done to lessen the need for extra recruiting to fill JTPA spots; if the control group was one-half of the sample, JTPA operators would have had to recruit twice the usual number of volunteers. The resulting study was conducted in 16 sites, and admitted applicants between November 1987 and September 1989.

The sample construction of this study follows EHL97 and EHL02. As noted above, we examine adult women recommended for the Classroom Training (CT) set of services. Women who entered the program prior to January 1988 are excluded from our sample. At the baseline, individuals were either in the middle of a nonemployment spell or employment spell. Following EHL97, we consider only those in the middle of a nonemployment spell since women who were employed at the baseline consist of only about 15% of this group, and hence there are insufficient observations to analyze the effect of training participation on left-censored employment spells. We exclude women for whom the following information was missing: the start dates of their left-censored nonemployment spells in progress at the baseline; employment outcome data between the baseline month and month 18; site information; age or race. We use imputation methods to fill in missing responses for years of schooling, high school dropout status, and number of children under the age of four, married with spouse present, and never married.

The resulting sample contains 1,940 women. As noted above, at the baseline, each individual was in the middle of a nonemployment spell (i.e., a left-censored nonemployment spell hereafter). This initial left-censored nonemployment spell was either censored at the end of the sample period or was followed by subsequent employment and nonemployment spells. (These spells beginning after the baseline will be referred to as fresh employment or fresh nonemployment spells hereafter.) About 77% of the sample has a fresh employment spell and about 55% of the sample has a fresh nonemployment spell.

An important test of the correct implementation of any social experiment is a check on whether the baseline characteristics of the treatments and controls are balanced. As shown by the first three columns of Table 3.1, the means of all characteristics are not statistically significantly different between the two groups, except for two characteristics – the fraction of never married and the fraction of having received cash welfare (AFDC) for at least two years prior to the baseline.[6]

3.2.2 The Effect of Being Assigned to the Treatment Group on Employment Rates and Evidence of Balance in the Characteristics of the Treatments and Controls

The effect of being assigned to the treatment group of JTPA is seen in Figure 3.1 as the difference between the post-baseline employment rates of the treatments and controls. We observe in Figure 3.1 that the treatments and controls employment rates begin to diverge around the 6^{th} month after the baseline, and by the 18^{th} month (the last month after the baseline for which we have data on all participants), the treatment effect on the employment rates is about 6 percentage points (or about 15% of the employment rates of the controls).

Next, we compare the training participation and employment of the treatment and control groups. Because of treatment no-shows and early dropouts, as well as control group substitution, at month 18, only about 71% of the treatments had received any training, whereas about 41% of the controls had received similar training, with a difference in participation of 30 percentage points between the two groups. As is now standard in the program evaluation literature, at month 18, the estimated local average treatment effect (LATE) is the difference in employment rate divided by the difference in training participation rate between the treatment and control group. The LATE estimate suggests that participating in training increases employment rate by 20 percentage points among compliers (individuals who choose to participate in training only if they are assigned to the treatment group).[7]

6. The AFDC program was the program through which disadvantaged families received cash welfare assistance up until 1996, when it was replaced by the TANF (Temporary Assistance for Needy Families) program.

7. Under a more general assumption of heterogeneous treatment effects, this LATE parameter may be different from the average treatment effect parameter. In our nonlinear model that follows, we assume homogeneous treatment effects to keep our problem manageable.

This is a large estimate since the sample average of the employment rate at month 18 was 48% .

Turning to the last three columns of Table 3.1, we compare the mean characteristics of the trainees and non-trainees. Since entering training is a choice variable, there is no reason for these characteristics to balance. For this table, we define a trainee as any member of the sample (treatment or control) who enrolled in a training program before the end of the sample period. A non-trainee is anyone who did not enroll in training during the sample period. Notice that the pattern in the last three columns of Table 3.1 is different from that in the first three columns of Table 3.1. Among the non-trainees African-Americans and never married individuals are significantly over-represented and Hispanics are significantly under-represented. Of course, there is also the more important issue that trainees and non-trainees may have different unobservables.[8] We will investigate this issue again when we estimate the model and incorporate endogenous training by allowing for dependence between the unobservable in the training hazard function and the unobservables in the labor market hazard functions.

3.2.3 Kaplan-Meier Entry into Training Hazards: Treatments versus Controls

We begin by focusing on dynamic selection for the trainees and non-trainees into the fresh spells. The first three columns of Table 3.2 show the means of the explanatory variables for the trainees and non-trainees, as well as the differences in these means, for those who make it to a fresh employment spell. The next three columns of Table 3.2 show the corresponding information for the trainees and non-trainees who have a fresh nonemployment spell. At least in terms of observables, there do not seem to be large differences between trainees and

8. Since we control for observable characteristics in estimation, differences in observed characteristics are important solely because they may signal differences in unobservables across the trainees and non-trainees.

non-trainees (except for that African-Americans are significantly under-represented among trainees who make it to a fresh employment spell) in either type of fresh spells, as indicated by columns (3) and (6) of Table 3.2.

We are also interested in comparing the dynamics of training entry of the treatments and controls. As shown by Figure 3.2, during the first four months after the baseline, the transition rate into training was substantially higher for the treatments than it was for the controls. We would expect a higher participation in training for the treatments because being randomly assigned access to (JTPA) job training effectively lowers the cost of participation in training for the treatments relative to the controls, so this result for the first four months is not surprising.

Notice, however, that by the fifth month following the baseline, the training hazards for the two groups converge. This is a puzzle since one might expect that the treatments would always have higher entry rates at any duration given that training is less expensive (in terms of effort) for them. However, several forms of selection are taking place by month 4. First, the treatments that are most eager to enter training have already done so. Second, the most able controls have left nonemployment already (although they can and still do participate in training after leaving nonemployment). Third, it may be harder for the controls to enter training in months 1-4 because it takes time for them to arrange it through another agency. It is difficult to disentangle these different effects to determine why treatments and controls have similar entry rates to training after month 4 without an explicit econometric model, such as the one we describe in the next Section.

We further compare individuals who enter training early versus those who enter late in Table 3.3. Columns (1), (2), and (4) of Table3.3 show the mean characteristics of individuals (treatments and controls combined) who entered training within the first three months, between months 4 and 18, and between months 19 and 33 following the baseline, respectively. Column (6) presents the mean characteristics for participants who never enter training during the sample period. The mean differences between the participants who entered training within the first three months and the three other groups are presented in columns (3), (5), and (7) respectively.

Column (3) shows that the women who entered training in months 1-3 after the baseline clearly have observable characteristics that make them more employable than those who entered training in months 4-18. Column (5) shows large differences in the observable characteristics of those who entered training in months 1-3 after the baseline versus those who entered training 19+ months after the baseline, but many of these differences are not statistically significant because only 400 women were followed beyond month 18. Finally, when we examine the observable differences between those who entered training in months 1-3 with all the non-trainees, it is clear that those who entered training in months 1-3 have characteristics that make them more employable. Thus, it is important to control for not only *who* eventually obtains training, but *when* they obtain training; we control for both *who* and *when* in our econometric model in Section 3.3.

3.2.4 Kaplan-Meier Nonemployment and Employment Hazards for Participants and Nonparticipants in Training

Next, we compare the trainees' and non-trainees' empirical transition rates out of the three types of spells: i) the left-censored nonemployment spells, ii) the fresh employment spells, and iii) the fresh nonemployment spells. If a woman enters training in period t after the baseline then she is treated as a non-trainee in the first t months after the baseline, and a trainee afterwards till the end of the sample period. Of course, these comparisons are meant to be descriptive since they do not account for dynamic selection into fresh spells and endogenous training participation. Figure 3.3 shows the Kaplan-Meier hazards of the trainees and non-trainees out of left-censored nonemployment spells. For the first three months after the baseline, the transition rates out of left-censored nonemployment spells are substantially larger for the non-trainees than the trainees. However, the trainees do better from month 4 onward.

In Figure 3.4, we plot the Kaplan-Meier hazards from the left-censored unemployment spell for trainees who started training within three months of the baseline and those who started more than three months after the baseline. Notice that these hazards are substantially larger during the three-month period after the baseline for those women who started training later. This pattern indicates that it is important to allow for training entry after the left-censored spell ends because many of those who quickly left the left-censored unemployment spell underwent training in later months.

In Figure 3.5, we further divide the participants into four groups based on whether they entered training during the sample period, and by whether they were unemployed for more than two years at the baseline. Beginning with the Kaplan-Meier hazards of trainees (the two solid lines), the transitions out of left-censored nonemployment spells are greater among those trainees whose spells were shorter at the baseline (being unemployed less than two years). The same pattern holds for non-trainees. This pattern persists mainly for the first six months after the baseline for non-trainees but up to 12 months after the baseline for trainees.

Finally, Figures 3.6 and 3.7 provide the Kaplan-Meier hazards rates out of fresh employment and fresh nonemployment spells for the trainees and non-trainees, respectively. These figures reveal that the trainees have greater transition rates out of employment than do the non-trainees in our sample.[9] However, it is also true that the rates out of fresh nonemployment spells tend to be higher for the trainees than the non-trainees. Note that, as shown by Figure 3.6 the hazard of both trainees and non-trainees for the fresh employment spells rises for the first three months and then begins to fall. By contrast, as shown by Figure 3.7, the hazard for the fresh nonemployment spells, especially for the trainees, exhibits negative

9. Interestingly, we do not find significant effect of training participation on fresh employment hazards when controlling for dynamic selection and endogenous training.

duration dependence throughout their spells.

Because relative to the controls, the treatments had significantly higher training participation (Figure 3.2) and employment rates (Figure 3.1), we know that training "worked" (Bloom et al. [1997]).[10] The transition rates out of the three types of spells must "add up" in the sense that when we simulate the model, we need to be able to replicate this effect. While it is possible for all three types of spells to be associated with higher transition rates for the trainees, we expect higher employment rates to be generated by greater hazard rates out of nonemployment and (possibly) lower hazard rates out of employment.[11] A plausible explanation for the observed greater hazards out of employment by trainees is dynamic selection. In other words, the trainees who are observed in fresh employment spells are less able (because of their unobservables) than the non-trainees, since training helps them exit the left-censored nonemployment spell. Thus, we would reiterate our argument that the results in this Section are descriptive and not causal. We will turn to causal effects of training participation when we present our econometric estimates in Section 3.6.

3.3 Econometric Issues that Occur in Analyzing in the JTPA Data using Duration Models

In this Section, we first review the HL96 model for estimating the effect of random assignment on labor market hazards in the presence of potential dynamic selection. We then consider the EHL97 model for estimating the effect of (endogenous) participation in training on the labor market hazards in the presence of potential dynamic selection. Given the complexity of our model, we assume that both parameters are constant across individuals. Finally we discuss the use of simulated annealing to maximize our likelihood functions and obtain our

10. Formal evidence that treatment has significantly increased training participation can be found in the treatment status coefficients (and standard errors) in column (4) of Tables 3 and 4 below.

11. We say "possibly" because, at least with regard to job search assistance, JTPA was aimed more at finding a job than keeping it.

estimates.

3.3.1 Accounting for Selection into Future Spells When Estimating Dynamic Treatment Effects in Duration Models

All individuals in our sample are in the middle of a left-censored nonemployment spell at the baseline, and we know the (pre-baseline) starting times of these spells. The spells either end with the sample period (i.e. are also right censored) or are followed by subsequent fresh employment and nonemployment spells. In the JTPA study, individuals could enter training any month during the sample period, and the act of entering training is assumed to affect a woman's labor market hazard functions in the months after entry. Note that since training is part-time and averages about 10 hours a week so individuals can be in training and in employment in the same month. As we discussed above, there was substantial non-compliance in JTPA. This motivates our focus on the effect on training participation in this paper.[12]

We need to estimate the left-censored spells jointly with the fresh spells. Otherwise if JTPA training helps the treatments get out of the left-censored unemployment spell, we might spuriously estimate that being assigned to the JTPA treatment group spuriously increased transitions out of employment, because on average the treatments getting to the employment spells are of lower quality. Although our focus here is on the effect of participating in JTPA training, as opposed to the effect of JTPA treatment status, it is helpful, nonetheless, for expositional purposes, for us to first discuss the EHL97 approach to dealing with dynamic selection when looking at the effect of JTPA treatment status.

Since our data are monthly, we begin by specifying discrete time hazard functions for the fresh employment and nonemployment spells. The probability of individual i leaving a fresh

12. By contrast, in the NSW, there was an 85% participation rate by the treatments and 0% participation by the controls, so it was not necessary to go beyond the effect of being assigned to the treatment group.

spell at duration t , conditional on being in this spell in the previous $t-1$ periods is given as follows[13]

$$\lambda_{ki}\left(t|X_i\left(c_{ki}+t\right),RA_i,\theta_{ki}\right)=$$

$$\frac{1}{1+\exp\left(-h_k\left(t\right)-X_i\left(c_{ki}+t\right)\beta_k-\gamma_k RA_i-\theta_{ki}\right)},\ k=U,E\ (1)$$

In (1), U and E denote fresh nonemployment and employment spells respectively, t denotes current spell duration, $h_k\left(t\right)$ denotes current duration dependence, c_{ki} is the calendar time at the start of the current spell for individual i , so $c_{ki}+t$ is the calendar time at duration t , $X_i\left(c_{ki}+t\right)$ denotes a vector of (possibly) time changing explanatory variables at calendar time $c_{ki}+t$, RA_i denotes a dummy variable coded 1 if the individual is in the treatment group and 0 otherwise, and θ_{ki} denotes an unobserved heterogeneity term. Note that random assignment ensures that RA_i is independent of θ_{ki} and the explanatory variables at the baseline, and we follow the literature by assuming that θ_{ki} is independent of the other explanatory variables at the baseline.

Next, consider the left-censored nonemployment spells. In principle, the hazard function for these spells will be a complicated function of the start date of the spell (if known), pre-baseline values of the explanatory variables, and the parameters in (1), including those determining the joint distribution function of the unobserved heterogeneity terms. Rather than deriving this hazard function, Heckman and Singer [1984b] suggest defining and estimating a new hazard function for the left-censored spells. Thus, our hazard function for the

13. We choose the discrete time logit form of the hazard rate for several reasons. First, it has been widely used in empirical work starting with Nickell [1979]. The logit form was also used in two important Monte Carlo studies: Baker and Melino [2000]and Li and Smith [2015]. Second, Lancaster [1979] show that the logit hazard converges to its widely used continuous time analogue, the mixed proportional hazard rate, as the time period goes to zero. Finally, we find it most fruitful to think of labor market transitions happening in discrete time. We could have equally well specified a continuous time hazard and integrated up to a discrete time hazard function as in .Meyer [1990] Which approach a researcher chooses is essentially a matter of taste and likely depends on whether he or she views events in the labor market as taking place in continuous or discrete time.

left-censored spells is as follows

$$\lambda_{U'i}\left(t|X_i\left(c_{0i}+t\right),RA_i,\theta_{U'i}\right)= \frac{1}{1+\exp\left(-h_{U'}\left(t\right)-X_i\left(c_{0i}+t\right)\beta_{U'}-\gamma_{U'}RA_i-\theta_{U'i}\right)} \quad (2)$$

where the subscript U' denotes a left-censored nonemployment spell, c_{0i} is calendar time at the baseline for individual i, and t is duration *since baseline*. The remaining variables are defined analogously to those in (1).

Since we know pre-baseline duration in the left-censored spell τ_i , the issue arises as to whether we should use this information in estimation; note that by random assignment, RA_i will be independent of τ_i To use τ_i in estimation, we rewrite (2) as[14]

$$\lambda_{U'i}\left(t+\tau_i|X_i\left(c_{0i}+t\right),RA_i,\theta_{U'i}\right)= \frac{1}{1+\exp\left(-h_{U'}\left(t+\tau_i\right)-X_i\left(c_{0i}+t\right)\beta_{U'}-\gamma_{U'}RA_i-\theta_{U'i}\right)} \quad (3)$$

EHL97 and HL96 find that replacing (2) with (3) increases the precision of the coefficients but does not qualitatively, or quantitatively, affect the size and sign of the estimated treatment effects.[15] Thus we estimate the model in terms of (1) and (3) by maximum likelihood, where the distribution function of the unobserved heterogeneity is given by $G\left(\theta_{U'},\theta_E,\theta_U\right)$.

Instead of presenting a general form of the likelihood function, we have found it most useful to consider an example. Consider the labor market history in Figure 3.8, and, for now, focus only on the three nonemployment and employment spells, ignoring the non-training

14. Note that strictly speaking we should use new notation for coefficients and the unobservables in (3) since they are now conditional on , but we skip doing this for ease of exposition.

15. We are unaware of any other evidence on the desirability of estimating (2) versus estimation of (3).

spell on the top of the graph.[16] The figure indicates that this woman has been unemployed at the baseline for τ months. She finds a job $t_{U'}$ months after the baseline. She then experiences an employment spell lasting t_E months, followed by an nonemployment spell that is censored (at the end of the sample period) after t_U months. Her contribution to the likelihood function is as follows

$$L = \iiint_{\theta_{U'},\theta_U,\theta_E} \begin{bmatrix} \begin{bmatrix} \lambda_{U'}\left(t_{U'}+\tau|X\left(c_0+t_{U'}\right),RA,\theta_{U'}\right)\cdot \\ \prod_{r=1}^{t_{U'}-1}\left(1-\lambda_{U'}\left(r+\tau|X\left(c_0+r\right),RA,\theta_{U'}\right)\right) \end{bmatrix}\cdot \\ \begin{bmatrix} \lambda_E\left(t_E|X\left(c_E+t_E\right),RA,\theta_E\right)\cdot \\ \prod_{r=1}^{t_{E-1}}\left(1-\lambda_E\left(r|X\left(c_E+r\right),RA,\theta_E\right)\right) \end{bmatrix}\cdot \\ \left[\prod_{r=1}^{t_U}\left(1-\lambda_U\left(r|X\left(c_U+r\right),RA,\theta_U\right)\right)\right] \end{bmatrix} dG\left(\theta_{U'},\theta_U,\theta_E\right) \quad (4)$$

We use the McCall [1996] multivariate generalization of the Heckman and Singer [1984a] approach, where $G\left(\theta_{U'},\theta_E,\theta_U\right)$ follows a discrete distribution given by

$\theta_1 = \left(\theta_{U_1'},\theta_{E_1},\theta_{U_1}\right)$ with probability p_1

$\theta_2 = \left(\theta_{U_2'},\theta_{E_2},\theta_{U_2}\right)$ with probability p_2

.....

$\theta_{J-1} = \left(\theta_{U_{J-1}'},\theta_{E_{J-1}},\theta_{U_{J-1}}\right)$ with probability p_{J-1} and

$\theta_J = \left(\theta_{U_J'},\theta_{E_J},\theta_{U_J}\right)$ with probability $p_J = 1-\sum_{j=1}^{J-1} p_j$.

By allowing dependency across the unobserved heterogeneity terms for the labor market hazard functions, we address the dynamic selection into the fresh spells (i.e. the loss of random assignment in the fresh spells) through our specification of the θ vector. We could estimate the treatment effect for the left-censored spells without considering the fresh spells, but we must consider the unobserved heterogeneity term $\theta_{U'}$. If not, we will underestimate

16. This example is taken from EHL97 but we have changed the notation to make it consistent with that in this paper.

the training effect in week 2 and beyond, because the unobserved heterogeneity will be negatively correlated with treatment status.[17] As a result, the difference between the empirical (Kaplan-Meier) hazard functions in the second period and beyond will underestimate the effect of being in the treatment group.[18]

3.3.2 *Accounting for Selection into Future Spells and the Endogeneity of Training Participation When Estimating Training Effects*

In this subSection, we turn to our main interest: estimating the effect of participation in training on the labor market hazards. Specifically, let $D_i\left(c_{ki}+t\right)$ denote a dummy variable coded to 1 if the individual has entered training by calendar time, $c_{ki}+t$, and 0 otherwise. Now the probability of individual i leaving a fresh spell at duration t , conditional on being in the current spell for the previous $t-1$ periods, is as follows

$$\lambda_{ki}\left(t|X_i\left(c_{ki}+t\right), D_i\left(c_{ki}+t\right), \theta_{ki}\right) = \frac{1}{1+\exp\left(-h_k\left(t\right)-X_i\left(c_{ki}+t\right)\beta_k-\gamma_k D_i\left(c_{ki}+t\right)-\theta_{ki}\right)},\ k=E,U \quad (5)$$

Further, the probability of individual i leaving a left-censored spell t periods after the baseline, conditional on being in the spell for the previous $t-1$ periods, is as follows

$$\lambda_{U'i}\left(t+\tau_i|X_i\left(c_{0i}+t\right), D_i\left(c_{0i}+t\right), \theta_{U'i}\right) =$$

17. Suppose being in the treatment group helps in exiting from the left-censored spell. Then, in the first period the average quality of the treatments who leave will be lower than the average quality of the controls who leave. Therefore, in the second period among individuals left in the left-censored spell the controls will be of higher quality, on average, than the treatments in terms of their unobserved ability. Thus, we will no longer have random assignment in the second period and later.

18. However, one could consistently estimate separate empirical survivor functions for the treatments and controls in each period, since survivor functions are measured over the full sample of treatments and controls in each period.

$$\frac{1}{1+\exp\left(-h_{U'}\left(t+\tau_i\right)-X_i\left(c_{0i}+t\right)\beta_{U'}-\gamma_{U'}D_i\left(c_{ki}+t\right)-\theta_{U'i}\right)} \quad (6)$$

We model the post-baseline training entry process jointly with the labor market hazard functions. The conditional probability of individual i entering training in post-baseline period r , conditional on having not entered by period $r-1$, is as follows

$$\lambda_{Si}\left(r|X_i\left(c_{0i}+r\right),RA_i,\theta_{Si}\right)=$$

$$\frac{1}{1+\exp\left(-h_S\left(r\right)-X_i\left(c_{0i}+r\right)\beta_S-\delta_S f(RA_i)-\theta_{Si}\right)} \quad (7)$$

here the subscript "S" denotes training, and $f\left(RA_i\right)$ is a function of RA_i and of interactions between RA_i and other variables or duration. Note that our approach can be viewed as the nonlinear analogue of estimating endogenous training participation effects using two-stage least squares where the $f\left(RA_i\right)$ terms are the excluded instruments.[19]

We denote the distribution function for the unobserved heterogeneity as $G\left(\theta_S,\theta_{U'},\theta_E,\theta_U\right)$, and again use the McCall [1996] approach described above to model this function. Note that random assignment ensures that RA_i is independent of $\left(\theta_S,\theta_{U'},\theta_E,\theta_U\right)$, as well as with the explanatory variables, at the baseline. We address the endogenous training problem by allowing θ_S and $\left(\theta_{U'},\theta_E,\theta_U\right)$ to be dependent.

Again, we use the labor market history in Figure 3.8 to illustrate the contribution to the likelihood for a representative trainee, where we now include the training entry spell into the likelihood function. She enters training S periods after the baseline while she is still in her left-censored nonemployment spell. Her likelihood contribution is as follows

19. Since we have three hazard functions, we will have three overidentifying restrictions. Our specification of is the same as that used by EHL97 and EHL2002.

$$L = \iiint_{\theta_{U'},\theta_U,\theta_E} \int_{\theta_S} \Gamma dG\left(\theta_{U'}, \theta_U, \theta_E, \theta_S\right) \quad (8)$$

where

$$\Gamma = \begin{bmatrix} \lambda_S\left(S|X\left(c_0+S\right), RA, \theta_S\right) \cdot \\ \prod_{r=1}^{S-1}\left(1-\lambda_S\left(r|X\left(c_0+r\right), RA, \theta_S\right)\right) \end{bmatrix} \cdot \begin{bmatrix} \lambda_{U'}\left(t_{U'}+\tau|X\left(c_0+t_{U'}\right) D\left(c_0+t_{U'}\right), \theta_{U'}\right) \cdot \\ \prod_{r=1}^{t_{U'}-1}\left(1-\lambda_{U'}\left(r+\tau|X\left(c_0+r\right), D\left(c_0+r\right), \theta_{U'}\right)\right) \end{bmatrix} \cdot \begin{bmatrix} \lambda_E\left(t_E|X\left(c_E+t_E\right), D\left(c_E+t_E\right), \theta_E\right) \cdot \\ \prod_{r=1}^{t_E-1}\left(1-\lambda_E\left(r|X\left(c_E+r\right), D\left(c_E+r\right), \theta_E\right)\right) \end{bmatrix} \cdot \left[\prod_{r=1}^{t_U}\left(1-\lambda_U\left(r|X\left(c_U+r\right), D\left(c_U+r\right), \theta_U\right)\right)\right]$$

Determining the number of support points, J , for the unobserved heterogeneity is a nontrivial task, since this is a non-standard testing situation; the additional support point is not identified under the null hypothesis that the probability associated with it is zero. Hence the likelihood ratio statistic does not have a standard Chi-square limiting distribution. However, the Monte Carlo results in Li and Smith [2015] (LS15 hereafter) suggest that given a flexible specification of duration dependence and the Heckman-Singer specification of the unobserved heterogeneity distribution, a good rule of thumb is to use a likelihood ratio test in the standard manner to determine the number of support points.[20] Hence for each model, we start by assuming no unobserved heterogeneity and then continue adding support points and keep the model with the fewest points of support that is not rejected by a standard likelihood ratio test.

In duration studies, since the issues of identification and the asymptotic distribution of the estimators often come up, we believe the following discussion and review of related theoretical

20. By standard manner, we mean use the Central Chi-Square distribution, with degrees of freedom equal to the number of additional parameters that would be added with the extra support point. In our model corresponding to (8), adding another support point means adding five parameters – the probability and the four components of the vector – so the test has five degrees of freedom.

research are informative. First, we assume that the duration dependence and the unobserved heterogeneity depend on a fixed number of parameters as the sample grows large. Thus, we have a standard nonlinear simultaneous equation model with one endogenous variable $D(.)$ and two excluded instruments from $f(RA_i)$ in each labor market hazard function, and by the properties of maximum likelihood, our estimators will have $\sqrt{n}$ convergence and be asymptotically normally distributed.[21]

Establishing identification of a version of this model with unrestricted duration dependence and unobserved heterogeneity is beyond the scope of this paper. To the best of our knowledge, the model's identification has not been investigated when the unobserved heterogeneity terms are dependent.[22] Further, Hausman and Woutersen [2014]is the only study to investigate the $\sqrt{n}$ convergence and asymptotic normality of the coefficients estimated from a duration model with unrestricted duration dependence and unrestricted unobserved heterogeneity. For the functional forms of the hazard they consider, they can establish these properties, in either continuous or discrete time, only when the model contains time changing explanatory variables.

3.3.3 Maximizing the Likelihood Function Using Simulated Annealing

The likelihood function given by equation (8) presents a difficult numerical optimization problem. The semiparametric mixture models resulting from the Heckman-Singer specification of unobserved heterogeneity often have multiple local optima. Moreover, these models can behave in complicated ways at the boundaries of the parameter space. Derivative-based algorithms provide few safeguards against terminating at inferior local optima or being trapped near the boundaries of the parameter space trying to find the maximum of

21. This, of course, only checks the order condition, and we are implicitly assuming that the corresponding rank condition also holds.

22. If all of the heterogeneity terms are independent, the log likelihood function separates into four parts, one for each type of spell, and the results of Ridder [1990] can be used to establish identification for each type of spell.

an asymptote of the likelihood function. Numerical issues associated with duration model estimation have been repeatedly documented in the literature; see for example Meyer [1990], Heckman and Walker [1990], Baker and Rea [1998], and Stevens [1999].

LS15 conducted a Monte Carlo study and showed that in estimating single-spell single-state duration models, an alternative numerical optimization algorithm, simulated annealing (SA), clearly dominates derivative-based algorithms such as DFP, a routine commonly used by economists. In the current study we used SA to obtain our maximum likelihood estimates of the models presented in Section 3.3. SA is a derivative-free algorithm first suggested by Corona et al. [1987].[23] The SA algorithm explores the surface of the likelihood function by moving both uphill and downhill. While the algorithm always retains uphill (improving) moves, it also allows for temporary acceptance of downhill (worsening) moves. Ultimately, this feature enables SA to escape from local optima and flat parts of likelihood functions. SA progresses through a sequence of stages. At each stage, several random searches of parameter space are initiated. As the algorithm moves through its stages, downhill moves become less likely to be accepted, and the search region of the parameter space is systematically refined.

In this study, we encountered derivative-based search algorithms terminating close to parameter boundaries in one specification (details described below). As mentioned above, this is a very common problem in estimating duration model. We estimated this model using three derivative-based algorithms including DFP and used multiple sets of starting values in order to eliminate the numerical problem. None of the attempts by derivative-based algorithms escaped the boundary region and none converged to the (better and not close to the parameter boundaries) optimum reached by SA. A close look at the termination values reached by all three derivative-based algorithms revealed that they terminated close to each other: the optima contain one component of θ (the unobserved heterogeneity parameter in the above hazard specification) estimates in very large absolute value with huge standard

23. A referee suggested that we review SA for readers who are unfamiliar with it.

error associated with it, and some other hazard coefficients exhibit larger standard errors compared to the optimum reached by the SA algorithm. As a result, the estimated hazards were close to either 1 or 0.[24]

Overall, our experience suggests that SA, as a search algorithm, has clear advantages in estimating duration models with Heckman-Singer type specification of unobserved heterogeneity.[25]

3.4 Interpreting the Estimated Hazard Functions

Since the estimated parameters of the hazard functions are difficult to interpret, we use two measures (and provide confidence intervals for them) that are much more intuitive. The first measure is the effect of changing an independent variable (such as going from never entering training to entering training as soon as possible) on the expected duration of each type of spells. The second is the effect of changing an explanatory variable on the percentage of time spent employed over a given time horizon (specified by the researcher). Below we describe the estimators corresponding to the two measures. EHL02 look at these effects for being assigned to the treatment group in JTPA, but do not calculate them for the effect of training participation.

24. The boundary region is probably an asymptote: once a derivative-based algorithm ventures along it, the algorithm cannot escape. Because SA does not rely on derivatives and conducts random search both uphill and downhill, it can escape such a trap. However, such a case can pose problem for SA. We have to control search to make it progress slowly (by a parameter of the SA routine). Goffe, Ferrier, and Rogers (1994) provide very helpful guidelines on implementing the SA algorithm, and LS15 offers Monte Carlo evidence of using SA to estimate duration models.

25. One possible drawback of SA algorithm is that an individual run of SA is more time consuming than an individual run of gradient-based method. But this difference is not dramatic, since the typical computational time for a single SA estimation run of our model specified in (8) (with about 80 parameters) is about 12 hours on an Intel® Core™ i7-6700 @ 3.40ghz processor.

3.4.0.1 Using Expected Durations to Understand the Parameter Estimates

The effect of changing an independent variable on the expected duration of each type of spell has a natural policy interpretation, since it is analogous to the coefficient on an independent variable when the dependent variable is the duration of a spell of a given type. Moreover, the effect of changing an independent variable on the expected duration has the advantage that it is comparable across continuous time hazards and discrete time hazards, as well as being comparable across weekly, monthly or quarterly discrete time hazards; this is not true for the hazard function parameter estimates. Further, hazard coefficient estimates can be quite sensitive to the specification of duration dependence and unobserved heterogeneity, while the Monte Carlo results in both Mroz and Zayats [2008] and LS15 show that this is less likely to be true for the expected duration. Looking at expected durations dates back at least to Ham and Rea [1987] and has been used in Ham et al. [1998][26], EHL02, Ham et al. [2016] (hereafter HLSS16), and Crépon, Ferracci, Jolivet and van den Berg (2016); only HLSS16 provide confidence interval for the expected duration.

We now consider calculating the expected duration of for example, a fresh employment spell for a given individual,[27] and drop the individual subscript for expositional ease. Conditional on the unobserved heterogeneity, θ_{Ej}, the probability that a fresh employment spell lasts t months is given by the density function

$$g_E\left(t|\theta_{Ej}\right) = \lambda_E\left(t|\theta_{Ej}\right)\prod_{r=1}^{t-1}\left(1-\lambda_E\left(r|\theta_{Ej}\right)\right) \quad (9)$$

Again, conditional on the unobserved heterogeneity θ_{Ej}, the expected duration for a fresh employment spell is therefore

26. Ham et al. [1998] carry out a nonlinear Oaxaca-type decomposition for the expected durations.

27. The calculations for the left-censored and fresh unemployment spells are analogous and omitted to save space.

$$ED_E\left(\theta_{Ej}\right) = \sum_{t=1}^{\infty} t \cdot g_E\left(t|\theta_{Ej}\right) \quad (10)$$

Thus the overall expected duration is given as follows

$$ED_E = \sum_{j=1}^{J} p_j \cdot ED_E\left(\theta_{Ej}\right) \quad (11)$$

Since there is no guarantee that this expected duration will be finite, we instead calculate a truncated expected duration conditional on θ_{Ej} as follows

$$ED_E^{trun}\left(\theta_{Ej}\right) = \sum_{t}^{T^*-1} t \cdot g_E\left(t|\theta_{Ej}\right) + T^* S_E\left(T^*|\theta_{Ej}\right) \quad (12)$$

where $S_E\left(T^*|\theta_{Ej}\right) = \prod_{r=1}^{T^*}\left(1 - \lambda_E\left(r|\theta_{Ej}\right)\right)$ and we set $T^* = 60$ for this study. Finally, we have for the truncated expected duration

$$ED_E^{trun} = \sum_{j=1}^{J} p_j \cdot ED_E^{trun}\left(\theta_{Ej}\right) \quad (13)$$

We can calculate (13) for each individual, given her explanatory variables, and then take the (sample) average over all individuals. For each individual, within the sample, the calculations are based on the observed characteristics; out of the sample, the last period variable values are used. For example, if we observe an individual for 20 months, then we use observed characteristics for the first 20 months, and for months 21 to 60, we use the month 20 values for this individual. To prevent the out-of-sample durations from having a disproportionate impact on the estimated truncated expected duration, we also follow EHL02 and freeze the duration dependence term for long durations in this calculation. Specifically, we freeze the duration function for durations longer than 50 months at 50 months for left-censored nonemployment spells, freeze the duration function for fresh employment spells at 15 months, and freeze the duration function for fresh nonemployment spells at 10 months.

To conduct the counterfactual analysis for each type of spell k , we first set an explanatory variable of interest at a given level, e.g. entering training immediately, for each person, while leaving the other variables at their (individual) actual values and calculate the sample average truncated expected duration, which we denote by $ED1$. Next, we set this explanatory variable at a different level, e.g. never entering training, for each person, while again leaving the other variables at the individual actual values and make the analogous calculation, which we denote by $ED0$. Then the difference $ED1 - ED0$ is the estimated effect of the change in training status. The corresponding 95% confidence intervals are estimated using the WCS bootstrap approach of Woutersen and Ham [2016], discussed in Appendix 3.9.[28]

One might think that it would be simpler to use a regression model to explain expected duration, especially for the left-censored spell at the beginning of the sample.[29] The problem with this approach is that, unlike the duration model, it cannot accommodate the fact that some of the explanatory variables, e.g. the state unemployment rate and training participation status, change over the spell. This leaves one with the problem in the regression approach of whether training participation status should equal 1 if the individual enters training during the spell, or equal one 1 only if the woman enters training in the first half of the spell, and so on. Further, should one use the state unemployment rate at the beginning of the spell, when training took place, or when the spell ends? Not only will there be considerable arbitrariness in defining these variables, but any independent variable based on

28. Implementing the WCS bootstrap is discussed in detail in . Appendix 3.9 As Appendix 3.9 indicates, the researcher must set the number of simulations in calculating the WCS confidence intervals; we used 50,000 repetitions here. Note that we also considered using the delta method since the function is differentiable, but it was difficult to determine whether the derivatives were bounded and nonzero, which the delta method requires. The WCS bootstrap is applicable if derivatives are zero or unbounded. Further, the WCS bootstrap is asymptotically equivalent to the delta method when the latter is valid, so there should be no loss in efficiency in using the WCS bootstrap. To the best of our knowledge, HLSS16 is the only paper other than ours to estimate a consistent confidence interval for this estimated effect.

29. To take care of right censoring in spells in progress at the end of the sampling frame, one can replace the regression model by a Tobit-type model. Note that this regression or Tobit approach will not be valid for the fresh spells because these spells suffer from dynamic selection.

spell length will be endogenous.[30] Thus, looking at the effect of a variable on the expected duration eliminates this arbitrariness while offering the simplicity of something analogous to a regression coefficient. Second, we can also calculate the change in the expected duration when we allow the variable of interest to change over the spell, i.e. increasing the state unemployment rate by two percentage points in each month, as compared to the actual path of the unemployment rate. This, of course, would not be possible with a regression approach.

3.4.0.2 Understanding the Parameter Estimates by Simulating the Effects of Changing in an Explanatory Variable on the Percentage of Time Spent Employed

Our second counterfactual measure is the effect of training participation on the expected percentage of time that an individual is employed over a given time horizon (60 months for this study). This measure combines the estimated effects of training participation on the employment and nonemployment hazard rates in a sensible way. We consider three counterfactual scenarios, entering training immediately after the baseline, entering training in month 19 and never entering.The percentage of time spent in employment under each scenario ($ER1$, $ER0$ and $ER19$) is simulated by setting training status variable accordingly.[31] For example, to simulate the percentage of time spent employed if entering training in month 19, we set the training participation dummy variable (D_i in hazard equations (5) and (6)) to zero for months 1 to 19 and set it to one from month 20 to month 60.[32]The differences among the three measures represent the effect of training participation at different times on

30. See Flinn and Heckman [1983]

31. This approach (without standard errors) has been used in HL96 and EHL02.

32. Within the sample period, the simulations are based on individual observed characteristics in each period, while for periods out of the sample, the last period variable values are used. As for duration dependence, we freeze the duration function for durations longer than 50, 15, and 10 months, respectively, for the left-censored nonemployment, fresh employment, and fresh nonemployment spells. HLSS16 offers the details of simulating fraction of time spent employed.

percentage of time employed.

As we cannot obtain a close form solution for the expected percentage of time in employment, we estimate this using a simulation. Further, since our simulated employment percentages are discontinuous functions of the parameter estimates, the delta method cannot be used to calculate consistent estimates of the relevant confidence intervals. For such a case, researchers often form $(1-\alpha)$ confidence intervals by taking a large number of draws from the asymptotic distribution of the parameter estimates, simulating the model using each draw of the parameter estimates, and then trimming the bottom and top 0.5α of the function values (the estimated employment percentages in our case).We refer to this approach as the AD-bootstrap. However, Woutersen and Ham [2016]note that there is no consistency proof for the AD bootstrap, and show that it produces confidence intervals that are too small in several examples.[33] Woutersen and Ham [2016] propose two alternatives to the AD bootstrap, both of which produce consistent confidence intervals: the CS bootstrap and the WCS bootstrap. They find that the WCS bootstrap produces smaller standard errors using the HLSS16 parameter estimates and data.[34]

Again, one could ask whether there is a simple, regression-based method of proceeding here, where, e.g., the dependent variable is the percentage of time employed at different time horizons and we focus on some function of the monthly unemployment rate and training participation as our explanatory variables, instrumenting the latter with the Random Assignment variable. However, the same problem of arbitrariness in defining the unemployment conditions and training status arises as in the regression-based approach to estimating expected duration effects, except that there is no longer an endogeneity problem because the

33. EHL02 incorrectly suggest that the AD bootstrap should be used to estimate confidence intervals for the employment effects. Note that several recent studies calculating couter-factuals for duration models use the AD bootstrap.

34. Here we use 100,000 repetitions since the expected percentage of time in employment is a more complicated function than the truncated expected duration. Implimenting the WCS bootstrap is discussed in detail in Appendix 3.9.

researcher sets the time horizon.

3.5 Problems with the Exclusion Restrictions When Training is not Undertaken Immediately

As noted in Section 3.3, our approach to estimating the impact of endogenous participating in training is to model training entry jointly with labor market transitions.[35] If some individuals do not enter training immediately, as in our case, the assumption that, conditional on participation in training, the random assignment variable does not affect the outcome of interest may be violated. One manifestation of this would occur in a health insurance experiment where people are randomly made eligible for insurance but do not necessarily take it up right away. Then it could be the case that those eligible for, but not enrolled in, the insurance program take poorer care of their health because they know that if they become seriously ill, they can enroll in the insurance program and get free medical attention. In this case it would not be valid to run a monthly regression of, say, current mental health on current insurance coverage, using treatment status as an instrument for current coverage, since the exclusion restriction would be violated.

In our case, a similar problem would occur if the treatments, who had not yet participated in training, searched less intensely because they knew they could obtain free easy-access training in the future. In this case, the exclusion restriction that treatment status does not enter the structural hazard functions defined by equations (5) and (6) is violated – individuals

35. This Section was inspired by Abbring and Van den Berg [2003]'s (p.1494) comparison of their identification strategy to that used by EHL97 "We do not impose exclusion restrictions on covariates, i.e. we do not require that the data contain a variable that affects the treatment assignment but does not affect the outcome of interest other than by way of the treatment. A variable that is observed by the analyst is often also observable to the individuals under consideration. If such a variable affects the treatment process, then a rational individual will take this variable into account in determining his optimal strategy. This behavior affects the rate at which the individual leaves the state of interest. As a result, exclusion restrictions are difficult to justify and instrumental variable methods of inference are not likely to be of help." It is worth noting that they need the different, but strong, assumption that individuals do not know the date the will be assigned to training, or if individuals know this information, they do not act on it.

belonging to treatment and control groups behave differently in the labor market even before they participate in training.

At a more general level, having the option of free easy-access training will raise the treatments' expected lifetime income and affect them even before they participate in training, and again treatment status will affect behavior even if a woman has not yet entered training.To show this formally by considering the dynamic optimization problem for our treatments and controls is beyond the scope of this paper. However, in Appendix 3.10 we consider a simple intertemporal utility maximizing problem that captures this idea. Specifically, there we consider the problem of individuals optimally choosing their consumption of two goods conditional on the prices in each period and the discounted present value of their income stream. Then the individuals are randomly assigned to a control group and a treatment group, and the latter receives a subsidy in terms of the price of one of the goods in future periods. If we consider a treatment and control who have identical income streams, before the experiment, they will have identical consumption in each period. However, once the experiment is introduced, they will have different consumption bundles in all periods, i.e. including during those periods before the subsidy takes effect.

We investigate this problem empirically by testing the three overidentifying restrictions that excludes the random assignment indicator from the labor market hazards. However, it is worth mentioning that the effect on pre-training behavior from being assigned to the treatment group will depend on how much those individuals valued free easy-access training, versus having to find another social agency to provide the training (as the control group had to do). If they did not value it very highly, the effect on their behavior in periods before training may be too small to be detected in our test of the overidentifying restrictions, but, in this case, this impact is also likely to be too small to substantially bias our estimate of the training participation effect.

3.6 Empirical Specification, Parameter Estimates, and Estimated Counterfactual Effects

This Section presents our empirical results. All duration models are estimated by maximum likelihood using the simulated annealing approach discussed in Section 3.3.3. Section 3.6.1 provides the hazard coefficient estimates from the model with endogenous training participation corresponding to hazard specifications (5), (6) and (7) in Section 3.3.2. In Section 3.6.2 we carry out the tests of our overidentifying restrictions that we discussed in Section 3.5. In Section 3.6.3we present our estimated counterfactual effects that were discussed in Section 3.4. These counterfactual estimates are policy relevant and easy-to-interpret measures of training participation effects.

3.6.1 Empirical Specification and the Training Participation Coefficient Estimates

In this Section, we focus on the effects of training participation on the labor market hazards based on the model defined in Section 3.3.2. While EHL97 used a factor model as a multivariate generalization of the Heckman and Singer [1984a] (hereafter HS) approach for unobserved heterogeneity, here we use McCall [1996] multivariate extension (hereafter MUH96) of the HS approach. For two points of support, the EHL97's factor model specification produces results that are identical to the MUH96 approach, but for three points of support or more, the MUH96 approach will be more general than the factor model.[36] By allowing for dependence between all of the unobserved heterogeneity terms, our approach corrects for dynamic selection bias and endogenous training.

In the hazard functions specified by (5), (6), and (7), the training participation status is the variable of interest, which is coded 1 in duration t if the individual enters training

36. In practice, the two approaches lead to very similar estimates.

before or at duration $t - 1$, and 0 otherwise. For control variables, we use the following time-constant explanatory variables (measured at the baseline): years of schooling, a dummy variable that equals 1 for a high school dropout and 0 otherwise, a dummy variable that equals 1 if the woman has any children under 4 years old, a dummy variable that equals 1 for a never married woman, a dummy variable that equals 1 if the individual is married and living with her spouse,[37] a dummy variable indicating whether the individual is African American, and a dummy variable indicating whether the individual is Hispanic. In terms of time-varying explanatory variables, we have the state unemployment rates. Our last explanatory variable is age. Since current age equals age at the beginning of the spell plus duration so it does not act as a time varying variable within a spell. However, initial age will vary across the spells, which helps us deal with the dynamic selection problem.

In the training entry hazard specified by (7), in addition to the control variables described above, we have two treatment (random) assignment variables: a dummy variable coded 1 if the individual is assigned to the treatment group and a dummy variable indicating both being assigned to the treatment group and being in the first three months of the spell.[38] Both variables are statistically significant in the training hazard, so we do not have the analogue of a 'weak' instruments problem.

We use two specifications for duration dependence: EHL97's fourth order polynomial in log duration and a step function. We slightly modify the fourth order polynomial specification by including a dummy variable for the first month after baseline in the left-censored nonemployment hazard, and by including a dummy variable if the individual is in the first three months in the training entry hazard. This specification exactly replicates the one in

37. Individuals who are divorced or separated from their spouse form the base category in the model.

38. One can ask why did we separate, for the treatments, the first three months instead of the first two or the first four months? We limited the grouping to the first 3 months after the baseline because the transition rates into training were so much larger for the treatments than for the controls during the first 3 months from the baseline. There was also a month 4 difference, but it is relatively small compared to the first 3 months. Thus, we chose this break solely on empirical grounds.

EHL97. In specifying the step function, we follow LS15 by requiring that each step covers at least 10% of the transitions to avoid overparameterizing the model, and this rule leads to a different number of steps across the labor market hazard functions.[39]

The results for the model with a fourth order polynomial in log duration and dependent unobserved heterogeneity are presented in Table 3.4. As noted above, these estimates control for both dynamic selection into the fresh spells and endogenous training in all of the labor market hazards, and are comparable to the estimates in Table A3 of EHL97, except for our use of the MUH96 instead of their factor model for the unobserved heterogeneity.[40]

To help focus on the training participation effects across models, we have copied the training participation coefficients in the employment and nonemployment hazards from Table 3.4 to Table 3.7, A.1 under Panel A. The estimates (standard errors) corresponding to the left-censored nonemployment spells, fresh nonemployment spells, and fresh employment spells, 0.450 (0.078), 0.153 (0.135), and (0.130), respectively, indicate a significant training effect in left-censored spells only. To investigate the importance of considering endogenous training entry, we also estimate a model treating the unobservable for training as independent of the unobservables in the labor market hazards, i.e. treating training participation as exogenous. The estimates of training participation coefficients for this case are contained in B.1 under Panel B, Table 3.7. The coefficient estimates the left-censored nonemployment spells, fresh nonemployment spells, and fresh employment spells, are 0.286 (0.066), 0.243 (0.107), and 0.102 (0.080) respectively. We compare these estimates using a Hausman test,

39. This rule is a modification of the requirement, in ,Ham and Rea [1987] that each step in the duration dependence, cover at least 3% of the transitions. The idea is that the estimate in each step is a function of the number of transitions in that duration range. LS15 showed that once they impose this type of rule for the step function, they are able to simultaneously estimate Heckman-Singer heterogeneity, in contrast to the previous literature.

40. When we tried to replicate the results from EHL97, we found an error with one of the models. The model with a quartic in the duration dependence and with three points of support was affected by a programming error where we inadvertently left out 164 trainees from the time until training hazard but included them in all of the labor market transition hazards, This error caused us to overstate the effect of training on the duration of the left-censored and the fresh nonemployment spells.

and reject (at the 5% level) the null hypothesis that the coefficients are the same only for the left-censored nonemployment spell, indicating a significant endogeneity problem in the left-censored nonemployment spells.[41]

Next, we consider the estimates from the model when we use a step function for duration dependence and allow all unobservables to be dependent. The full set of parameter estimates is presented in Table 3.5. Again we focus on the estimated training participation coefficients in the three labor market hazard functions, which we have placed in Table 3.7, A.2 under Panel A. The coefficients (standard errors) for the training participation effects for the left-censored nonemployment spells, fresh nonemployment spells, and fresh employment spells are 0.391 (0.071), 0.201 (0.121), and 0.007 (0.077) respectively. The major qualitative difference with the estimates from the polynomial specification in A.1 of Table 3.7 is that the coefficient in the fresh nonemployment spells is somewhat larger and now is (marginally) statistically significant. The estimates when we treat the training participation as exogenous and use a step function for duration dependence are contained in Table 3.7, B.2 under Panel B. Now we estimate the following coefficients (standard errors) for the left-censored nonemployment spells, fresh nonemployment spells, and fresh employment spells: 0.329 (0.068), 0.238 (0.109), and 0.083 (0.070) repectively. Again, we can compare these estimates using a Hausman test. When we do this, we reject (at the 5% level) the null hypothesis that the coefficients are the same for the left-censored nonemployment spells only. Thus we see evidence of a significant endogeneity problem in the left-censored nonemployment spells independently of how we model duration dependence.

41. Let and denote the estimate when we do, and do not treat, the unobserable in training entry hazard as independent of the unobservables in the labor market hazards respectively. Under the null hypothesis that the unobserable in training entry is independent of the unobservables in the labor market, is efficient and is merely consistent. Hausman [1978]shows that the variance .

3.6.2 Tests of the Overidentifying Restrictions

We next turn to testing the overidentifying restrictions. Recall that the issue here is whether it is valid to exclude any variable based on random assignment from the labor market hazard functions conditional on training participation. Our hazard functions for entering training shown in Tables 3.4 and 3.5 contain the random assignment dummy and the interaction of this dummy with a dummy variable coded 1 for the first three months after the baseline. It is important to note that our only time changing variable, the unemployment rate, does not significantly affect the training entry hazard.[42] Thus, we cannot identify the training effect from variation over time in the unemployment rate, and unless we want identification to come from nonlinearities, we cannot test whether *both* random assignment variables affect the labor market hazard rates conditional on (endogenous) training participation. However, we can test an overidentifying restriction in each hazard function by putting the treatment dummy in the labor market hazard functions which include training participation, while (exactly) identifying the model through the interaction of the treatment dummy and the first three months' dummy, which only enters the training entry hazard function. The estimates of the coefficients for training and treatment status are in Panel C of Table 3.7. The treatment status dummy is never statistically significant (even at the 10% level) in any labor market hazard function, independent of whether we use a polynomial or a step function for duration dependence. These results suggest that our estimates have not been biased by assuming that the random assignment variable does not enter the labor market hazard functions conditional on training participation.

42. EHL97 and Hausman and Woutersen (2014) note that time changing explanatory variables serve as an exclusion restriction in the sense that past values of the variable affect training status but do not affect the current labor market hazard functions.

3.6.3 Estimated Counterfactual Effects of Training Participation

All simulated counterfactual estimates presented in this Section are based on the models in Section 3.2 with endogenous training entry. Table 3.8 presents the training participation effects on the truncated expected durations. Consider the participation effects for the left-censored nonemployment spells in Panel A. We provide the average truncated expected duration across the sample for two scenarios: entering training immediately ($ED1$) and never entering ($ED0$) . The estimated training participation effect $ED1-ED0$ is presented in the third row. Panels B and C provide the corresponding estimates for the fresh employment and nonemployment spells respectively. Within each panel, the estimates based on the model with a polynomial specification for duration dependence (polynomial model hereafter) are shown in the left column, and those based on the model with a step function specification for duration dependence (step function model hereafter) are shown in the right column. Consistent 95% confidence intervals based on WCS Bootstrap procedure are presented in square brackets below each estimate.

For the left-censored nonemployment spells, the polynomial estimates in Panel A indicate that entering training immediately versus never entering training reduces the expected duration [confidence interval] by 6.162 [8.480, 3.699] months. (For the remainder of this Section we present estimated effects followed by their 95% confidence intervals in square brackets.) This reduction in the expected duration implies a large percentage effect of 31.3% , given that the truncated expected duration for someone who never enters training is 19.174 [16.290, 26.009] months. To the best of our knowledge, these are the first consistent estimates of the confidence intervals for the effect of participating in training on expected durations. The corresponding estimates from the step function model are quite similar. Entering training immediately after the baseline versus never entering reduces the expected duration by 5.270 [7.177, 3.293] months, which implies a drop of 28% , given that the expected length of a left-censored nonemployment spell in the absence of training is 18.760 [16.939, 20.578] months.

From Panel B of Table 3.8, we see that there are no (close to) statistically significant training participation effects on the fresh employment spells for either specification of duration dependence, which is what we would expect given the quite small and insignificant estimated training coefficients for these spells. The corresponding estimates for the fresh nonemployment spells from both models in Panel C of Table 3.8 show a larger effect of reducing the expected duration by 2 to 4 months, but neither estimate is significant at the 5% level. The latter results are in line with the insignificant or marginally significant training coefficients for the fresh nonemployment spells.[43]

Our second counterfactual estimates show the effect of training participation on the percentage of time spent in employment over a certain period (which we choose to be 60 months.) As noted above, the simulated expected percentage of time spent in employment will be a function of all three estimated labor market hazards and the estimated entering training hazard function. These counterfactual estimates are presented in Table 3.9, which consists of three panels. Panels A, B and C contain counterfactual estimates of entering training immediately after the baseline versus never entering, entering training in month 19 versus never entering, and entering training immediately after the baseline versus entering in month 19, respectively. In each panel, we present the estimates from the polynomial model first, and then those from the step function model second. Again, the 95% confidence intervals based on the WCS bootstrap procedure are in square brackets under each estimate.

The polynomial model estimates in Panel A suggest that entering training immediately versus never being in training raises the average percentage of time employed by 8.20 [1.64, 15.42] percentage points on a base of 43.89 [35.91, 49.38] percentage points. The step function estimates indicate this effect to be 7.62 [1.66, 13.21] percentage points on a base of 43.08 [37.86, 48.56] percentage points. Again, both effects are significant at the 5% level.

43. Recall that all confidence intervals are at the 95% level, while in the step function specification the training coefficient for fresh nonemployment spells in Table 3.5is only significant at the 90% level.

Next, we consider the percentage of time spent in employment when a woman delays entering training until month 19, and compare this to the percentage of time spent in employment for someone who never participates in training. The estimates in Panel B indicate that the difference between the delayed entry and never participating is 3.80 [, 8.61] and 3.82 [,7.45] when we use the polynomial and step function respectively. These effects are on the edge of being significant at the 5% level. Finally, Panel C indicates that the difference between entering training immediately versus entering training in month 19 in percentage of time employed is around 4 percentage points based on both models, and these estimates are significant at the 5% level. Thus, women lose about half the beneficial effects of training participation on the expected percentage of time employed over a sixty-month horizon by delaying training entry for 18 months.[44]

In closing we make two observations. First, we feel that these counterfactuals are much more informative for both researchers and policy makers than the magnitude and statistical significance of hazard function coefficients. Second, the training participation coefficients and estimated counterfactual effects are very similar between the two specifications of duration dependence.

3.7 A Brief Comparison between Our Approach and European Approaches for Estimating Participation Effects

There is a now quite large literature on the impact of training on labor market outcomes in Europe. One stream of research is based on hazard function estimation, which models the unobservables and uses maximum likelihood estimation to adjust for selection and the endogeneity of training participation. Another stream is based on static and dynamic matching

44. Our choice of a counterfactual based on someone entering training in the 19th month is interesting in that it implies that a woman gets half the benefit from training even with such a long delay in entry. On the other hand, relatively few of the women in our sample wait this long to enter training, so in future work we will also do this counterfactual for shorter delays in training entry.

(e.g. Lechner [2002, 2009]), i.e. finding a set of conditioning variables and relying on selection on observables assumption to solve the problem of endogeneity in training participation. A comparison of all of this work and the approach we take is beyond the scope of this paper. Instead, we briefly highlight some of the similarities and differences between our approach and the literature on hazard function estimation and the part of the matching literature that is closest to our approach: the papers by Sianesi [2004] and Fredriksson and Johansson [2008]. We would emphasize that we are not, in any way, suggesting the rest of literature that we do not cover is not important or interesting.

It is important here to consider the institutional and other differences between the JTPA programs and participants and the European programs and participants. One difference is that in JTPA women start in the midst of a left-censored nonemployment spell, while in Europe individuals usually are starting a fresh nonemployment spell. Another difference is that none of our women are covered by US unemployment insurance, while individuals in the European data are covered by unemployment insurance. Moreover, our subjects are much more disadvantaged than those studied in Europe.

From an econometric point-of-view, we believe the most important difference is how individuals come to participate in a program. In JTPA, a woman voluntarily enters training at any point over the sample period. In the European data, individuals generally enter training when a social worker assigns them to training during their fresh unemployment spell. In other words, individuals are only eligible for program participation if they are still unemployed. Thus, the longer her unemployment spell, the more likely an individual is to participate in a program. This introduces a type of 'length biased sampling' that arises even if the unobservables determining the labor market hazard rates are independent of the unobservable that affects program participation. Hence, in the European setting, estimating transition rates is a more difficult econometric problem since researchers potentially must control for i) endogenous training participation ii) dynamic selection (in the more recent papers that follow individuals past the first unemployment spell) and iii) the 'length bias sampling' problem described above.

The early hazard function stream of the European literature consists of estimating the continuous time analogues to the hazard functions defined by (5) and (7) by closely following the structure of Abbring and Van den Berg [2003]. In addition to working in continuous time, these studies did not use time changing explanatory variables, considered only the first fresh unemployment spell, and focused on the coefficient on the program participation variable; all of this is very different from the approach in EHL97. However, in recent years both the European work and our work have become closer. For example, Crépon et al. [2016] considered time changing explanatory variables and looked at expected durations. Further, Tatsiramos [2010] and Osikominu [2013] analyzed multiple spells and multiple states, and Osikominu [2013] also simulated the effect of training participation on the truncated expected duration and different percentiles of the estimated survivor function. Finally, both Fitzenberger and Paul [2016] and Brewer and Cribb [2017] worked in discrete time, had time changing explanatory variables, had multiple states and multiple spells, and simulated the effect of participation on the fraction of time employed. Brewer and Cribb [2017] obtained consistent confidence intervals for these effects using the CS bootstrap.

Our discussion of the second stream of the European literature that we consider is based on Sianesi [2004] and Fredriksson and Johansson [2008]. These papers took a nonparametric approach to estimating the effect on individuals in the midst of a fresh unemployment spell of being assigned to a program at each duration. Specifically for each duration t , they compared the outcome of being assigned at duration t versus not being assigned at t (with the possibility of being assigned to a program in the future). Their key insight was that for individuals assigned to a program at duration t in their unemployment spell, a natural comparison group consists of individuals who reach duration t but have not yet been assigned to a program, i.e., it is reasonable to assume that the conditional independence assumption holds given current duration. Sianesi [2004] looked at the effect of being assigned to a Swedish labor market program at different duration t on a variety of labor market outcomes over the next five years, including a future spell of nonemployment covered by unemployment insurance. Fredriksson and Johansson [2008] considered the effect of participating in a Swedish employment subsidy program at duration t on the Kaplan-Meier survival rate for durations greater than t (in the current spell). Both papers calculated overall effects by aggregating over different durations.

By comparing individuals with the same duration, both studies above solved the "length biased sampling problem" that individuals must still be in their unemployment spell to be assigned to a program. In both studies, they can measure the outcomes of interest for their entire treatment and comparison groups (at each duration). Hence, they do not have a dynamic selection problem, as we do, e.g. in terms of only observing fresh employment spells for a subset of the treatments and controls.

The paper by Biewen et al. [2014] can be thought of as a significant extension of the Sianesi [2004] and Fredriksson and Johansson [2008] approach to German training data. It contrasts dynamic and static matching estimates of training effects by matching on the duration of unemployment at program entry.[45]

The above papers explicitly assumed that conditional on current duration, entry to training is random. We believe that this is a more reasonable assumption in their context than for JTPA participants, since the latter volunteer for training, while in the Swedish system, the majority of individuals are assigned to a program by a social worker, and face sanctions if they do not accept the assignment.

45. The paper investigates the sensitivity of treatment effect estimates with regard to methodological choices. The study also motivates the plausibility of the conditional independence assumption for the institutional setting where caseworkers decide upon the assignment to training.

3.8 Conclusion

This paper makes several contributions. First, we conduct a much more detailed descriptive analysis of the relationship between labor market transitions and training participation in JTPA than has appeared in the literature. JTPA was a unique training program in that it was part-time in nature, and therefore individuals could receive training while working. Second, we show that the EHL97 results for the effects of training participation are robust to a substantial generalization of the specification of their model.

Third, we consider the potential problem that arises when individuals postpone their entry to training. One version of this problem is that the treatment group may behave differently from the control group before training entry, because the treatments know they can easily obtain free easy-access training in the future. We also consider a more general example when a price subsidy in future periods affects behavior in the non-subsidy periods, to capture the idea that having subsidized training in the future may affect the behavior of the treatments today. Both examples suggest that random assignment may affect the labor market hazards conditional on training participation, invalidating our identification strategy. We propose and implement a test of the overidentifying restrictions to shed light on this problem, and conclude that in our context this identification problem is not important.

Fourth, we use our results to estimate and obtain consistent confidence intervals for the effects of changing training participation from never participating to immediately participating on i) on the expected duration of each spell type and ii) the effect of changing an explanatory on the percentage of time over a given time horizon. We also consider the effect of an intermediate case where an individual postpones training for a certain period. Ours are the first consistent confidence intervals for the effect of entering training on the expected durations and among the first consistent confidence intervals for the effect of entering training on the expected average percentage of time spent in employment. We argue that policy makers will find these estimated effects much more useful than simply focusing on hazard function coefficients, especially now that we can provide consistent confidence intervals for the effects. Previously, these effects were available in JTPA only for being assigned to the treatment group, as opposed for participating in training, and confidence intervals were not provided for these treatment estimates.

Fifth, we find that participating in JTPA training: i) reduces the expected duration of a left-censored unemployment spell by about six months, which is a large effect given that the expected duration of a non-trainee for such a spell is about 19 months and ii) increases the average expected percentage of time spent in employment by about 8 percentage points on a base of about 43 percentage points for a non-trainee. Hence, it would complement any program that increases employment duration, such as an NSW type program. Finally, we use the Hausman test for testing the null hypothesis that the unobservable in the training hazards is independent of the unobservables in the labor market hazard functions for each specification of duration dependence. We reject this null hypothesis for both specifications of duration dependence in the left-censored nonemployment hazard, and find that we substantially underestimate the training participation coefficient when we treat it as exogenous in this hazard function.

Table 3.1: Comparison of means of the treatments, controls, trainees, and nontrainees: characteristics at the baseline

	Treatments (1)	Controls (2)	Differences (1)-(2) (3)	Ever Trained=1 (4)	Ever Trained=0 (5)	Differences (4)-(5) (6)
Age	31.71	31.14	0.57	31.27	32.00	-0.73
	(8.07)	(7.57)	(0.39)	(7.61)	(8.42)	(0.37)
Years of schooling	11.32	11.27	0.05	11.32	11.27	0.04
	(1.73)	(1.59)	(0.08)	(1.69)	(1.68)	(0.08)
High school dropout	0.50	0.51	-0.02	0.49	0.52	0.03
	(0.49)	(0.49)	(0.02)	(0.49)	(0.49)	(0.02)
Has children under 4 years old	0.36	0.34	0.01	0.36	0.34	0.02
	(0.46)	(0.46)	(0.02)	(0.46)	(0.46)	(0.02)
Married with spouse present	0.21	0.19	0.02	0.21	0.18	0.03
	(0.39)	(0.38)	(0.02)	(0.40)	(0.37)	(0.02)
Never married	0.35	0.42	-0.07***	0.35	0.41	-0.06***
	(0.46)	(0.48)	(0.02)	(0.46)	(0.48)	(0.02)
African American	0.35	0.37	-0.01	0.32	0.42	-0.10***
	(0.48)	(0.48)	(0.02)	(0.47)	(0.49)	(0.02)
Hispanic	0.14	0.15	0.00	0.16	0.11	.05***
	(0.35)	(0.35)	(0.02)	(0.37)	(0.32)	(0.02)
Fraction of the weeks worked previous year	0.25	0.23	0.02	0.24	0.25	0.00
	(0.31)	(0.31)	(0.02)	(0.31)	(0.31)	(0.01)
Has been on AFDC for the past 2 years	0.36	0.42	-0.07***	0.37	0.40	-0.03
	(0.48)	(0.49)	(0.02)	(0.48)	(0.49)	(0.02)
Has never been on AFDC	0.36	0.34	0.02	0.35	0.35	0.00
	(0.48)	(0.47)	(0.02)	(0.48)	(0.48)	(0.02)
Number of women in sample	1,324	616		1,247	693	

Notes: Columns 1, 2, 4, and 5 report sample means and standard deviations (in parentheses). Columns 3 and 6 report the differences in means and the associated standard errors (in parentheses). AFDC 5 Aid for Families with Dependent Children. *p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 3.2: Comparison of means of the trainees and nontrainees who reach fresh employment and nonemployment spells: characteristics at the baseline

	Fresh Employment Spells			Fresh Nonemployment Spells		
			Differences			Differences
	Trainees	Nontrainees	(1)-(2)	Trainees	Nontrainees	(4)-(5)
	(1)	(2)	(3)	(4)	(5)	(6)
Age	31.79	32.69	-0.90**	31.47	32.26	-0.79
	(7.64)	(8.61)	(0.45)	(7.34)	(8.11)	(0.59)
Years of schooling	11.33	11.34	-0.01	11.19	11.09	0.10
	(1.70)	(1.70)	(0.09)	(1.74)	(1.70)	(0.13)
High school dropout	0.49	0.51	-0.02	0.55	0.56	-0.01
	(0.49)	(0.49)	(0.03)	(0.49)	(0.49)	(0.04)
Has children under 4 years old	0.35	0.32	0.03	0.35	0.33	0.02
	(0.46)	(0.46)	(0.03)	(0.46)	(0.46)	(0.03)
Married with spouse present	0.21	0.19	0.02	0.19	0.21	-0.02
	(0.39)	(0.39)	(0.02)	(0.38)	(0.40)	(0.03)
Never married	0.34	0.39	-0.05	0.37	0.39	-0.02
	(0.46)	(0.48)	(0.03)	(0.47)	(0.48)	(0.04)
African American	0.31	0.4	-0.09***	0.35	0.42	-0.07
	(0.46)	(0.49)	(0.03)	(0.48)	(0.49)	(0.04)
Hispanic	0.14	0.12	0.02	0.14	0.10	.04**
	(0.35)	(0.32)	(0.02)	(0.34)	(0.31)	(0.02)
Fraction of the weeks worked previous year	0.28	0.29	-0.01	0.27	0.26	0.01
	(0.32)	(0.32)	(0.02)	(0.31)	(0.31)	(0.02)
Has been on AFDC for the past 2 years	0.32	0.36	-0.04	0.36	0.39	-0.03
	(0.47)	(0.48)	(0.03)	(0.48)	(0.49)	(0.04)
Has never been on AFDC	0.38	0.38	0.00	0.36	0.36	0.00
	(0.49)	(0.49)	(0.03)	(0.48)	(0.48)	(0.04)
Number of women in sample	952	510		502	269	

Notes: Columns 1, 2, 4, and 5 report sample means and standard deviations (in parentheses). Columns 3 and 6 report the differences in means and the associated standard errors (in parentheses). AFDC 5 Aid for Families with Dependent Children. *p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 3.3: Participants by training entry month after the baseline mean comparison: characteristics at the baseline

	Entry Month	Entry Month	Differences	Entry Month	Differences	Never Trained	Difference
	1-3	4-18	(1)-(2)	19-33	(1)-(4)	(4)-(5)	(1)-(6)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	31.58	30.71	0.87	29.38	2.20**	32.00	-0.42
	(7.76)	(7.24)	(0.51)	(7.00)	(1.05)	(8.42)	(0.41)
Years of schooling	11.33	11.26	0.07	11.40	-0.06	11.27	0.06
	(1.68)	(1.69)	(0.11)	(1.81)	(0.23)	(1.68)	(0.09)
High school dropout	0.48	0.50	-0.02	0.59	-0.10	0.52	-0.04
	(0.49)	(0.49)	(0.03)	(0.50)	(0.07)	(0.49)	(0.02)
Has children under 4 years old	0.35	0.36	-0.01	0.38	-0.02	0.34	0.01
	(0.46)	(0.46)	(0.03)	(0.45)	(0.06)	(0.46)	(0.02)
Married with spouse present	0.24	0.15	.09***	0.16	0.07	0.18	0.06***
	(0.41)	(0.34)	(0.03)	(0.36)	(0.06)	(0.37)	(0.02)
Never married	0.31	0.43	-0.12***	0.51	2.19***	0.41	2.10***
	(0.45)	(0.48)	(0.03)	(0.49)	(0.06)	(0.48)	(0.02)
African American	0.30	0.36	2.07**	0.47	2.17***	0.42	2.13***
	(0.46)	(0.48)	(0.03)	(0.50)	(0.06)	(0.49)	(0.02)
Hispanic	0.16	0.16	0.00	0.17	-0.01	0.11	.04**
	(0.37)	(0.37)	(0.02)	(0.38)	(0.05)	(0.32)	(0.02)
Fraction of the weeks worked previous year	0.25	0.23	0.02	0.25	0.00	0.25	0.00
	(0.31)	(0.30)	(0.02)	(0.29)	(0.04)	(0.31)	(0.02)
Has been on AFDC for the past 2 years	0.35	0.42	2.07**	0.40	-0.05	0.40	-0.05
	(0.47)	(0.49)	(0.03)	(0.48)	(0.06)	(0.49)	(0.02)
Has never been on AFDC	0.37	0.30	.07**	0.33	0.04	0.35	0.02
	(0.48)	(0.46)	(0.03)	(0.47)	(0.07)	(0.48)	(0.02)
Number of women in sample	890	299		58		693	

Notes: Columns 1, 2, and 4 show the mean characteristics of individuals (treatments and controls combined) who entered training within the first 3 months, between months 4 and 18, and between months 19 and 33 following the baseline, respectively. Column 6 presents the mean characteristics for participants who never enter training during the sample period. The mean differences between the participants who entered training within the first 3 months and the three other groups are presented in cols. 3, 5, and 7, respectively. AFDC 5 Assistance for Families with Dependent Children. Columns 1, 2, 4, and 6 report the standard deviations in parentheses. Columns 3, 5, and 7 report the standard errors in parentheses. *p-value $<$ 0.10, **p-value $<$ 0.05, *** p-value $<$ 0.01.

Table 3.4: Parameter estimates when the hazard functions depend on training participation and a fourth-order polynomial in log duration

	Nonemployment Spells		Employment Spells	Training
	Left-Censored	Fresh	Fresh	Entry
	(1)	(2)	(3)	(4)
Training participation status	0.450***	0.153	-0.040	
	(0.078)	(0.135)	(0.130)	
First month	0.473***			
	(0.141)			
Years of schooling	0.037	0.049	-0.091***	0.025
	(0.026)	(0.039)	(0.033)	(0.027)
High school dropout	0.091	-0.079	0.271**	-0.100
	(0.083)	(0.134)	(0.114)	(0.092)
Has children under 4 years old	-0.198**	-0.167	0.033	-0.071
	(0.081)	(0.124)	(0.100)	(0.085)
Never married	-0.125	-0.149	0.074	-0.113
	(0.083)	(0.142)	(0.111)	(0.092)
Married with spouse present	-0.037	0.097	-0.008	0.101
	(0.088)	(0.138)	(0.116)	(0.096)
African American	-0.038	-0.139	0.211**	-0.300***
	(0.077)	(0.135)	(0.103)	(0.086)
Hispanic	-0.157	-0.041	0.018	0.145
	(0.100)	(0.169)	(0.135)	(0.107)
Age	-0.089***	-0.052	-0.019	0.066*
	(0.031)	(0.053)	(0.042)	(0.035)
Age squared/100	0.108***	0.049	0.001	-0.111**
	(0.041)	(0.070)	(0.056)	(0.047)
State unemployment rate	-0.062***	-0.077**	-0.018	0.009
	(0.018)	(0.030)	(0.023)	(0.018)
Log duration	0.247	0.345	1.952***	-0.353
	(0.522)	(0.818)	(0.574)	(0.586)
Log duration2	-0.358	-0.752	-1.377*	-1.898*
	(0.299)	(1.387)	(0.795)	(1.007)
Log duration3	0.066	0.302	0.397	0.944**
	(0.069)	(0.760)	(0.381)	(0.474)
Log duration4	-0.003	-0.047	-0.044	-0.132*
	(0.005)	(0.131)	(0.059)	(0.070)
Treatment status				0.413***
				(0.115)
First 3 months after baseline				-0.771***
				(0.281)
First 3 months x treatment Heterogeneity terms:				0.974***

Notes: See next page.

Table 3.4, continued

	Nonemployment Spells		Employment Spells	Training
	Left-Censored	Fresh	Fresh	Entry
	(1)	(2)	(3)	(4)
θ_1	-0.485	-1.298	-0.573	-1.920**
	(0.809)	(1.141)	(0.910)	(0.782)
θ_2	-0.328***	-1.427	-2.673***	-3.886***
	(0.839)	(1.156)	(0.908)	(0.936)
θ_3	-0.314	-0.888	-2.458***	-1.998***
	(0.748)	(1.105)	(0.882)	(0.749)
Probability θ_1	0.155***			
	(0.062)			
Probability θ_2	0.109***			
	(0.023)			
Probability θ_3	0.736			
Log likelihood	-14,127.949			

Notes: Results are based on 1,324 treatments and 616 controls. Standard errors are in parentheses. *p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 3.5: Parameter estimates when the hazard functions depend on training participation and a step function in duration

	Nonemployment Spells		Employment Spells	Training
	Left-Censored	Fresh	Fresh	Entry
	(1)	(2)	(3)	(4)
Training participation status	0.391***	0.201*	0.007	
	(0.071)	(0.121)	(0.077)	
Years of schooling	0.033	0.044	-0.069***	0.024
	(0.026)	(0.038)	(0.026)	(0.026)
High school dropout	0.090	-0.075	0.233***	-0.101
	(0.084)	(0.131)	(0.089)	(0.089)
Has children under 4 years old	-0.192**	-0.190	0.037	-0.077
	(0.081)	(0.122)	(0.078)	(0.083)
Never married	-0.137	-0.152	0.072	-0.106
	(0.084)	(0.141)	(0.087)	(0.090)
Married with spouse present	-0.049	0.083	0.019	0.107
	(0.089)	(0.136)	(0.092)	(0.093)
African American	-0.047	-0.108	0.173**	-0.295***
	(0.077)	(0.130)	(0.081)	(0.084)
Hispanic	-0.161	-0.049	0.028	0.149
	(0.102)	(0.167)	(0.107)	(0.104)
Age	-0.093***	-0.060	-0.017	0.066*
	(0.031)	(0.052)	(0.034)	(0.034)
Age squared/100	0.114***	0.058	0.001	-0.111**
	(0.041)	(0.069)	(0.045)	(0.046)
Unemployment rate	-0.062***	-0.076***	-0.021	0.005
	(0.018)	(0.029)	(0.019)	(0.018)
Treatment status				0.432***
				(0.105)
First 3 months x treatment				0.912***
				(0.13)
Heterogeneity terms:				
θ_1	-0.020	-0.844	-2.082***	-2.710***
	(0.679)	(1.091)	(0.696)	(0.684)
θ_2	3.545***	-1.009	-2.493***	-4.092***
	(0.740)	(1.112)	(0.725)	(0.725)
Probability θ_1	0.884***			
	(0.015)			
Probability θ_2	0.116			
Log likelihood	-14,139.380			

Notes: See next page.

Table 3.6: Parameter estimates when the hazard functions depend on training participation and a step function in duration (Continued)

Nonemployment Spells				Employment Spells			
Left-Censored		Fresh		Fresh		Training Entry	
(1)		(2)		(3)		(4)	
Month	Coefficient	Month	Coefficient	Month	Coefficient	Month	Coefficient
4–6	-0.417**	2–3	-0.089	2	0.674***	2	-0.854***
	(0.169)		(0.143)		(0.162)		(0.096)
7–10	-0.674***	4–5	-0.531***	3	0.899***	3–4	-1.613***
	(0.167)		(0.168)		(0.158)		(0.117)
11–18	-0.890***	6–8	-0.563***	4	0.546***	5–9	-1.929***
	(0.158)		(0.166)		(0.175)		(0.138)
19–33	-1.298***	> 8	-1.069***	5	0.645***	10–20	-2.581***
	(0.162)		(0.167)		(0.171)		(0.139)
34–100	-1.509***			6–7	0.484***	> 20	-2.667***
	(0.165)				(0.157)		(0.181)
> 100	–1.724***			8–9	0.339**		
	(0.159)				(0.168)		
				10–11	0.420**		
					(0.172)		
				12–15	0.152		
					(0.169)		
				> 15	-0.021		
					(0.175)		

Notes: Results are based on 1,324 treatments and 616 controls. Standard errors are in parentheses. *p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 3.7: Estimated training participation effects for different model specifications

	Nonemployment Spells		Employment Spells
	Left-Censored	Fresh	Fresh
	(1)	(2)	(3)
A. Endogenous participation effects:			
1. Polynomial:			
Training participation status	0.450***	0.153	-0.04
	(0.078)	(0.135)	(0.130)
2. Step function:			
Training participation status	0.391***	0.201*	0.007
	(0.071)	(0.121)	(0.077)
B. Exogenous participation effects:			
1. Polynomial:			
Training participation status	0.286***	0.243**	0.102
	(0.066)	(0.107)	(0.080)
2. Step function:			
Training participation status	0.329***	0.238**	0.083
	(0.068)	(0.109)	(0.070)
C. Tests of the overidentifying restrictions:			
1. Polynomial:			
Training participation status	0.464***	0.056	0.021
	(0.083)	(0.148)	(0.148)
Treatment status	-0.029	0.204	-0.087
	(0.074)	(0.129)	(0.111)
2. Step function:			
Training participation status	0.395***	0.147	0.029
	(0.074)	(0.130)	(0.084)
Treatment status	-0.020	0.171	-0.064
	(0.071)	(0.123)	(0.079)

Notes: Under the null hypothesis that the model is correct, the coefficients on treatment status in panel C should not be significantly different from zero. Standard errors are in parentheses. *p-value < 0.10, **p-value < 0.05, *** p-value < 0.01.

Table 3.8: The effect of entering training immediately (ED1) versus never entering training (ED0) on the expected durations

	Polynomial	Step Function
A. Left-censored nonemployment spells:		
ED0	19.174	18.760
	[16.290, 26.009]	[16.939, 20.578]
ED1	13.012	13.490
	[10.912, 20.088]	[12.260, 14.868]
ED1 – ED0	-6.162	-5.270
	[-8.480, -3.699]	[-7.177, -3.293]
B. Fresh employment spells:		
ED0	20.910	20.506
	[17.241, 24.360]	[18.106, 22.993]
ED1	21.498	20.396
	[18.516, 24.317]	[18.211, 22.410]
ED1 – ED0	0.588	-0.111
	[-3.519, 4.571]	[-2.696, 2.466]
C. Fresh nonemployment spells:		
ED0	17.058	19.988
	[13.041, 25.068]	[16.080, 24.325]
ED1	14.803	16.631
	[11.551, 20.487]	[14.277, 19.522]
ED1 – ED0	-2.255	-3.357
	[-6.870, 2.217]	[-7.578,0.924]

Notes: In brackets are the 95% confidence intervals based on the WCS bootstrap procedure with 50,000 repetitions.

Table 3.9: The effect of entering training at various stages on the average percentage of time employed during the first 60 months after the baseline

	A. The Effect of Entering Training Immediately (ER1) versus Never Entering Training (ER0)		
Model	ER1	ER0	ER1– ER0
Polynomial	52.09	43.89	8.20
	[43.70, 57.27]	[35.91, 49.38]	[1.64, 15.42]
Step function	50.70	43.08	7.62
	[46.01, 54.78]	[37.86, 48.56]	[1.66, 13.21]
	B. The Effect of Entering Training in Month 19 (ER19) versus Never Entering Training (ER0)		
Model	ER19	ER0	ER19–ER0
Polynomial	47.69	43.89	3.80
	[40.94, 51.80]	[35.91, 49.38]	[-0.60, 8.61]
Step function	46.90	43.08	3.82
	[42.34, 50.99]	[37.86, 48.56]	[-0.03, 7.45]
	C. The Effect of Entering Training Immediately (ER1) versus Entering Training in Month 19 (ER19)		
Model	ER1	ER19	ER1– ER19
Polynomial	52.09	47.69	4.40
	[43.70, 57.27]	[40.94, 51.80]	[1.57, 8.33]
Step function	50.70	46.90	3.80
	[46.01, 54.78]	[42.34, 50.99]	[1.11, 6.75]

Notes: In brackets are the 95% confidence intervals based on the WCS bootstrap procedure with 100,000 repetitions.

Figure 3.1: Employment rates by month since the baseline

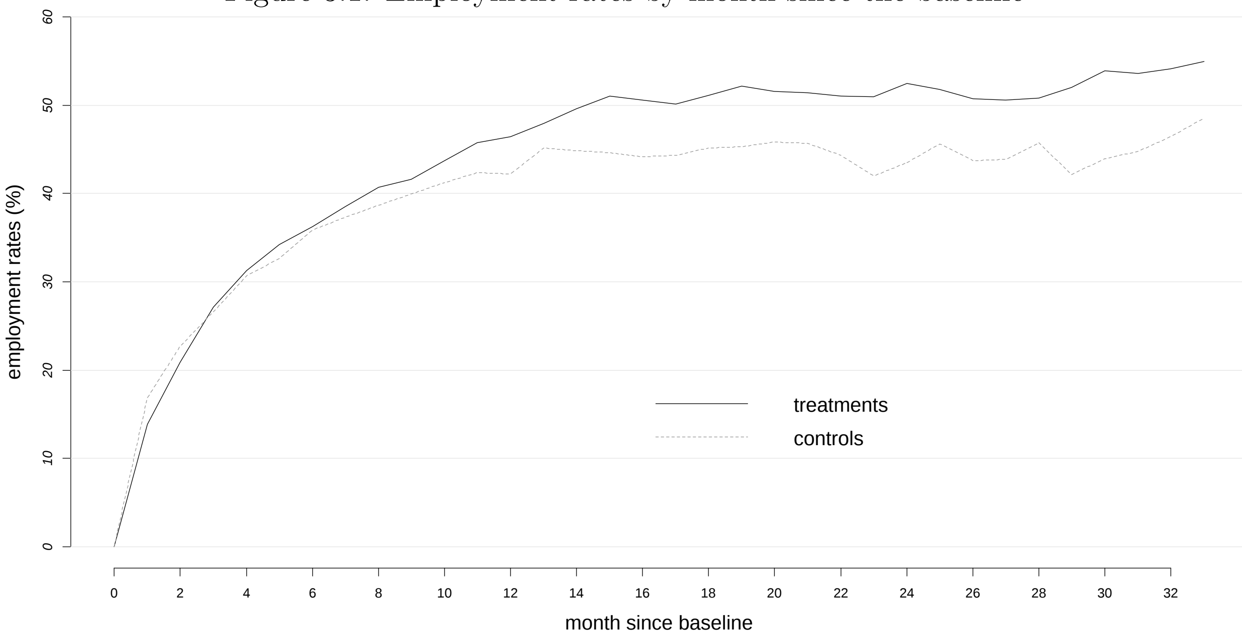


Figure 3.2: Kaplan-Meier hazards of entry into training spells for treatments and controls by month since the baseline.

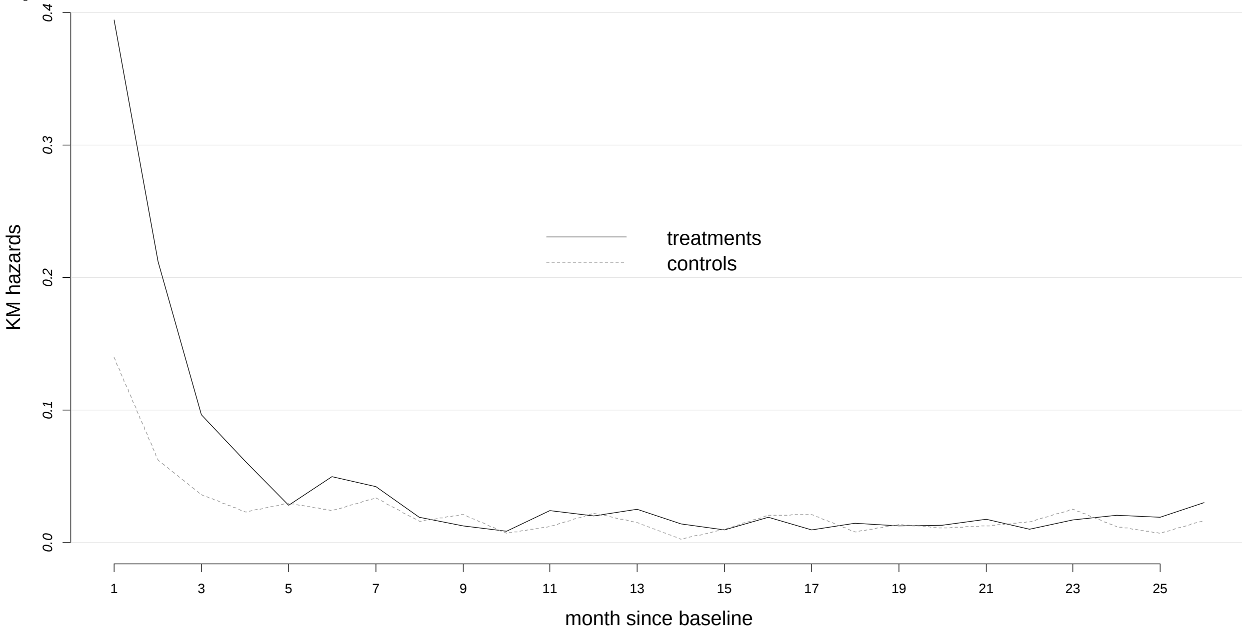


Figure 3.3: Kaplan-Meier hazards for the left-censored nonemployment spells of the trainees and nontrainees

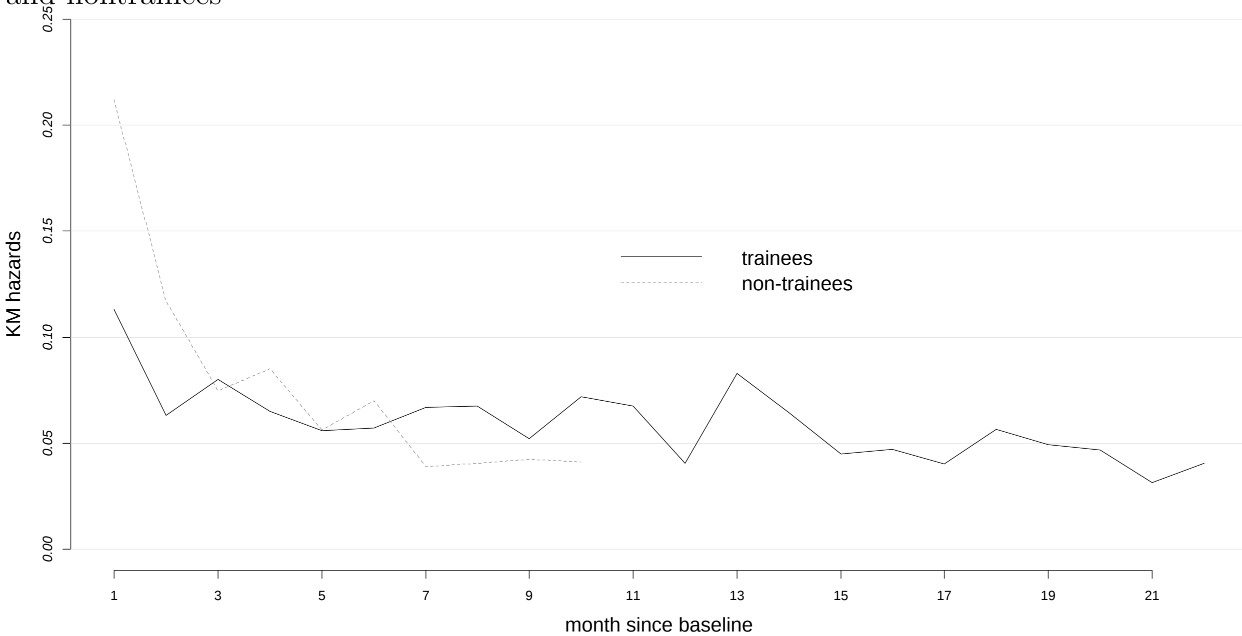


Figure 3.4: Kaplan-Meier hazards for the left-censored non-employment spells of the trainees who entered training within 3 months versus after 3 months, since the baseline

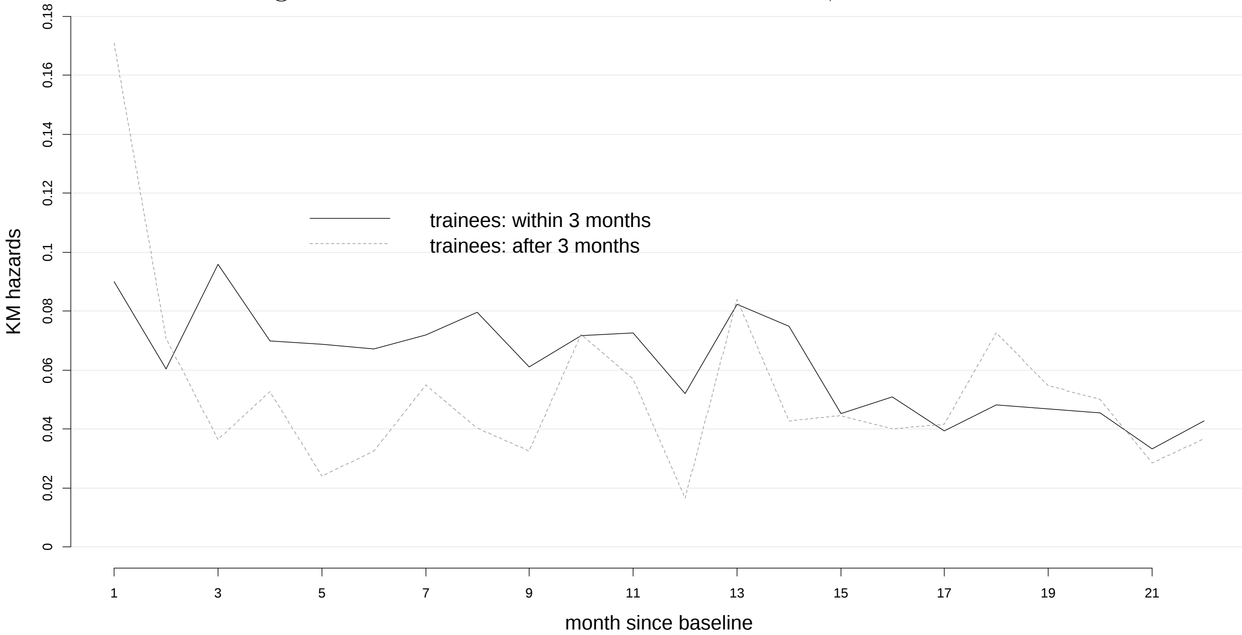


Figure 3.5: Kaplan-Meier hazards for the left-censored non- employment spells of the trainees and nontrainees by month since the baseline

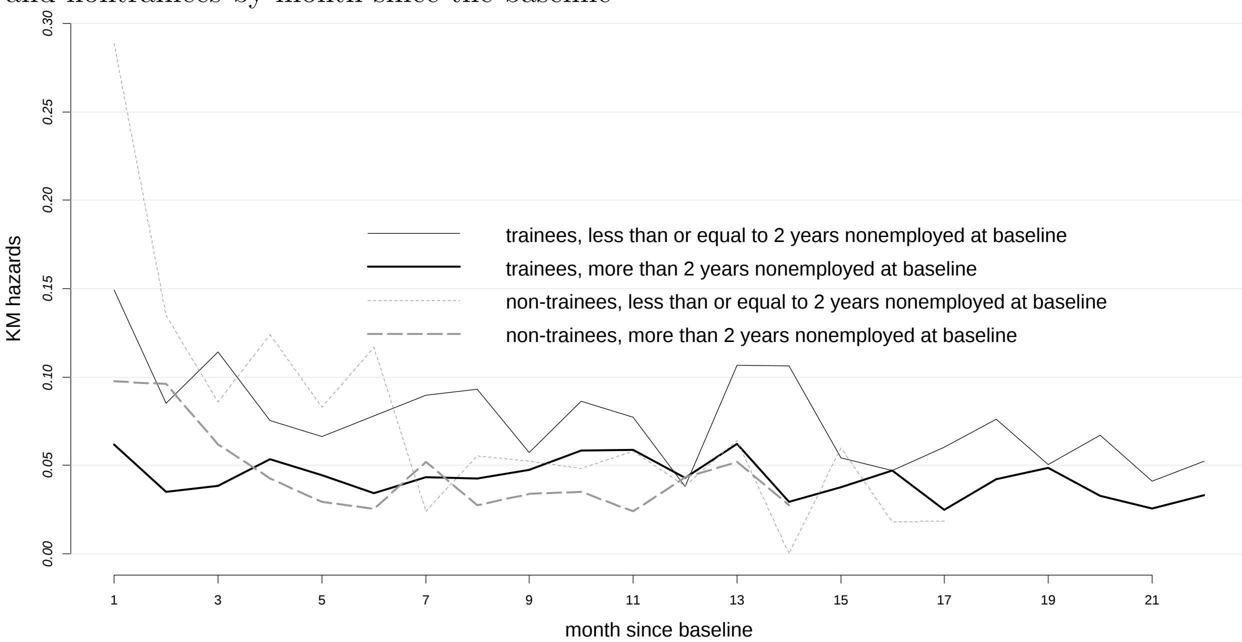


Figure 3.6: Kaplan-Meier hazards for the fresh employment spells of the trainees and non-trainees

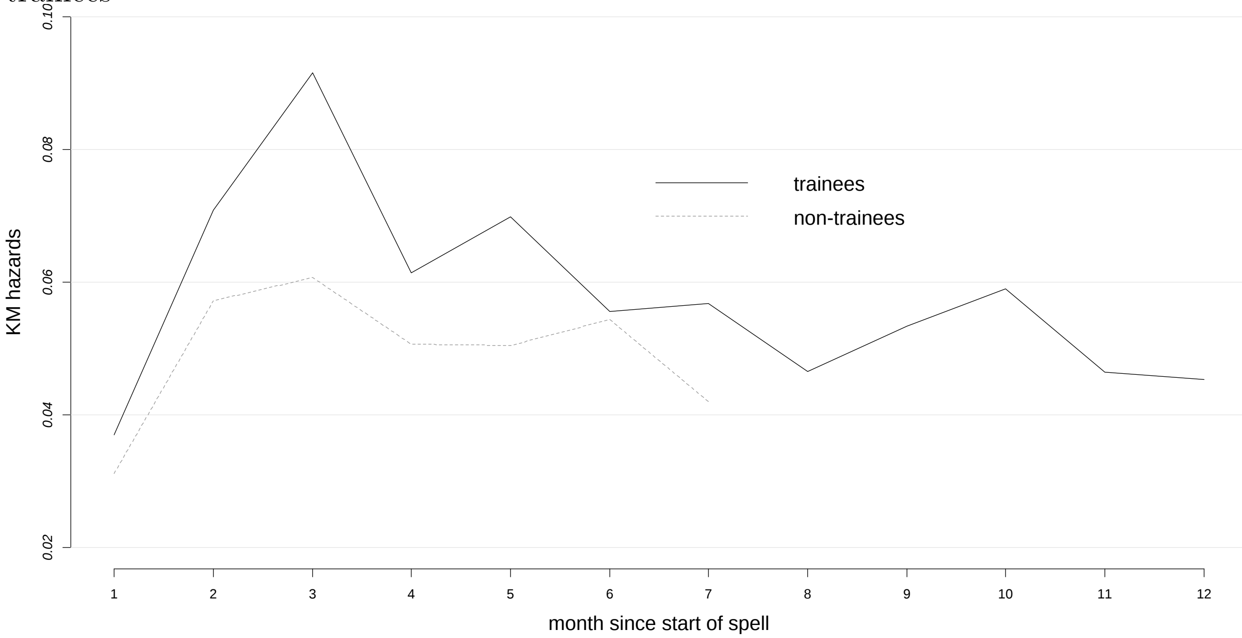


Figure 3.7: Kaplan-Meier hazards for the fresh unemployment spells of the trainees and nontrainees

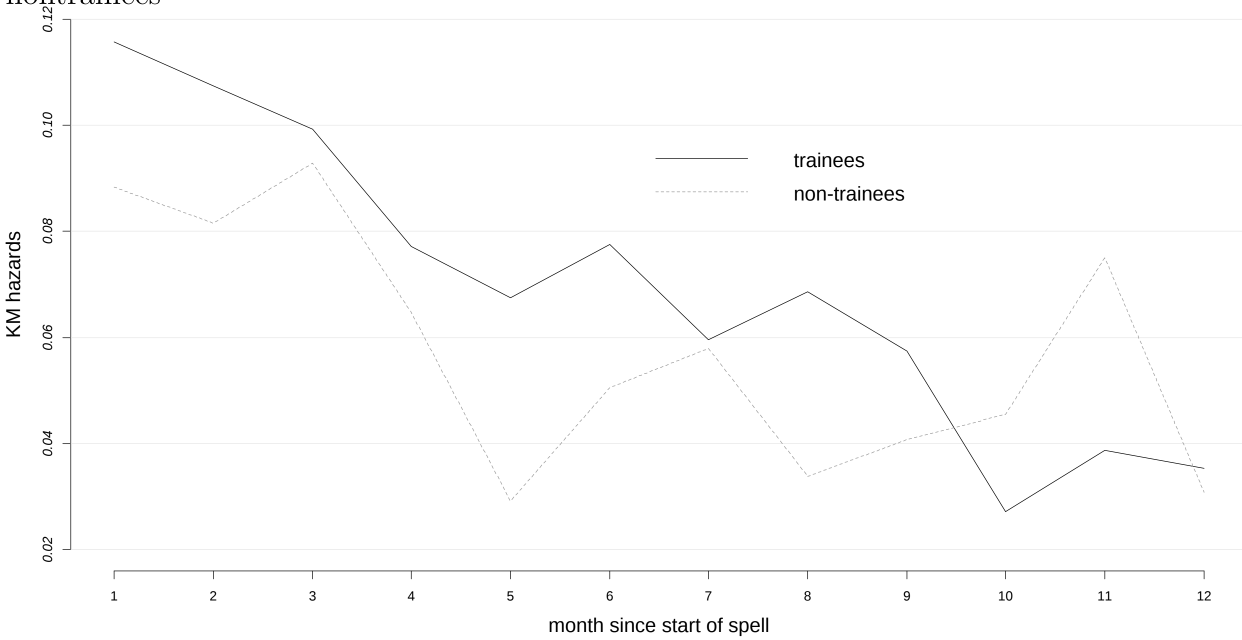
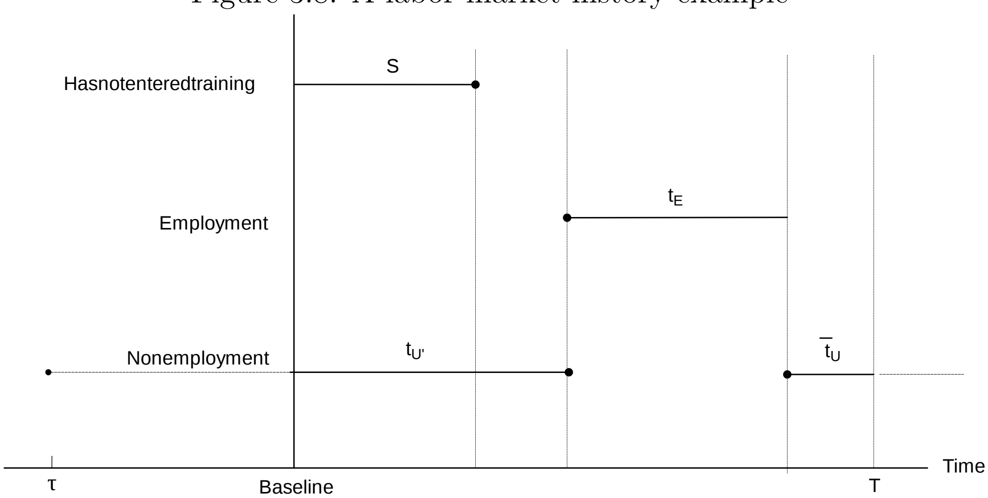


Figure 3.8: A labor market history example



3.9 Appendix A: The WCS Bootstrap Procedure

As discussed in Section 4, the effect of training participation on the expected duration, and the percentage of time that an individual is employed are highly relevant parameters for policy makers because they are easy to interpret. We describe our expected duration estimators in Section 4.1 (with the corresponding estimates presented in Table 3.8) and our simulation process for percentage of time employed in Section 4.2 (with the corresponding estimates presented in Table 3.9). For each estimate in Tables 6 and 7, we need to provide a 95% confidence interval for each estimate from the WCS bootstrap proposed by Woutersen and Ham [2016]. Here we give a step by step guide to implementing the WCS bootstrap procedure.[46]

Let θ contain all the parameters in our baseline polynomial model (corresponding to 77 point estimates in Table 3.4) including the unobserved heterogeneity parameters. Let θ denote our MLE estimate of θ , and let Ω denote the estimated variance-covariance matrix. Further, let $h(\theta)$ denote the function of θ that we are interested in. In this study, $h(\theta)$ is either the expected duration or the average percentage of time employed over the first 60 months after the baseline. To obtain a 95% confidence interval for $h(\theta)$ we carry out steps $1-4$ below.

- *Step 1:* Generate 100,000 parameter draws $\theta_j, j = 1, \ldots, 100, 00$ from the asymptotic distribution of θ using Ω . Calculate $h\left(\theta_j\right)$ for $j = 1, \ldots, 100, 00$.

- *Step 2:* Estimate a weight vector of length L (total number of parameters in the model) by regressing $h\left(\theta_j\right)$ on an intercept and θ_j .(The regression will have 100,000 'observations'. The estimated weight vector w consists of the slope coefficient estimates from the regression.)

46. Woutersen and Ham [2016] present some refinements of their procedure. We ignore them here because these refinements make little difference for the confidence interval.

- *Step 3:* Choose parameter draws θ_j and corresponding $h\left(\theta_j\right)$ to calculate the 95% confidence interval for $h(\theta)$. Among the 100,000 draws, we keep the θ_j 's with the 95,000 lowest values of $\left(\theta_j-\theta\right)' w\left[w'\Omega w\right]^{-1} w\left(\theta_j-\theta\right)$. The intuition if that we only want to keep a draw θ_j if $\theta_j' w$ is not too far from $\theta_j' w$.

- *Step 4:* Use the minimum and maximum of the 95,000 values of $h\left(\theta_j\right)$ associated with the 95,000 θ_j 's chosen in step 3 as the lower bound and upper bound estimates of the 95% confidence interval.

3.10 Appendix B: An Informative Simple Life-Cycle Consumption Problem where Treatment Takes the Form of a Price Subsidy in the Later Periods

Here we show that a subsidy given in later periods affects consumption in every period, i.e. we could not restrict the subsidy to only have effects when it is given. Consider an individual who lives for T periods in a world of perfect certainty, and consumes X_{1k} and X_{2k} in each period k . Her life-time utility function takes the Stone-Geary form[47]

$$V=\sum_{t=1}^{T}\frac{U\left(X_{1t},X_{2t}\right)}{(1+\rho)^t}=\sum_{t=1}^{T}\frac{\beta_1 ln\left(X_{1t}-\gamma_1\right)+\beta_2 ln\left(X_{2t}-\gamma_2\right)}{(1+\rho)^t}$$

$$V=\sum_{t=1}^{T}\beta_{1t}^{*} ln\left(X_{1t}-\gamma_1\right)+\beta_{2t}^{*} ln\left(X_{2t}-\gamma_2\right)$$

where $\beta_{1t}^{*}=\frac{\beta_1}{(1+\rho)^t}$, $\beta_{2t}^{*}=\frac{\beta_2}{(1+\rho)^t}$, and we employ the normalization $\sum_{t=1}^{T}\left(\beta_{1t}^{*}+\beta_{2t}^{*}\right)=1$. Her lifetime budget constraint is given by

47. Ashenfelter and Ham [1979] used an intertemporal Stone-Geary utility function to study life-cycle labor supply.

$$LFBC = \sum_{t=1}^{T} \left[\frac{y_t - P_{1t} X_{1t} - P_{2t} X_{2t}}{(1+r)^t} \right] = \sum_{t=1}^{T} [y_t^* - P_{1t}^* X_{1t} - P_{2t}^* X_{2t}] = 0$$

where $y_t^* = \frac{y_t}{(1+r)^t}$, $P_{1t}^* = \frac{P_{1t}}{(1+r)^t}$, and $P_{2t}^* = \frac{P_{2t}}{(1+r)^t}$.

We make the standard assumption that β_1 , β_2 , ρ , r , γ_1 , and γ_2 are all positive. Further, we can interpret γ_1 as the minimum feasible consumption of good 1 in each period and γ_2 as the minimum feasible consumption of good 2 in each period. Her Lagrange multiplier problem is

$$L = \sum_{t=1}^{T} \beta_{1t}^* ln\,(X_{1t} - \gamma_1) + \beta_{2t}^* ln\,(X_{2t} - \gamma_2) + \mu \{ \sum_{t=1}^{T} [y_t^* - P_{1t}^* X_{1t} - P_{2t}^* X_{2t}] \}$$

Her first order conditions are given by $\frac{\beta_{1t}^*}{(X_{1t}-\gamma_1)} = \mu P_{1t}^*$ and $\frac{\beta_{2t}^*}{(X_{2t}-\gamma_2)} = \mu P_{2t}^*$ or

$$\frac{\beta_{1t}^*}{\mu} = P_{1t}^* (X_{1t} - \gamma_1) \ (B1)$$

$$\frac{\beta_{2t}^*}{\mu} = P_{2t}^* (X_{2t} - \gamma_2) \ (B2)$$

Sum (B1) and (B2) over time to obtain

$$\frac{1}{\mu} \sum_{t=1}^{T} \beta_{1t}^* = \sum_{t=1}^{T} P_{1t}^* (X_{1t} - \gamma_1) = \sum_{t=1}^{T} P_{1t}^* X_{1t} - \sum_{t=1}^{T} P_{1t}^* \gamma_1 \ (B3)$$

$$\frac{1}{\mu} \sum_{t=1}^{T} \beta_{2t}^* = \sum_{t=1}^{T} P_{2t}^* (X_{2t} - \gamma_2) = \sum_{t=1}^{T} P_{2t}^* X_{2t} - \sum_{t=1}^{T} P_{2t}^* \gamma_2 \ (B4)$$

Now add (B3) and (B4) to obtain

$$\frac{1}{\mu}\sum_{t=1}^{T}(\beta_{1t}^{*}+\beta_{2t}^{*})=\sum_{t=1}^{T}P_{1t}^{*}X_{1t}+\sum_{t=1}^{T}P_{2t}^{*}X_{2t}-\sum_{t=1}^{T}P_{1t}^{*}\gamma_{1}-\sum_{t=1}^{T}P_{2t}^{*}\gamma_{2}\ (B5)$$

Recall that

$$\sum_{t=1}^{T}(\beta_{1t}^{*}+\beta_{2t}^{*})=1$$

and

$$\sum_{t=1}^{T}P_{1t}^{*}X_{1t}+\sum_{t=1}^{T}P_{2t}^{*}X_{2t}=\sum_{t=1}^{T}y_{t}^{*}\ (B6)$$

Using the expressions in (B6) to simplify (B5) yields

$$\frac{1}{\mu}=\sum_{t=1}^{T}y_{t}^{*}-\sum_{t=1}^{T}P_{1t}^{*}\gamma_{1}-\sum_{t=1}^{T}P_{2t}^{*}\gamma_{2}$$

$$\mu=\left[\sum_{t=1}^{T}y_{t}^{*}-\sum_{t=1}^{T}P_{1t}^{*}\gamma_{1}-\sum_{t=1}^{T}P_{2t}^{*}\gamma_{2}\right]^{-1}\quad(B7)$$

Equation (B7) illustrates the most attractive feature of using the Stone-Geary specification: One obtains a closed form solution for the marginal utility of life-time income, μ . Substituting (B7) into (B1) and (B2) yields the intertemporal demand functions

$$X_{1t}=\gamma_{1}+\frac{\beta_{1t}^{*}}{P_{1t}^{*}}\left[\sum_{t=1}^{T}y_{t}^{*}-\sum_{t=1}^{T}P_{1t}^{*}\gamma_{1}-\sum_{t=1}^{T}P_{2t}^{*}\gamma_{2}\right]\ (B8)$$

$$X_{2t}=\gamma_{2}+\frac{\beta_{2t}^{*}}{P_{2t}^{*}}\left[\sum_{t=1}^{T}y_{t}^{*}-\sum_{t=1}^{T}P_{1t}^{*}\gamma_{1}-\sum_{t=1}^{T}P_{2t}^{*}\gamma_{2}\right]\ (B9)$$

Now consider two individuals who have the same path for y_t and hence the same consumption of X_1 and X_2 in each period, and randomly assign one as a treatment and one as a control. Now assume that we subsidize the consumption of X_1 for the treatment, such that for $t \geq \bar{t}$ she pays $(1-\tau)\, P_{1t}^{*}$ (where $0 < \tau < 1$)for X_1 The control's demand functions continue to be given by (B8) and (B9) but for $t \geq \bar{t}$, the treatment's demand functions will change to

$$X_{1t}^{TR} = \gamma_1 + \frac{\beta_{1t}^{*}}{(1-\tau)\, P_{1t}^{*}} \left[\sum_{t=1}^{T} y_t^{*} - \sum_{t=1}^{t-1} P_{1t}^{*}\gamma_1 - \sum_{t=t}^{T} (1-\tau)\, P_{1t}^{*}\gamma_1 - \sum_{t=1}^{T} P_{2t}^{*}\gamma_2 \right] \; (B10)$$

$$X_{2t}^{TR} = \gamma_2 + \frac{\beta_{2t}^{*}}{P_{2t}^{*}} \left[\sum_{t=1}^{T} y_t^{*} - \sum_{t=1}^{t-1} P_{1t}^{*}\gamma_1 - \sum_{t=t}^{T} (1-\tau)\, P_{1t}^{*}\gamma_1 - \sum_{t=1}^{T} P_{2t}^{*}\gamma_2 \right] \; (B11)$$

Thus, there is a clear treatment effect for $t \geq \bar{t}$, as the arguments in the treatment's demand functions B(10) and B(11) differ from those of the control's demand functions (B8) and (B9) in two ways. First, they differ in terms of the current price of first good, which is $(1-\tau)\, P_{1t}^{*}$ for the treatment versus P_{1t}^{*} for the control. This change will create an intertemporal substitution effect. Second, the term inside the square brackets, reflecting the (inverse of the) marginal utility of lifetime income, now also differs between the treatment and the control. For the treatment, this term is

$$\left[\sum_{t=1}^{T} y_t^{*} - \sum_{t=1}^{t-1} P_{1t}^{*}\gamma_1 - \sum_{t=t}^{T} (1-\tau)\, P_{1t}^{*}\gamma_1 - \sum_{t=1}^{T} P_{2t}^{*}\gamma_2 \right] \; (B12)$$

while for the control it is

$$\left[\sum_{t=1}^{T} y_t^{*} - \sum_{t=1}^{T} P_{1t}^{*}\gamma_1 - \sum_{t=1}^{T} P_{2t}^{*}\gamma_2 \right] \; (B13)$$

The subsidy creates a positive 'income effect' for the treatment, since (B12) is larger than (B13). For periods $t < t$ when the subsidy is not given to the treatment, her demand functions are

$$X_{1t}^{TR} = \gamma_1 + \frac{\beta_{1t}^*}{P_{1t}^*}\left[\sum_{t=1}^{T} y_t - \sum_{t=1}^{t-1} P_{1t}^*\gamma_1 - \sum_{t=t}^{T}(1-\tau)\,P_{1t}^*\gamma_1 - \sum_{t=1}^{T} P_{2t}^*\gamma_2\right] \quad (B14)$$

$$X_{2t}^{TR} = \gamma_2 + \frac{\beta_{2t}^*}{P_{2t}^*}\left[\sum_{t=1}^{T} y_t - \sum_{t=1}^{t-1} P_{1t}^*\gamma_1 - \sum_{t=t}^{T}(1-\tau)\,P_{1t}^*\gamma_1 - \sum_{t=1}^{T} P_{2t}^*\gamma_2\right] \quad (B15)$$

Note that these also differ from the control's demand functions in (B8) and (B9) for $t < \bar{t}$ because the terms inside the square brackets continue to be different, as in (B12) versus (B13). In other words, the 'income effect' carries over to the treatment's consumption of both goods in periods $t < \bar{t}$. In fact, it is straightforward to show that the treatment has higher consumption of both X_1 and X_2 in all periods than the control does.

In this example, there is a treatment effect even in periods when subsidy is not operating, i.e. treatment status affects consumption even in the periods that the treatment is not participating in the subsidy program. In terms of the JTPA programs, those taking training in the future will face a subsidized price for this training relative to the controls. This subsidy will affect them in all periods, including the periods before they enter training, just as in the above consumption subsidy example. Hence the treatments are not comparable to the controls in the periods before they enter training.[48]

48. Recall that we are assuming a world of perfect certainty, i.e. individuals know whether they will participate in training in the future. If there is uncertainty over which treatments will take training in the future, any treatment with a positive probability of taking training will be affected by the option of receiving free training in the future.

REFERENCES

Jaap H Abbring and Gerard J Van den Berg. The nonparametric identification of treatment effects in duration models. *Econometrica*, 71(5):1491–1517, 2003.

ACLU. Less lethal force: Proposed standards for massachusetts law enforcement agencies, nd.

Geoffrey P Alpert and Roger G Dunham. Policy and training recommendations related to police use of ceds: Overview of findings from a comprehensive national study. *Police quarterly*, 13(3):235–259, 2010.

Nejat Anbarci and Jungmin Lee. Detecting racial bias in speed discounting: Evidence from speeding tickets in boston. *International Review of Law and Economics*, 38:11–24, 2014.

Shamena Anwar and Hanming Fang. An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review*, 96(1):127–151, March 2006.

Orley Ashenfelter and John Ham. Education, Unemployment, and Earnings. *Journal of Political Economy*, 87(5):99–116, October 1979.

Bocar A. Ba, John C. Ham, Robert J. LaLonde, and Xianghong Li. Estimating (easily interpreted) dynamic training effects from experimental data. *Journal of Labor Economics*, 35(S1):S149–S200, 2017.

Michael Baker and Angelo Melino. Duration dependence and nonparametric heterogeneity: A Monte Carlo study. *Journal of Econometrics*, 96(2):357–393, June 2000.

Michael Baker and Samuel A. Rea. Employment Spells And Unemployment Insurance Eligibility Requirements. *The Review of Economics and Statistics*, 80(1):80–94, February 1998.

Gary S Becker. Crime and punishment: An economic approach. In *The Economic Dimensions of Crime*, pages 13–68. Springer, 1968.

Martin Biewen, Bernd Fitzenberger, Aderonke Osikominu, and Marie Paul. The effectiveness of public-sponsored training revisited: The importance of data and methodological choices. *Journal of Labor Economics*, 32(4):837–897, 2014.

Dan A. Black, Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South. *American Economic Review*, 105(2):477–503, February 2015.

Howard S Bloom, Larry L Orr, Stephen H Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M Bos. The benefits and costs of jtpa title ii-a programs: Key findings from the national job training partnership act study. *Journal of human resources*, pages 549–576, 1997.

Howard S Bloom et al. The national jtpa study. title ii-a impacts on earnings and employment at 18 months. 1993.

William P Bozeman, William E Hauda, Joseph J Heck, Derrel D Graham, Brian P Martin, and James E Winslow. Safety and injury profile of conducted electrical weapons used by law enforcement officers against criminal suspects. *Annals of emergency medicine*, 53(4): 480–489, 2009.

TR Braidwood. Why? the robert dziekanski tragedy: Braidwood commission on the death of robert dziekanski (bc), 2010.

Mike Brewer and Jonathan Cribb. Lone parents, time-limited in-work credits and the dynamics of work and welfare. ISER Working Paper Series 2017-01, Institute for Social and Economic Research, January 2017.

Alexander A Bustamante. Overview of less-lethal force tools and deployment, 2017. URL `http://www.lapdpolicecom.lacity.org/022817/BPC_17-0057.pdf`.

Cory Capps, David Dranove, and Mark Satterthwaite. Competition and market power in option demand markets. *RAND Journal of Economics*, pages 737–763, 2003.

David Card. Using geographic variation in college proximity to estimate the return to schooling. *in Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp, ed. by Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky. Toronto: University of Toronto Press, 201-222*, 1995.

Aaron Chalfin and Justin McCrary. Criminal deterrence: A review of the literature. *Journal of Economic Literature*, 55(1):5–48, 2017.

A Corona, M Marchesi, C Martini, and S Ridella. Minimizing multimodal functions of continuous variables with the simulated annealing algorithm. *ACM Transactions on Mathematical Software*, 13(3):262–280, 1987.

Bruno Crépon, Marc Ferracci, Grégory Jolivet, and Gerard J van den Berg. Information shocks and the empirical evaluation of training programs during unemployment spells. Technical report, Unpublished manuscript, CREST- INSEE, 2016.

Gordon B Dahl. Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica*, 70(6):2367–2420, 2002.

Donald R. Davis, Jonathan I. Dingel, Joan Monras, and Eduardo Morales. How segregated is urban consumption. Technical report, Technical report, Columbia University, 2016.

Chicago Police Department. Tactical response report - form preparation instructions, cpd-63.467, 467, 2016a.

Chicago Police Department. Tactical response report (vol. cpd-11.377), 2016b.

Seattle Police Department. Special report on the m26 advanced taser (pt . 1). seattle, 2003.

Manasi Deshpande and Yue Li. Who is screened out? application costs and the targeting of disability programs. Technical report, National Bureau of Economic Research, 2017.

Matthew Desmond, Andrew V Papachristos, and David S Kirk. Police violence and citizen crime reporting in the black community. *American Sociological Review*, 81(5):857–876, 2016.

Esther Duflo. Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, 91(4): 795–813, September 2001.

Curtis Eberwein, John C Ham, and Robert J LaLonde. The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: Evidence from experimental data. *The Review of Economic Studies*, 64(4):655–682, 1997.

Curtis Eberwein, John C Ham, and Robert J LaLonde. Alternative methods of estimating program effects in event history models. *Labour Economics*, 9(2):249–278, 2002.

Liran Einav, Amy Finkelstein, and Heidi Williams. Paying on the Margin for Medical Care: Evidence from Breast Cancer Treatments. *American Economic Journal: Economic Policy*, 8(1):52–79, February 2016.

Christine Eith and Matthew R Durose. Contacts between police and the public, 2008. *Washington, DC*, 2011.

Zusha Elinson and Dan Frosch. Cost of police-misconduct cases soars in big us cities. *The Wall Street Journal*, 2015.

Amy Finkelstein, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, and Katherine Baicker. The Oregon Health Insurance Experiment:

Evidence from the First Year. *The Quarterly Journal of Economics*, 127(3):1057–1106, 2012.

Aderonke Osikominu Fitzenberger, Bernd and Marie Paul. The heterogeneous effects of training incidence and duration on labor market transitions. Technical report, Seminar slides for a presentation at the National University of Singapore, March 10, 2016.

Chris Flinn and James J Heckman. The likelihood function for the multistate-multiepisode model in models for the analysis of labor force dynamics. *Advances in econometrics*, 2: 225–231, 1983.

Peter Fredriksson and Per Johansson. Dynamic treatment assignment: The consequences for evaluations using observational data. *Journal of Business Economic Statistics*, 26(4): 435–445, 2008.

Roland G. Fryer. An empirical analysis of racial differences in police use of force, 2016.

Andrew Gelman, Jeffrey Fagan, and Alex Kiss. An analysis of the new york city police department's "stop-and-frisk" policy in the context of claims of racial bias. *Journal of the American Statistical Association*, 102(479):813–823, 2007.

Edward L Glaeser. An overview of crime and punishment. *Washington: World Bank. Mimeographed*, 1999.

Sharad Goel, Justin M Rao, Ravi Shroff, et al. Precinct or prejudice? understanding racial disparities in new york cityÕs stop-and-frisk policy. *The Annals of Applied Statistics*, 10 (1):365–394, 2016.

Felipe Goncalves and Steve Mello. A few bad apples? racial bias in policing. 2017.

Gautam Gowrisankaran, Aviv Nevo, and Robert Town. Mergers when prices are negotiated:

Evidence from the hospital industry. *The American Economic Review*, 105(1):172–203, 2014.

Jeffrey Grogger and Greg Ridgeway. Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association*, 101(475):878–887, 2007.

John C Ham and Robert J LaLonde. The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training. *Econometrica: Journal of the Econometric Society*, pages 175–205, 1996.

John C Ham and Jr Rea, Samuel A. Unemployment Insurance and Male Unemployment Duration in Canada. *Journal of Labor Economics*, 5(3):325–353, July 1987.

John C Ham, Jan Svejnar, and Katherine Terrell. Unemployment and the Social Safety Net during Transitions to a Market Economy: Evidence from the Czech and Slovak Republics. *American Economic Review*, 88(5):1117–1142, December 1998.

John C. Ham, Xianghong Li, and Lara D. Shore-Sheppard. The Employment Dynamics of Disadvantaged Women: Evidence from the SIPP. *Journal of Labor Economics*, 34(4): 899–944, 2016.

Jerry Hausman and Tiemen Woutersen. Estimating the derivative function and counterfactuals in duration models with heterogeneity. *Econometric Reviews*, 33(5-6):472–496, 2014.

Jerry A. Hausman. Specification tests in econometrics. *Econometrica*, 46(6):1251–1271, 1978.

Paul Heaton. Understanding the effects of antiprofiling policies. *The Journal of Law and Economics*, 53(1):29–64, 2010.

James Heckman and Burton Singer. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society*, pages 271–320, 1984a.

James Heckman, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. Characterizing selection bias using experimental data. *Econometrica*, 66(5):1017–1098, 1998.

James J Heckman and Burton Singer. Econometric duration analysis. *Journal of Econometrics*, 24(1):63–132, 1984b.

James J. Heckman and Jeffrey Smith. The Sensitivity of Experimental Impact Estimates (Evidence from the National JTPA Study). In *Youth Employment and Joblessness in Advanced Countries*, NBER Chapters, pages 331–356. National Bureau of Economic Research, Inc, May 2000.

James J Heckman and Jeffrey A Smith. The pre-programme earnings dip and the determinants of participation in a social programme. implications for simple programme evaluation strategies. *The Economic Journal*, 109(457):313–348, 1999.

James J Heckman and Jeffrey A Smith. The determinants of participation in a social program: Evidence from a prototypical job training program. *Journal of Labor Economics*, 22(2):243–298, 2004.

James J Heckman and James R Walker. The relationship between wages and income and the timing and spacing of births: evidence from swedish longitudinal data. *Econometrica: journal of the Econometric Society*, pages 1411–1441, 1990.

James J Heckman, Hidehiko Ichimura, and Petra E Todd. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, 64(4):605–654, 1997.

Matthew J Hickman and Jane E Poore. National data on citizen complaints about police use of force: Data quality concerns and the potential (mis) use of statistical evidence to address police agency conduct. *Criminal Justice Policy Review*, 27(5):455–479, 2016.

Dan Hinkel and Jennifer Smith Richards. Chicago cops get more tasers, but red flags remain, 2017. URL `http://www.chicagotribune.com/news/local/breaking/ct-chicago-police-tasers-met-20170825-story.html`.

Taser International. User certification course taser x26 electronic control device, 2010.

Taser International. User certification course taser x26 electronic control device, 2011.

Emma Jenkinson, Clare Neeson, and Anthony Bleetman. The relative risk of police use-of-force options: evaluating the potential for deployment of electronic weaponry. *Journal of clinical forensic medicine*, 13(5):229–241, 2006.

Jesse Kalinowski, Stephen L Ross, Matthew B Ross, et al. Endogenous driving behavior in veil of darkness tests for racial profiling. Technical report, 2017.

Robert J Kaminski, Robin S Engel, Jeff Rojek, Michael R Smith, and Geoffrey Alpert. A quantum of force: The consequences of counting routine conducted energy weapon punctures as injuries. *Justice Quarterly*, 32(4):598–625, 2015.

Robert J Kane. The social ecology of police misconduct*. *Criminology*, 40(4):867–896, 2002.

John Knowles, Nicola Persico, and Petra Todd. Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, 109(1):203–229, 2001.

LLC LAAW International. Outline of partial selected cew research and information, 2017.

Tony Lancaster. Econometric methods for the duration of unemployment. *Econometrica: Journal of the Econometric Society*, pages 939–956, 1979.

Michael Lechner. Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *Review of Economics and Statistics*, 84(2): 205–220, 2002.

Michael Lechner. Long-run labour market and health effects of individual sports activities. *Journal of health economics*, 28(4):839–854, 2009.

Xianghong Li and Barry Smith. Diagnostic analysis and computational strategies for estimating discrete time duration models: A monte carlo study. *Journal of Econometrics*, 187 (1):275–292, 2015.

Yu-Sheng Lin and Tonisha R Jones. Electronic control devices and use of force outcomesincidence and severity of use of force, and frequency of injuries to arrestees and police officers. *Policing: An international journal of police strategies & management*, 33(1):152–178, 2010.

Cynthia Lum and Daniel S Nagin. Reinventing american policing. *Crime and justice*, 46(1): 339–393, 2017.

John M MacDonald, Robert J Kaminski, and Michael R Smith. The effect of less-lethal weapons on injuries in police use-of-force events. *American journal of public health*, 99 (12):2268–2274, 2009.

Charles F Manski and Daniel S Nagin. Assessing benefits, costs, and disparate racial impacts of confrontational proactive policing. *Proceedings of the National Academy of Sciences*, page 201707215, 2017.

Brian P McCall. Unemployment insurance rules, joblessness, and part-time work. *Econometrica: Journal of the Econometric Society*, pages 647–682, 1996.

Bruce D Meyer. Unemployment Insurance and Unemployment Spells. *Econometrica*, 58(4): 757–82, July 1990.

Thomas A. Mroz and Yaraslau V. Zayats. Arbitrarily Normalized Coefficients, Information Sets, and False Reports of "Biases" in Binary Outcome Models. *The Review of Economics and Statistics*, 90(3):406–413, August 2008.

Jena Neuscheler and Akiva Freidlin. Report on electronic control weapons (ecws) submitted to the city of berkeley, 2015. URL https://sanfranciscopolice.org/sites/default/files/Documents/PoliceCommission/2015-StanfordStudy_ECWfromDPA.pdf.

Stephen J Nickell. Estimating the Probability of Leaving Unemployment. *Econometrica*, 47 (5):1249–1266, September 1979.

NIJ. Monitoring police use of conducted energy devices, 2008. URL https://www.nij.gov/topics/technology/less-lethal/pages/monitoring-ced-use.aspx.

NIJ. Police use of force, tasers and other less-lethal weapons, 2011. URL Retrievedfromhttp://www.ncjrs.gov/pdffiles1/nij/grants/231176.pdf.

Aderonke Osikominu. Quick Job Entry or Long-Term Human Capital Development? The Dynamic Effects of Alternative Training Schemes. *Review of Economic Studies*, 80(1): 313–342, 2013.

Eugene A Paoline III, William Terrill, and Jason R Ingram. Police use of force and officer injuries: Comparing conducted energy devices (ceds) to hands-and weapon-based tactics. *Police Quarterly*, 15(2):115–136, 2012.

Canice Prendergast. Selection and oversight in the public sector, with the los angeles police department as an example. Technical report, National Bureau of Economic Research, 2001.

Canice Prendergast. The response of the lapd to increased oversight. Technical report, 2002.

Canice Prendergast. The limits of bureaucratic efficiency. *Journal of Political Economy*, 111 (5):929–958, 2003.

Charles H. Ramsey and Laurie O Robinson. Final report of the president's task force on 21st century policing. 2015.

Geert Ridder. The non-parametric identification of generalized accelerated failure-time models. *The Review of Economic Studies*, 57(2):167–181, 1990.

Greg Ridgeway and John M MacDonald. Doubly robust internal benchmarking and false discovery rates for detecting racial bias in police stops. *Journal of the American Statistical Association*, 104(486):661–668, 2009.

JA Ritter and D Bael. Detecting racial profiling in minneapolis traffic stops: A new approach. cura reporter, 11–17, 2009.

Joanna C Schwartz. Police indemnification. 2014.

Joanna C Schwartz. How governments pay: Lawsuits, budgets, and police reform. *UCLA L. Rev*, 63:1144, 2016.

Lan Shi. The limit of oversight in policing: Evidence from the 2001 cincinnati riot. *Journal of Public Economics*, 93(1):99–113, 2009.

Barbara Sianesi. An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. *The Review of Economics and Statistics*, 86(1):133–155, February 2004.

Kenneth A Small and Harvey S Rosen. Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society*, pages 105–130, 1981.

Michael R Smith, Robert J Kaminski, Jeffrey Rojek, Geoffrey P Alpert, and Jason Mathis. The impact of conducted energy devices and other types of force and resistance on officer and suspect injuries. *Policing: An International Journal of Police Strategies & Management*, 30(3):423–446, 2007.

Michael R Smith, Robert J Kaminski, Geoffrey P Alpert, Lorie A Fridell, John MacDonald, and Bruce Kubu. A multi-method evaluation of police use of force outcomes: Final report to the national institute of justice. *SURVEY METHODOLOGY*, 3(1), 2009.

William Sousa, Justin Ready, and Michael Ault. The impact of tasers on police use-of-force decisions: Findings from a randomized field-training experiment. *Journal of Experimental Criminology*, 6(1):35–55, 2010.

Ann Huff Stevens. Climbing out of poverty, falling back in: Measuring the persistence of poverty over multiple spells. *The Journal of Human Resources*, 34(3):557–588, 1999.

Annie Sweeney and Kristen Schorsch. Chicago police expanding taser use, 2010. URL `http://articles.chicagotribune.com/2010-03-10/news/ct-met-police-expand-taser-use-20100310_1_taser-chicago-police-department-stun`.

Konstantinos Tatsiramos. Job displacement and the transitions to re-employment and early retirement for non-employed older workers. *European Economic Review*, 54(4):517–535, 2010.

Bruce Taylor and Daniel J Woods. Injuries to officers and suspects in police use-of-force cases: A quasi-experimental evaluation. *Police Quarterly*, 13(3):260–289, 2010.

Bruce Taylor, Daniel Woods, Bruce Kubu, Chris Koper, Bill Tegeler, Jason Cheney, Mary Martinez, James Cronin, and Kristin Kappelman. Comparing safety outcomes in police use-of-force cases for law enforcement agencies that have deployed conducted energy devices and a matched comparison group that have not: a quasi-experimental evaluation. In *Police Executive Research Forum*, 2009.

William Terrill and Eugene A Paoline III. Conducted energy devices (ceds) and citizen injuries: The shocking empirical reality. *Justice Quarterly*, 29(2):153–182, 2012.

Tom R Tyler. Enhancing police legitimacy. *The annals of the American academy of political and social science*, 593(1):84–99, 2004.

Tom R Tyler and Jeffrey Fagan. Legitimacy and cooperation: Why do people help the police fight crime in their communities. *Ohio St. J. Crim. L*, 6:231, 2008.

Maarten Van Craen and Wesley G Skogan. Trust in the belgian police: The importance of responsiveness. *European journal of criminology*, 12(2):129–150, 2015.

Samuel Walker and Morgan Macdonald. An alternative remedy for police misconduct: A model state pattern for practice statute. *Geo. Mason UCRLJ*, 19:479, 2008.

Ronald Weitzer. Citizens' perceptions of police misconduct: Race and neighborhood context. *Justice Quarterly*, 16(4):819–846, 1999.

Ronald Weitzer and Steven A Tuch. Race and perceptions of police misconduct. *Social problems*, 51(3):305–325, 2004.

Jeremy West. Racial bias in police investigations, 2015.

Tiemen Woutersen and John C. Ham. Calculating standard errors for discontinuous functions of parameter estimates. Technical report, 2016.