UNIVERSITY OF CALIFORNIA
RIVERSIDE

Homo Economicus Goes to Prison:
Individual and Group Behavior in Prison

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Jonathan Lucas Kurzfeld

September 2018

Dissertation Committee:

Dr. Richard James Arnott, Chairperson
Dr. Steven Clark
Dr. Joseph Cummins
Dr. Urmee Khan

The Dissertation of Jonathan Lucas Kurzfeld is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am ever grateful to my advisor, Dr. Richard Arnott. I cannot imagine a more knowledgable guiding influence to lead one through the tribulations of scholarship. It is only with his support and kind encouragement that I have brought these efforts to fruition. Dr. Arnott's patient mentorship has helped me develop my skills in writing, communication, and critical thinking. Most importantly, he has been a role model that engendered my passion and deep respect for the dual role of researcher and educator.

I am also deeply grateful to the other members of my dissertation committee, past and present, each of whom have played critical and unique roles in my own development and that of my research. In particular, I am grateful to: Dr. Steven Clark for his enthusiastic support of my research interests and teaching me about both interdisciplinary collaboration and public sector collaboration. Dr. Joseph Cummins for helping me develop a critical eye for empirical challenges and for role modeling an unrivaled passion for research that I have found deeply inspiring. Dr. Urmee Khan for her instructions in contract theory and information economics; for her patience and guidance in developing my instructional skills; and for her helpful comments and suggestions on my research.

I am thankful to my classmates and friends in the Department of Economics at UC Riverside for their help and support during the last six years. I would also like to express gratitude towards Ms. Damaris Carols, Ms. Tanya Wine, and, in particular, Mr. Gary Kuzas for their exceptional administrative support.

I am immensely grateful to all my family and friends. Their continued support and encouragement made it possible for me to be where I am today.

To my parents and family, whose support and faith in me has ever been absolute.

ABSTRACT OF THE DISSERTATION

Homo Economicus Goes to Prison:
Individual and Group Behavior in Prison

by

Jonathan Lucas Kurzfeld

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, September 2018
Dr. Richard James Arnott, Chairperson

This dissertation presents three independent research projects. The first studies the effect of prison crowding on violent misconduct, the second presents a model of prison gangs as profit-maximizing suppliers of illicit goods, and the third is an impact analysis of eyewitness identification protocols.

Chapter 1 provides a brief overview of each research project.

The first research project, Chapter 2, seeks to estimate the causal relationship between prison crowding and violent behavior. This study exploits exogenous variation in California prison populations, resulting from a Supreme Court mandate to reduce prison crowding, to estimate the effect on violence. Using both difference-in-differences and instrumental variables identification strategies, a significant positive relationship is identified that is robust to a variety of model specifications. These are the first empirical estimates showing a causal link between crowding and violence, suggesting that reducing prison crowding by 10 percentage points leads to a reduction in the rate of assault and battery of approximately 15%. In addition, differential reductions in the rates of violence between population types

is presented as evidence of a compositional effect associated with shocks to prison crowding, which poses a threat to the validity of empirical estimates of the link between crowding and violence.

Chapter 3 synthesizes existing research on prison gangs into an explicit modeling framework that treats gangs as profit-maximizing suppliers and sources of informal governance in an illicit marketplace. The model offers broad policy implications for prison enforcement and highlights the futility of certain policy approaches that don't account for the profit motive underlying gang activity.

In Chapter 4, we test for the presence of an identifiable impact on police clearance rates from the implementation of statewide reforms that adopt the sequential lineup process. We find insufficient evidence to identify an average effect for all reform states, but evidence this is the result of heterogeneous effects. We are also able to bound an the possible effect on clearance rates to rule out concerns that reforms lead to large reductions in positive identifications.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

This dissertation presents three independent research projects. The first project studies the effect of prison crowding on violent misconduct, the second presents a model of prison gangs as profit-maximizing suppliers of illicit goods, and the third is an impact analysis of eyewitness identification reforms.

The objective of Chapter 2 is to estimate the causal relationship between prison crowding and violent behavior. This study exploits exogenous variation in California prison populations, resulting from a Supreme Court mandate to reduce prison crowding, and estimates the effect on violence. Using both difference-in-differences and instrumental variables identification strategies, a significant positive relationship is identified that is robust to a variety of model specifications. These are the first empirical estimates showing a causal link between crowding and violence. They suggest that reducing prison crowding by 10 percentage points leads to a reduction in the rate of assault and battery of approximately 15%. In addition, differential reductions in the rates of violence between population types is presented as evidence of a compositional effect associated with shocks to prison crowding, which poses a threat to the validity of empirical estimates of the link between crowding and violence.

Chapter 3 explores the motives and violent behavior of prison gangs. Prison gangs are often credited as central to creating a "culture of violence" in U.S. prisons and jails. Yet mounting research and evidence suggests that prison gangs, in pursuit of profits from illicit market activity, also act as a check on the violent behavior of the broader prison population. This paper synthesizes existing research on prison gangs into an explicit modeling framework that treats gangs as profit-maximizing suppliers, and sources of informal governance, in an illicit marketplace. The model offers broad policy implications for prison enforcement and highlights the futility of certain policy approaches that don't account for the profit motive underlying gang activity.

The final chapter presents new evidence on the effectiveness of reforms to eyewitness lineup procedures in criminal investigations. Eyewitness testimony has long been a cornerstone of criminal justice practices in the United States. However, the proper processes and procedures of generating reliable eyewitness identifications are are a subject of vociferous academic debate. A particular point of contention is on the alleged superiority of a blindly administered sequential presentation of an eyewitness lineup to the traditional (non-blind) simultaneous presentation. In this research, we test for the presence of an identifiable impact on police clearance rates from the implementation of statewide reforms to eyewitness procedures. We find insufficient evidence to identify an average effect for all reform states, but do find evidence that this is the result of heterogeneous effects among the reform states. We are also able to statistically bound the possible effect on clearance rates to rule out concerns that reforms lead to large reductions in positive identifications.

CHAPTER 2

Prison Crowding and Violent Misconduct: Evidence from the California Public Safety Realignment [1]

2.1 Introduction

The issue of interpersonal violence within prisons is endemic to the administration of justice. The fundamental nature of the problem is evidenced by the fact that it can be observed throughout history and across numerous cultures, ideologies, and nations. A study in the U.K. found that roughly half of inmates reported having been both bully and victim while incarcerated (South and Wood, 2006). Meanwhile a simple web search for the world's most violent prisons will yield numerous lists that include prisons on every continent and in nations that are rich and poor, developing and industrialized, and of every religious association.[2] Given the function of prisons within a system of justice, the pervasive nature of violence is no great surprise. Nevertheless, it is generally held that violence within the prison setting is not an intended element of the sanctions imposed by the

[1] I am grateful to Richard Arnott, Steven Clark, Joseph Cummins, and Urmee Khan for constant guidance and support throughout this project. This paper has benefited from discussions with seminar participants at the Applied Economics seminar at UC Riverside, APPAM Regional student conference, the Conference of Empirical Legal Studies, Western Economics Association International annual conference, the French Economics Association annual meeting, and the UC Riverside brownbag workshops. All errors are my own.

[2] Some examples of such lists are found at www.criminaljusticedegreehub.com/most-violent-prisons-in-the-world and list25.com/25-most-brutal-prisons-in-the-world.

judiciary. This position has been affirmed on several occasions by the U.S. Supreme Court and legal scholars have further argued that violence in prison should be viewed as an infringement on the human rights of those subjected to incarceration (White, 2008). The preceding implies that a civilized society must take measures to minimize the presence of prison violence, even in the absence of direct institutional benefits from doing so.

Yet there are clear fiscal benefits to effectively managing prison violence. This is especially true in the United States, which is well known to be the most incarcerated nation in the world. Recent estimates have the U.S. spending approximately $80 billion per year on incarceration and as much as $8.2 billion[3] on prison healthcare alone (Glaze and Herberman, 2013). Although only a fraction of the latter cost is a direct result of violence, there are a host of additional costs associated with violence, from penal practices like solitary confinement to disability for injured guards and the cost of extending sentences for inmates convicted of new crimes.

Not surprisingly prison violence has drawn a great deal of attention from academic researchers seeking to identify individual and institutional characteristics that correlate with violent behavior. Prison crowding is among the most common institutional characteristics examined but is not a consistent correlate of violence (Franklin, Franklin, and Pratt, 2006). This paper is the first in such literature to both adopt a quasi-experimental design and directly estimate the relationship between prison crowding and violent behavior. The two questions posed in this research are, is there a causal relationship between crowding and prison violence? And, if so, why has previous research struggled to provide consistent evidence of such a relationship?

The quasi-experimental design used in this paper relies on a court mandated reduction in the overall level of crowding in California prisons. On May 23, 2011, California was placed under court order by the U.S. Supreme Court to reduce its prison population to 137.5% of design capacity or less within a two-year period.[4] This resulted in the enactment of new legislation that drastically reduced the flow of inmates into California correctional facilities, creating a plausibly exogenous shock to the

---

[3] Report by the PEW Trusts and MacArthur Foundation found at www.pewtrusts.org.

[4] The full text of the decision can be found at http://www.cdcr.ca.gov/News/docs/USSC-Plata-opinion09-1233.pdf

levels of crowding in California prisons. This legislation is referred to as California Public Safety Realignment or simply AB 109 (henceforth as the latter).

The exogeneity of the shock to crowding is critical to identifying a causal relationship between crowding and violence because, above and beyond the typical omitted variable concerns, there remains a plausible simultaneity problem in this setting. Suppose that administrators believe the crowdedness of a facility does lead to more violent behavior from the inmates, and therefore new inmates are directed away from facilities that have had recent issues with violence. Then more crowdedness may lead to more violence, but more violence also leads to less crowdedness. This simultaneity will depress naive estimates of the effect of crowding on violence and therefore cause a bias towards zero, assuming the true causal relationship is indeed positive.

Two estimation strategies are used to exploit the exogenous variation created by AB 109, both using monthly observations of 30 California prisons over more than four years. The first is a straightforward difference-in-differences (DD) strategy, with the added dimension that several treatments are estimated simultaneously to account for the distinct way in which AB 109 impacted different types of inmate populations. The second strategy is an instrumental variables (IV) approach that uses the time intensity of the policy's effect on crowding, in conjunction with differences in the mix of population types prior to the shock, to predict changes in crowding. Those predictions are then used in the second-stage regression to estimate the marginal effect of crowding on the rate of assaults. The two approaches provide robust, statistically significant estimates showing a strong positive relationship between crowding and violence, particularly among security level 2 populations.

It is a widely accepted doctrine among correctional practitioners, as well as philosophers on the topic, that crowding causes increased violence. Despite the research attention that has been paid to this topic, the empirical literature has thus far failed to present consistent evidence of such a relationship, providing null estimates as often as finding any positive correlation. A second contribution of this paper is descriptive evidence supporting a hypothesis that helps explain the puzzling disconnect between evidence and conventional wisdom. A summary of the hypothesis is that changes in the level of crowding in a prison are generally accompanied by changes to the composition of the population. This can confound estimates because compositional changes in individual propensities for violence are distinct from the direct effect of crowding that the researcher wishes to estimate. The presence of such a compositional effect is discussed in Section 2.4 and formally presented in a theoretical framework in Appendix A.1.

The results of the empirical estimation suggest that a ten percentage point decrease in the level of crowding (e.g. from 160% of capacity to 150% of capacity) is associated with an approximate 15% to 22% decrease in the rate of assault. To grasp the magnitude of this estimate consider that a 15% decrease in assaults for the entire state prison population would amount to 106 fewer assaults *per month* in the state of California. In a more specific example, Avenal State Prison's total inmate population fell from $5,766$ to $4,946$ between September 2011 and September 2012, and the prison was designed for less than $3,000$ inmates. That is about a 27 percentage point decrease in crowding, from 192% of capacity down to 165%. Avenal is a prison with a relatively low base rate of violence, approximately half of the statewide mean, yet this decrease in crowding is still estimated to result in a monthly decrease of 4.5 assaults.

This paper proceeds as follows: Section 2 provides background and context on the study of violence in contemporary prisons. A detailed history and description of California Public Safety Realignment is given in Section 2.3. Section 2.4 defines compositional effects and their implication for the literature and this research. Section 4.3.1 provides a description of the data that is used in this research and the challenges inherent to that data. Section 2.6 presents the two empirical strategies and the resulting estimates. A brief conclusion is presented in Section 2.7.

2.2 Prison Violence

Researchers have long sought to understand and predict the behavior of those who are incarcerated. As it stands, the body of literature studying inmate misconduct largely exists within the ambit of sociology, criminology, and psychology. This literature categorizes the determinants of misconduct into two groups: individual and institutional. Individual level covariates include both inherent characteristics – such as race, gender, and age – and historical characteristics – such as educational attainment and criminal record. Institution level covariates can include numerous environmental factors, a variety of security measures, and population density or "crowding".

A major challenge in evaluating the existing evidence in the field is inconsistency in the measurement of violence itself. Many studies use dependent variables that aggregate observations of violent misconduct with drug-related and other forms of non-violent misconduct (Goetting and Howsen, 1986; Ruback and Carr, 1993; Wooldredge et al., 2001). This implies the restrictive assumption that the marginal effect of a covariate is stable across different types of misconduct, to which other research has shown contradictory evidence (Camp et al., 2003; Steiner and Wooldredge, 2013). Steiner and Wooldredge present the best available evidence on this issue, suggesting that there are significant differences in correlations of most covariates with the different types of misconduct. Accordingly, this paper narrows the focus to physical violence or the direct threat thereof, referred to in practice as *assault* and *battery*.[5]

Understanding the primary determinants of violent behavior is foundational to improving management and enhancing safety within correctional institutions (DiIulio, 1990; Bottoms, 1999). This is true at every level of management and decision making in the correctional system, from the daily choices of correctional officers in the prison yard up to the strategic planning and policy decisions of wardens and legislators. Even the architectural design of correctional facilities has been associated with some forms of misconduct (Morris and Worrall, 2014).

The importance of identifying causal relationships with violence is dependent upon the motivation for studying violent behavior. We can assume three relevant motivations for the study of prison violence, one a purely academic motive and the others policy oriented. The first, an academic desire to better understand violent behavior, would imply an interest in the causal effect of each

[5] The legal definition of *assault* is the credible threat of bodily harm to another. The legal definition of *battery* is to cause bodily harm to another.

potential covariate, individual or institutional, as well as any interactions between them. The second motive is to accurately identify high-risk individuals, for which it is sufficient to simply identify correlations between individual characteristics and the likelihood of violent misconduct.[6] The last motive is to better assess the costs and benefits of policy and/or management options, which demands identification of the causal effect of institutional characteristics. To base such policy decisions on simple correlations, without consideration for causality, poses a risk of unexpected and potentially adverse outcomes. Hence in studying the relationship between crowding and violence, causality plays an important role in the value of the research.

This research aims to enhance existing evidence of a causal link between crowding and violent misconduct. Existing research often lacks any form of quasi-experiment or other source of exogeneity by which to make a claim for causality. Some studies estimate multivariate regressions, with cross-sectional or panel data, and show that crowding is associated with higher rates of violence (Megargee, 1977; Gaes and McGuire, 1985). Yet other research contradicts that conclusion, such as a 2003 study by Camp and several coauthors. Despite being one of very few studies that use individual inmate data, they do not find consistent evidence of a correlation between crowding and violence (Camp et al., 2003).

Of the research that does exploit some sort of quasi-experimental design, the sources of exogeneity that are used do not apply to the relationship between crowding and violence. For example, Chen and Shapiro use regression discontinuity design to exploit the mechanism by which inmates are assigned to different security levels, granting "as good as random" variation with respect to security classification but not crowding (Chen and Shapiro, 2007). Another study has inmates with the same security classification and randomly selects a portion of them to serve their sentence in a lower security level facility, then examines the implications for misconduct (Camp and Gaes, 2005). As with the previous example, Camp and Gaes use random variation with respect to security level

[6] In identifying "high risk" inmates, there are notable social justice and equity concerns to be raised if individuals are ascribed differential treatment based on inherent characteristics. Discussion of these concerns is omitted since that motivation does not apply to this research.

rather than crowding. Additionally, Camp and Gaes note that the correlation between crowding and violence is not statistically significant in their research design.

In recent theoretical research on the topic, Blevins et al. (2010) propose a theory synthesizing the several disparate hypotheses of what determines violent behavior in prisons. They acknowledge crowding as the most common "noxious stimuli" repeatedly linked to prison assaults and overall misconduct. The precept that crowding leads to increased violence is also the prevailing view among practitioners in the field. However, these views seem at odds with the fact that empirical evidence of such a link is wholly unconvincing. This puzzle motivates the main objective of this research, to provide evidence of a causal effect of prison crowding on violent behavior, if such a link does in fact exist. It equally motivates the secondary objective, which is to help explain why it is so challenging to identify a consistent relationship in the data.

2.3 California Public Safety Realignment

On May 23rd, 2011, the United States Supreme Court upheld a lower court ruling (Brown v. Plata) which had determined that the level of overcrowding in California prisons was so severe that the Eighth Amendment rights of prisoners (freedom from "cruel and unusual punishment") were being systematically violated. The Court ordered California to reduce its prison population to 137.5% of design capacity on or before June 27th, 2013. Given the prison population at the time of this order, the required reduction amounted to approximately one quarter of the existing California prison population. According to the California Department of Corrections and Rehabilitation (CDCR) population reports[7] the total prison population in California in January 2011 was over 156,000. This number fell to 135,000 in January 2012, and further to 124,000 in January 2013. Although the CDCR did not quite achieve the full reduction demanded by the courts,[8] the total prison population declined more swiftly and significantly than any large U.S. prison population has in recent history.

[7] Available at www.cdcr.ca.gov/Reports_Research/Offender_Information_Services_Branch/Population_Reports.

[8] The target reduction was later achieved after the implementation of Proposition 47, which changed sentences for a set of minor drug and theft offenses from felony to misdemeanor

The response to the court mandate came in the form of a new law, AB 109, implementing dramatic sentencing reform. A rich literature has already arisen around this law, including a panoply of policy papers and numerous research articles. Of the research undertaken, authors examine the effect of AB 109 on various outcomes. Among these are the effect on county jail populations, crime rates, recidivism, and whether the policy achieved its goals (Lofstrom and Raphael, 2013, 2015; Lofstrom et al., 2014; Petersilia and Snyder, 2013; Sundt et al., 2016).

Because the law that resulted from the court mandate was strict sentencing reform, it is important to first understand the process by which individuals enter the California prison system. There are two possible channels by which a new admit is referred to a California prison – either through the courts following conviction or through the parole system for violation of the parolee's conditions of parole. In either case, the new inmate is first sent to one of the select prisons known as reception centers, of which there were nine in 2011. Inmates are typically held for several months at the reception center awaiting a classification hearing. This hearing determines which security classification the inmate should be assigned to, after which the same committee selects a suitable prison for the inmate to serve the remainder of their sentence. The inmate is then transferred to the assigned prison, where they are housed with inmates of the same security classification.

Within a prison, prisoners of one security classification do not generally interact with prisoners of other security classifications. Each California prison has several different facilities within it and these are designed to operate independent of one another. Thus a prisoner that is classified as security level 2 will be held in a level 2 facility, which is completely separate from the housing and recreational areas of other security levels. Reception center populations and special needs populations[9] are also held in separate facilities. This segregation within prisons creates the possibility for greater levels of crowding in some areas than in others.

[9] "Special needs" in the CDCR is what would commonly be known as protective custody.

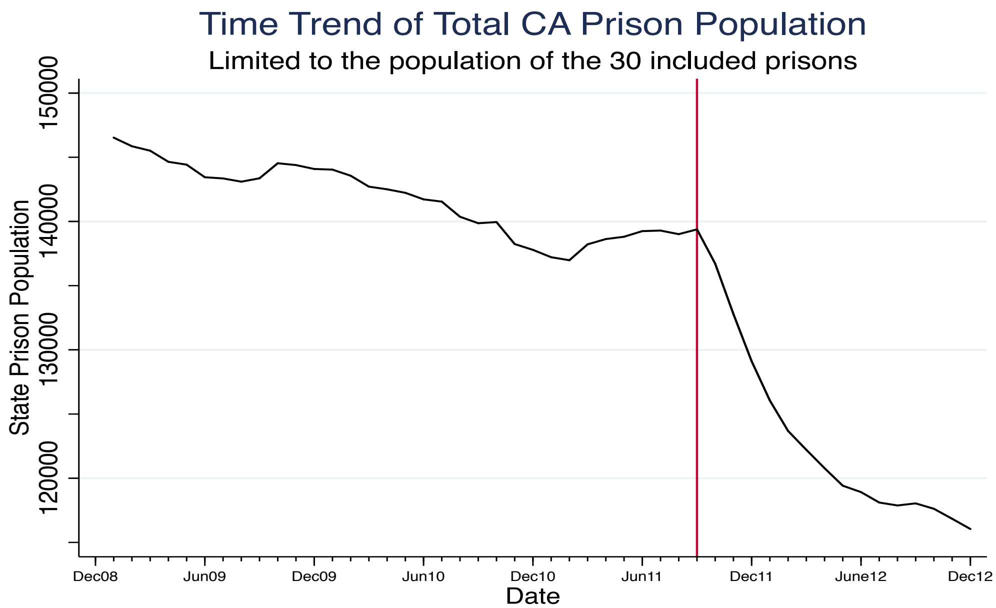


Figure 2.1: *CA Prison Population Time Series. Total population of male prisoners incarcerated in the state of California, observed monthly. The vertical line denotes the last observation prior to implementation of AB 109. Source: Generated from CDCR CompStat reporting data.*

2.3.1 AB 109

California achieved the massive reduction in population mentioned above by introducing Assembly Bill 109 (AB 109) and signing it into law. The new law took effect on October 1st, 2011, making two major changes to how inmates are handled in California. First, technical parole violators who were previously taken back into state custody are thereafter sent to county jails, with a few exceptions for very serious violent and sexual offenders. This is a shift from the aforementioned practice wherein such parole violators were remitted to a nearby state prison for a term of up to 12 months. The second major element of the reform defined a set of non-violent, non-sexual, non-serious felony offenses for which the sentences were to be served in county jails rather than state prisons.

AB 109 did not commute any existing sentences and no prisoners held in state prison prior to implementation of the law were transferred to county jails; the law only changed who would take custody of new admissions. Despite the intervention being a change in flow rather than stock, the impact on the prison population was rapid and distinct. This is depicted in Figure 2.1. Essentially, California prisons had a high rate of "churn" in their populations and the new law caused a sharp decrease at the front end of that churn without immediately affecting the back end. Hence the true nature of the shock is that the "vacancies" left by released prisoners are no longer being filled at the same rate as they were prior to the new law.

Note that the shock created by AB 109 was selective by nature, designed to target non-violent offenders[10] for diversion. The population decrease can therefore be expected to concentrate among two subpopulations. The first, rather obvious given that they handle all new admissions, is the entirety of the shock is channeled through the relatively few reception center facilities in the state. Parole violations made up a significant proportion of admissions prior to AB 109, so the dramatic impact on reception centers was predictable. After passing through reception centers the residual impact is concentrated among security level 2 populations. This is because the point system by which inmates are assigned a security classification makes it unlikely to begin at the lowest classification (level 1) and also unlikely that any inmate will accumulate enough points for level 3 unless they have a long criminal record or are convicted of a very serious felony. Given that the law targets lower level felonies, it follows that we should expect to see the secondary effects focused among the security level 2 populations of California prisons. Rather than a liability, this selective compositional feature of the shock provides a measure of variation that is important to the empirical strategies that follow. This will be discussed further in the data and empirical sections.

Figures 2.2 and 2.3 show the time trends of the different subpopulations for the entire state. It is apparent that the largest decreases are indeed among the reception and security level 2 populations depicted in Figure 2.2. More specifically, the effect on the reception center populations is immediate, very sharp, and appears to stabilize again rather quickly. The shock to the level 2 population lags by one month, not surprisingly, then there is a sharp decline and it takes longer (into the eighth or

[10] Parole violators may or may not have been originally convicted of a violent offense, however Orrick and Morris (2015) show that technical parole violators are significantly less likely to engage in misconduct than inmates admitted to prison for new offenses.

ninth month) to stabilize at what appears to be a new equilibrium. Figure 2.3 clearly demonstrates that the trends of each of the other subpopulations are not sensitive to the implementation of the new law. In total, averaging over the six months prior to implementation and the first six months of 2012, the reception center population falls by about 45% and the level 2 population falls by a more modest 12%. There are also some moderate decreases in other subpopulations, but none exceed 6% and these can just as likely be attributed to long term downward trends.

2.3.2 Reception Adjustment

It was not unexpected that AB 109 would greatly reduce reception center populations. All else constant, this would have resulted in prisons with reception facilities that were suddenly near or below their design capacity while other facilities, both at other prisons and within the same prison, remained severely overcrowded. The Reception Adjustment[11] (RA) was a reclassification of certain reception facilities to serve an alternate subpopulation. This did not mean that any given prison was no longer a reception center. Each California prison has between three and nine individually defined facilities within its organizational structure and most reception centers had two or three of these dedicated to reception populations. So when RA occurs reception populations are simply consolidated into fewer facilities.

RA complicates the impact of AB 109 in two ways. First, it makes the true decrease in crowding for reception center populations less dramatic than implied by the statewide population reduction. Although the statewide reception population is approximately halved by AB 109, to a smaller degree RA also decreases the design capacity dedicated to that population. Therefore RA diminishes the degree of the AB 109 shock to crowding in reception populations. Simultaneously, it decreases crowding for some other population by increasing the total design capacity devoted to them in the state. In nearly every case the repurposed facilities were filled with level 3 inmates. Although some of the new inmates were likely transferred from within the same prison, the overall transfers

[11] "Reception Adjustment" is the author's term and does not represent official CDCR language.

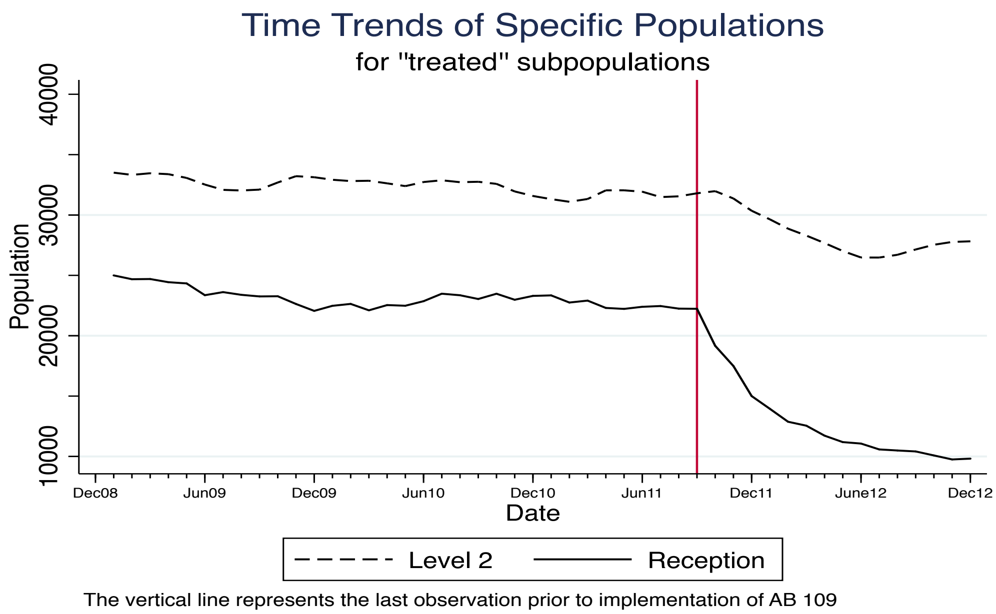


Figure 2.2: *Time Trends of Specific Populations – "Treated". Trends for the two subpopulations among which the shock from AB 109 was concentrated, summed across all 30 California prisons included in this study. The vertical line denotes the last observation prior to implementation of AB 109. Source: Generated from CDCR CompStat reporting data.*

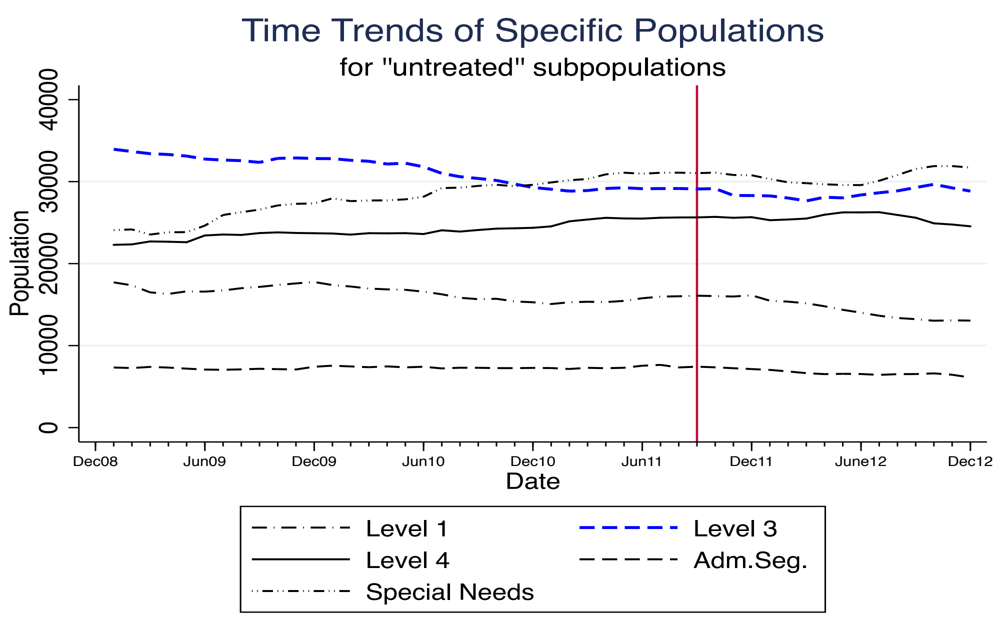


Figure 2.3: *Time Trends of Specific Populations – "Untreated". Trends for the major subpopulations which were less impacted by AB 109, summed across all 30 California prisons included in this study. The vertical line denotes the last observation prior to implementation of AB 109. Source: Generated from CDCR CompStat reporting data.*

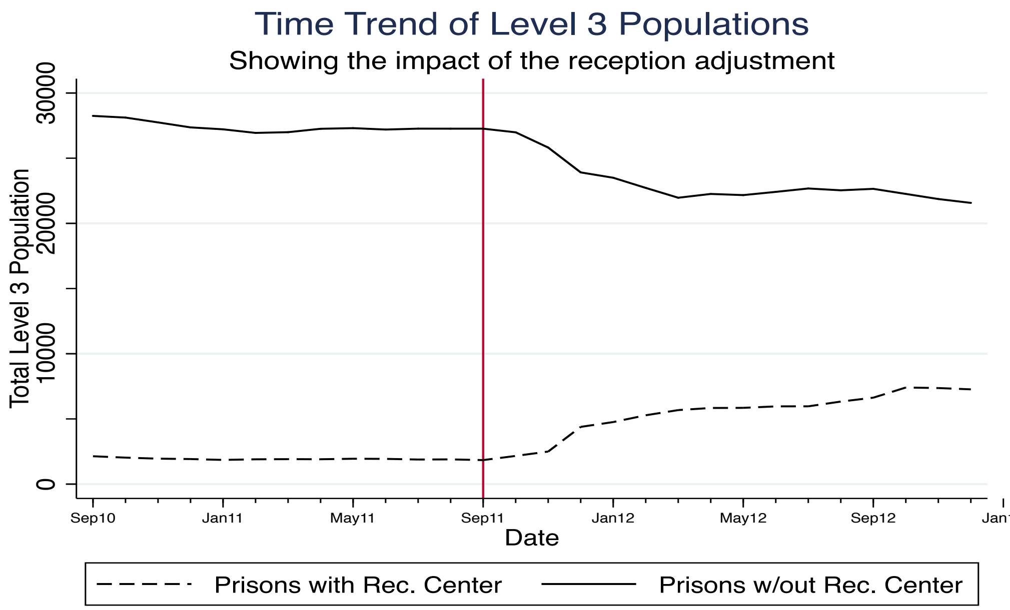


Figure 2.4: *Time Series Showing Reception Adjustment. Trends highlighting the impact of the Reception Adjustment, which closely followed implementation of AB 109. The time trends are for sum of security level 3 populations parsed by whether the prison has a reception center facility or not. The vertical line denotes the last observation prior to implementation of AB 109. Source: Generated from CDCR CompStat reporting data.*

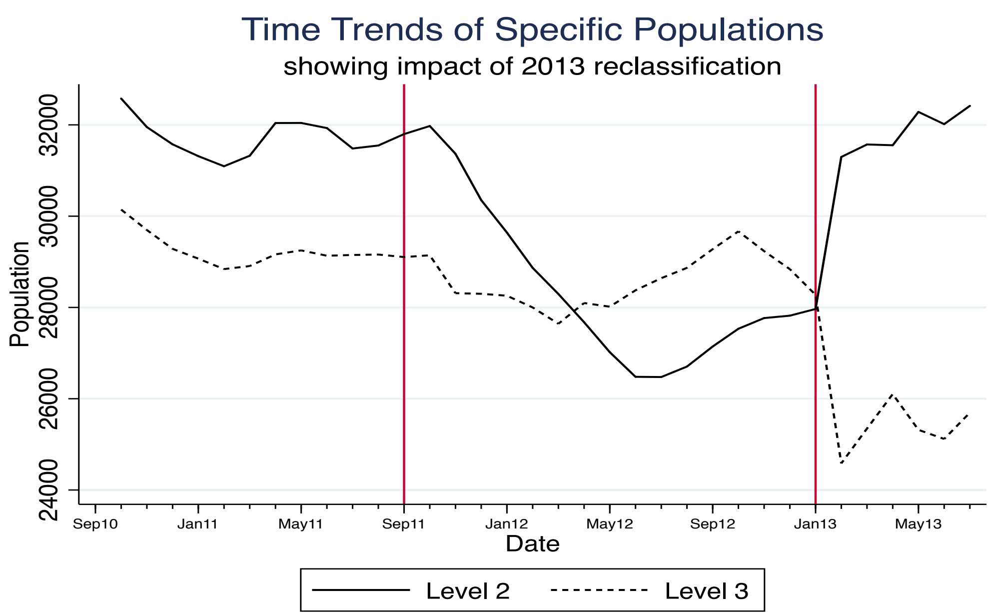


Figure 2.5: *Time Series Showing Reclassification. Population Level 2 and level 3 population trends through implementation of both AB 109 and the new classification system. The vertical lines denote the last observation prior to implementation of AB 109 and January 2013. Source: Generated from CDCR CompStat reporting data.*

were statewide.[12] So while AB 109 does not significantly decrease the statewide level 3 population, evidenced in Figure 2.3, it does result in decreased levels of crowding for that population.

The evidence of decreased crowding for level 3 populations is given in Figure 2.4. Looking only at the number of level 3 inmates in reception centers, the dashed line shows that there were very few prior to AB 109 and that number begins to climb shortly after implementation of the law. On the other hand, the total level 3 population in all other facilities, the solid line, decreases in near perfect synchronicity with the increases at reception centers. The percent decrease in non-reception center level 3 population is of a magnitude similar to the overall impact of AB 109 on the statewide level 2 population. Similar figures are provided in Appendix A.2 for each of the other subpopulations, showing that none have a distinct shift like the one seen here for the level 3 population.

2.3.3 Reclassification

This time period of the following analysis is truncated prior to January 2013 because of another policy adjustment that was implemented around that time. Likely in response to the statewide decrease in level 2 inmates following AB 109, the CDCR commissioned a review of their classification system. This review and the resulting adjustment in the classification system changed the point thresholds for certain security classifications. These changes led to a shift of inmates from level 3 classification to level 2 classification. Figure 2.5 shows the time trends of statewide level 2 and level 3 populations into 2013 and the impact of this policy change is very evident. It is not clear in the available data what portion of these inmates were transferred between facilities or if some facilities were repurposed (similar to the RA repurposing) to accommodate the greater number of level 2 inmates.

Although this second policy shock provides opportunity to explore some interesting research questions about composition and crowding, the variation is quite different from that induced by the AB 109 shock. As such the current research is limited to the months preceding the reclassification.

[12] Many reception centers did not have any level 3 population prior to RA, so some degree of statewide rebalancing is a given.

2.4 Compositional Change: A Theoretical and Empirical Oversight

In any conception of the sequence of events that leads to a violent assault, individually rational behavior requires that the decisions of each participant be informed by the characteristics and choices of the other involved parties. As the saying goes, it takes two to "tango". Given any degree of heterogeneity among inmates, this implies that changes in the composition of a prison's population can influence the rate of violence in that prison. Since any change to the level of crowding is likely to be accompanied by a change in the composition of the prison population,[13] it is a potentially critical oversight to not consider the effect of compositional change in any study of prison crowding and violence.

This section provides a description of three simultaneous channels, or mechanisms, by which prison crowdedness may effect the rate of violence in a prison and then summarizes the implications for this paper and other empirical work in this area. The three mechanisms represent broad conceptual channels meant to capture all the reasonable means by which crowding may be correlated with violent behavior. Appendix A.1 develops a model that formalizes the three mechanisms and presents relevant derivations.

1. ***Structural Mechanism:*** Increased crowding leads to increasingly limited personal space and more individuals sharing a fixed set of available resources – such as basketball courts, payphones, and restroom facilities – which in turn leads to a higher frequency of potentially contentious encounters between inmates. Therefore increased crowding can cause each individual to experience a greater number of potentially violent confrontations with other inmates in a given time period. This mechanism is likely to be reflected by a positive association between crowding and violence.

[13] Increasing or decreasing crowding is most often done by increasing or decreasing the population size, which can be expected to change the composition of the population unless the selection by which changes are made is as good as random. Such policy or administrative changes tend to be distinctly non-random, as in the case of AB 109 which specifically targeted non-violent criminals.

2. ***Behavioral Mechanism:*** Crowding exacerbates the individual's lack of access to personal freedoms, amenities, and basic necessities, potentially leading to a behavioral increase in an individual's willingness to resort to violent behavior. This, in turn, increases the likelihood that any given contentious interaction between inmates becomes violent.[14] Sleep is a classic example of a basic human need whose deprivation has been shown to increase aggressive behavior (Kamphuis et al., 2012). This mechanism is also likely to be reflected by a positive association between crowding and violence.

3. ***Compositional Mechanism:*** The manner in which a particular change in the level of crowding changes the composition of the inmate population, especially with respect to individual propensities for violent behavior. Depending on the specific policy design, this can result in a population that is either more or less prone to violent behavior, on average. This mechanism could therefore result in either a positive or negative association between crowding and violence.

The first two mechanisms can be thought of as "pure" crowding mechanisms, representing direct effects of crowding itself. The compositional mechanism is distinct in that it occurs as a result of correlation between changes in crowding and population composition, where the actual impact on the rate of violence is derived from the compositional change rather than the change in crowding. In addition, the nature of compositional change, and thus the resulting effect of the compositional mechanism on violence, is dependent upon the case-specific process of selection by which the population is increased or decreased.

In the case of AB 109, as well as nearly any other policy intervention meant to decrease crowding, the new law decreases crowding by reducing the portion of the population least likely to be prone to violent behavior. After implementation of the law, offenders convicted of non-violent, non-serious, and non-sexual felonies are no longer sentenced to serve time in state prisons; this effectively redirected many of the least violent new admissions away from California prisons.[15] Thus

[14] This mechanism may be viewed as an emotional response wherein the individual becomes more irritable or unstable, or as a rational response where the individual recognizes a lower opportunity cost associated with violent behavior. The rational response assumes that violent behavior results in punishment with some positive probability and the opportunity cost is then the expected utility of avoiding the implied sanctions. That expected utility is reduced if prison life without sanctions grants access to fewer resources or less freedom due to crowding.

[15] Research shows that violent crime and past criminal records are strong predictors of future violent misconduct (Walters and Crawford, 2013).

the resulting distribution prisoners is expected to be more prone to violent behavior, on average. So while the structural and behavioral mechanisms may decrease the rate of violence after AB 109, the compositional mechanism is expected to do the opposite, increasing the rate of violence.

The three mechanisms can be conceptually contrasted with the example of a prisoner using the payphone in the prison yard. Consider a relatively well-behaved inmate, Richard, at a maximum-security prison, who calls his mother from the prison payphone every Sunday. AB 109 takes effect and the prison population decreases. As the prison becomes less crowded, it becomes less likely that another inmate will seek to use the phone while Richard is using it, harassing him to cut his call short. This is an example of the structural mechanism since such an interaction could have resulted in violence. At the same time, Richard gets more sleep now that the prison has reverted from triple to double occupancy in each cell. He is therefore marginally more patient and less prone to losing his temper. This constitutes an element of the behavioral mechanism. On the other hand, the diminished population in the prison is due to fewer non-violent offenders who are less likely to resort to violence than certain other types of offenders. As a result when Richard is harassed it is more likely to be by an inmate who is aggressive and prone to violence than prior to AB 109. This is the effect of the compositional mechanism.

The implications of these mechanisms for the empirical study presented in this paper can be understood through Equation 2.1. The elasticities in this equation are derived from an identity (Equation A.1 in the appendix) stating that the total violence in a prison will equal the average probability that any pairwise interaction results in violence multiplied by the total number of such contentious interactions in the prison per unit time. The full model and derivation is available in Appendix A.1.

The term on the left side of Equation 2.1 is the elasticity of aggregate violence, V, with respect to a policy shift parameter, λ. The shift parameter is defined such that prison crowding, c, exhibits unit elasticity with respect to it (a policy that *decreases* crowding by 10% is represented by a 10% decrease in λ). The first term on the right side of the equation is the elasticity of violence with respect to crowding. This term encompasses both of the "pure" crowding mechanisms from above – structural and behavioral – and represents the impact of crowding that researchers typically seek to estimate. Yet, without any direct measurement of compositional changes, the latter elasticity is inevitably confounded with the "compositional elasticity" represented by the final term in the equation. This last elasticity measures how the probability of any particular encounter becoming violent, π, responds to the policy represented by λ. $E_{\pi:\lambda} < 0$ implies that as the policy increases crowding the population becomes less prone to violence or conversely if the policy is decreasing crowding ($\Delta\lambda < 0$) then the remaining population, on average, becomes more prone to violence. As explained previously, the latter is exactly what is expected in the case of AB 109.

Derived Elasticity of Aggregate Prison Violence

$$E_{V:\lambda} = E_{V:c} + E_{\pi:\lambda}. \tag{2.1}$$

Equation 2.1 provides a mathematical representation of what is proposed in the earlier description of the mechanisms. Assuming there is some validity to the structural or behavioral mechanisms, it is the case that $E_{V:c} > 1$ and therefore the rate of violence increases with the level of crowding. However, when empirical estimates actually represent $E_{V:\lambda}$ then the desired $E_{V:c}$ is conflated with $E_{\pi:\lambda}$. Although the sign of $E_{\pi:\lambda}$ depends on policy design, there are many cases where the obvious expectation is $E_{\pi:\lambda} < 0$ and this implies a downward bias in estimates meant to capture the $E_{V:c}$ relationship.

Unfortunately, California prison data on individual inmate misconduct is not currently available. Nor is there data measuring capacity or rates of misconduct for the individual facilities within each prison. This means that it is not feasible with existing data to separately identify the two crowding mechanisms of this model nor to precisely isolate the compositional mechanism. Therefore the empirical strategies used in this paper do not attempt to directly identify the compositional effect of AB 109. Instead, because the AB 109 population shock is concentrated among three distinct subpopulations and at least two of those interact in predictably different ways with the compositional element of the policy, the potential for a compositional effect is used to analyze between-group differences in the point estimates. Reflectively, the differences in point estimates can also be taken as supportive evidence of an existing compositional mechanism. In addition, the potential presence of the compositional mechanism contributes to the empirical analysis the implication that the estimated coefficients are in fact lower bounds on what would be considered the true effect of crowding on violence.

2.5 CompStat Data

Panel data used for this research has been taken from the California Department of Corrections and Rehabilitation (CDCR) CompStat Reports. These contain monthly observations for each adult correctional facility in the state of California. There are 35 currently operational prisons in California, five of which are excluded from the analysis. The excluded prisons are either medical facilities, female detention centers, or were not operational at the time of the policy shock. The time period of analysis is limited to July 2008 through December 2012, due to irregularities in the earlier data and the January 2013 reclassification that was discussed in section 2.3.3. Given that AB 109 was implemented at the beginning of October 2011, the data in use spans three years prior to and fifteen months post implementation of the law.

The variable of interest is the level of overcrowding in each prison, which is constructed in this analysis from two variables in the CompStat data. *Total population* is a simple count of the total number of inmates held in a prison in a given month. *Design beds* is reportedly the number of beds that a given facility was designed to hold. The institutional definition of *design beds* for a facility is a single bunk per cell and single level bunks in dorm housing.[16] However, there is minor month-to-month variation in the CompStat measure of *design beds*. This variation is not consistent

[16] This definition of capacity could be seen as reasonably conservative, which may explain why the target given in the court mandate was for the CDCR to reduce overcrowding to only 137.5% of design capacity.

with the idea that prisons have a fixed capacity – notwithstanding new construction or demolition, which would not be characterized by such frequent and minor changes. Therefore *design beds* for each prison is averaged over the six months preceding implementation of AB 109 and this is taken as the *fixed design capacity* of each prison. Table 2.1 shows that these two measure of capacity are quite similar, *fixed design capacity* having only slightly less variation. Prison crowding ($crowd_{it}$) is defined in this research as the ratio of *total population* to *fixed design capacity*[17].

The CompStat data includes a number of measures of misconduct, broadly defined as either disciplinaries or incidents. Disciplinaries are individual reports of misconduct for each prisoner, which are included in their personal files. There are several different types of disciplinary, ranging from simple conduct or cell phone possession to assault and battery or murder. Incidents are recorded in slightly more detail (such as the type of drug confiscated) but with less frequency, suggesting that a disciplinary can be issued without having to write a full incident report or that each incident may involve an unspecified number of individuals.

The dependent variable used for this research is the monthly sum of disciplinary reports for assault on an inmate, assault on staff, battery on an inmate, and battery on staff.[18] The inclusion of murder and attempted murder in this sum does not noticeably impact the estimates reported below, which is expected since such attacks are relatively infrequent. Although assault and battery are measured separately in the more recent CompStat reports, they were not in the earlier years and are therefore summed into a single variable, $assaults$, in the statistics below.

In Table 2.1 assault statistics are given for all assaults and then categorized by whether the victim was a staff member or an inmate. Assaults on inmates are the most frequent, averaging more than 18 per month in each prison; however, at just below five per month, assaults on staff members are also quite common. The other categories of violent misconduct occur with less frequency and

[17] This measure of prison crowding differs slightly from the official measures used by the CDCR, likely due to the manner in which *fixed design capacity* is constructed. The measure defined here does closely track the official measure and has been determined to be the most appropriate for this research given the timing of the policy shock being studied.

[18] As defined previously, the legal distinction between an assault and a battery rests in the credible threat of bodily harm versus actually causing bodily harm.

Table 2.1: Simple Summary Statistics

	count	mean	sd	min	max
Measures of Violence					
Rate of Assaults	1620	.537083	.3496848	0	2.777778
Total Assaults	1620	23.45247	14.4966	0	133
Assaults on Inmates	1620	18.72531	12.75872	0	126
Assaults on Staff	1620	4.72716	4.829611	0	58
Murders and Attempts	1620	.5197531	2.433345	0	67
Rioters	900	10.43556	28.4308	0	466
Resisting Staff	900	3.455556	5.287425	0	49
Possessions of a Weapon	899	5.292547	5.950058	0	46
Population & Crowding					
Crowding (P/K)	1620	1.872671	.2652445	1.155107	2.44784
Total Population (P)	1620	4566.968	1084.344	2212	7179
Design Beds	1620	2451.812	612.4285	1234	3880
Fixed Design Capacity (K)	1620	2466.111	608.9742	1557	3789.333
Population Shares					
Security Level 1	1620	.1094138	.1547788	0	.685277
Security Level 2	1620	.2131347	.2962029	0	.9995067
Security Level 3	1620	.2329096	.2615549	0	.9088176
Security Level 4	1620	.1967216	.2658214	0	.8865633
Reception Center	1620	.1383862	.2605312	0	.9434235
Special Needs	1620	.2042753	.2238708	0	.9279493
Admin. Segregation	1620	.0543071	.0263553	0	.1202622
ADA Inmates	1620	.0637266	.0553348	.0031173	.4041008
CCCMS Inmates	1620	.1969473	.1090689	0	.4756164
Single Inmate Cells	1620	.0473859	.0853995	0	.4496124
Program Enrollment					
Prison Industries	1620	148.5117	162.1272	0	616
Academic	1620	413.923	306.4613	0	1687
Non-PIA Work	1620	2251.761	1171.362	0	5591
Subst. Abuse	1620	103.7352	229.5731	0	1818
Subst. Abuse Waitlist	1620	56.23951	103.6996	0	603
Observations	1620				

The dependent and key explanatory variables are indicated in italics.

most were not reported in the early years of CompStat, which is noticeable in Table 2.1 by the fewer numbers of observations for these types of misconduct. Assaults were selected as the most appropriate category of violence for this study because, in addition to being a reasonable area in which to expect responsiveness to crowded conditions, their high frequency indicates they are the most present threat to the safety of inmates and staff. Further, assaults are a distinct and well-defined form of misconduct, making their measure less susceptible to misreporting by the prison staff.

The data also include population counts for each prison subpopulation, which are used to generate population shares of each. There are six major subpopulations: the four levels of security classification, special needs, and reception center populations. Each of these populations reside in separate facilities. Administrative segregation units constitute an additional type of facility, but these are smaller and serve a more temporary purpose. The security classifications are ranked in ascending level of security threat, one to four. Security level 1 prisoners are eligible for housing units that are not within the prison fences and are also permitted to have prison jobs that allow relatively free movement within the facilities. Security level 2 and 3 require much closer supervision and are not permitted to be outside secure areas, the major difference between the two levels typically being dormitory housing vs. cells. Security level 4 is reserved for the most disruptive and violent prisoners, although a sufficiently heinous crime can result in this classification without any record of institutional misconduct. The special needs and administrative segregation populations are held apart from the rest, each for their own reasons: one for long-term protective custody and the other a temporary punitive or safety measure, respectively. There are several other subpopulations that are not housed exclusively, such as Americans with Disabilities (ADA) inmates, single-cell inmates, and the Correctional Clinical Case Management Services (CCCMS) population. The shares of these are included as controls in the empirical strategies. Finally, the CompStat data includes counts of inmates with life sentences with and without possibility of parole. Unfortunately, there is a significant amount of missing data for these two variables so they have been excluded from the analyses.

All of these data are observed at the prison level. However, AB 109 affects crowding, and thus misconduct, at the facility level. Therefore observed rates of assault are averages across the facilities within a prison, which makes the population shares of each subpopulation both necessary to the identification strategy and critical as direct controls for the effect of changing shares. The latter is due to an observable element of compositional change that occurs across the whole prison. Suppose a prison has a level 2 facility and a level 4 facility, each of equal capacity and population size. The AB 109 shock reduces the level 2 population but not the level 4 population. Then, above and beyond any differences because of reduced crowding, the rate of violence in this prison will have increased due to the fact that the population of the level 4 facility comprises a greater share of the total population and level 4 facilities have much higher rates of violence, on average, than level 2 facilities. Thus without controlling for population shares, the implementation of AB 109 would be correlated with changes in rates of violence that were not due to the effect of diminished crowding.

There are no demographic controls available in these data (such as age and race statistics for the prison population) but there is fairly detailed information on program participation in academic, vocational, work, and drug rehabilitation programs. Tables 2.1 and 2.2 include variables for the degree of enrollment in these programs. Prison Industries (PIA) is a work program that produces marketable products, with higher skill positions that pay relatively high wages. Non-PIA work positions are the more common prison positions such as maintenance and food service. Each of these variables is a simple count of enrollment, except for the substance abuse program for which both enrollment and waitlist are included.

Table 2.2 divides the summary statistics into two periods: pre- and post-implementation of AB 109. There was a slight downward trend in total population over the full time period, so the difference in the population means reported in the table slightly exaggerates the actual policy impact on population and crowding. Table 2.2 illustrates that there were moderate decreases in all measures of assault and objectively larger decreases in *total population* and *crowding*. One notable point in this table is the decrease in *design beds*, minor though it is. This decrease implies that if

Table 2.2: Pre and Post Statistics

	Pre	Post
Measures of Violence		
Rate of Assaults	0.55	0.50
	(0.35)	(0.36)
Total Assaults	25.00	19.43
	(14.62)	(13.36)
Assaults on Inmates	19.95	15.53
	(12.91)	(11.78)
Assaults on Staff	5.05	3.90
	(4.83)	(4.74)
Population & Crowding		
Crowding (P/K)	1.95	1.67
	(0.24)	(0.20)
Total Population (P)	4755.79	4076.02
	(1070.81)	(959.11)
Design Beds	2480.08	2378.33
	(613.68)	(603.67)
Subpopulations		
Level 1	556.89	481.57
	(846.82)	(821.05)
Level 2	1090.50	944.70
	(1537.78)	(1332.32)
Level 3	1059.96	952.60
	(1180.91)	(938.93)
Level 4	792.95	852.42
	(1076.58)	(1109.62)
Reception	778.33	413.69
	(1403.34)	(1017.92)
Special Needs	915.41	1021.08
	(1033.05)	(1039.39)
Program Enrollment		
Prison Industries (PIA)	154.38	133.25
	(165.36)	(152.53)
Academic	436.19	356.03
	(334.99)	(204.55)
Non-PIA Work	2344.31	2011.14
	(1219.25)	(998.36)
Subst. Abuse	128.54	39.24
	(263.25)	(61.89)
Subst. Abuse Waitlist	51.91	67.62
	(94.22)	(124.48)
Observations	1170	450

Means reported. Standard deviations are in parentheses.

one believes *design beds* to be a more appropriate measure of capacity than the fixed measure used in this research, then *crowding* as currently defined actually exaggerates the true impact of AB 109 on California prison crowding. Exaggerating the impact on crowding in this way is of little concern since it would only result in attenuation bias, which means the estimated coefficients are smaller and less significant than they would be otherwise.

Table 2.2 also highlights the dramatic decrease in reception center populations following AB 109. Excessive focus on this impact to reception populations is discouraged, since the effect on crowding was significantly countered by the reception adjustment. An extrapolation based on the observed decrease in reception populations and increase in level 3 populations at reception centers suggests the true decrease in crowding at reception facilities was approximately 40 percentage points, only slightly greater than the decrease at level 2 facilities. This approximation is further supported by the population means presented by Table 2.4 in the next section.

2.6 Empirical Strategies and Results

The simplest and most common approach to estimating the effect of a policy shock such as AB 109 is a difference-in-differences (DD) strategy. This section begins with a DD approach and then develops a more sophisticated instrumental variables (IV) identification strategy. The IV strategy builds on the same source of variation as the DD strategy by better capturing the varying time intensity of treatment across the treated populations. The dependent variable, Y_{it}, in both the DD and IV strategies is the natural log of the rate of assaults in prison i during month t.[19] Results in Appendix A.2 show there is no meaningful change when the dependent variable is altered to include other violence, such as murder and attempted murder, or to exclude assaults and batteries on staff members. The log-linear form[20] is an intuitive way of incorporating the expectation of a nonlinear

[19] In this literature, rates of violence are measured per 100 inmates.

[20] The log-linear form transforms the dependent variable so that marginal changes are approximations of the percent change in the original variable. Thus a 0.1 increase in $ln(Y_{it})$ is an approximate 10% increase in Y_{it}.

relationship between crowding and violence. Such expected nonlinearities are a natural conclusion of the frequent conjecture that "violence begets violence".

The log-linear form, with the $crowd_{it}$ variable constructed as it is, makes the coefficients reported below semi-elasticities. These are interpreted as the *percent* change in the rate of assaults associated with a *percentage point* change in crowding. For example, the 0.6 point estimate in column 3 of Table 2.3 asserts that a *ten percentage point* increase in crowding is associated with a *six percent* increase in the rate of assaults. The percentage point changes in crowding are with respect to the percent by which the population exceeds the prison's design capacity.

Table 2.3 provides estimates from a basic ordinary least squares (OLS) model with prison fixed effects. These estimates give an idea of the baseline correlation between crowding and violence in these data, absent the quasi-experimental design used in later estimates. The model is estimated using a flexible time trend or time fixed effects and the final two columns restrict the time period to only include observations prior to implementation of AB 109. The coefficients on crowding are only marginally significant if at all and suggest a semi-elasticity of approximately $0.5 - 0.6$. The weak statistical significance of these correlations is consistent with the overall state of existing empirical evidence on this topic, as discussed in Section 2.2.

The coefficients for the shares of major subpopulations are also reported in Table 2.3. These coefficients preview the importance of the share of security level 2 population and its dependence upon the variation that occurs as a result of AB 109. Note that because the shares of these subpopulations are generally very stable over time, most of the correlation between assaults and each share is absorbed by the prison fixed effect. Table A.1 in Appendix A.2 shows the raw correlations between each population share and the rate of assaults. The coefficients in Table 2.3, on the other hand, reflect the effects, or lack thereof, of time variation in the shares. However, AB 109 is responsible for the bulk of such variation and, among the treated populations, the effect of this is only apparent in the level 2 coefficient. Furthermore, even the effect for level 2 share dissipates entirely in columns 3 and 4 when the months following AB 109 are excluded from the regressions. This suggests that

Table 2.3: OLS Regression with Prison Fixed Effects

Dependent Variable: Log Rate of Assaults				
	(1)	(2)	(3)	(4)
VARIABLES	Trend	TimeFE	Trend	TimeFE
Crowding (P/K)	0.624	0.483	0.601	0.621
	(0.306)	(0.342)	(0.443)	(0.470)
Security Level 2	-1.527	-1.724	-0.0466	-0.0920
	(0.784)	(0.776)	(1.112)	(1.137)
Security Level 3	-0.0361	-0.329	0.119	0.103
	(1.021)	(0.989)	(1.247)	(1.235)
Security Level 4	-0.101	-0.442	0.113	0.0402
	(1.262)	(1.218)	(1.296)	(1.263)
Reception Center	-0.131	-0.377	0.338	0.168
	(0.797)	(0.792)	(0.926)	(0.927)
Observations	1,440	1,440	1,140	1,140
Controls	X	X	X	X
Period	Full	Full	Pre-AB109	Pre-AB109

Robust standard errors in parentheses

implementation of AB 109 is the source of the only significant variation in the rate of assaults that is correlated with changing population shares.

Figure 2.6 emphasizes the importance of the treated subpopulations in identifying the impacts of AB 109. The figure depicts the policy-induced variation that is exploited by the empirical strategies in the remainder of this section, the change in prison crowding and the change in the rate of assault. For the figure, these changes are calculated as the average in the six months preceding AB 109 differenced from the average in the fourth through ninth months following AB 109. The prison groupings depicted are defined as any prison whose total population is made up of at least 20% of the given subtype and the control group is the set of prisons that have less than 20% share of each of the treated subpopulations. Note that these groups are not mutually exclusive. Also, there is one observation point excluded from Figure 2.6. One prison in the reception group is such a outlier that it radically distorts the fitted line for that group. That prison was excluded from the figure for presentational purposes. However, an otherwise identical figure with the outlier included is shown in Section 2.6.3 with a brief discussion of the issue.

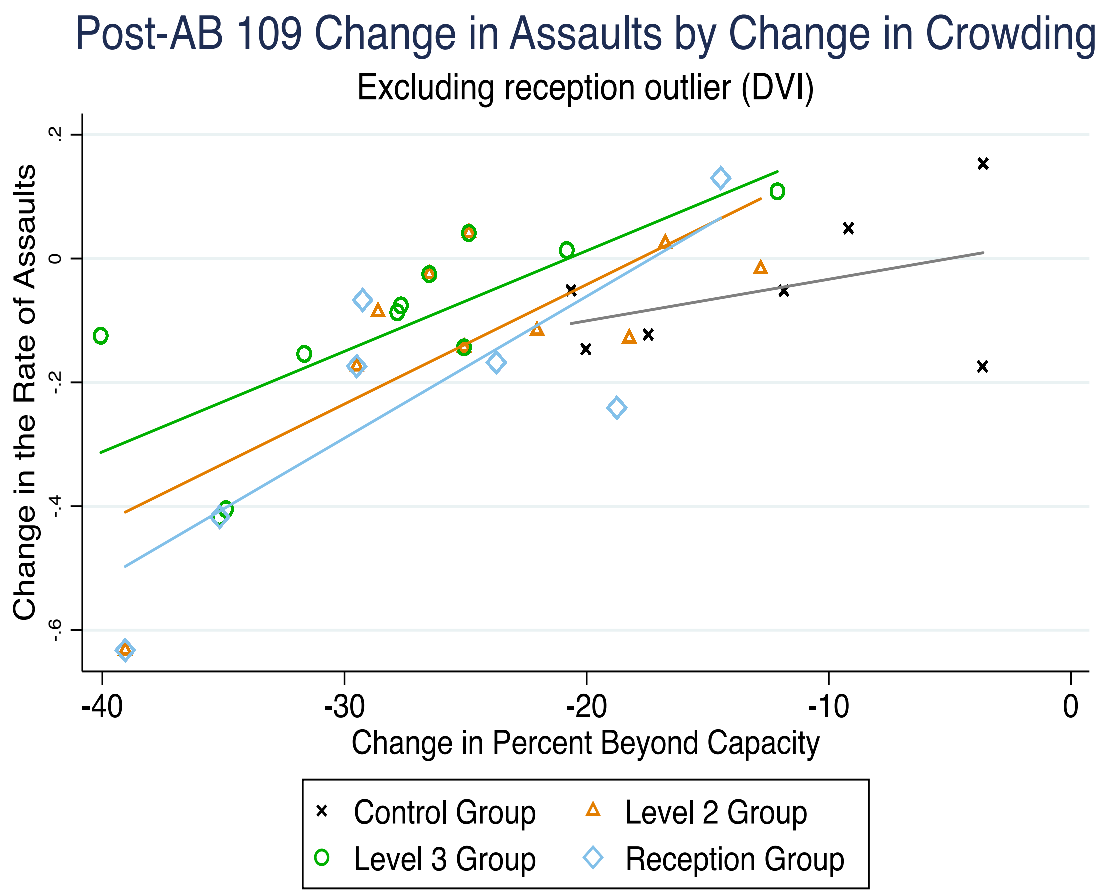


Figure 2.6: *Correlations between the change in crowding and the change in the rate of assault following implementation of AB 109, separated by "treatment groups". Changes are calculated between the 6 month average just prior to AB 109 (Apr11 – Sep11) and first half of 2012 (Jan12 – Jun12). The outlier observation for Deuel Vocational Institute (DVI), a member of the reception group, has been omitted from this figure. Source: Generated from CDCR CompStat reporting data.*

Figure 2.6 illustrates how the overall impact of AB 109 on crowding and associated changes in violence relate to each of the treated subpopulations. Reductions in crowding are much greater for all of the treatment groups relative to the control group and those reductions are clearly associated with reduced violence. In addition, the gradient by which violence changes is notably steeper for the treatment groups than for the control group.

2.6.1 Difference-in-Differences Strategy (DD)

$$Y_{it} = \beta_0 Post_t + \beta_1 Post_t * Treat_i + \beta_2 X_{it} + \delta_i + \epsilon_{it}. \tag{2.2}$$

The concept of the difference-in-differences (DD) strategy in this setting is that a subset of prisons are “treated” with an exogenous reduction in their degree of crowding, while other prisons remain untreated. In such a case, differencing the post-shock change in violent behavior for the treated prisons with that of the untreated prisons provides an unbiased estimate of the causal effect of crowding on violence. The reality of the quasi-experiment provided by AB 109 deviates from such a straightforward DD setting in two important ways. First, there are three different subpopulations for which crowding is significantly decreased by AB 109. Each of these subpopulations experiences a different degree or form of treatment and therefore must be accounted for as three simultaneous but distinct treatments. Second, the unit of observation in the data, a prison, is not the same as the treatment unit, a facility. This means that each observational unit is subject to partial treatment by each of the three treatment types, although no one prison has a large proportion (greater than 5%) of more than two treated population types.

The basic form of the DD estimation is represented in Equation 4.3.2. The variable $Post_t$ is an indicator for whether the observation is post-implementation of AB 109. $Treat_i$ is a vector of variables for the three treatment groups: Reception, Level 2, and Level 3; each is equal to the share of prison i’s total population that is of the given type, averaged over the six months immediately preceding implementation of AB 109. X_{it} is a vector of control variables, which at a minimum includes indicators for when the reception adjustment occurs at a given reception center. In the full set of controls (used for most specifications and indicated in the tables by an X in the row labeled *controls*) are all variables from Table 2.1 listed under "Population Shares" and "Program Enrollment". δ_i and ϵ_{it} are the prison fixed effect and an *iid* error term, respectively. Columns 3 and 4 of Table 2.5 also include either time fixed effects or a flexible time trend in the specification.

This estimation strategy relies on the identifying assumption that $E(\epsilon_{it}|t, Treat_i, X_{it}) = 0$. The intuitive interpretation of that assumption, as it pertains to AB 109, is that any systematic variation in the rate of violence pre- and post-implementation, beyond that induced by the shock to crowding, is uncorrelated with the pre-implementation shares of the treatment populations. However, the assumption in this case must allow for at least some minimal control variables, X_{it}, which is necessary to account for the reception adjustment and the fact that changing population shares are correlated with implementation of AB 109.

Table 2.4 provides a view of the data categorized to show before and after treatment for each of the different treated populations. The groups are the same as those in Figure 2.6 and thus do not align directly with the model in Equation 4.3.2, since the latter uses shares rather than a discrete indicator of treatment. Table 2.4 illustrates a number of observations about the different treatments. All three treatments see a decrease in the average rate of assault in the post periods, but only marginally so for the reception group. In addition, there are notable differences in the baseline rates of assault between the groups, which follow a predictable pattern. Assaults are most common in the control group because most of the state's security level 4 population is incarcerated in those prisons. Reception centers are also prone to higher levels of misconduct, ostensibly because the perpetual turnover is disruptive to mechanisms of informal governance. Average differences in the rates of assault across these groups are generally stable over time, with the exception that reception centers do have greater time variation than other populations.

The highlighted *Crowding* row in Table 2.4 shows that all three treatment groups experience a similar decrease in crowding and it is much larger than the decrease in the control prisons. This

Table 2.4: Pre and Post Statistics by Treatment Group

	Control Group		Reception Group		Level 2 Group		Level 3 Group	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Violence								
Rate of Assaults	0.71	0.71	0.70	0.66	0.32	0.27	0.51	0.40
	(0.41)	(0.44)	(0.29)	(0.32)	(0.21)	(0.23)	(0.32)	(0.24)
Total Assaults	27.94	25.17	34.79	27.47	17.89	13.01	22.55	15.08
	(14.62)	(14.30)	(14.76)	(13.76)	(12.02)	(11.41)	(13.45)	(8.53)
Assaults on Inmates	21.81	18.97	27.39	23.12	14.61	10.69	17.95	12.01
	(13.06)	(12.28)	(13.80)	(13.18)	(10.13)	(9.82)	(12.27)	(7.74)
Assaults on Staff	6.13	6.20	7.40	4.34	3.28	2.32	4.60	3.07
	(4.96)	(7.59)	(5.25)	(2.81)	(3.49)	(2.51)	(5.20)	(3.41)
Crowding								
Crowding (P/K)	1.77	1.61	2.06	1.70	1.96	1.66	1.99	1.68
	(0.25)	(0.23)	(0.21)	(0.20)	(0.20)	(0.21)	(0.20)	(0.17)
Total Population (P)	4132.62	3740.85	5065.20	4213.28	5493.77	4671.10	4691.94	3985.47
	(816.58)	(670.22)	(684.79)	(817.66)	(1057.11)	(970.41)	(1084.07)	(1040.11)
Design Beds	2390.71	2288.49	2462.17	2395.36	2853.50	2749.36	2404.84	2286.36
	(609.32)	(568.37)	(414.08)	(417.62)	(623.83)	(591.01)	(609.38)	(643.45)
Subpopulations								
Level 1	718.81	619.01	580.73	521.27	607.61	548.91	352.21	263.08
	(1133.89)	(1043.51)	(624.99)	(720.92)	(948.91)	(909.85)	(264.07)	(220.14)
Level 2	219.69	137.55	496.30	571.23	2907.98	2540.53	842.06	706.93
	(357.78)	(319.39)	(653.22)	(745.82)	(1405.96)	(1147.00)	(1148.23)	(1032.20)
Level 3	298.98	175.03	236.28	622.53	976.08	920.69	2432.76	2040.53
	(336.73)	(271.48)	(327.86)	(530.68)	(1171.54)	(1025.45)	(773.24)	(557.59)
Level 4	2183.19	2158.75	356.33	450.18	230.17	190.08	423.16	533.76
	(1165.56)	(1125.70)	(571.30)	(927.09)	(578.40)	(572.22)	(461.50)	(514.26)
Reception	72.35	26.65	2855.43	1528.01	372.41	118.45	174.23	56.50
	(178.73)	(87.92)	(1212.16)	(1481.51)	(767.12)	(315.42)	(551.91)	(232.59)
Special Needs	683.94	628.47	578.55	906.38	1302.55	1421.83	907.52	1076.85
	(653.53)	(614.05)	(826.60)	(1054.63)	(1177.03)	(1249.39)	(1115.88)	(1059.35)
Programs								
Prison Industries	22.44	7.80	167.78	122.53	264.08	230.67	242.29	221.54
	(38.10)	(14.02)	(104.32)	(77.21)	(185.10)	(172.39)	(190.25)	(176.44)
Academic	406.55	340.47	172.55	162.53	639.61	480.46	480.33	401.21
	(222.53)	(141.21)	(184.05)	(134.68)	(391.83)	(199.87)	(286.15)	(180.21)
Non-PIA Work	1997.30	1724.30	1346.78	1195.29	3364.15	2839.77	2536.09	2177.30
	(1014.45)	(921.60)	(818.91)	(617.45)	(1096.89)	(875.37)	(882.35)	(784.79)
Subst. Abuse	13.10	0.00	94.13	32.35	294.83	104.55	92.94	30.92
	(54.45)	(0.00)	(137.12)	(56.70)	(379.89)	(58.44)	(136.25)	(50.87)
Subst. Abuse Waitlist	8.50	0.00	67.12	72.66	121.84	156.62	42.87	67.18
	(37.34)	(0.00)	(117.82)	(159.85)	(113.53)	(115.08)	(85.39)	(118.51)
Observations	273	105	312	120	390	150	429	165
Number of IDs	7	7	8	8	10	10	11	11

Means reported. Standard deviations are in parenthesis. Treatment groups defined by > 20% share of the given population.

further supports the earlier claim that the reception adjustment was a sufficient response to diminish the reception center impact of AB 109 to a level similar to that in level 2 facilities and simultaneously create a similar magnitude shock to crowding in level 3 facilities. The effect on level 3 populations is also visible in the other highlighted sections of the table, which show the level 3 population experience an approximately equivalent increase and decrease in the reception group and level 3 group, respectively.

To review, Table 2.4 and figures in Section 2.3 demonstrate that each of the three AB 109 treatments provide the crucial variation in crowding necessary to identify its relationship with violence. Meanwhile the compositional mechanism proposed in Section 2.4 implies a form of omitted variable bias for any estimates relying on these reductions in crowding. However, it is further implied that the expected bias in the level 2 treatment should be minimal, or possibly even absent, while the expected bias for the reception treatment is larger and potentially quite significant. Minimal bias for the level 2 treatment relies on similarities between the the base population and that which is targeted by AB 109, which requires some efficacy to the selection by which security classification is determined (evidence of this is clear in Table A.1). The implication for the level 3 treatment is not immediately obvious since nothing is known about the selection process by which prisoners were chosen for transfer to repurposed reception facilities. Indeed it is quite possible that very different selection criteria were used by officials at different prisons, as opposed to the very uniform selection criteria that AB 109 implemented for reducing the overall population. Uncertainty about the exact form of selection in the level 3 treatment indicates that the estimates for this group will not be particularly informative, but nonetheless need to be included in the identification strategy to control for the fact that these populations were subject to a simultaneous treatment.

DD Estimates

Table 2.5 shows the β_1 coefficient for each of the three AB 109 treatments. Indicators for the timing of reception adjustment are always included as controls and X represents a full set of controls

Table 2.5: Difference-in-Differences Estimation

Dependent Variable: Log Rate of Assaults				
	(1)	(2)	(3)	(4)
VARIABLES	No Controls	Main	TimeFE	3mo.Gap
ShareLv2*Post ($\hat{\beta}_{12}$)	-0.529	-0.388	-0.390	-0.446
	(0.176)	(0.142)	(0.145)	(0.159)
ShareLv3*Post ($\hat{\beta}_{13}$)	-0.303	-0.211	-0.206	-0.306
	(0.279)	(0.271)	(0.261)	(0.310)
ShareRec*Post ($\hat{\beta}_{1R}$)	0.252	0.272	0.218	0.471
	(0.191)	(0.179)	(0.173)	(0.127)
Observations	1,470	1,470	1,470	1,380
Controls	RA only	X	X	X
Hypothesis tests	-	-	-	-
P-value $\beta_{12} = \beta_{13}$	0.471	0.549	0.522	0.664
P-value $\beta_{12} = \beta_{1R}$	0.003	0.005	0.009	0.000
P-value $\beta_{13} = \beta_{1R}$	0.081	0.136	0.166	0.026

Robust standard errors in parentheses

including population shares and program participation. Columns (3) and (4) are two different extensions of the specification in column (2). Column (3) adds time fixed effects while column (4) adds a flexible time trend and omits the first three months of post-treatment observations. In all specifications standard errors are clustered at the prison level and each observation is weighted by the average population size of that prison measured over the six months prior to AB 109 implementation.

The relationship between the pre-treatment share of level 2 population and violence is consistently negative and quite large. On the other hand, the coefficients for the other two treatments are either not significantly different from zero or significantly positive. Yet the positive coefficient on the reception treatment is consistent with a large compositional effect driving an increase in assaults that overwhelms any decrease from reduced crowding. Examining columns (2) and (4), which are effectively the same specifications reveals that omitting the few months of AB 109 prior to the reception adjustment increases both the magnitude and significance of the point estimate for the reception treatment. This also aligns with the implications of the model in Section 2.4 since the omitted months in column (4) are the months for which the crowding effect (decreasing violence via

$E_{V;c} > 0$) as well as the compositional effect (increasing violence via $E_{\pi;\lambda} < 0$) can both be expected to be quite large. By contrast, during the later months the reception adjustment offsets much of the decreased crowding but the change in composition is sustained, making the expected compositional effect stronger relative to the crowding effect in the column (4) analysis relative to the column (2) analysis.

Note that although the point estimates for the level 3 treatment are reported, the informativeness of these estimates with regard to the effect of crowding on misconduct is extremely limited. As shown in previous sections, there is certainly an impact of AB 109 on the level 3 population. However, without any prior expectation on what type of selection bias may have been present in the manner by which inmates were chosen for relocation, it is not possible to present a fair assessment of how much of the observed effect on violence in any given facility is due to the reduced crowding. At most it is fair to expect an overall reduction in violence across all facilities, given a true causal effect of crowding on violence, because there is no change in the statewide composition of the level 3 population. Yet even if that is so, administrative selection in the redistribution of level 3 inmates is likely to result in greater variation in the effect on violence between facilities and thereby lead to inflated standard errors for this treatment estimate.

Absent the theory of compositional change, the estimates in Table 2.5 suggest that the decrease in crowding due to AB 109 led to a decrease of approximately 40% in the rate of assaults at level 2 facilities. The level 2 facilities saw crowding fall about 30 percentage points from an initial point of almost 200% of design capacity, implying a semi-elasticity of approximately 1.3 for this specific type of prison population. On the other hand, there is no clear effect and possibly an increase in assaults for the reception center populations. It is possible that this is due to some fundamental difference between reception centers and level 2 facilities, either with regard to the populations themselves or the housing and security protocols. Yet it is also true that the difference in estimates for the level 2 and reception populations comport well with the assumption that there is a compositional element to the population shock generated by AB 109, because such compositional change would be necessarily more significant among the reception population. These estimates are therefore interpreted as descriptive evidence supporting the presence of such a compositional effect.

2.6.2 Instrumental Variable (IV)

$$\hat{C}_{it} = \alpha_0 Months_t + \alpha_1 Months_t^2 + \alpha_2 Months_t * S_i^n + \alpha_3 X_{it} + \alpha_4 f(t) + \gamma_i + u_{it}. \tag{2.3}$$

$$Y_{it} = \beta_0 + \beta_1 \hat{C}_{it} + \beta_2 X_{it} + \beta_3 f(t) + \delta_i + \epsilon_{it}. \tag{2.4}$$

Equations 2.3 and 2.4 present the basic structural form of the IV strategy. Equation 2.3 is the first-stage estimating equation wherein crowding is estimated as a function of the number of months since the implementation of the policy, $Months_t$, interacted with the pre-implementation shares, S_i^n, of population type n. In the baseline IV, only initial population shares for the treated subpopulations are used as instruments. Specifications tested with an expanded set of instruments find only minor adjustments to the coefficient of interest. Equation 2.4 is the second-stage estimation, which is a fixed effects regression of the rate of assaults on the predicted values of the crowding variable. The estimation includes the same control variables included in the main DD specification.[21] A polynomial time trend is included in each stage of estimation.

In effect, the IV strategy uses the time since implementation of AB 109 and the initial share of each population type to predict the change in crowding at each prison.[22] This approach exploits the same exogenous variation as the DD approach but allows flexibility in modeling variation in the intensity of treatment over time. The IV strategy also has the benefit of added flexibility in modeling the impact of the reception adjustment. Specifically, the "Interact RA" specification in Table 2.6 allows the degree to which pre-shock reception share predicts changes in crowding to be diminished as the reception adjustment occurs in the given prison. The major shortcoming of the IV

[21] These include indicators for the timing of the reception adjustment and the population share and program enrollment variables included in Table 2.1.

[22] These population shares are defined as the population of type n divided by the total prison population, averaged over the six months preceding implementation of AB 109.

specification is that it necessarily conflates any compositional bias from any of the three treatments into the estimated effect of crowding on the rate of assaults.

The exclusion restriction for this IV strategy requires that the pre-shock composition of prison populations is uncorrelated with future variation in the rate of assault and battery, other than through its correlation with changes in the level of crowding due to the policy design. The only clear challenge to the validity of this restriction is the aforementioned fact that pre-AB 109 population shares are correlated with the post-AB 109 changes in those shares. For this reason, all specifications of the IV model include a full set of controls that include contemporaneous population shares for each subpopulation.

IV Estimates

Table 2.6 provides the estimated $\hat{\beta}_1$ for a number of variations on the IV model. A benefit of the IV model is the straightforward interpretation of the $\hat{\beta}_1$ coefficient. It is a semi-elasticity showing the marginal effect of crowding on the rate of assaults, where crowding is measured in percentage point changes and the rate of assault in percent changes. The first column of the table is the base specification of the model, exactly as presented in Equations 2.3 and 2.4. The other three columns each represent a different variation from the base specification. The second column includes an additional term in the excluded instruments, which interacts the RA_{it} variable[23] with the $Months_t * S_i^n$ term of the reception treatment. Column (3) is a robustness check that removes the $Months$ terms that are not interacted with population shares from the IV exclusions, allowing that there may be some statewide correlation between the policy timing and assaultive behavior. The standard error inflates slightly with this change, but the coefficient remains large and statistically significant. In the final column the three-month period just following implementation and preceding full-scale reception adjustment is excluded from the analysis. This is an alternate means of accounting for the effects of the reception adjustment, so the RA control variable is excluded from the specification in column (4).

[23] RA_{it} is an indicator that turns on in the month that prison i has the reception adjustment.

Table 2.6: IV Model with Prison Fixed Effects

Dep. Variable: Log of Assaults and Batteries per 100 inmates				
	(1)	(2)	(3)	(4)
VARIABLES	Base IV	Interact RA	Exclusion	Drop 3mo.
Crowding (P/K)	2.117	1.548	1.718	2.289
	(0.627)	(0.497)	(0.767)	(0.629)
Observations	1,440	1,440	1,440	1,350
Number of ID	30	30	30	30
Controls	X	X	X+Months	X
Exclusions	Base	Months*S*RA	Months*S	Base
F-test IVs	10.34	26.81	10.05	10.59

Robust standard errors in parentheses

A few points are immediately observable from these results. First, a comparison of these estimates with the earlier OLS estimates shows evidence of significant downward bias in the OLS estimates, likely due to endogeneity as discussed earlier. The estimates from each of the IV specifications are large, statistically significant, and reasonably stable. These estimates are possibly subject to significant attenuation due to compositional bias, yet still exhibit semi-elasticities as high as 2.2. The more conservative estimate in column (2) suggests that decreasing crowding by 10 percentage points can decrease the rate of assault by as much as 15 percent.

The difference between the columns (1) and (2) estimates is also of interest. The two estimations are identical beyond the one addition of the interaction $Months_t * S_i^R * RA_{it}$. This interaction provides a greater ability to predict changes in crowding using the pre-shock share of reception population. In the absence of this interaction, the controls for administrative response only allow for a level shift in crowding once there has been a response. In the first stage regression coefficients (available in Table A.2 in Appendix A.2), the point estimate for the coefficient on this interaction is almost precisely the same magnitude as the $\hat{\alpha}_2$ for reception share, which is the negative of the former. This implies that the pre-shock share of reception population correctly predicts decreases in crowding when AB 109 begins and up until the reception adjustment occurs, at which point it ceases to have any predictive power at all because the two coefficients cancel each other out.

This improved fit in the first-stage modeling of crowding via the reception share together with the diminished coefficient in Table 2.6, aligns with the results from the DD strategy that reception centers do not see assaults decrease to the degree that the other treated facilities do.

2.6.3 Tests and Robustness

AB 109 provides rationale for the position that the estimates in this paper are unbiased evidence of the causal effect of crowding on violent behavior. Yet a plausible source of bias remains, namely the compositional effect defined in Section 2.4. However, the compositional effect is reasonably expected to bias estimates towards zero and thus at worst the estimates in this paper would be a lower bound on the true value. Still the limitations of these data and the exogeneity of the AB 109 intervention do warrant some further discussion.

The first concern pertains to the validity of the channel by which the instruments and DD design identify the effect of crowding on violence. Although it is certainly the true that AB 109 reduced crowding, it remains a possibility that there is some spurious artifact of the assault data driving the results or some unknown features to the implementation of the new law that were the real causes of reduced violence. For example, the inability to control for staffing changes, due to limitations of these data, is a potential concern. Furthermore, even if staffing changes were insignificant,[24] it is possible that staff were simply able to operate more effectively once there were fewer inmates crowding the facilities. Yet if the policy-induced variation in assaults were due to enforcement gains from improved staffing ratios or effectiveness of staff then the estimated effects should be present for other forms of misconduct as well.

Table 2.7 confronts the above concerns by repeating the IV specification with different measures of misconduct. Some of these measures were not recorded in the early years of CompStat data, so the number of pre-AB 109 observations in this table is limited to 15 months. Column (1) repeats the IV

[24] Available data on staffing is incomplete and poses other issues that make it unsuitable to include staffing variables in this research; nevertheless, a general read of the existing data suggests that staffing changes due to AB 109 were minor.

Table 2.7: IV Model: Testing Other Measures of Misconduct

Dep. Variable: Log-rate of the given form of misconduct.				
	(1)	(2)	(3)	(4)
VARIABLES	Assaults	Incidents	Drugs	Cellphone
Crowding (P/K)	1.522	1.406	0.418	-0.0583
	(0.558)	(0.312)	(0.761)	(0.810)
Observations	750	750	750	750
Number of ID	30	30	30	30
Controls	X	X	X	X
Exclusions	Base	Base	Base	Base
F-test IVs	16.91	16.91	16.91	16.91

Robust standard errors in parentheses

specification from column (2) of Table 2.6 for the limited time period of the other specifications here, showing a qualitatively identical estimate to those in Table 2.6. Column (2) replaces the measure of assaults, previously disciplinaries, with the "incidents" measure of the same type of violation. The point estimate is not sensitive to which measure of assaults is used. In stark contrast, columns (3) and (4) prove there is no identifiable impact on other forms of misconduct that occur with similarly high frequency to that of assaults.

The validity of using the variation in crowding from AB 109 is further tested in Table 2.8. Column (1) is a replication of column (1) from Table 2.6 and each of the following columns repeats the model with the policy implementation beginning the specified number of months earlier than the true implementation date of AB 109. To maintain comparability of the estimates, the data is truncated in each specification so the number of post-implementation months remains constant. Since there was no equivalent in these periods to the reception adjustment that occurred following AB 109, the *RA* variable is omitted from the lagged specifications. The results are no different if the *RA* control is reintroduced into the estimating equations. None of the resulting estimates are close to statistical significance in either the first- or second-stage regressions.

Tables 2.7 and 2.8 both rely on the IV strategy to test their respective alternate hypotheses. This approach is more direct and has the benefit of reporting fewer coefficients, whereas the multiple

Table 2.8: IV Model: Placebo Tests with Policy Implementation at Alternate Dates

Dep. Variable: Log of Assaults and Batteries per 100 inmates

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Base IV	9 Month	15 Month	21 Month	27 Month
Crowding (P/K)	2.117***	1.891	0.791	0.356	0.195
	(0.627)	(1.998)	(1.097)	(0.901)	(2.196)
Observations	1,440	1,140	960	780	600
Number of ID	30	30	30	30	30
Controls	X	X	X	X	X
Exclusions	Basic	Basic	Basic	Basic	Basic
F test IVs	10.34	2.874	5.248	2.761	2.054

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

coefficients reported in the DD strategy increases the risk of spurious statistical significance. In addition, the less refined modeling of time variation in the DD strategy is more likely to show correlations from any of several earlier, less dramatic, policy changes that occurred in the years leading up to AB 109.

It is also possible that while overall variation in crowding from AB 109 is appropriately exogenous, the subpopulations used to identify the intensity of crowding reductions are not the proper channel. The likely alternative is that prison administrators were able to manipulate new prisoner classifications to reduce populations in facilities that were the most crowded, rather than those naturally targeted by reducing the inflow of non-violent offenders. The result of this would be that initial crowding should do a better job predicting the post-AB 109 reductions in crowding than the pre-shock population shares do.

Figures 2.7 and 2.8 depict the relationship between the initial level of crowding and change in crowding, with Figure 2.8 dividing the trends into the same rough treatment groups used in Table 2.4 and Figure 2.6. If the above concern were valid, then the correlation apparent in Figure 2.7 would remain mostly intact in Figure 2.8, with the treated populations simply having systematically higher levels of initial crowding (which would then induce the larger reductions seen for those groups in

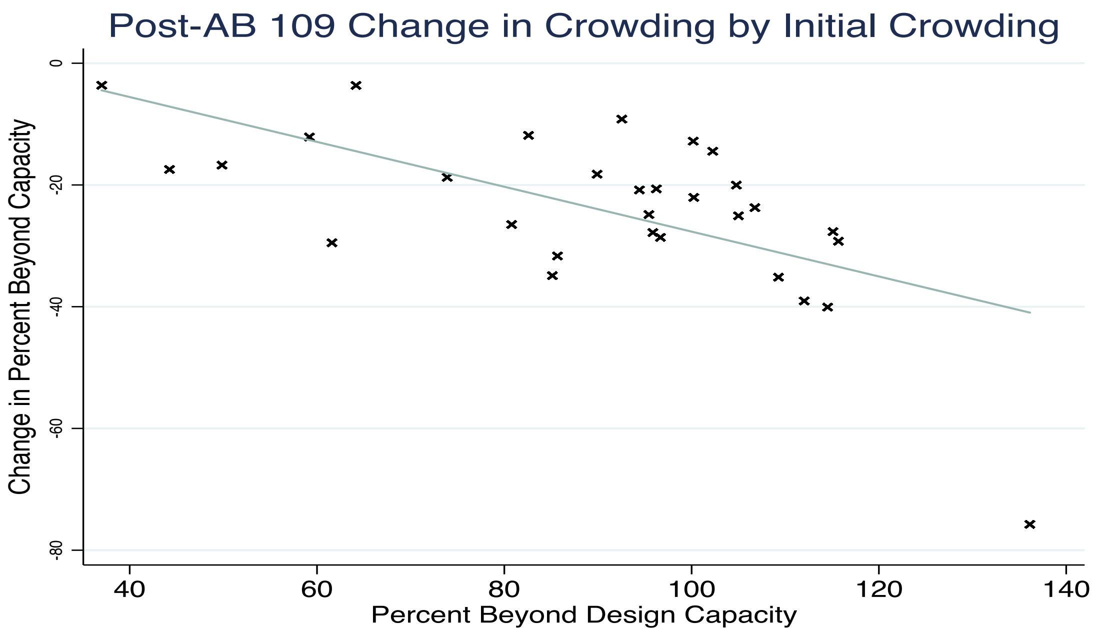


Figure 2.7: *Change in crowding by initial crowding. Showing correlation between initial crowding and the decrease in crowding after implementation of AB 109.*

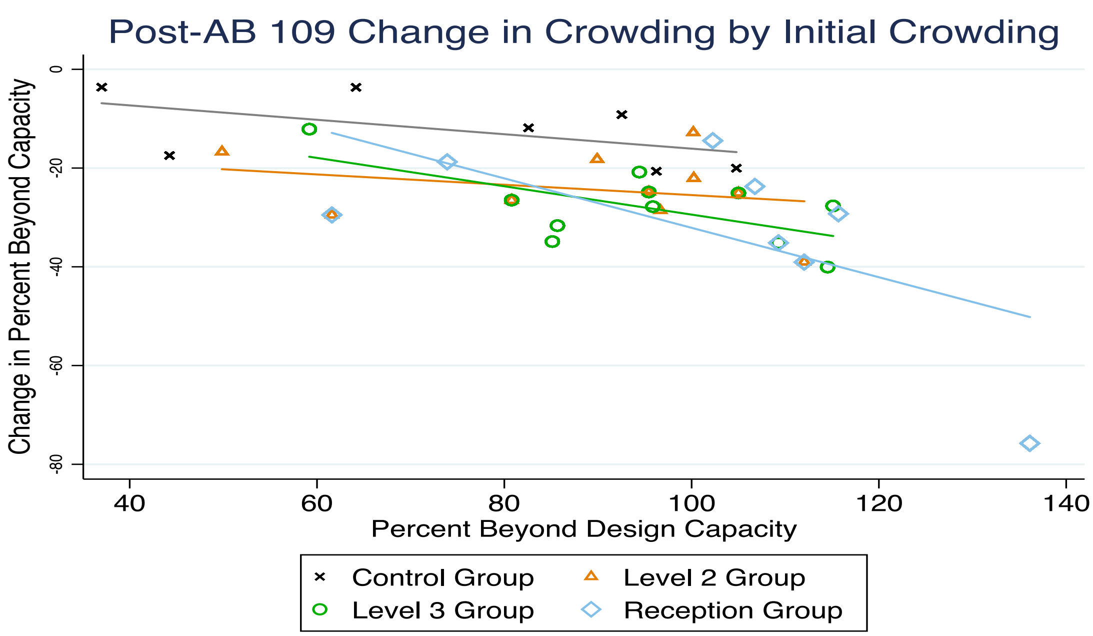


Figure 2.8: *Grouped change in crowding by initial crowding. Showing correlations between initial crowding and the decrease in crowding after implementation of AB 109 by "treatment groups". This figure indicates that the apparent correlation in Figure 2.7 is due to the fact that the treated populations were, on average, those that were more crowded at the start. Source: Generated from CDCR CompStat reporting data.*

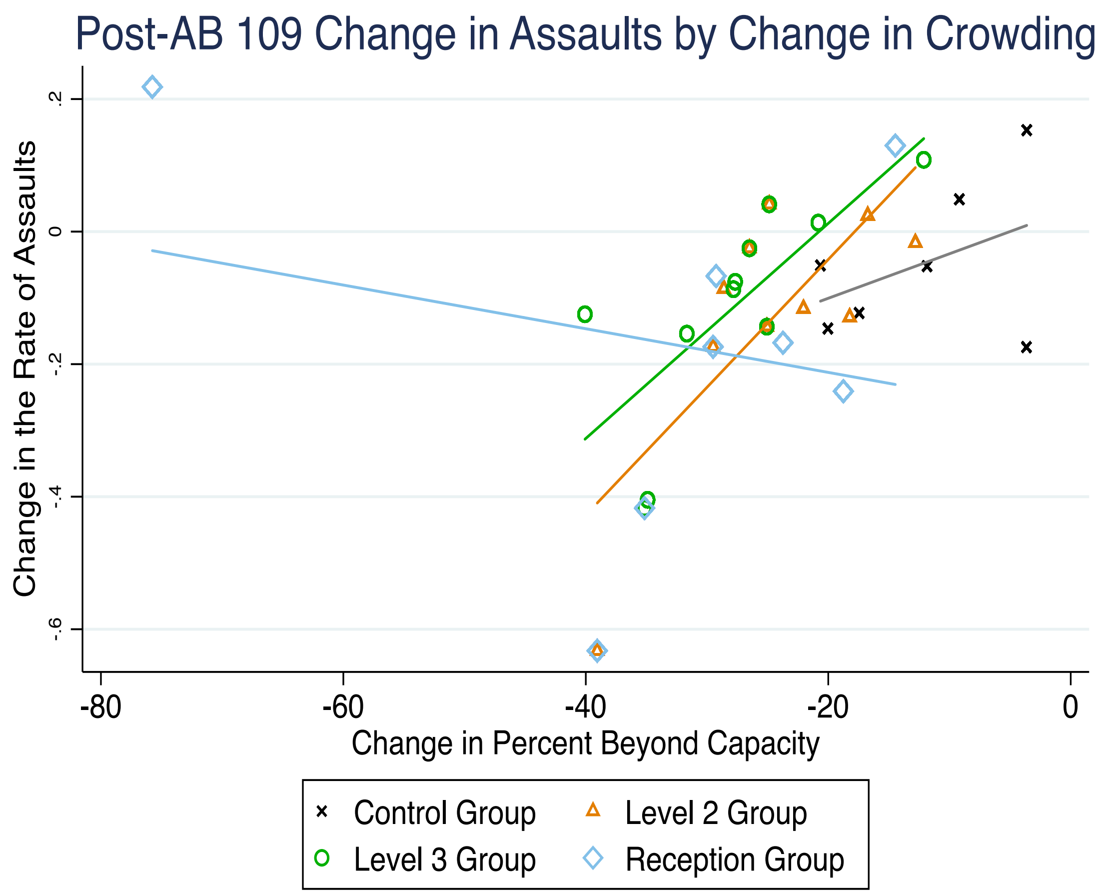


Figure 2.9: *Correlations between change in crowding and the change in the rate of assault following implementation of AB 109, separated by "treatment groups". Source: Generated from CDCR CompStat reporting data.*

Table 2.4). Instead the fitted lines for each group in Figure 2.8 have relatively shallow gradients and those observations in the treatment groups with less crowding saw larger reductions than comparable prisons in the untreated group. The exception to the shallow slopes in Figure 2.8 is the line for the reception group, for which the fitted line retains a slope similar to that in Figure 2.7. However, note that the slope for the reception group is driven by the single outlier in the bottom right of the figure.[25] With the outlier removed, the reception line has a slope as shallow as any of the others. In sum, when treatment groups are accounted for there is still a weak correlation between initial crowding and the reduction induced by AB 109, but the treated subpopulations appear to be an appropriate predictor of changes in crowding.

A final issue for discussion is the omitted outlier from Figure 2.6. The omission has a significant impact on the presentation of that figure, which is shown by Figure 2.9. The outlier, DVI, experiences the largest reduction in crowding among the prisons but also experiences the largest *increase* in assaults, which dramatically alters the fitted line for the reception group. A brief investigation did not reveal a clear reason for the distinct experience of DVI relative to the other prisons. As such, DVI was not excluded from any of the empirical specifications in this section, only the earlier figure to better illuminate the consistent relationship in the data. However, it is also of note that exclusion of DVI does not qualitatively change any of the empirical estimates. Table A.3, in Appendix A.2, demonstrates this by replicating the full set of DD specifications with DVI excluded from the analysis.

2.7 Conclusion

Violent behavior in prisons may be caused by a variety of mechanisms. In line with much of the literature, this paper examines the role of prison crowding as a determinant of violence. This relationship plays an important role in well-informed policy decisions and its study is particularly apropos given recent political movements to reduce mass incarceration in the United States.

[25] The outlier in this figure is again DVI, the same prison that was omitted from Figure 2.6.

Although theory and popular opinion have long held prison crowding as a key determinant of violent behavior, empirical estimates in the existing literature have been surprisingly inconsistent (Franklin et al., 2006). The estimates in this paper are the first to use a quasi-experimental design that specifically targets the effect of crowding on violence. These estimates represent persuasive evidence of a causal effect of crowding on violent misconduct that is positive and qualitatively large. They suggest a 10 percentage point decrease in prison crowding results in a decrease in the rate of violent assaults by approximately 15%, which could imply significant cost savings associated with reductions in crowding. A careful review of the empirical results also provides evidence that the compositional effect defined in Section 2.4 is indeed a factor in the outcomes from the AB 109 shock to crowding. This further implies that the compositional effect is a plausible factor to be considered with regard to previous and future empirical work on crowding and violence. The compositional effect, as a new source of potential bias, adds a nuanced explanation for conflicting evidence in existing empirical research on crowding and violence.

There are some natural limits to the implications of this research in policy application. Foremost among these is external validity when considering dissimilar prison populations. Although the IV strategy incorporates some statewide variation in crowding across prisons, the estimates are largely driven by variation in security level 2 populations. These populations have relatively low rates of violent misconduct and it is possible that prison populations with a higher propensity for violence could be either more or less responsive to changes in crowding. The DD estimates for the other treated populations do little to better inform this issue since both are expected to be subject to greater compositional bias than the level 2 treatment. It is also important to recognize that the identifying variation for this research is from very high levels of overcrowding, many of the affected facilities beginning well above 200% capacity prior to implementation of AB 109.[26] It is reasonable to expect some variation in the marginal impact of crowding on violent behavior when examining the relationship in much less crowded facilities.

[26] The initial levels of crowding can be observed along the horizontal axis in Figures 2.7 and 2.8.

In summary, this research offers several unique contributions to the literature. It presents the first empirical evidence of a strong causal effect of prison crowding on violent misconduct. It also introduces the idea that there is a compositional element to policy interventions that reduce crowding, which can lead to bias in empirical estimates. Although this concept is not necessarily novel, this paper is the first to formally discuss the idea and its role in studying prison crowding and violence. Finally, this work highlights many areas of opportunity for subsequent work studying prison violence.

A particular topic for future research is heterogeneous effects among different types of prison populations. The estimates in this paper are mostly pertinent to prison populations that are relatively less prone to violence and prison settings that are subject to extreme levels of crowding. Maximum security facilities are of particular interest in this regard, both because they have the highest baseline rates of violence[27] and because housing and security protocols in such facilities tend to be quite different from other prison facilities.

Another important extension of this research will be to access improved data to allow an in-depth analysis of the interaction of the compositional and crowding effects from changes in prison populations. With sufficiently detailed individual inmate data, a full decomposition of the respective mechanisms is possible and will allow a much more nuanced understanding of policy impacts on violent misconduct. This information could be an invaluable contribution to social, political, and administrative insights regarding the costs and benefits of a broad variety of justice-related policy reforms.

[27] Security level 4 facilities in California have significantly higher rates of violence than other facilities, show in Table A.1.

CHAPTER 3

Anarchy or Strategy? – Prison Violence as a Means to Informal Governance and Rent Extraction[1]

3.1 Introduction

Violent behavior is an ubiquitous phenomenon in the prison setting and gangs are often viewed as the root cause of this issue. Yet the motivation for the culture of violence fostered by these gangs need not rely on a fundamental preference for violence. Neither does correlation between the rate of violence and gang prevalence necessarily imply that the absence of a gang presence would lead to less violence. In fact, it has been suggested that even in their own use of violence prison gangs act to limit the anarchic behavior of the prison population as a whole (Skarbek, 2012). This research distills existing theory on the motives of prison gangs, and other informal governance institutions, into a modeling framework to better understand prison gang interactions and inform efficient use of prison resources in the management of inmate violence. Comparative statics of the model suggest that, under reasonable conditions, optimal enforcement involves a mix of punitive and market-based strategies and that the responsiveness of gang violence to any policy is highly dependent upon

[1] Special thanks to Richard Arnott and Urmee Khan for their invaluable contributions to this work. Additional thanks to Joseph Cummins, Mindy Marks, Michael Bates, Gregory DeAngelo, Steven Clark, and other colleagues whose comments and support made this work possible.

the informal governance role of gangs. The proposed framework also provides a starting point for nuanced extensions of the model and, ultimately, for systematic evaluation of the overall costs, and potential benefits, of gang prevalence in the correctional systems.

The model presented in this research relies on two broad assumptions about the prison setting. First, there exists demand for a variety of illicit goods and services[2], the trade of which is officially prohibited by the prison staff. Second, in the absence of official channels that can be relied upon for enforcement, trade of such illicit goods requires the existence of some informal source of enforcement for property rights and market norms. Together these assumptions imply that there are market rents available from the trade of illicit goods, but a source of informal governance is necessary to facilitate rent extraction. Although decentralized self-governance regimes were observed in early 20th century U.S. prisons, those have given way to prison gangs as the main source of governance for all forms of illicit market activity.

In determining the "optimal" supply of violence and illicit goods, this work assumes that prison gangs are profit-maximizing entities that utilize violence as a means to regulate market behavior and compete for market share. In contrast with the limiting assumption that gang violence arises from a direct preference for violence, the presence of a profit motive expands the set of potential policy tools by which prison staff may intervene to suppress violent behavior. To be sure, it is not unreasonable to assume that there are individuals in the prison system with what is, perhaps, a preference for violence. However, the profit motive provides a more likely explanation for the coordinated behavior of individuals in the gangs that exert wide-spread influence over violence in prisons. To explore the policy implications of this view of prison gang priorities, comparative statics of the model are examined in the context of several types of policy action, including the severity of punishment for violence but also less direct measures such as drug rehabilitation and efforts to suppress smuggling of illicit goods into the prison.

[2] Henceforth, "illicit goods" refers to both good and services.

Rationalizing behavior in the prison setting is an important step towards designing effective, evidence-based policy and practice. Prison officials have expressed frustration at the limited effectiveness of enforcement efforts to eliminate violence, which typically rely on increased likelihood and severity of punishment for infractions.[3] Meanwhile reformists argue that environmental factors and rehabilitation programs are the key to reducing misconduct. Yet empirical evidence has struggled to provide consistent support for even the most widely accepted determinants of violent behavior, such as population density (Franklin, Franklin, and Pratt, 2006). More conclusive evidence regarding prison behavior and misconduct is difficult to come by because of the observational nature of existing data and human rights concerns that raise ethical questions about even the most subtle experimental intervention. Even when data is available or experimental intervention is possible, researchers need a theoretical basis to draw on in order to ask the right questions. Work by Skarbek, proposing the informal governance role of prison gangs, provides a clear example of this need (Skarbek 2010, 2011, 2012). As discussed below, researchers have shown that greater gang prevalence is associated with higher rates of misconduct, but Skarbek's work raises the question of whether the presence of the gang results in misconduct, or the misconduct itself leads to the presence of the gang. If the latter is the case then it is possible, even in light of existing evidence, that prison gangs actually suppress misconduct and a policy that manages to disrupt formation of such gangs could actually result in higher rates of misconduct. This work contributes a formal model to the existing body of theory regarding the behavior of prison gangs.

3.2 Literature

This work draws on and contributes to several branches of existing literature. First among these is the aforementioned literature exploring the role of prison gangs as a source of informal governance in the prison setting (Skarbek 2010, 2012, Butler et al. 2018, Symkovich 2017). Skarbek explains

[3] E.g. see https://nicic.gov/evidence-based-practices-ebp

that U.S. prison populations prior to 1950 were small enough that reputation based enforcement of social norms was sufficient to provide a stable system of property rights and informal governance. As prison populations grew larger, this decentralized enforcement mechanism became less reliable and over the span of a few decades the presence of prison gangs across the United States grew to fill the need for effective governance of illicit market transactions and other behavior that prison officials cannot or will not regulate. Skarbek also emphasizes the point that, regardless of the original motivation for the establishment of a gang, each prison gang operates as a well-defined organization that exhibits rent-seeking behavior and employs violence in pursuit of those objectives.

The specific branch of literature examining the prison setting adds to a rich body of work that looks more broadly at self-governance institutions throughout history. These works examine the many ways individuals who have no recourse for oversight via an organized government, or who eschew the laws of formal governance, have developed social rules for private enforcement (Leeson, 2014). The private emergence of laws for prevention of violence, theft, and other socially destructive activities has been observed in a great variety of anarchic settings, including pre-colonial West African traders, pirate ships, the Wild West, early California merchants, World War II prison camps, and all types of criminal organizations (Anderson and Hill 2002, Clay 1997, Cordingly 2006, Leeson 2007a, 2007b, 2010, 2014, Peden 1977, Radford 1945, Skaperdas 2001, Skaperdas and Syropoulos 1995). These self-governance regimes vary from decentralized enforcement of social norms to very structured governance mechanisms, as in the case of criminal constitutions employed by pirates, prison gangs, and mafias (Leeson 2014, Skarbek 2010, 2012, Leeson and Skarbek 2010). This body of work supports Skarbek's hypotheses as to why prison gangs eventually supplanted decentralized self-governance in U.S. prisons as the population sizes swelled, insofar as decentralized self-governance is mostly observed in close-knit, homogenous communities with few barriers to information transmission regarding individual transgressions.

A third branch of literature is the body of empirical work examining the relationship between gang prevalence and prison violence or misconduct. As a whole, this area of research has been limited by access to data, often relying on self-reported survey data due to the illicit nature of gang activity. Illustrating this limitation, the prevalence of gang membership itself is a contentious issue with the estimated rate of membership as low as 5% or as high as 25% of the U.S. prison population, depending on which study is referenced (Trulson et al., 2006; Knox 2005).[4] However prevalent gang membership truly is in prisons, there is a preponderance of empirical support for the notion that gang membership or affiliation is a strong predictor of prison misconduct (Cunningham and Sorenson, 2007; Drury and DeLisi, 2011; DeLisi et al., 2004; Gaes et al., 2002; Griffin, 2007; Griffin and Hepburn, 2006; Jiang and Fisher-Giorlando, 2002; Reisig, 2002; Trulson, 2007). There are also a few interesting subtleties to this correlation. For instance, Drury and DeLisi (2011) show that although inmates convicted of homicide are less prone to violent misconduct, those that are both gang affiliated and convicted of homicide are more prone to violence. Gaes and colleagues (2002), on the other hand, show that gang affiliated inmates serving longer sentences are less prone to violent misconduct than those serving shorter sentences. These findings are suggestive of the many subtleties underlying the unquestionable link between gang activity and violent behavior.

The research of Skarbek and others extends existing work on self-governing institutions to understand the governance role of gangs in the inmate social system, which is unique in that inmates are inescapably subject to the authority of formal governance but also experience social coordination problems that exist outside of that governance. This research takes that contribution to a new level by introducing it into a flexible modeling framework. Furthermore, with respect to the empirical relationship between gang prevalence and violence, this research provides a systematic approach to evaluating how that relationship may interact with policy choices, in light of the core motives of prison gangs with policy choices.

[4] One reason for the difficulty in estimating the prevalence of gang membership is that, in almost all prison facilities in the U.S., every inmate must associate with a racial or ethnic group (Dolovich, 2011). For example, an hispanic inmate from Northern California without any gang ties will still affiliate with and follow the rules of the Norteño gang while incarcerated in a California prison. Accurately accounting for different degrees of gang affiliation presents a serious challenge in measuring prevalence.

3.3 The Model Setting

We begin with a setting in which N gangs operate as suppliers in the illicit market for a single composite good. However, rather than engaging in price competition, as we would expect of firms in a traditional marketplace, each gang uses force to capture a share of the market and operate as a monopolist within the portion they control. This follows the model of organized crime often observed among mobs, biker gangs, and street gangs, which capture "territory" within which they are able to exclude competitors from supplying the drugs and other goods they supply (Campana, 2011). The prison setting differs insofar as gangs have limited ability to claim physical territory, relying in part on less visible forms of market segmentation such as racial division and control of supply chains, but the general strategy remains the same.[5] Although in reality there is likely some measure of price competition for some of the goods offered in prison black markets, it is certain that each gang is able to exert market power and the assumption of monopoly power is a straightforward way of incorporating the profit incentive to control a share of the market.

Due to the illicit nature of the market, each gang must also act as a regulatory presence, enforcing a set of social norms that are conducive to their rent-seeking activities. Thus each gang supplies violence as a means to enforce property rights and suppress behaviors that disrupt market activity, including the idiosyncratic violence of individual inmates. For example, if an aggressive inmate frequently causes physical altercations that lead to general lockdowns, his behavior will likely not be tolerated for long by the gangs in that facility. This aligns with anecdotal evidence provided by inmates who report that, despite not being gang members, their behavior is monitored by the gang[s] associated with their racial group and there are harsh consequences for violence that is not properly justified (Skarbek 2014). Although the responsibility for this regulatory behavior largely falls along the racial lines of gang affiliations, the model also allows it to increase with the market

[5] There are a number of potential explanations for the racial segregation that occurs in prisons. One of these is that racial lines provide a convenient means of market division and enforcement authority (Skarbek, 2014).

share controlled by the gang.[6] Thus a gang's increasing market share leads to (weakly) increasing regulatory responsibility.

The other form of violence supplied by gangs in the model is deterrent, or combative, violence. This is the type of violence used to capture the gang's market share in the prison economy. However, we can think of this violence as deterrent because the model is concerned only with steady state equilibria, so there is no active conflict between gangs. Instead, this violence is meant to represent strategic cases of assault that are meant to maintain the credible threat of aggressive retaliation for any infringement upon the gang's authority. It may be more accurate to think a gang chooses its *capacity* for violence, implied by the number of members recruited as "soldiers".[7] However, a given capacity for violence is expected to result in a monotonically increasing amount of realized violence, which will come in the form of either strategic displays of the gang's willingness to defend their territory or as, perhaps less strategic, assaults by individual gang members reasserting their tendency for violence.[8] Regardless of the specific interpretation, a particular gang's market share is increasing in their own deterrent violence and decreasing in the deterrent violence of all other gangs. Furthermore, it is necessary for a gang to utilize some positive amount of deterrent violence in order to control a non-zero share of the market.

To clarify, this paper models two forms of violence directly supplied by prison gangs and excludes other violence, which may be subject to some influence by gangs but is not itself an action taken by any gang. Regulatory violence is supplied by each gang in their capacity as informal governance for the prison social and market setting. That is, regulatory violence is that which enforces property rights and market norms with the scope of a gang's authority, which increases with its market share. Deterrent violence, on the other hand, is that which is used by a gang to deter their rivals from encroaching upon their territory and therefore determines the size of gang's market share.

[6] There are interesting policy implications from this dimension of the inmate social system's self-governance. As implied in the example, lockdowns in response to serious infractions could induce greater efforts by gangs to suppress the related offending behaviors. The author is unaware of literature regarding such enforcement strategy, but doubts that this would be an entirely novel idea to most correctional staff.

[7] See Skarbek (2010) for a detailed discussion of the generic structuring of prison gangs.

[8] This assertion is consistent with empirical evidence that gang membership is associated with higher rates of violent misconduct (Drury and DeLisi 2011).

Gangs are modeled as straightforward profit-maximizers that earn revenues from market activity, incur direct production costs, and additional costs deriving from their supply of violence. Regulatory violence bears the risk of resulting in lockdowns that disrupt market activity and/or gang members suffering solitary confinement or other punitive action. The latter implies the loss, at least for a limited time, of a productive member of the gang and thus an opportunity cost. Regarding deterrent violence, increasing capacity implies expanding the gang's membership. Prison gangs operate with contracts, whether explicit or implicit, granting members access to resources, many of which are rival goods and thus imply an explicit cost of expanding membership. Additionally, the realized violence associated with deterrent violence bears similar, and perhaps greater, risks of punitive costs for the productivity of the gang. Therefore each type of violence has separate, possibly independent, marginal costs to the gang.

Policy dynamics are considered in the model setting with regards to how they impact factors that enter the prison gang's objective function. So punitive severity for violent misconduct is examined in terms of increasing the marginal cost of violence, whereas punitive policy regarding drug trafficking related offenses would be expected to impact the marginal cost of production. The third channel for policy to impact the optimal choices for gangs is through the demand for illicit goods, by such means as changing sanctions for possession of contraband or altering the total prison population size. By parameterizing shocks to these elements of the gang's optimization problem, comparative statics of the model are left open to interpretation for a variety of policy ideas.

3.4 The Base Model

Notation

Gang Variables

v_i Level of deterrent violence employed by gang i.

$\bar{v}_{-i}$ Sum of deterrent violence from rival gangs, i.e. $\sum_{j \neq i} v_j$.

P_i Market price offered by gang i.

r_i Level of regulatory violence employed by gang i.

c_i Marginal cost of illicit good production for gang i.

Policy Variables

θ Parameterized severity and likelihood of punishment for violent misconduct.

ρ Scale parameter for policy impacts on the demand for illicit goods.

Model Equations

$s_i = s(v_i, \bar{v}_{-i})$ Share of the market controlled by gang i. This share is strictly decreasing in the violence of rival gangs, $\bar{v}_{-i}$. For a given $\bar{v}_{-i}$, s_i is an increasing and concave function of v_i with $s(0, \bar{v}_{-i}) = 0$.

$q_i(P_i, \rho) = s(v_i, \bar{v}_{-i})Q(P_i, \rho)$ Demand for illicit goods from gang i, given s_i, at price P_i.

$f(v_i, r_i, \theta)$ Total cost of all violence employed by gang i.

The basic form of the model allows each gang in the prison to select their own supply of violence and a price at which to offer illicit goods in the portion of the market under their control. As mentioned previously, the gang's market share is an increasing function of their own violence, v_i, and a decreasing function of others' violence, $\bar{v}_{-i}$. The individual demand function subject to which a gang maximizes their profit is defined as a share of an aggregated demand function, $Q(P_i, \rho)$, which is the demand for illicit goods if there were a single supplier in the market. The effective assumption inherent to this is simply that each gang is a price-maker with regard to the goods that it supplies.

Proposition 1 assumes that the optimal amount of regulatory violence is an increasing and convex function of the market volume served by the gang, hence increasing in both market share and the overall market scale. The convexity of this function is justifiable via the observation that as a gang

Assumption 1 *The optimal supply of regulatory violence is dependent on the share of the market that is regulated and total volume of market activity, but independent of other factors. In addition, the marginal increase in the optimal regulatory violence as market share increases is positive and increasing. It follows that a gang's regulatory violence is an implicit function of deterrent violence, both its own and that of rival gangs.*

a) Regulatory violence: $r_i = r(s_i, \rho)$

b) Convexity: $\frac{\partial r_i}{\partial s_i} \geq 0$ *and* $\frac{\partial^2 r_i}{\partial s_i^2} \geq 0$.

c) Implicit Function: $\frac{\partial r_i}{\partial v_i} = \frac{\partial r_i}{\partial s_i}\frac{\partial s_i}{\partial v_i} \geq 0$ *and* $\frac{\partial r_i}{\partial \bar{v}_{-i}} = \frac{\partial r_i}{\partial s_i}\frac{\partial s_i}{\partial \bar{v}_{-i}} \leq 0$.

controls a larger market share they are necessarily responsible for regulating a subset of the prison population that is both larger and increasingly diverse. It is assumed that an increasingly diverse population demands greater efforts to maintain order due to the expanding variety of cultural norms and values of those subject to the gang's authority.

Under Assumption 1, v_i and P_i are the two choice variables of an individual gang. In the stationary setting provided here each gang simultaneously determines their own price and supply of violence. The concept of Nash equilibrium is utilized, requiring that the choices of each gang are a best response given the choices of all other players. The model distills the market incentives for the use of violence by prison gangs into a simple set of equations, which allow basic analysis of how policy parameters influence the supply of such violence.

3.4.1 The Policy Dimension

The two exogenous policy parameters, θ and ρ, as well as the marginal cost of production, c_i, are each able to represent a variety of different policy changes. There is not sufficient existing evidence to impose assumptions about the strength nor second order relationships of these parameters with any given policy. For example, assuming drug rehabilitation is effective in reducing the treated inmates' drug dependence, expanding the drug rehab program will reduce the illicit demand within the prison ($\Delta\rho < 0$). On the other hand, if food shipments are a major pipeline for drugs into the prison, $\Delta c_i > 0$ might be achieved by having shipments checked by drug-sniffing canine units.

However, the author is aware of neither anecdotal nor empirical evidence sufficient to justify any assumptions regarding the functional forms that any such policy effects may take. It is consequently left to subjective analysis which of these parameters a policy change will impact, while the model analytics illuminate the equilibrium effects of a given change in any particular parameter.

The marginal cost of production, c_i, is indexed by the subscript i denoting a single gang for two reasons; baseline marginal production costs can differ between gangs and the impact of a policy intervention may differentially impact the marginal production costs of gangs. There are a variety of logical rationale by which to expect the former, such as a gang whose racial identity aligns with a larger share of the total prison population may have easier access to a variety of smuggling options. Alternately, looking beyond the static nature of this model, a gang that established itself earlier than its competitors may have secured the most efficient means of smuggling or illicit production and thus claim a form of first-mover advantage with regard to production costs. The potential for differential policy effects on this variable is a natural extension of the previous points regarding the potential for baseline heterogeneity.

Heterogeneity in marginal production costs raises questions about the relative impacts of anti-competitive ($\Delta|c_i - c_j| > 0$) and pro-competitive ($\Delta|c_i - c_j| < 0$) policy interventions. Lacking a model with insights into the optimal responses of each gang, a prison administration may select a course of action based on cost minimization, or some other esoteric objective, and consequently allocate prison resources in an inefficient manner. For example, enforcement actions that target the smuggling activity of an individual gang that is the largest and most active in the prison may produce an unintended surge of illicit market activity and violent behavior from the other gangs.

Policy Variables – *The marginal cost of production, c_i, and the two policy parameters, ρ and θ, are exogenous factors by which policy intervention is possible. These relationship are defined as follows (with examples of relevant policy levers):*

1. *Market scale (ρ): This is a scale parameter representing any policy change that increases or decreases illicit market demand. Specifically, at a given price, illicit demand is strictly increasing in ρ.*
   - *Notation:* $Q(P,\rho) < Q(P,\rho'),\ \forall\ \rho < \rho'$. *Equivalently,* $\frac{\partial q_i(\cdot)}{\partial \rho} > 0$.
   - *Example 1: Expanded drug rehabilitation program* $\implies \Delta\rho < 0$.
   - *Example 2: Prison population growth* $\implies \Delta\rho > 0$.
2. *Punitive Index (θ): This parameterizes the expected cost of violent behavior, thus the marginal costs of both deterrent and regulatory violence are strictly increasing in θ. This can be representative of policy that changes the severity of punishment for misconduct, such as the length of solitary confinement, the likelihood of punishment, or lost revenues due to security protocols, such as lockdowns.*
   - *Notation:* $\frac{\partial^2 f(\cdot)}{\partial v_i \partial \theta} > 0$ *and* $\frac{\partial^2 f(\cdot)}{\partial r_i \partial \theta} > 0$. *(Surveillance technology, lockdowns).*
   - *Example 1: Decreased maximum stays in solitary confinement* $\implies \Delta\theta < 0$.
   - *Example 2: New surveillance technology* $\implies \Delta\theta > 0$.
3. *Marginal cost of production (c_i): The marginal production cost of a gang represents smuggling costs for illicit goods that are brought into the prison and the direct productions costs of illicit goods and services that are actually produced by inmates. Increasing c_i decreases the profit margin available to that gang. Since gangs often have distinct supply channels, and supply different illicit goods, policy impacts on c_i are permitted to vary between gangs.*
   - *Example 1: Increased number of external service providers* $\implies \Delta c_i < 0$.
   - *Example 2: New screening technology in visitation center* $\implies \Delta c_i > 0$.

3.4.2 Optimization

Notation for the Costs of Violence

Since the cost function for violence, $f(v_i, r_i, \theta)$, is a function of both types of violence and regulatory violence is an implicit function of deterrent violence, the marginal cost of deterrent violence is notationally cumbersome. Let $f_n(v_i, r_i, \theta)$ be the derivative of $f(v_i, r_i, \theta)$ with respect to its n^{th} argument, $f_{nn}(v_i, r_i, \theta)$ be the second derivative with respect to the n^{th} argument, and $f_{nk}(v_i, r_i, \theta)$ be the cross-partial with respect to the n^{th} and k^{th} arguments. Then the marginal cost of violence for gang i and its derivative are as follows:

- $\frac{\partial f(v_i,r_i,\theta)}{\partial v_i} = f_1(v_i, r_i, \theta) + f_2(v_i, r_i, \theta)\frac{\partial r_i}{\partial v_i}$
- $\frac{\partial^2 f(v_i,r_i,\theta)}{\partial v_i^2} = f_{11}(v_i, r_i, \theta) + f_{22}(v_i, r_i, \theta)(\frac{\partial r_i}{\partial v_i})^2 + 2(f_{21}(v_i, r_i, \theta)\frac{\partial r_i}{\partial v_i}) + f_2(v_i, r_i, \theta)\frac{\partial^2 r_i}{\partial v_i^2}$

Where possible, both of the above derivatives will be represented with the more concise term on the lefthand side of their respective equality.

The optimal choices, v_i^* and P_i^*, for gang i maximize the profit function shown in Equation 3.1. The profit function is a simple per unit profit margin multiplied by the quantity of goods sold, less the total cost of violence supplied to capture and regulate the gang's market share. Although there are implicit non-negativity constraints on the choice variables, these are redundant under very basic conditions, defined below, on illicit demand and the costs of violence. Assumption 2 introduces the basic conditions on the costs of violence that permit analysis of the first order conditions derived from Equation 3.1.

Recall that $q_i(P_i, \rho) = s_i(v_i, \bar{v}_{-i})Q(P_i, \rho)$, so v_i enters both terms of the profit function but P_i only enters the first term. This allows the notably simplified form of $FOC(P_i)$ presented below, where market share has been factored out and eliminated. The resulting $FOC(P_i)$ is the generic optimality condition for a monopolist subject to the demand function $Q(P_i, \rho)$. It follows that P_i^*

Assumption 2 *The marginal cost with respect to each type of violence,* $f_1(v_i, r_i, \theta)$ *and* $f_2(v_i, r_i, \theta)$*, is strictly positive and increasing. These partials are also non-decreasing with respect to the alternate form of violence. In notation:*

a) $f_1(v_i, r_i, \theta) > 0$ *and* $f_{11}(v_i, r_i, \theta) > 0$.

b) $f_2(v_i, r_i, \theta) > 0$ *and* $f_{22}(v_i, r_i, \theta) > 0$.

c) Cross-partials: $f_{12}(v_i, r_i, \theta) = f_{21}(v_i, r_i, \theta) \geq 0$.

is independent of the choice of v_i and can be treated as a determinant and increasing function of ρ. This result is subject to the condition that $\bar{P}_i > c_i$ where $Q(\bar{P}_i, \rho) = 0$, which is sufficient to guarantee an interior solution for P_i^*. Put simply, the maximum willingness to pay for illicit goods must exceed a gang's marginal production cost in order for that gang to participate in the market.

Profit Function:

$$\pi_i(v_i, P_i) = (P_i - c_i)q_i(P_i, \rho, s_i) - f(v_i, r_i, \theta) \tag{3.1}$$

First Order Conditions:

$$FOC(P_i): \qquad 0 = (P_i - c_i)\frac{\partial Q(P_i, \rho)}{\partial P_i} + Q(P_i, \rho) \tag{3.2}$$

$$FOC(v_i): \qquad 0 = (P_i - c_i)Q(P_i, \rho)\frac{\partial s_i}{\partial v_i} - \frac{\partial f(v_i, r_i, \theta)}{\partial v_i} \tag{3.3}$$

The first order condition on violence equates the marginal benefit of increasing market share with the marginal cost of additional violence. The latter combines both the direct cost of deterrent violence and the cost of the resulting increase in regulatory burden. The crux of the model implications lie within $FOC(v_i)$ and how the resulting v_i^* responds to any changes in the model.

It is not possible to explicitly solve $FOC(v_i)$ for the optimal v_i^*, unless the model is further restricted by imposing functional forms on $f(\cdot)$, $s(\cdot)$, and $Q(\cdot)$. However, the existence of $v_i^* > 0$, as well as the comparative statics with respect to $\bar{v}_{-i}$ and the exogenous policy parameters, can be analyzed without an explicit functional form. For this we define the RHS of $FOC(v_i)$ as the function

ϕ, shown in equation 3.4. This function is strictly decreasing in v_i, which implies the existence of $v_i^* > 0$ if and only if $\lim_{x\to 0} \phi(x, \bar{v}_{-i}, c_i, \theta, \rho) > 0$ and $\lim_{x\to\infty} \phi(x, \bar{v}_{-i}, c_i, \theta, \rho) < 0$. It is sufficient to require that $f(\cdot)$ is strictly convex in v_i, i.e. $f_{11} > 0$, and $\lim_{x\to 0} f_1(x, r_i, \theta) = 0$, which is a reasonable assumption given the setting despite being more restrictive than necessary.[9]

$$\phi(v_i, \bar{v}_{-i}, c_i, \theta, \rho) = (P_i - c_i)Q(P_i, \rho)\frac{\partial s_i}{\partial v_i} - \frac{\partial f(v_i, r_i, \theta)}{\partial v_i} \tag{3.4}$$

As noted below, ϕ is strictly decreasing in v_i and thus a shock that increases ϕ will lead to an increase in v_i^*. The derivative of ϕ with respect to deterrent violence from the gang's rivals varies depending on the functional form of $s(v_i, \bar{v}_{-i})$ and the relative magnitudes of the two inputs. Generally there is a "fight zone" for a mid-range of values for v_i^* where ϕ is increasing in $\bar{v}_{-i}$, with "flight" zones on the tails where the gang optimally responds to an increase in $\bar{v}_{-i}$ by decreasing its own deterrent violence. The relationships of ϕ and v_i^* with the policy parameters will be discussed in the duopolistic setting of section 3.5.

Partial Derivatives of ϕ.*

- Own violence: $\frac{\partial\phi}{\partial v_i} < 0$
- Rival violence: $\frac{\partial\phi}{\partial \bar{v}_{-i}} =$ indeterminate
- Production cost: $\frac{\partial\phi}{\partial c_i} < 0$
- Punitive costs: $\frac{\partial\phi}{\partial \theta} < 0$
- Market scale: $\frac{\partial\phi}{\partial \rho} > 0$

*Full derivations are shown in the technical appendix.

The variability of $\partial\phi/\partial\bar{v}_{-i}$ has important implications for understanding the behavior of prison gangs in response to changes in the prison setting. Consider the earlier policy example of increasing the marginal production costs for a single gang and the stated possibility of other gangs responding by increasing their use of violence. The targeted gang recognizes fewer benefits to market activity

9 Assuming the marginal cost of deterrent violence goes to zero is reasonable on the basis that the founding members of the gang are capable of supplying some amount of violence themselves and sufficiently small acts of violence can be committed without fear of detection and punishment.

and accordingly decreases its use of violence, as indicated by $\partial\phi/\partial c_i < 0$, but the response of rival gangs could be to increase or decrease their use violence, depending on whether or not their prior v_i^* falls in the competitive zone. Thus the endogeneity of the optimal v_i between gangs can suppress or compound the effects of a policy change or other exogenous shock. The next section takes a closer look at these implications by considering the duopolistic case of just two gangs.

3.5 Baseline Analysis: Duopolist Gangs

Notation

- $i \in \{1, 2\}$
- $\phi^1 = \phi^1(v_1, v_2, c_1, c_2, \theta, \rho)$
- $\phi^2 = \phi^2(v_1, v_2, c_1, c_2, \theta, \rho)$
- Derivatives of ϕ^i are denoted with subscripts in the same manner as $f(\cdot)$.
  - E.g. $\phi_2^1 = \frac{\partial \phi^1}{\partial v_2}$ and $\phi_2^2 = \frac{\partial \phi^2}{\partial v_2}$

Note that the v_i inputs now have a fixed order regardless of which ϕ^i is considered.

Consider the case in which there are only two gangs in a prison. Aside from any difference in their marginal costs of production, c_1 and c_2, they face symmetric optimization problems. The notation is adapted above to clearly identify the ϕ generated by the $FOC(v_i)$ of each gang. The system of equations represented by ϕ^1 and ϕ^2 is evaluated using the Hessian matrix of their partial derivatives. Equation 3.7 is the total differentiation of equations 3.5 and 3.6, with each set equal to zero to satisfy the individual gangs' optimality conditions.

$$\phi^1(v_1, v_2, c_1, c_2, \theta, \rho) = (P_1^* - c_1)Q(P_1^*, \rho)\frac{\partial s_1}{\partial v_1} - \frac{\partial f(v_1, r_1, \theta)}{\partial v_1} \tag{3.5}$$

$$\phi^2(v_1, v_2, c_1, c_2, \theta, \rho) = (P_2^* - c_2)Q(P_2^*, \rho)\frac{\partial s_2}{\partial v_2} - \frac{\partial f(v_2, r_2, \theta)}{\partial v_2} \tag{3.6}$$

$$\begin{bmatrix} \Phi_1^1 & \Phi_2^1 \\ \\ \Phi_1^2 & \Phi_2^2 \end{bmatrix} \begin{bmatrix} \partial v_1 \\ \\ \partial v_2 \end{bmatrix} + \begin{bmatrix} \Phi_3^1 \\ \\ \Phi_3^2 \end{bmatrix} \partial c_1 + \begin{bmatrix} \Phi_4^1 \\ \\ \Phi_4^2 \end{bmatrix} \partial c_2 + \begin{bmatrix} \Phi_5^1 \\ \\ \Phi_5^2 \end{bmatrix} \partial \theta + \begin{bmatrix} \Phi_6^1 \\ \\ \Phi_6^2 \end{bmatrix} \partial \rho = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix} \tag{3.7}$$

Define the first matrix in equation 3.7, the partial derivatives of ϕ with respect to violence, as matrix $\mathcal{A}$. Then the existence of a stable equilibrium requires that $det\mathcal{A} \geq 0$. This condition implies three separate equilibrium cases, each dependent upon the signs of the off-diagonal elements of $\mathcal{A}$. Those off-diagonal elements are the $\frac{\partial \phi}{\partial \bar{v}_{-i}}$ from the previous section, which are positive for some non-zero range of v_i and *potentially* negative for other values of v_i. A fourth case is possible, where $\Phi_2^1 \cdot \Phi_1^2 > \Phi_1^1 \cdot \Phi_2^2$ causes the model to diverge and there is no equilibrium. However, this requires extreme conditions on the functional forms, the intuition for which would imply that a gang's marginal net benefits are more responsive to changes in their rival's violence than their own. Furthermore, regardless of functional form assumptions, it is the comparative statics within the vicinity of existing equilibria that are of policy relevance. Thus the following discussion focuses on the three cases with a stable equilibrium defined by $\{v_1^*, P_1^*, v_2^*, P_2^*\}$ and the pertinent questions of how v_1^* and v_2^* react to shocks to the system.

The comparative statics with respect to each policy parameter can be evaluated by applying Cramer's Rule to equation 3.7, which results in equations 3.8 through 3.11. Allow Δ to represent the determinant of $\mathcal{A}$, which is positive. This set of equations only show the effect of each parameter on v_1^*. However, the symmetry of the model implies the same signs with respect to v_2^*, with the exception that equations 3.8 and 3.9 are then swapped. It can be assumed, without loss of generality, that $c_1 \geq c_2$ (which implies $v_1^* \leq v_2^*$).The individual partials of ϕ are easily signed by referring to the highlighted box at the end of last section, bearing in mind that ϕ_3^2 and ϕ_4^1 are the partials with respect to the *rival* gang's marginal production cost and are therefore equal to zero.

$$\frac{\partial v_1^*}{\partial c_1} = \frac{1}{\Delta}\Big(- \phi_3^1 \cdot \phi_2^2 - (-\phi_3^2) \cdot \phi_2^1 \Big) < 0 \tag{3.8}$$

$$\frac{\partial v_1^*}{\partial c_2} = \frac{1}{\Delta}\Big(- \phi_4^1 \cdot \phi_2^2 - (-\phi_4^2) \cdot \phi_2^1 \Big) \sim ?? \tag{3.9}$$

$$\frac{\partial v_1^*}{\partial \theta} = \frac{1}{\Delta}\Big(- \phi_5^1 \cdot \phi_2^2 - (-\phi_5^2) \cdot \phi_2^1 \Big) <^* 0 \tag{3.10}$$

$$\frac{\partial v_1^*}{\partial \rho} = \frac{1}{\Delta}\Big(- \phi_6^1 \cdot \phi_2^2 - (-\phi_6^2) \cdot \phi_2^1 \Big) >^* 0 \tag{3.11}$$

Unsurprisingly, a gang's optimal violence is decreasing in their own marginal production cost. Also, given that $\phi_4^1 = 0$, the sign of $\partial v_1^* / \partial c_2$ comes back to whether or not v_1^* is in the fight zone where the gang will respond to aggression with further aggression of its own, which is signified by $\phi_2^1 \geq 0$. This derives from the intuitive fact that c_2 influences v_1^* solely through its effect on v_2^*.

The dynamics of equations 3.10 and 3.11 are a bit more subtle. The inequalities in each of these equations include an asterisk to signify that the sign is accurate for a broad range of values but can be violated in very extreme cases, namely if ϕ_2^1 is negative and larger in magnitude than ϕ_2^2. Note that if this condition is true for both v_1^* and v_2^* then this is the fourth case when the model becomes divergent.[10] In the fight zone case when the cross partials with respect to violence are positive, the given signs are accurate and the feedback from the rival's response increases the magnitude of $\partial v_1^* / \partial \theta$ and $\partial v_1^* / \partial \rho$. For cases where v_1^* falls outside of the fight zone, it remains likely outcome is that the sign is negative for equation 3.10 and positive for equation 3.11, but the response to the policy change is muted by the endogeneity of each gang's choice of violence.

The basic comparative statics of the model illustrate an interesting point regarding policy interventions in this setting. The effectiveness of an intervention depends on the competitive nature of

[10] It follows that the signs given in equations 3.10 and 3.11 can only be violated for a single gang in the vicinity of an equilibrium.

gang interactions. When conditions are such that gangs respond to aggression of their rivals with further aggression of their own, policy changes are more likely to be effective in reducing the violence employed by these gangs. Alternately, under the same conditions, changes such as population growth are expected to cause significant increases in the use of violence. However, non-competitive conditions can suppress response to policy change. It is also possible for administrators or policy makers to move conditions towards the more competitive setting by targeting the marginal production cost, c_i, of a particular gang. The conditions of a particular v_1^* and v_2^* are determined by which of the following three cases they fall in.

Case 1: Strategic Complementarity – v_1^* and v_2^* such that $\phi_2^1 \geq 0$ and $\phi_1^2 \geq 0$.

Both gangs will respond in kind if their rival decides to supply more, or less, violence. Thus v_1^* and v_2^* behave like complements and a shock that increases c_2 will cause both v_2^* and v_1^* to decrease, albeit a more moderate response from v_1^*. This implies that the endogeneity of gang violence will intensify the individual response to any given policy shock. In other words, the overall response to increasing θ is greater than suggested by the partials with respect to it, ϕ_5^1 and ϕ_5^2. The competitive case indicates conditions that generally encourage high rates of violence, but this simultaneously means there are higher returns to efforts to discourage violence.

Case 2: Strategic Substitution – v_1^* and v_2^* such that $\phi_2^1 \leq 0$ and $\phi_1^2 \leq 0$.

Deterrent violence from one gang leads to a decrease in the marginal net benefits of the rival gang, causing the rival to decrease its use of violence. Naturally, the reverse is true as well. This provides to the aforementioned example where the decrease in v_2^* from increasing c_2 is partially offset by the response of v_1^*. The conditions of the Substitution case cause muted policy effects in the vicinity of such an equilibrium.[11] On the other hand, these conditions also may imply decreasing returns to scale with regards to population growth and the rate of violence.

[11] The Substitution case is also when it is feasible, if unlikely, that the signs flip on $\frac{\partial v_i^*}{\partial \theta}$ and $\frac{\partial v_i^*}{\partial \rho}$, but this can only be true for one of the i.

*Case 3: Split Response – either $\phi_2^1 \geq 0$ and $\phi_1^2 < 0$ **or** $\phi_2^1 < 0$ and $\phi_1^2 \geq 0$.*

One gang responds in kind to its rival's choices, while the other does the opposite. This case is likely to exist for a narrow bandwidth of v^* between Case 1 and Case 2, or for sufficiently different values of v_1^* and v_2^*. The comparative statics are split in this case, akin to Case 1 or 2 for each gang depending upon the sign of their ϕ_{-i}^i, with a net policy effect that is also a moderated version of the intensified or muted responses described above.

Unfortunately, it is difficult to meaningfully bound the conditions that determine each of these cases in the current state of the model. This is due to the number of margins that that shift simultaneously when a rival alters their choice of violence, each margin relating to an undefined functional form. The ambiguity is largely driven by two factors. First, the sign of $\frac{\partial^2 s_1}{\partial v_1 \partial v_2}$ is ambiguous and determines the sign of several terms in ϕ_2^1. In other words, it is possible for the rival's violence to increase or decrease the marginal effectiveness of v_1 in capturing market share. The other key factor is the relative importance of deterrent violence vs. regulatory violence in marginal cost of violence for the gang. To better illustrate the role of these, the next section considers the case of a common function for the division of market shares.

3.5.1 A Special Case: The Tullock Function

The Tullock Function is commonly used in contest theory for determining probabilities of success based on relative inputs of the of the interested parties. It is also a common approach to awarding shares of a fixed prize. The intuition is straightforward: each party's share, or probability of success, is determined by the effort exerted by that party relative to the total effort exerted by all parties. In the duopolistic case of two prison gangs, effort takes the form of deterrent violence and the prize is market share. Thus market shares are divided by the simple function in equation 3.12.

The derivatives of $s_1(\cdot)$ are highlighted in the box below. Continue to assume, without loss of generality, that $v_1^* \leq v_2^*$. Note that the cross partial is equal to zero when $v_1 = v_2$.

Duopoly with Tullock Market Shares

Functional form:

$$s_1(v_1, v_2) = \frac{v_1}{v_1 + v_2} \tag{3.12}$$

First and second partial derivatives:

- $\frac{\partial s_1}{\partial v_1} = \frac{v_2}{(v_1+v_2)^2} > 0$
- $\frac{\partial s_1}{\partial v_2} = \frac{-v_1}{(v_1+v_2)^2} < 0$
- $\frac{\partial^2 s_1}{\partial v_1^2} = \frac{-2v_2}{(v_1+v_2)^3} < 0$
- $\frac{\partial^2 s_1}{\partial v_1 \partial v_2} = \frac{v_1^2 - v_2^2}{(v_1+v_2)^4} \leq 0$ if $v_1 \leq v_2$.

As with the generic form of $s(\cdot)$, the cross-partial derivative cannot be globally signed.

$$\phi_2^1 = (P_1^* - c_1)Q(P_1^*, \rho)\frac{\partial^2 s_1}{\partial v_1 \partial v_2} - f_2(v_i, r_i, \theta)\frac{\partial^2 r_1\big(s_1(v_1, v_2)\big)}{\partial v_1 \partial v_2} - \Big(f_{12} + f_{22}\frac{\partial r_1}{\partial s_1}\frac{\partial s_1}{\partial v_1}\Big)\frac{\partial r_1}{\partial s_1}\frac{\partial s_1}{\partial v_2} \tag{3.13}$$

Equation 3.13 shows ϕ_2^1 in some detail. An even more detailed breakdown of equation 3.13, as well as discussion of when it is positive or negative, is provided in the technical appendix. Note here that when $v_1 = v_2$, the first two terms of equation 3.13 are equal to zero. The third term in the equation is strictly positive and therefore $\phi_2^1 > 0$ when $v_1 = v_2$. For some bandwidth around $v_1 = v_2$ the model must therefore exhibit Case 1 (Competition) from the previous section. As v_1 decreases, the first term of ϕ_2^1 becomes negative and the second positive. Meanwhile, the reverse occurs for the same terms in ϕ_1^2.

The other crucial point that we can take from equation 3.13 is that the second two terms are representative of how the the change in the gang's *regulatory* violence, due to its change in market share, affects the marginal cost of violence. It follows that if regulatory violence is a relatively small portion of the violence employed by gangs, the first term of equation 3.13 will play a larger role in determining how they respond to the violence of their rival. In such a setting, as we consider $v_1 < v_2$, we will quickly switch from Competition case to the Divided case, where $\phi_2^1 < 0$ and $\phi_1^2 > 0$. On the other hand, when regulatory violence is a major contributor to the marginal costs of violence then a gang's response is more likely to be determined by the last two terms of the equation. The

Proposition 1 *Marginal regulatory costs and equilibrium characteristics, subject to $N \geq 3$ and the Tullock function for market shares. The role of regulatory violence determines the equilibrium behavior of gangs in the following way:*

1. *The greater the association between regulatory violence and market share, the greater the likelihood that gangs will respond to violence from their rivals with increased violence of their own.*
2. *The lesser the association between regulatory violence and market share, the greater the likelihood that gangs will respond to violence from their rivals by decreasing their own violence.*

outcome in such a setting is less clear, since even when the first term of equation 3.13 is dominated by the other two, the sign of ϕ_1^2 will depend on the relative magnitudes of those last two terms.

At this stage the more interesting implications come when the duopolistic assumption is discarded. With $N \geq 3$ and the Tullock function in equation 3.12 adjusted properly, the numerator of the cross-partial derivative becomes $v_1^2 - \bar{v}_{-i}^2$. Therefore, excluding the case when a single gang dominates at least half of the market, we now have $\frac{\partial^2 s_i}{\partial v_i \partial \bar{v}_{-i}} < 0,\ \forall i$. Then, for each gang, the first term of equation 3.13 is negative and the other two are positive. This results in much clearer implications when considered in conjunction with the relevance of regulatory violence in determining marginal costs, which are detailed in Proposition 1.

In either of the cases laid out in Proposition 1, if a single gang gains sufficient comparative advantage ($c_i < c_j,\ \forall j$) to be significantly dominant in the markets, it is possible for that gang to violate the behavior suggested in the proposition. This would then be the Divided Case from Section 3.5, rather than the Competition Case or the Substitution Case.

3.5.2 Threats to the Tullock Function

The Tullock function is a commonly applied functional form because it has the intuitively pleasing property of apportioning shares to each party proportionately to their share of the total efforts supplied. At the same time, the greatest threat to its application in this paper may be the very fact that effort, or violence, is equally productive for all parties. In its basic form, as applied to the gang setting above, no one gang has an advantage over another. Yet there to two rationale by which such an assumption can be challenged.

First, it has been established that prison gangs largely define their respective "jurisdictions" along racial lines. This directly impacts the regulatory responsibility of a gang. However, it is also reasonable to speculate that having the racial composition of the prison lean in its direction may increase the effectiveness of a particular gang's efforts to capture market share. If such is the case, then the behavior of gangs in prisons with imbalanced racial composition may not fit well with the implications provided under the assumption of the Tullock function. However, this concern is at least partially mitigated by the observation that in prisons with large populations of a single racial identity it is common for gangs to further divide into more narrowly defined associations. For example, hispanic prison gangs in California are well known to be divided by southern and northern hispanic origins, respectively Sureños and Norteños.

The second threat regards the influence of correctional officers. It is natural to expect that, despite maintaining a contrary public image, a gang's leadership may seek mutually beneficial cooperation with correctional officers. Although such arrangements are most likely to characterized as leading to heterogeneity in marginal production costs or even in the marginal costs of violence, it is feasible for a gang to use this to leverage an advantage in capturing market share. Such an advantage could also lead to outcomes that do not align well with the model implications.

While there are prison settings in which the merits of the Tullock function may be questionable, the general characteristics of the function are taken to be broadly applicable. Furthermore, the results discussed in Proposition 1 are truly only dependent upon the expectation that, within the relevant range of equilibria, the marginal effectiveness of deterrent violence in capturing market share is decreasing in the deterrent violence of rival gangs (i.e. $\frac{\partial^2 s_i}{\partial v_i^* \partial \bar{v}_{-i}} < 0$).

3.6 Discussion

The model presented in this paper provides a new framework for understanding the portion of prison violence that is a direct result of gang activity. The novel element of the model, derived from qualitative research by Skarbek (2010, 2012), is the treatment of prison gangs as profit-maximizing suppliers of illicit goods and informal governance, the latter as a consequence of the former. The broad intent of this is to provide a starting point from which future research can extend the model to study the profit motives and behavior of gangs in the wide variety of specific settings that arise in heterogeneous prison environments across the United States. Yet even the basic framework presented here provides some interesting insights and policy implications.

The model makes an explicit connection between opportunity for profit and the violent behavior of prison gangs, on the assumption that gangs are profit-maximizing entities. Although this result is expected, it highlights the potential effectiveness of a variety of policies that decrease the profit margin available to gangs as a means to limiting their violent behavior. This can include very direct action, like the targeting of smuggling operations to increase production costs, or much more obscure administrative action, such as expanding the drug rehabilitation program. Removing products from the list of contraband, pornographic media for example, and allowing them to be offered by the prison commissary is another way to reduce demand in illicit markets and therefore lead to a marginal decrease in gang motivated violence, as demonstrated by $\frac{\partial \phi}{\partial \rho} > 0$. A related area that has seen recent controversy is the market for cellular phones in prisons. Prisoner rights groups have suggested that cell phones should simply be permitted, claiming this would help prevent the violence that has been attributed to their trade. Meanwhile some prisons are considering a very different approach to the problem, using new technology to block all cellular signals in areas of the prisons.[12] In either case, the model provides a framework for considering the effect of such policy changes on the market incentives of the prison gangs.

[12] See https://www.nij.gov/topics/corrections/institutional/contraband/Pages/cell-phones.aspx

A measure of caution is necessary in the interpretation of the results derived from this modeling environment. Any particular policy considered is likely to have other dimensions of merit, or demerit, for consideration, as is illustrated in the pornography and cell phone examples above. It is certainly expected that a prison administrator considering implementation of a new policy to deter violence would be keenly aware of the multidimensional impacts that may ripple through a prison environment that includes a starkly closed economy with unique social and behavioral elements. Alternately, the potential policy effects on illicit market incentives that we derive from this model may themselves be tertiary effects of a new policy, which, in the absence of this framework, might be overlooked in evaluating that policy. Consider a new prison warden who wishes to crack down on correctional staff who transport contraband into the prison. It is likely that his or her motivation in this regard has little to do with the rate of violence in the prison, but the additional benefit of reduced violence, implied by $\frac{\partial v_i^*}{\partial c_i} < 0$, should still play a role in the decision to devote scarce prison resources to such an effort.

The possibility of equilibria with very different comparative statics leads to interesting policy ramifications. Recall the three cases in section 3.5, where the response of each gang to the violence of their rival was related to the overall effectiveness of new policy. It was noted that gang violence is expected to be highly responsive to policy interventions in the Strategic Complements case, but policy response is muted in the Strategic Substitutes case. Furthermore, under likely conditions for the division of market shares, Proposition 1 puts forth that decreasing association between regulatory violence and market share leads to the Substitution case and muted policy effects. This indicates that in prison populations where inmates require less coercion to abide by market norms, such as minimum-security facilities, gang violence will be less responsive to policy change than it is among more unruly populations, such as maximum-security facilities.

The relevance of these research findings is evident in recent criminal justice news from California. Officials at the California Department of Corrections and Rehabilitation (CDCR) announced their intent to phase out the use of Sensitive Needs (SNY) facilities as a means of segregating vulnerable populations from the general prison population[13]. The stated reason for this decision is that the SNY facilities have not resulted in the expected reductions in violence, nor have officials been successful in preventing gangs from developing within the SNY facilities. However, under the assumptions of this framework, creating a distinctly safer prison environment through segregation alone is unlikely to be successful unless the segregated population has very little demand for illicit goods. Although preferences may vary across populations, it is difficult to imagine that the SNY population would exhibit significantly less demand than the general prison population of the state. Even in the previous example of segregation in this model, comparing minimum- and maximum-security populations, there may be lower rates of regulatory violence in the Strategic Substitution outcome of minimum-security facilities but rent-seeking behavior will still motivate gang activity and deterrent violence. The most direct inference of the model suggests that suppressing gang activity in a diverse prison population, beyond the point where punitive efforts become ineffective, must entail disrupting the illicit demand for goods or make the production/smuggling of illicit goods prohibitively costly[14].

A meaningful comparison of the relative merits of θ-policy and ρ-policy is not readily available in the current state of model. What can be said of both policy parameters, with regards to decreasing gang violence and under the assumption of basic regularity conditions on the derivatives of each parameter, is that each exhibits a form of decreasing returns to scale. In the case of θ, the strict convexity of the cost function means that proportional ratcheting up of marginal costs has a decreasing effect on overall violence as v^* grows smaller. With regard to ρ, the proportionate scaling down of illicit demand has a similarly diminishing marginal effect on v^*. It follows that to achieve a given reduction in gang motivated prison violence, if that reduction is sufficiently large then it is never optimal to focus on one type of policy intervention, θ or ρ, to the exclusion of the other.

As mentioned in the setup of the model, marginal production costs are the key source of heterogeneity between gangs. With regards to the discussion of suppressing gang activity, given the

[13] Article from the Sacramento Bee available at http://www.sacbee.com/news/local/crime/article211942034.html

[14] Making production and smuggling costly is straightforward with some goods but can be exceedingly difficult with others, particularly services typically offered such as protection and the sex trade.

assumption of constant marginal production costs, it is technically possible to raise c_i sufficiently high to eliminate all illicit market activity. However, such a case is unrealistic and consequently uninteresting. On the other hand, there are many interesting implications of targeted policy and the influence of dominant gangs. For example, in the case of Strategic Substitution there is a potential persistence issue where, once reaching a sufficiently dominant market share, a dominant gang effectively becomes embedded. That is, the behavior of the dominant gang switches from substitution to responding with aggression, whereas the other gangs still respond with less aggression. This creates a structural resistance to the dominant gang losing market share. When targeting the marginal production costs of individual gangs is possible, raising the c_i of the dominant gang can help prevent them from reaching that level of market dominance. Conversely, allowing a dominant prison gang to further decrease its own c_i, perhaps through coordination with compromised prison staff, has the potential to both increase the influence of that gang and further reinforce its continuing control of the marketplace.

Testable Implications

Testable implications are a challenge in this setting. The first concern is the limited granularity of existing data and the second is the very nature of prison violence itself. For the latter concern, note that what is regarded in this research as gang violence constitutes only a portion of the total violence that occurs in the prison environment. Fundamentally, this research codifies violent behavior according to its motivation, so as to consistently model the determinants of such behavior, but motivation is not directly observable. Just as gang violence is distinguished by two separate motivations in this paper, individually motivated violent behavior could also be demarcated according to a set of characteristics or the underlying motivation.[15] For simplicity, consider there to be just three types of violence: regulatory gang violence, deterrent gang violence, and individual violence. Each of these are distinguished by their motivation rather than an easily observable characteristic and therein lies the challenging interaction with the state of existing prison data. As mentioned previously, even

[15] In Kurzfeld (2018) a distinction is made between extemporaneous individual violence and personal vendettas.

consistent measurement of gang membership remains questionable in practice. Existing data does not provide the necessary detail to consistently collate incidents into gang violence and individual violence, much less the finer detail needed separate regulatory violence from deterrent violence.

Two likely paths are available to increase the feasibility of testing the model implications. The obvious path is improving granularity in the data available from prison administrations. At a minimum, the ability to consistently determine whether assailants are gang affiliated would greatly improve the prospect of related hypothesis testing. A second path is development of a model for individually motivated violent behavior, to then be paired with this model. The current issue is that a policy intervention expected to decrease the violence of prison gangs might simultaneous have a similar, or inverse, impact on individual violence and the two cannot be distinguished from one another. A comprehensive framework will improve the likelihood of developing hypotheses that are testable even in the absence of the ability to distinguish between different types of violence.

Entry and Exit

An interesting dimension to the strategic interaction of prison gangs, not directly explored in this paper, is the possibility of entry and exit. This could be introduced to the model via a fixed cost of entry, perhaps due to a discontinuous increase in regulatory and/or deterrent violence upon formation of a gang. The obvious question would be whether fewer gangs, meaning less competition over market rents, leads to more or less violence in the prison. The one clear indication shown in this paper was that moving from the duopoly case to $N \geq 2$ increases the likelihood of a negative sign on the cross-partial derivative of market share. However, it is still dependent upon the importance of regulatory violence to determine whether this results in the case of Strategic Complements or Strategic Substitutes. The former would imply a likely increase in total violence and the latter a likely decrease. Given that there are several margins of the optimality condition shifting simultaneously, a full treatment of the comparative statics and some additional model assumptions are necessary in order to report definite results on this topic.

A further extension of the model, to thoroughly examine entry and exit, could adapt this framework to a dynamic setting and even allow deterrent violence to function more like a stock of capital. Insight into the dynamic interaction of gangs, such as strategic exclusion, could be highly valuable to prison administrators. Unfortunately, the complexity of this extension is quite cumbersome, particularly given that it would entail accounting for the dynamics of open conflict between gangs.

3.7 Conclusion

The model introduced in this paper provides the framework for conceptualizing prison gangs as profit-maximizing entities that utilize violence in their role of providing informal governance to the marketplace in which they operate. It characterizes several dimensions of the ways in which prison gangs simultaneously interact with one another and the prison administration. Several conclusions regarding prison policy are drawn from the model and its comparative statics. These conclusions are indicative of general characteristics of effective policy, such as balancing punitive policy with market-based approaches, and also highlight the potential for great variation in policy responsiveness across prison environments. The latter refers to the cases of Strategic Substitution, when gangs will respond to an exogenous decrease in their rivals' violence by increasing their own, and Strategic Complementarity, when gangs will instead respond by also decreasing their violence. Clearly any given policy will be more effective in the case of Strategic Complementarity, although that same setting is likely to be characterized, ceteris paribus, by higher overall rates of violence. The general framework is also shown to be useful in conceptualizing the potential connections between superficially unrelated policy and the conduct of prison gangs, as in the example where relaxing constraints on pornography might lead to a decrease in gang violence.

Within the breadth of research on the behavior of prison gangs, the intent of this work is to offer a modest first step in applying the concept of the prison gang as a profit-maximizer to an explicit modeling framework. As such, the model as presented here invites further analysis, adaptation, and a variety of extensions. Dynamic optimization and the possibility of entry and exit were discussed in Section 4.4. Another direct extension is inclusion of a social planner to model the objectives and constraints of the prison administrator. Other possibilities include relaxing the assumptions on regulatory violence and instead modeling it as an additional choice variable; allowing price competition by relaxing the monopolistic market power assumption; incorporating the goods available through the prison commissary as a numeraire market good that is a substitute for illicit goods. Each of these offer potential advancements to the understanding of behavior in the prison environment, which is critical to the development of informed correctional policy and justice reform efforts.

The fundamental contribution of this work is summarized in the following statement. If we accept the idea that rent-seeking behavior is the core motivation of gang activity, then efforts to eliminate gang activity without confronting the rent-seeking opportunities in the prison are, at best, ill-conceived. As in the case of the CDCR's Sensitive Needs facilities, if market rents remain accessible then some entrepreneurial group of inmates will inevitably organize to capture those rents. At its core, this is no different than the multitude of settings in which economists study rent-seeking behavior. The unconventional means of rent-seeking in prison are a result of an unconventional market setting and therefore require a unique framework for analysis.

CHAPTER 4

Blind and Unbiased: An Impact Analysis and Discussion of Eyewitness Reforms[1]

4.1 Introduction

Eyewitness testimony has long been a cornerstone of criminal justice practices in the United States. To this day, the value of firsthand testimony in determining criminal culpability is without dispute. Yet the proper processes and procedures for generating reliable eyewitness identifications are themselves quite disputable. The concern at the center of this debate is that eyewitnesses make mistakes, and these mistakes can send innocent people to prison. Notably, this concern has been validated via evidence provided by exonerations (Gross, Jacoby, Matheson, Montgomery, and Patil, 2005). According to the National Registry of Exonerations, there were 37 exonerations that involved a false eyewitness identification in 2017 alone.[2] In an effort to alleviate the risk of such false convictions, researchers have recommended a number of reforms to improve the objectivity of law enforcement's eyewitness procedures. There is abundant experimental evidence that the recommended reforms significantly reduce the risk of false identifications, with the caveat that this is accompanied by an

[1] Special thanks to the Presley Center for Crime and Justice Studies and the University of California Consortium on Social Science and Law for their generous support of this research. I am also grateful to Steven Clark, Molly Moreland, and Melissa Blackford for their invaluable contributions to this research. All errors are my own.

[2] Report available at http://www.law.umich.edu/special/exoneration/Documents/ExonerationsIn2017.pdf.

increased risk of non-identification of guilty suspects (Clark, 2012). Application of the recommended reforms is already well underway in many states and counties around the country. In this study, we use nationwide crime data from the FBI Uniform Crime Reports to examine the effect of such eyewitness reforms on police clearance rates.[3] We find evidence of heterogenous reform effects across the states that implemented reforms, and we are also able to place a conservative lower bound on the possible reduction in clearance rates in the "worst case" reform scenario. The former contribution illustrates the need for further field research on this topic, while the latter helps to eliminate fears that reforms could lead to a dramatic increase in the number of guilty suspects who escape prosecution.

The reforms recommended by researchers include a number of distinct components. It was already widely acknowledged that presenting the witness with a full lineup of individuals is preferable to presenting the suspect to them alone, a procedure known as a "showup." The classic presentation of a lineup entails showing the suspect to the witness among a group of "fillers," also referred to as "foils," and asking if the person the witness saw is present. The lineup can be presented in-person or in a photo array of mugshots. The reforms include four major recommendations, each intended to reduce bias and the possibility of external influence: (a) the individual administering the lineup should be unaware of the identity of the suspect; (b) the lineup should be composed in such a way that the suspect does not stand out from the foils; (c) the witness should be given unbiased instructions, including the fact that the perpetrator of the crime may not be present in the lineup; (d) the administration of the lineup should use a sequential procedure rather that presenting all individuals simultaneously. In addition, some reforms include specific instructions for the handling of multiple witnesses and/or call for a statement of confidence from the witness.

The common objective of all four major reform components is to reduce the risk of false identifications. However, each component reduces risk in one of two subtly distinct manners. The first two, blind administration and lineup composition, reduce the risk that external factors will influence the

[3] Clearance rate refers to the rate at which reported crimes are "cleared" by an arrest being made.

witness to "correctly" select the suspect, regardless of the suspect's actual guilt. On the other hand, unbiased instructions and sequential presentation are recommendations that aim to mitigate the tendency of witnesses to assume that the perpetrator is present in the lineup. Such beliefs can lead witnesses to make *relative judgements*, wherein they select the individual that, relative to the other members of the lineup, *most* resembles their recollection of the perpetrator and they are unwilling to reject a lineup that they might have otherwise (Wells, 1984). In each case, the reform can lead to reductions in both correct identifications and false identifications, but unbiased instructions and sequential presentation are more likely to increase non-identifications. Note that all of these reform components should lead to a decrease in clearance rates, assuming that eyewitness evidence increases the likelihood of prosecution.

The debate regarding eyewitness reform efforts has mainly centered on the adoption of sequential lineups rather than simultaneous lineups. One explanation for the focus of controversy around this one reform component is that the other components can be reasonably justified as increasing the objectivity of the eyewitness procedure. On the other hand, simultaneous and sequential presentation of the lineup are simply two different ways of framing the test, one of which tilts the outcome towards a false positive and the other towards a false negative. This issue is analogous to the classic tradeoff in statistical methodology between type 1 and type 2 errors. In effect, by reducing the likelihood of one type of error (false identifications) we inadvertently increase the likelihood of another type of error (non-identification of a guilty suspect). The subjective view of which type of error is more egregious has more bearing on the value of sequential lineups because it is not preceded by an preference for objectivity.

The sequential lineup procedure is proposed by Lindsay and Wells (1985) as a means of minimizing the risk of the relative judgements that are a threat in the simultaneous procedure. Rather than presenting all of the potential suspects at once, the witness is informed that they will view a number of potential suspects whom they will be shown one at a time. The witness is asked whether each lineup member is the person they saw, before proceeding to the next. The witness is thus induced to make an up or down decision on each potential suspect without the ability to make direct comparisons with the others. There is very little dispute about whether the sequential process reduces false identifications (at least in laboratory conditions), but there is a highly debated concern regarding the potential reduction in correct identifications. A rich literature debates whether the risk to correct identifications is sufficient to warrant questioning the superiority of the sequential method over the simultaneous method (Clark, Howell, and Davey, 2008; McQuiston-Surret et al., 2006; Steblay et al., 2001; Steblay et al., 2011). The debate over sequential superiority is informed by numerous laboratory experiments that have generated a wealth of data. This research adds to what is presently a very limited body of evidence that draws on data from reforms to actual law enforcement practices.

The connection between laboratory evidence and the expected outcomes of eyewitness reforms in the field is confounded by a number of factors. First, there is a distinct informational asymmetry between the laboratory setting and the law enforcement setting. There are two possible cases for any given lineup: either the perpetrator is present, or the perpetrator is absent. Researchers in the lab know this information with certainty, whereas investigators gathering eyewitness evidence presumably have widely varying degrees of confidence regarding the guilt of suspects they bring before an eyewitness. The aforementioned exonerations of individuals convicted on the basis of eyewitness evidence demonstrate that there is a significant risk of an innocent suspect being falsely accused. Yet a suspect may also be falsely identified and never exonerated, or falsely identified but cleared by new evidence prior to prosecution. The witness may also identify one of the fillers in the lineup who is already known to be innocent of the crime in question. Hence false identifications pose a threat to the just application of the law, but the exact benefit of a laboratory-implicated decrease in the rate of false identifications is heavily dependent on other factors in the system of justice. Furthermore, there are analogous concerns with the interpretation of an implied decrease in the rate of positive identifications.

Another factor that distinguishes the field setting is an inability to distinguish the effects of individual reform components. Although the elements of a reform can be individually studied in the laboratory setting, the reforms adopted by police departments and other law enforcement agencies invariably confound the implementation of multiple reform components. This issue could be circumvented in regression analysis if there were sufficient variation in the composition of reforms. However, most jurisdictions that adopt reforms include all or most of the major reform components mentioned above. Table 4.3 shows the limited variation among the reforms used in this research. Thankfully, each reform component shares the common objective of reducing the risk of false identifications and also carries the same potential cost of reducing correct identifications. It is therefore a reasonable starting point in empirical analysis of field data to treat all eyewitness reforms uniformly.

To date, there has been very little empirical research studying the effects of eyewitness reforms in the field. In addition, the existing research tends to come in the form of case studies that examine localized reforms and compare outcomes in terms of suspect identifications, filler identifications, and non-identifications (Klobucher, Steblay, and Caliguiri, 2006; Mecklenburg, Bailey, and Larson, 2008; Wells, Steblay, and Dysart, 2015; Wixted et al., 2016). In this study we take a broad-based approach, examining the impacts of numerous statewide reforms using nationwide longitudinal data from the FBI Uniform Crime Reports. Rather than looking at the rates and types of identifications, the primary outcome of interest in this study is the clearance rate of the robbery crime category, where clearance rate is defined as the rate at which verified criminal offenses are "cleared" by an arrest being made. The limitations of using an outcome that is one step removed from the identification itself is counterbalanced by providing a within-jurisdiction counterfactual for post-reform outcomes by comparing the robbery clearance rate with that of crimes that are not likely to have an eyewitness, such as burglary and auto theft. Using dual-identification strategies and treating reform states individually or as a treated group, we show that statewide reforms do not appear to lead to a significant decrease in robbery clearance rates. On the surface, this is most consistent with a "low-cost" perspective of sequential superiority. However, we discuss several potential explanations that also account for the limited identifiable impact of these reforms and emphasize the need for further field research in this area.

This paper proceeds as follows: Section 2 provides further background on existing research and also frames the challenges inherent to the empirical exercise of this and other field research. The empirical strategies and results are described in Section 4.3. Lastly, section 4.4 provides a discussion of the empirical estimates, their implications for future work, and a brief conclusion.

4.2 Framing the Issue

Early modern legal institutions adopted the use of simultaneous lineups, ostensibly to increase the objectivity of eyewitness evidence. The alternative, now known as "showups," was to simply show the suspect to the witness and ask if they are the individual that the witness saw. Showups are still occasionally used in the field, where law enforcement argue that they can be conducted within minutes while the witness' memories remain fresh, rather than waiting for a lineup to be composed (Calandra and Carey, 2005; Clark, 2012). However, legal experts and court rulings have roundly criticized showups as susceptible to bias and manipulation, particularly where there is a risk of racial bias (Stovall v. Denno, 1967; Wagenaar and Veefkind, 1992; Wisconsin v. Dubose, 2005; Yarmey, Yarmey, and Yarmey, 1996).

The simultaneous lineup procedure first began to face critical examination after Wells (1984) hypothesized that the procedure would induce bias. Wells posits that witnesses are prone to the belief that the police have found the perpetrator and their role is to identify the *correct* person in the lineup shown to them, as opposed to informing the officers *if* the person they saw is present in the lineup. This hypothesis implies a form of relative judgement that biases the witness towards making an identification regardless of whether the perpetrator is actually present in the lineup. In practice, there are five potential outcomes of an eyewitness lineup. The witness may correctly identify the perpetrator of the crime (a Correct ID), identify an innocent suspect (a False ID), identify an innocent filler from the lineup (or a Foil ID), or reject the lineup when the suspect is innocent or guilty (a Correct Non-ID and a False Non-ID, respectively). The form of relative judgement hypothesized by Wells implies higher rates of all ID outcomes and lower rates of all Non-ID outcomes. Lindsay and Wells (1985) propose the sequential lineup procedure as an alternative that is resistant to the bias of relative judgements because the witness must select or reject each member of the lineup before viewing the next. The proposed sequential lineup inspired dozens of experimental comparisons of the two lineup procedures in psychology labs across the United States. Although the lab evidence roundly supports the hypothesis that sequential lineups reduce false identifications, relative to their simultaneous counterparts, it also gave rise to a lively debate over the associated cost in terms of correct positive identifications (Clark et al., 2008; McQuiston-Surret et al., 2006; Steblay et al., 2001; Steblay et al., 2011).

4.2.1 The Sequential Superiority Debate

The debate over sequential superiority arose over early claims that the sequential lineup is a "no-cost" improvement over the simultaneous procedure. Lindsay and Wells' (1985) own lab tests reported that sequential lineups resulted in a "reduction of inaccurate identifications without loss of accurate identifications", and this finding was echoed in several later studies (e.g., American Bar Association, 2004; Devenport, Penrod, and Cutler, 1997; Lindsay et al., 1991). However, the data from those very same studies does show reductions in the rate of accurate identifications, albeit smaller in magnitude than the reductions in the rate of false identifications.

More recent research frames the controversy more in terms of "low-cost" sequential superiority, wherein there is a reduction in accurate identifications that is negligible relative to gains in reducing false identifications, versus a view that the tradeoff bears further investigation before arriving at a firm conclusion. Drawing on a significant accumulation of laboratory evidence, one meta-analysis acknowledges a 0.15 decrease in the rate of accurate identifications, but dismisses this cost as being small or non-existent under "the most realistic" simulations of crime and police procedure (Steblay et al., 2001). Some of the same authors later adjust their position by acknowledging the cost to accurate identifications, which they show fall by only 0.08 in an expanded meta-analysis, but they argue for sequential superiority based on a strictly higher posterior probability of guilt when a suspect is identified (Steblay et al., 2011). That is, regardless of the probability that the perpetrator is present in the lineup, Steblay et al. (2011) find that a suspect identified from a sequential lineup has a higher probability of guilt than one identified in a simultaneous lineup.

However, several other meta-analyses have found the magnitude of the decrease in accurate identifications to be sensitive to the selection of evidence included in the study. For example, McQuiston-Surret, Malpass, and Tredoux (2006) note that there is great deal of heterogeneity across studies in both research design and procedural elements for administering lineups. In their meta-analysis they demonstrate, by parsing the data several different moderators, that the reductions in both accurate and false identifications are sensitive to these design elements. Furthermore, in a study critical of the "no-cost" view of sequential lineups and other eyewitness reforms, Clark (2012) finds that a meta-analysis using a slightly different selection criteria from that used by Steblay et al. (2011) results in estimates for which the decrease in the rate of accurate identifications is on par with that of false identifications. Although we do not refute the possibility of "low-cost" sequential superiority, the variation in findings across these studies is certainly sufficient to question the most consistent approach to measuring costs and benefits. This motivates the approach of this study, which looks directly to legal outcomes when sequential lineup reforms are applied in the field.

4.2.2 Laboratory Versus Field Outcomes

A major issue in determining the value of laboratory evidence in this setting is the aforementioned difference in information available to the administrator. In a lab, the scientist is fully aware of whether a lineup includes the perpetrator or not, whereas investigators simply include their suspect in the lineup and may have a varying degree of confidence in the guilt of that suspect. Consequently, while laboratory outcomes can be classified as accurate (Correct IDs) or false identifications (False

IDs and Foil IDs), a field lineup results in one of three outcomes: the suspect is identified, which could be a Correct ID or a False ID; a filler is identified, which is a known Foil ID; or the witness is unable to identify any member of the lineup, which could be a Correct Non-ID or a False Non-ID. Notation for these outcomes is specified in the highlighted notation box below. Regardless of which outcome is realized, the value to the legal process lies in the informational content by which the probability of that suspect's guilt can then be updated. In theory, this posterior probability of guilt directly informs the decision to prosecute a suspect and later informs decisions of the judiciary. The direct link between lineup outcomes and the decision to prosecute is the basis of the empirical hypotheses tested later in this paper.

Steblay and colleagues (2011) base their argument for sequential superiority on an increasing posterior probability of guilt observed in their data, a claim which correctly identifies the posterior probability of guilt as a key factor in assessing the probitive value of an eyewitness procedure. However, they argue for sequential superiority based solely on the posterior probability of guilt *given identification* and ignore the half of the story when the suspect is not identified by the witness. Narrowing the focus to exclude the posterior probability when a witness does not identify the suspect fails to recognize the costs of declining to prosecute those suspects who are not identified despite being guilty of the alleged crime. The potential tradeoffs along these dimensions, as a result of eyewitness reforms, are illustrated in table 4.1, where a hypothetical eyewitness reform is presented with an escalating "cost" in terms of less success identifying guilty suspects.

Table 4.1 presents a hypothetical situation in which 50% of suspects placed in lineups are guilty of the alleged crime, and specifies identification rates and the consequent posterior probabilities in a number of cases. Since 50% of suspects are guilty, the prior probability of guilt, P_0, is equal to 0.5. An objective of collecting eyewitness testimony is to increase this level of certainty. The first column of the table imposes a set of probabilities in the top panel for the witness to make an identification given the guilt or innocence of the suspect. In the middle panel of the same column, we see that the posterior probability of guilt increases by 21 percentage points for suspects who

Notation: Outcomes and Probabilities for Field Lineups

Field Lineup Outcomes:

- **ID**: The suspect is identified by the witness.
  - $(ID|Guilty)$ is a Correct ID.
  - $(ID|Innocent)$ is a False ID.
- **F**: A filler is identified by the witness – Foil ID.
- **NID**: No identification by witness.
  - $(NID|Guilty)$ is a False Non-ID.
  - $(NID|Innocent)$ is a Correct Non-ID.
- **N**: Joint negative identification outcome implied by F or NID.

Probabilities

- $\mathbf{P_0}$ = Prior probability that a suspect is guilty.[a]
- $\mathbf{Prob(ID|Guilty)}$; Probability suspect is identified given that they are guilty.
- $\mathbf{Prob(ID|Innocent)}$; Probability suspect is identified given that they are innocent.
- $\mathbf{Prob(F)} = Prob(F|Guilty) = Prob(F|Innocent)$; Probability of a Foil ID.
- $\mathbf{Prob(ID)}$ = Prior probability the suspect is identified by the witness.[b]
- $\mathbf{Prob(G|ID)}$ = Posterior probability of suspect guilt given they've been identified in lineup.
- $\mathbf{Prob(G|N)}$ = Posterior probability of suspect guilt given they were *not* identified in lineup.

[a] In the lab setting this is determined by frequency with which lineups include the actual perpetrator of the crime. In the field, the probability of a suspect's guilt is determined by the accumulated evidence against him/her, or in an aggregate sense by the simple frequency that suspects are guilty of the intended crime.

[b] By definition, $Prob(ID) = P_0 \cdot Prob(ID|Guilty) + (1 - P_0) \cdot Prob(ID|Innocent)$.

Example: Eyewitness Reforms & Non-Identification Costs

Lineup Procedure	Classic	*No Cost* Reform	*Low Cost* Reform	*High Cost* Reform
ID Probabilities				
$Prob(ID \mid Guilty)$	0.5	0.5	0.4	0.2
$Prob(ID \mid Innocent)$	0.2	0.05	0.05	0.05
$Prob(F)$	0.3	0.1	0.1	0.1
Posterior Probabilities				
$Prob(G \mid ID)$	0.71	0.91	0.89	0.80
$Prob(G \mid N)$	0.38	0.34	0.39	0.46
Suspect IDs				
% Identified	0.35	0.275	0.225	0.125
guilty	0.25	0.25	0.20	0.10
innocent	0.10	0.025	0.025	0.025

Table 4.1: Eyewitness reforms and non-identification costs. *Prior probability of guilt = 0.5.* Cost in this example refers to the decrease in the probability of identifying a guilty suspect. The first panel presents the probabilities of Correct, False, and Foil IDs. The second panel shows the posterior probability of guilt given the lineup outcome. The third panel shows the percentage of all suspects that are identified under the given procedure and decomposes that percentage into those that are guilty and innocent. The probabilities in the first panel are imposed, all other values are derived from those and $P_0 = 0.5$.

are identified by the witness. While this increased degree of certainty is good, potential concerns regarding prosecution on such a basis are twofold. First, if all identified suspects are charged and 71% are guilty, then it follows that 29% are innocent. This may be viewed as an unacceptable degree of risk that prosecution will result in a false conviction. Second, 38% of suspects not identified are in fact guilty, but in the absence of further evidence, are unlikely to face prosecution. An alternate interpretation of the latter concern is that 50% of guilty suspects are likely to avoid prosecution in this setting.

Of the two concerns expressed above, it is the former for which sequential lineups and other reform components have shown demonstrable improvements in the laboratory setting. However, reforms could also exacerbate the latter concern by decreasing $Prob(ID|Guilty)$. In the example, the No Cost column shows the case when the reform reduces the probability of False and Foil IDs while maintaining the same probability of a Correct ID. The benefit of the reform, and sequential superiority, is obvious in this column. Relative to column one, the posterior probability of guilt increases for suspects that are identified and decreases for those that aren't. Furthermore, if we assume that only suspects who are identified are prosecuted, then there is no increase in the number of guilty suspects who escape prosecution, since the reduction in IDs comes strictly from False IDs.

The Low Cost and High Cost columns in table 4.1 maintain the probabilities of identifying an innocent suspect or a filler at the same level as in the No Cost reform case, while decreasing the probability of identifying a guilty suspect. In other words, the last two cases maintain the same degree of benefit from reforms, reducing inaccurate identifications, while progressively increasing the potential cost from fewer correct identifications. The bottom panels of the table clarify the problem posed by focusing exclusively on $Prob(G|ID)$ to determine the superior identification procedure. In each of the three sequential lineup cases, $Prob(G|ID)$ is strictly greater than for the simultaneous lineup case, but $Prob(G|N)$, the probability that a suspect who is not picked out of the lineup is actually guilty, rises as $Prob(ID|Guilty)$ falls. Note that these are relative to a prior probability of guilt that is just 50%, so in the High Cost sequential lineup case when a witness fails to identify the suspect, that suspect's probability of guilt falls by just 4 percentage points. In the final panel, we see that the end consequence is that the proportion of guilty suspects who are charged with the crime alleged, assuming witness identification is required by prosecution, may fall by up to 60%. Of course, this could be viewed as outrageously unacceptable or, just as easily, as an acceptable cost given the dramatic reduction in the probability of prosecuting innocent suspects. In terms of sequential lineups, the point illustrated in table 4.1 is that while sequential superiority is clear in the No Cost case, empirical evidence (e.g. Clark, 2012) suggests that some reduction in Correct IDs is more realistic, and despite universally higher $Prob(G|ID)$[4], the degree to which sequential lineups reduce those Correct IDs is critical to any responsible cost-benefit analysis.

The informational differences between the laboratory and the field further complicate evaluation because the observed field outcomes do not allow us to estimate the probabilities in table 4.1. Investigators only observe whether their suspect, a filler, or no one in the lineup was identified. As a result, $Prob(F)$ can be estimated from field data but $Prob(ID|Guilty)$ and $Prob(ID|Innocent)$ are confounded due to uncertainty about the suspects' guilt, uncertainty which persists even with the benefit of hindsight. The consequence of this reality is that there is no way to accurately calculate posterior probabilities from field data, nor to parse innocent and guilty suspects as was done in the bottom two rows of table 4.1.

Conversely, laboratory evidence rarely makes a distinction between False IDs and Foil IDs, further obscuring the relevance of lab estimates to field evidence. Consider a situation where a county switches from using simultaneous lineups to sequential lineups and keeps detailed information on the outcome of every lineup, finding that the rate of suspect identifications falls from 35% to 22.5%, and the rate of Foil IDs falls from 30% to 10%. This is identical to the Low Cost case in table 4.1, so those values are a possible explanation for the change. But researchers looking at the data are unable to verify the frequency at which suspects are guilty, P_0, nor can they verify false

[4] Steblay and colleagues point out that their posterior probability increases regardless of the prior probability of guilt (P_0). It is straightforward to show that $Prob(G|ID)$ is greater for any value of P_0 in each of the sequential lineup columns of table 4.1 than in the simultaneous lineup column.

identifications, and it is therefore equally possible that the reduction in suspect identifications was driven entirely by a decrease in Correct IDs. Alternately, the reduction could be driven entirely by fewer False IDs. Although each is an unlikely explanation, arguably the former more so than the latter, neither can be discounted entirely. Yet laboratory evidence is limited in its ability to help differentiate between these explanations because there is generally no distinction between the fillers in “perpetrator absent” lineups. The implicit assumption is that an innocent suspect is selected via the same mechanism as the lineup fillers, which is fundamentally inconsistent with the realities of law enforcement.

The natural response to the uncertainty around the rates of False IDs is to proxy for them with the rate of Foil IDs. This practice involves examining field evidence and comparing changes in suspect identifications with changes in Foil IDs (Klobuchar et al., 2006; Mecklenburg et al., 2008). The presumption is that a decreasing rate of Foil IDs is indicative of a decreasing rate of False IDs. However, more recent research recognizes the potential concerns with the assumed correlation between False IDs and Foil IDs, leading to a lively debate on the issue (Amendola and Wixted, 2015a, 2015b; Wells, Dysart, and Steblay, 2015; Wells, Steblay, and Dysart, 2015; Wixted et al., 2016). Although some degree of correlation is plausible, the validity of the inference is questionable. Our main concern, alluded to above, is that innocent suspects and lineup fillers are each sourced through distinct selection mechanisms. For fillers, the mechanism is relatively transparent; they are selected from a limited population of existing convicts that the police department has access to and the selection is based on similarity to the witness description of the perpetrator. However, the mechanism for finding suspects is less well-defined, less transparent, and could result in a variety of implications for eyewitness procedures. For instance, since suspects are drawn from a wider population of individuals, we could hypothesize that innocent suspects are prone to have a greater likeness to the suspect description than fillers and are therefore more likely to be mistaken for the perpetrator. On the other hand, suspects are also most often identified by some form of alternate evidence linking them to the crime. Thus perhaps they are less likely to bear a strong likeness to the witness description since they are not selected based solely on that description. Clearly the assumptions made with regards to this selection mechanism can influence both the strength and nature of the correlation between False and Foil IDs. At present we are unaware of any reliable source of evidence by which to justify limiting assumptions on this correlation.

4.2.3 Empirical Measures and Contribution

Taking account of the informational asymmetries between laboratory evidence and the realities of eyewitness lineups in the field, we agree with the assessment that the evidence supports the claim that eyewitness reforms reduce the risk of False IDs, and also that there is likely some cost in terms of decreasing the likelihood of Correct IDs. However, there is not sufficient depth nor detail in any existing longitudinal, criminal justice data to generate any reasonable estimates that differentiate between the impact on Correct and False IDs. With this in mind, we take a consequentialist approach to understanding the effects of eyewitness reforms that are already in place around the nation. We take advantage of statewide eyewitness reforms to estimate the impact on the proportion of reported crimes that are cleared by arrest. Our approach draws on the link between a suspect identification, the posterior probability of guilt, and the prosecutor's decision to press charges. As with other field evidence, this approach does not allow any direct differentiation between the effect on Correct and False ID rates. However, it does offer the opportunity to provide new pieces to the informational puzzle that has been slowly coalescing on this topic.

The logical underpinnings behind eyewitness reforms, as well as existing evidence, support the hypothesis that they will lead to lower clearance rates. Ideally, any observed reduction would be solely the result of eliminating False IDs, but such a direct attribution is not possible. Instead, we propose two sources of value from examining the impact of these reforms. The first is that there is directly inherent value in knowing whether there is a measurable impact of reforms on a meaningful criminal justice outcome such as clearance rates, even in the absence of any information about the accuracy of identifications. At the very least there are administrative implications for the allocation of law enforcement resources, not to mention academic interest. In addition, there is informational content that can be derived from the magnitude of any observed changes in clearance rates. Consider the High Cost case in table 4.1, where the proportion of suspects identified decreases from 35% to just 12.25%. Such a dramatic decrease in suspect identifications can only be explained by an exceptionally high rate of False IDs in pre-reform lineups, a significantly lower rate of Correct IDs in post-reform lineups, or some combination thereof. Since a 20% rate of False IDs seems unacceptably high, the example presented in the table is a reasonable combination of these two explanatory factors. In any case, a sizable reduction in the clearance rate is a negative signal.

It bears mentioning that an increase in clearance rates is not altogether infeasible. This possibility, remote though it is, arises from another distinguishing factor about the field setting with the potential to distort outcomes relative to what is inferred from laboratory evidence. This is the fact that investigators in the field have a degree of motivation and agency that cannot be expected of research assistants in lab tests. There is also a functional difference between observing whether a reported crime results in an arrest and observing actual lineup outcomes. Considering these points, the motivation of investigators may combine with the fact that after the reform, fewer witnesses discredit themselves by making a Foil ID, and investigators are therefore able to pursue more of their cases to the point of making an arrest. However, it strains our perception of feasibility to imagine that this effect could overwhelm the direct effect of fewer False and Correct IDs on clearance rates. A more reasonable expectation is that motivated investigators will moderate any decrease in clearance rates by continuing to pursue some portion of cases for which the suspect was not identified. The latter point is important to recognize because it indicates that an observed reduction in clearance rates understates the actual decrease in the rate of Correct and False IDs.

4.2.4 Reform Components and Hypothesis Testing

It is no surprise that when a reform addresses one concern regarding eyewitness procedure, it invariably implements a number of other recommendations as well. Consequently, each of the reforms examined in this research includes at least three of the reform components enumerated in section 4.1. Due to very little variation in the composition of these reforms, we make no effort to distinguish between the effects of individual reform components. This phenomenon of confounding the effects of multiple reform elements is well-documented and discussed throughout the literature. In a recent example, Mecklenburg and colleagues (2008) acknowledge the confound of sequential lineups and blind administration in the Illinois field experiment, arguing that it has bearing on the implications but does not diminish the value of the findings. Clark (2012) looks individually at each procedural recommendation in a meta-analysis of laboratory evidence, showing that each recommendation decreases both false and correct identification rates.

We note that each of the other potential reform elements has the same objective as sequential lineups, reducing inaccurate identifications that, in the case of False IDs, could lead to false convictions. Additionally, each of them bears the simultaneous possibility of reducing correct identifications. As a matter of consequence, confound is of little prior concern to our analysis. The same informational asymmetries between laboratory outcomes and field applications exist for each of the reform components, with perhaps mildly greater relevance to some than others. The broad question is whether there is a general effect of reforms on meaningful criminal justice outcomes. In effect, we simply adopt the approach of existing research and test our hypothesis on reforms that confound several reform elements, leaving discussion of which were the main drivers for the back end of the analysis and/or for future research. Thus, our hypothesis is that eyewitness reforms will lead to a decrease in the clearance rate for the crime of robbery in the reform state. This decrease should be due to fewer False and Correct IDs. Finally, larger reductions in the clearance rate would indicate a large reduction in Correct IDs because they are, we should hope, the greater proportion of the charges filed.

4.3 Empirical Strategies and Estimation

We test the hypothesis that eyewitness reforms lead to reduced robbery clearance rates using two empirical strategies. We focus on robberies as the key crime category for three reasons. First, it is a high frequency crime and it is always favorable to have the Law of Large Numbers on your side. Second, by it's very nature, robbery tends to leave eyewitnesses who are then likely to be called upon to identify the perpetrator. Finally, unlike crimes such as murder and rape, robbery is not likely to result in extensive forensic evidence, and therefore, the decision to press charges is expected to frequently rely on eyewitness testimony. We thus presume that robbery is the most likely crime category for clearance rates to respond to eyewitness reforms, and the least likely to be subject to confound from unrelated justice reforms or advances in forensic science.

Our first strategy is a Difference-in-Differences (DD) approach that has been extended into what might be recognized as a Triple Differences (DDD) approach. The DD portion of this method creates a counterfactual for clearance rates in reform states out of the average clearance rate adjustments among non-reform states. However, since we are concerned that eyewitness reforms may be correlated with other justice reforms, which may also impact clearance rates, we create a second in-state counterfactual by differencing the DD estimates for robbery with those of alternate crime categories that are very unlikely to be responsive to eyewitness reforms. In particular, we use both burglary and auto theft as two separate crime categories that can be used as a second counterfactual.

The other strategy we employ is the Synthetic Control method (SC) first presented by Abadie, Diamond, and Hainmueller (2010). In this approach we take an individual reform state and use the algorithm proposed by Abadie et al. (2010) to construct a "synthetic" counterfactual for the reform state from a weighted average of clearance rate time trends among untreated states. This approach allows a state-by-state test to identify the impact of their individual reforms. Given the data available, we are able to conduct an SC estimation for six states that adopted reforms sufficiently early in the analysis period. In our discussion of the results, we will focus on the earliest adopter, New Jersey, and the one state for which there does appear to be a significant decrease in the clearance rate for robberies.

4.3.1 Data

We use two sources of data for this research. The main body of data is drawn from the FBI Uniform Crime Reports dating from 1990 to 2014. These make up the observational data of crimes committed and cleared by arrest. The crime data is complemented with simple observations of the timing and implementation of statewide justice reforms in any of the fifty states and the District of Columbia. The observational data on reforms (henceforth reform data) was collated by researchers affiliated with this study and includes information on the dimensionality of each reform with respect to the eyewitness reform components discussed above. The details about the scope of each reform were used for some robustness checks and other tests, but the variable used in all of the main specifications is a simple indicator of whether a reform has occurred.

The FBI crime data includes every adult crime category recorded by law enforcement.[5] However, we exclude murder, manslaughter, and rape from this analysis. Although these crime categories do present the most compelling cases for consideration, it is the compelling nature of heinous crime that leads us to posit a risk of unknowable dimensions of investigative procedure, which may confound estimates in ways that we are unable to reasonably predict. Conversely, property crime is high frequency crime and, we presume, cases are pursued in a much more routine fashion. Petty theft has also been excluded due to widespread reporting failure in this crime category, despite it apparently constituting the difference between the *All Larceny* row of table 4.2 and the sum of the sub-categories. The remaining crime categories and subcategories are presented in table 4.2, with the state averages for crimes reported and crimes cleared by arrest. These averages have large standard deviations arising from state-to-state variation due to population size and urbanization disparities.

[5] Although juvenile crime data is also available in the original dataset, we are concerned about heterogeneity in the process of referring juveniles to prosecution. Merging the data into a single analysis would increase the risk of model mispecification.

Criminal Offenses and Crimes Cleared by Arrest (1990 - 2014)		
	Offenses	Cleared
Robberies		
*All Robbery**	8767.14	2204.18
	(15202.78)	(3813.58)
Robbery w/Gun	3401.88	679.43
	(5667.54)	(1119.04)
Robbery w/Knife	764.85	205.44
	(1702.34)	(442.51)
Strongarm Robbery	3372.45	947.90
	(6225.91)	(1804.66)
Other Robbery	866.97	219.36
	(1999.69)	(485.02)
Larcenies		
All Larceny	130263.76	24764.18
	(152978.52)	(27921.81)
Vehicular Larceny**	22661.38	2947.26
	(38009.35)	(4323.05)
Car Theft	16681.63	2231.99
	(28650.11)	(3237.48)
Truck Theft	3677.62	430.62
	(8113.62)	(841.71)
Other Larceny	1632.31	208.64
	(2303.60)	(316.79)
Burglaries		
*All Burglary***	43595.77	5562.64
	(56805.10)	(7486.51)
Burglary - Forced Entry	27370.02	3462.74
	(36520.15)	(5212.65)
Burglary - No Force	12587.71	1777.02
	(17681.37)	(2910.45)
Attempted Burglary	3044.98	379.61
	(3970.62)	(1714.12)
Observations	1225	1225

Means reported. Standard deviations are in parenthesis.

Raw data: Missing observations and outliers included.

Table 4.2: Mean offenses and crimes cleared. Statewide averages of verified criminal offenses and crimes cleared by arrest, broken down by crime category and sub-category. Averaged over all 25 years of data, for Washington D.C. and 48 states. Kansas and Illinois excluded due to missing data. *All Robbery** is the key crime of interest for which clearance rates are evaluated, while *Vehicular Larceny*** and *All Burglary*** are the crime categories used for comparison.

The raw FBI data provides monthly observations in crime reported by individual law enforcement agencies. However, there is a great deal of measurement error apparent at the granular level of the data. In particular, there is lumping across time periods, some of it regular and some irregular, and frequent gaps in reporting by individual agencies. Since reporting is not mandatory, we assume that agencies fail to report when staffing is low and that some agencies simply choose to submit their numbers less frequently, leading to the bunching. Aggregating the data to statewide annual observations eliminates any concern regarding the bunching since the aggregation is effectively just bunching the data up even further. The only way there could be a residual concern is if there were an irregular pattern of bunches crossing the January 1 threshold at times that correlated with eyewitness reforms, which we find no evidence of. The remaining missing data is assumed to exhibit classical measurement error with the exception of a few cases when all observations are missing for an entire state-year. Classical measurement error does not pose an estimation threat as far as bias in the estimates is concerned but it could lead to inflated standard errors, which we will reference in our discussion of results. The missing state-year observations are an issue we confronted directly. Kansas and Illinois were omitted from the study entirely due to a 6 year gap and 10 year gap, respectively, in their data for all crime categories. All other missing state-year observations were replaced with a simple average of the preceding and following years.

Aggregation of the FBI data also helps diminish the threat of temporal shift between the reporting of a crime and an arrest being made. The clearance rate variable, CR, used in each of the empirical strategies is constructed as the ratio of crimes cleared by arrest over verified crimes reported, where the crime is either robbery, auto theft, or burglary.[6] Thus it is possible, even likely, for observations to be shifted from the numerator of CR_t and instead show up in CR_{t+1}, based on the assumption that a law enforcement agency diligently reporting their numbers to the FBI every month will have some crimes reported in one month and not cleared until the following one. Assuming that the rate at which this occurs is constant over time, the only impact would be to delay by a single

[6] These three crimes are the categories of *All Robbery*, *Vehicular Larceny*, and *All Burglary* in table 4.2.

period our ability to observe some fraction of any real change that occurs in CR_t. Aggregating to annual observations effectively diminishes the fraction of clearances that occur in the following period because each period is longer. For example, if we suppose that 6% of crimes do not result in arrest until the following month, it follows than only 0.5% result in arrest in the following year. For the same reasons, we are not overly concerned about the possibility of correlation between temporal shift and implementation of eyewitness reforms. Reducing false identifications could lead to increased temporal shift, but the margins would be tiny relative to the actual impact on clearance rates. The most meaningful role of temporal shift in this data is that it explains the presence of a few observations, notably in states with very low crime numbers, where the value of CR_t approaches or even surpasses 1. We do not bound these values, but the weights used in the estimating equations limits the impact of such outliers on the estimates.[7]

The reform data was compiled by researchers who searched the internet for news releases, official reports, guidance memos, and other official documents that detailed new policy for state and local law enforcement agencies. Any inconsistencies or ambiguities in the documentation were clarified through direct email or telephone contact with the relevant state agencies. Reforms were only considered valid if the new procedures were explicitly mandated or rolled out as "best practice" guidelines to all jurisdictions in the state. Changes in policy and practice at the local level are potential moderators of the observed effects in this study, since the reform data available is not comprehensive with regard to local policy changes.

Table 4.3 shows the breakdown of reform elements for the 16 states in which we identify a statewide reform. Most notable here is that four of the states do not require sequential lineups as part of their new policies. However, since evidence suggests that the other major reform elements also reduce False IDs and Correct IDs, these "non-sequential" reform states are still coded as having experienced a reform for the main empirical specifications. The variable coded to indicate reforms is an indicator equal to zero in all years prior to reform (or in all years for non-reform states) and

[7] Observations are weighted by the number of crimes reported in that state-year.

Eyewitness Reforms

State	Sequential Lineups	Blind & Unbiased	Lineup Composition	Multiple Witnesses	Confidence Statement	Effective Date
Colorado	x	x	x	x	x	08/01/10
Conneticut	x	x	x	x	x	05/01/13
Florida		x	x		x	11/01/11
Hawaii	x	x	x	x	x	10/24/14
Massachuesetts	x	x			x	09/01/10
Montana		x	x	x	x	03/01/12
Nevada	x	x	x	x	x	11/20/14
New Jersey	x	x	x	x		04/18/01
New Mexico	x	x	x	x	x	12/01/14
North Carolina	x	x	x	x	x	03/01/08
Ohio		x	x		x	07/06/10
Rhode Island		x	x	x	x	11/01/11
Texas	x	x	x	x	x	09/01/11
Virginia	x	x	x	x	x	03/19/14
West Virginia	x	x	x	x	x	01/01/14
Wisconsin	x	x	x	x	x	01/01/06

Table 4.3: *Reform Data: This table shows the key components of the reforms adopted by each of 16 states that we identify as having implemented a statewide eyewitness reform prior to the end of 2014. 'Blind and Unbiased' refers to blind administration of lineups and guidelines for unbiased instructions to witnesses. The 'Lineup composition' and 'Multiple Witnesses' components specify procedures to ensure fillers bear a sufficient likeness to the suspect and contact between witnesses is prevented prior to recording their statement and identification. The final reform element is the recording of a 'Confidence Statement' after identifications, which we take note of but has no obvious bearing on this study.*

equal to one for all years after the effective date of the reform, shown in table 4.3. For the year in which the reform began, we code the variable zero or one such that the reform is only considered effective ($Ref = 1$) if the effective date is prior to July 1 (i.e. we assume that the reform had an impact during the year of implementation if, and only if, it was effective for at least half of the year). So in the use data, Virginia and West Virginia experience a reform in 2014, but Hawaii, Nevada, and New Mexico never experience a reform because the data ends in 2014, making the set of reform states 13 rather than the full 16 shown in table 4.3.[8]

In the appendix, the raw time trend data on clearance rates of each crime category used in this study is presented individually for each state. The figures also indicate the timing of the reform in states that had one. These figures do not reveal a clear pattern of change in clearance rates following reforms. States with low population tend to exhibit moderate yearly variation in clearance rates, for all crime categories, while high population states have pretty stable time trends. Most states have reasonably consistent long-term averages for each clearance rate category. Importantly, the time trends reveal irregularities in the data from a few states. The missing data from Kansas and Illinois are quite obvious and explain our exclusion of those states. The figures for Hawaii and New York show suspicious troughs where their clearance rates fall significantly for an extended period before rising again. And Maine has dramatic fluctuations that are likely a sign of some kind of measurement error that may be nonrandom. Although the latter three states are not excluded from the main empirical specifications of this study, we test whether the exclusion of these states alters the estimates and find no meaningful differences.

4.3.2 Difference-in-Differences Estimation

A straightforward difference in differences strategy (DD) would be to take every state that had a reform, look at their clearance rates for the crime category of interest - robbery - before and after reforms occurred, and compare those pre- and post-reform averages with those of states that had

[8] Several more states also implement similar statewide reforms in the years following 2014. Those states are not discussed here because our working FBI dataset only has the years 1990 through 2014.

no reform at all. It is a very direct and persuasive approach. Yet there may be concerns about the underlying assumption that clearance rates in the untreated states provide an unbiased estimate of what would have been happening to clearance rates in the treated states without the changes to witness identification protocols. The general concern is that other trends or unobservable factors may be differentially affecting states and leading to the choice to adopt new protocols, and those unobservables may influence law enforcement or crime, and thus clearance rates. Our data has no additional control variables to account for this concern. However, the clearance rates of other crime categories - burglary and auto theft - provide a persuasive within-state counterfactual for these reform state unobservables, so long as we can assume that clearance rates for those crime categories are not significantly impacted by the reforms. By estimating the same DD specification for the 'treated' and 'untreated' crime categories, we compare the two different estimates to get a triple-differences (DDD) style estimation. Since it is debatable whether burglary or auto theft cases are less reliant on eyewitness evidence, we generate estimates for both and use each for comparison.

DD Estimating Equation:

$$CR_{it} = \alpha_0 + \alpha_1 Ref_{it} + \gamma_i + \delta_t + u_{it} \tag{4.1}$$

DDD Estimate:

$$\hat{\beta} = \hat{\alpha}_1^R - \hat{\alpha}_1^C \tag{4.2}$$

Equation 4.1 is the basic functional form of the DD strategy used in this study, estimated separately for each crime category. In each case, CR_{it} is the clearance rate for that crime in state i and year t. Ref_{it} is an indicator equal to 1 if state i is a reform state and year t is post-implementation of the reform in that state. We use a standard fixed effects regression with γ_i and δ_t as the state and time fixed effects, respectively. Note that a separate time-invariant indicator for whether a state is a reform state, or a state-invariant $Post_t$, would be redundant because the state and time fixed effects subsume any variation that is time-invariant or state-invariant (i.e. any stable differences between

DD Model Estimation			
	(1)	(2)	(3)
VARIABLES	Robbery	Auto Theft	Burglary
Ref (α_1)	-0.0120	-0.0165	-0.00192
	(0.0176)	(0.0147)	(0.00744)
Observations	1,221	1,221	1,221
FE - State	Y	Y	Y
FE - Year	Y	Y	Y
$\hat{\beta}$		0.0045	-0.0101

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.4: *The table shows the point estimates of $\hat{\alpha}_1$ from equation 4.1 for each of the three crime categories. The dependent variable, CR_{it}, is the clearance rate for each crime. $\hat{\beta}$ is calculated separately for auto theft or burglary as the counterfactual to robbery, in neither case is the estimate statistically significant to any common threshold.*

states that will eventually adopt reforms and those that don't, or any nationwide shock that affects all states similarly). For all estimates, we cluster the error terms, u_{it}, at the state level.

Under the parallel trends assumption,[9] the point estimate for $\hat{\alpha}_1$ from equation 4.1 represents the average treatment effect of eyewitness reforms. However, as noted earlier, there are some potential concerns with the assumption of parallel trends in this setting. Furthermore, since reforms occur in several different years for each of the treated states, the typical graph demonstrating parallel trends prior to treatment is not readily available here. We are able to generate an approximate parallel trends figure, available in the appendix, by excluding the early adopting reform states, which shows reasonably similar trends but only enough to provide limited assurance about the parallel trends assumption. To help alleviate concern about this assumption, we make the within-state comparison with the same coefficient for a crime category that we do not expect to be significantly impacted by eyewitness reforms. That is, we take $\hat{\beta}$ from equation 4.2 as a more reliable estimate of the impact of reforms on the clearance rate for robbery. For this difference, $\hat{\alpha}_1^R$ is the DD estimate for robbery and $\hat{\alpha}_1^C$ is the DD estimate for either one of the counterfactual crime categories.

[9] The parallel trends assumption, in this context, is that the trends of robbery clearance rates in non-reform states offer a valid counterfactual for the changes that would be occurring in reform states if not for their reforms.

Table 4.4 shows the DD estimate (column 1) and the combined DDD estimates (the last row of columns 2 and 3) for the effect of eyewitness reforms on clearance rates for robbery. Each of the point estimates is small in magnitude, relative to robbery clearance rates that mostly live between 0.25 and 0.3, and do not approach any conventional level of statistical significance. Hence we do not find evidence of a significant reduction in clearance rates regardless of which estimate we prefer. However, it would be short-sighted to think that a failure to reject the null hypothesis that $\hat{\beta} = 0$ is the extent of the informational content available from the modeling of these data. In particular, with the reasonably tight standard errors on the estimates we can statistically rule out the type of dramatic reductions in the clearance rates seen in the High Cost column of table 4.1. Although the DDD model does not provide proof of a reduction in clearance rates, which one may have found doubtful to begin with, it does allow us to rule out the possibility of eyewitness reforms having very large negative effects and the possibility of moderate to large positive effects.

4.3.3 Synthetic Controls Strategy

In the second empirical approach we use the synthetic control (SC) strategy proposed by Abadie, Diamond, & Hainmueller (2010), which allows us to test states individually for an identifiable treatment effect. The thinking behind this is that while we can typically pinpoint an exact date when an official memo determines that eyewitness reforms become effective, there are a number of factors that may influence the rate of adoption by the various agencies across each state and thus lead to variability in the treatment effect of each reform. Additionally, local jurisdictions may have already adopted some of the recommended eyewitness practices prior to the state reform, which our data has limited ability to account for. It is therefore possible for the impact of reforms in some states to have been significantly moderated by pretreatment. This type of heterogeneity in treatment intensity would depress the coefficient estimate in cases like the DDD model where we are estimating an average treatment effect. The SC strategy provides an alternative because it attempts to identify a treatment effect for a single "treated" state, rather that averaging over a group of treated states.

The basic principle of the SC approach is to create a “synthetic” counterfactual for the treated state out of a set of comparable states. Take New Jersey, the earliest reform state, as an example. Ideally, there would exist a neighboring state that had no reform and is so alike to New Jersey that their clearance rates move in perfect sync prior to the reform. In other words, it is the perfect counterfactual for New Jersey and will tell us exactly what would have happened in New Jersey if not for their eyewitness reform. There is no such state, of course, but there are other states with similar time trends in crime and clearance rates. So with the SC strategy we use a packaged algorithm to create a “synthetic” New Jersey that is a weighted average of several other states. The weighting is optimized for minimal error between the time trends of the real New Jersey and its synthetic counterpart, over the entire period prior to treatment. If the algorithm is able to fit the synthetic trend closely to that of the treated state, then any deviation post-treatment is assumed to be a result of that treatment.

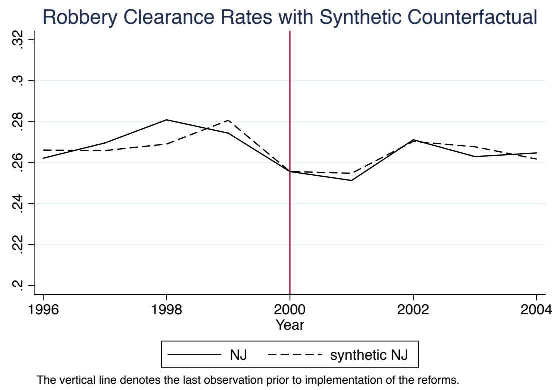
We run a SC analysis for each of the six earliest reform states. Six is the maximum number of reform states for which our data allowed at least 4 years of post-reform observations, to ensure that any isolated jumps in the data are not misinterpreted as changing trends. We use a uniform nine-year window for the analyses, shifting the window to fit with the reform date of each state. Additionally, the synthetic controls are matched based on both the clearance rate and the number of crimes reported, to increase the similarity of the states that weighted most heavily. The set of states from which the synthetic control is constructed includes every state, with the exception of Kansas and Illinois, that has not experienced a reform and will not have one within the nine-year window of the current treatment state.

Figure 4.1 has the graphical representation of all six SC trials. Figures 4.1a and 4.1b are presented first because they represent the most interesting two results, one suggesting no reform effects and the other, conversely, suggesting at least a moderate reduction in clearance rates. The remainder of the the trials, figures 4.1c - 4.1f, are informative mainly due to their incoherence. Each achieves at least a weak fit in the pre-reform trends, but in most cases the trends deviate prior to the reform and often reverse themselves in a later year. The most notable observation from these four trials is that there is a lot of variation in clearance rates that is difficult to consistently associate with the timing of eyewitness reforms.

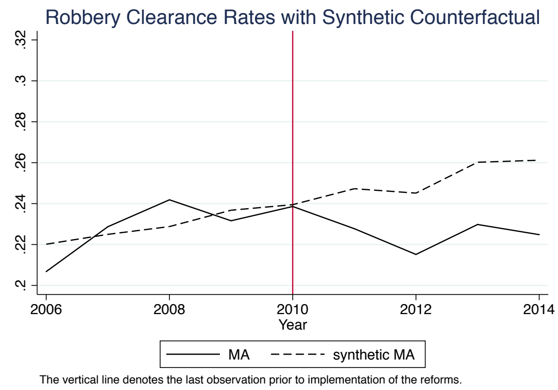
Figure 4.1a presents a much clearer picture, albeit a surprising one. The clearance rates of synthetic New Jersey are almost perfectly matched to the clearance rates of the actual state, both before and after the reform occurs, suggesting that the New Jersey reform had no impact on clearance rates at all. This result was unexpected because, to the best of our knowledge, the early adoption of eyewitness reforms in New Jersey was not so much the result of rising sentiment or concern about eyewitness procedures as the confluence of tangential political pressures and media attention. That would seem to indicate that the reform was unexpected and thus a greater shock to the status quo in law enforcement agencies. It had been our prior expectation that the unanticipated nature of the New Jersey reform would lead to a more significant impact on clearance rates. Two alternative narratives are more coherent with the evidence in figure 4.1a: either eyewitness reforms in New Jersey came at zero or very low cost in terms of reducing accurate witness identifications, or the lack of foreknowledge of and prior sentiment for eyewitness reform led to slow or poor uptake of the new protocols among law enforcement agencies.

Figure 4.1b offers a counterpoint to the New Jersey trial. The synthetic control for Massachusetts has a reasonable fit in the pre-reform years and shows a distinct divergence of trends after the eyewitness reforms become effective. Thus the Massachusetts reforms appear to have led to a notable reduction in clearance rates for the state, on the order of $10-15\%$. This provides supporting evidence for the heterogenous policy effects hypothesized at the beginning of this section. However, we also note that while the probability of picking up spurious correlation in such a specific form is quite low, the probability of spurious correlation in one of several trials increases with the number of trials. Hence, having run six SC trials, we cannot fully dismiss the possibility that the result of the Massachusetts trial, or the New Jersey trial, is in fact a spurious artifact of the data. Yet the degree of fit throughout the later post-reform periods, particularly in figure 4.1b where the time trends

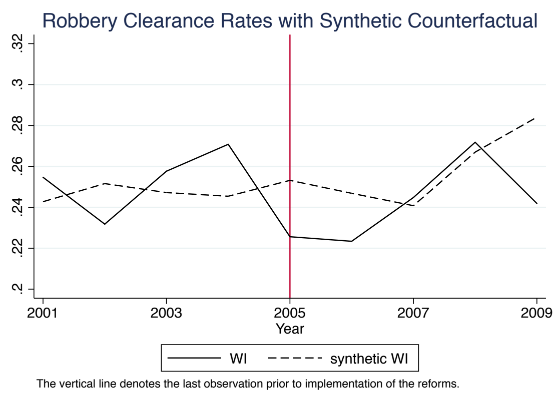
(a) New Jersey Reform



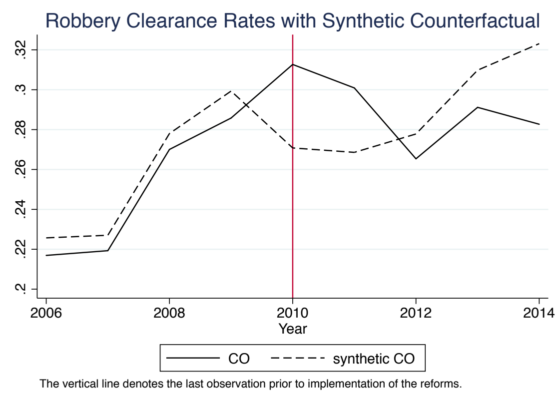
(b) Massachusetts Reform



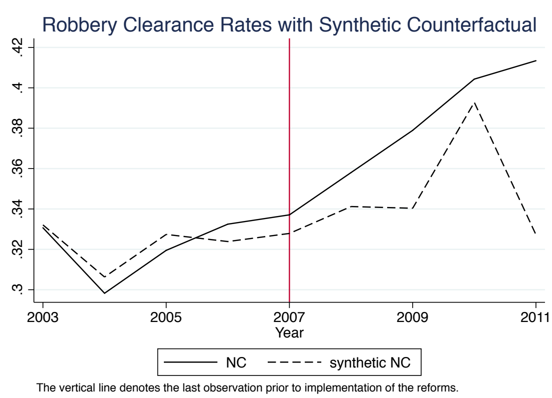
(c) Wisconsin Reform



(d) Colorado Reform



(e) North Carolina Reform



(f) Ohio Reform

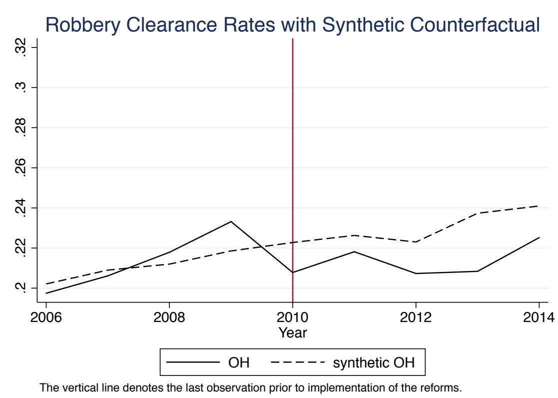


Figure 4.1: Synthetic control trends for six early reform states.

resume parallel movement after diverging, provides reassuring evidence that the synthetic versions of each state are in fact valid counterfactuals.

Altogether, the six trials of the SC strategy show a surprising degree of variation between the reform outcomes in different states. This reinforces the idea that eyewitness reforms may have heterogeneous effects, which could be dependent upon unobservable characteristics of the states themselves or the conditions under which the reforms were adopted. Furthermore, the heterogeneity observed in the SC results is also consistent with the inability of the DDD model to identify a statistically significant average treatment effect on reform states as a whole.

4.4 Discussion and Summary

In this paper we present empirical evidence that further informs an existing debate in the literature regarding eyewitness reforms, which are thought to be capable of reducing the risk of mistakes that can lead to false convictions. The existing debate relies heavily on laboratory evidence, which suggests that recommended reforms can reduce the risk of false identifications but may do so at a cost to the rate of accurate identifications (Clark, Howell, and Davey, 2008; Clark, Erickson, and Breneman, 2011; McQuiston-Surret et al., 2006; Steblay et al., 2011). In section 4.2.2 we discuss the informational differences between lab tests and eyewitness evidence in the field, which add to the uncertainty regarding the costliness of the recommended reforms. In particular, laboratory evidence relies on Foil IDs, those of individuals known to be innocent, as indicators of False IDs, which falsely identify the suspect of an alleged crime.[10] However, there is no evidence regarding the true nature of the correlation between the two types of false identification. In the absence of this information, the connection between the implications of laboratory evidence and the actual outcomes of eyewitness reforms is unclear. This does not diminish the importance of lab trials, instead it increases the need for field evidence, when available, to complement and validate findings from the lab.

[10] Existing field evidence, such as Klobuchar et al. (2006) and Mecklenburg et al. (2008), also relies on this same correlation between Foil and False IDs.

The empirical analysis presented in section 4.3 further reinforces the need for additional field studies of eyewitness reforms. Estimates from the DDD model show that the average treatment effect among all reform states in these data isn't sufficiently large to statistically differentiate from zero. However, in considering the varying results of the SC analysis for different states, we acknowledge that the results of both models are consistent with statewide eyewitness reforms having heterogeneous effects among the different states that have implemented such reforms. This could be a sign of varying levels of preemptive adoption of similar reforms in local jurisdictions throughout each state, varying attitudes among law enforcement communities across states regarding take-up of new policy, or other unobservable state level characteristics that moderate the response to these reforms. Thus we see two issues in understanding the effects of eyewitness reforms: the more persistent issue is the continuing challenge of differentiating between a reduction in False IDs or Correct IDs, the issue that is newer to this research is explaining the apparent heterogeneity in the response of clearance rates to eyewitness reforms. Each of these issues requires additional field research tailored to the specificity of the research questions.

Returning to the sequential superiority debate from section 4.2.1, our empirical estimates do not necessarily settle any portion of the debate and would not have been able to do so regardless of the outcome. However, the DDD estimates do help to rule out the possibility of a very high cost associated with sequential lineups, insofar as a 95% confidence interval around any of our point estimates rules out a decrease in excess of 30% from the baseline clearance rate. At the same time, the SC analysis affirms the position of those with continuing concerns about dismissing as negligible the cost of sequential reforms. The SC trial for Massachusetts shows a decrease in the clearance rate that, in the absence of a cost to Correct IDs, can only be explained if pre-reform simultaneous lineups were leading to false charges being brought in *at least* 15% of robbery cases. Such a high rate of error is difficult to accept knowing that, between plea deals and conviction rates, more than 90% of federal charges result in a conviction (Rakoff et al., 2014). It follows that there is evidence of a significant cost from sequential reforms but that cost does not manifest to the same degree in all reform states. In point of fact, this cost of sequential lineups does not manifest at all in the SC trial for New Jersey. Reliable identification of the cost of sequential reforms in the field requires access to richer data from law enforcement agencies, particularly given that our data suggests the cost varies depending on the characteristics of the reform setting.

We conclude that several types of field research are required on the topic of eyewitness evidence before any firm determination can be made about the effects of reforms. One important research focus is on understanding the risk of false identifications. This can be approached by careful study of regularities in the mechanism by which suspects are identified in the field, so that mechanism can be better replicated in the lab to simulate a realistic distinction between False IDs and Foil IDs. Alternately, new data sources could elicit a much clearer picture of the proportion of witness identifications in the field that are False IDs, and perhaps, the rate at which those lead to false convictions. Researchers should work with law enforcement and District Attorney's offices to track the rate at which suspects identified by a witness are later cleared of all suspicion by hard evidence. Even the ability to simply track which cases had a witness identification and which cases did not lead to a conviction would provide researchers with a rough basis to begin building an empirically-motivated understanding of False IDs in the field. Conversely, the same same data sources could be used to track cases demonstrating the exact risk of reforms like the sequential lineup, where a witness fails to identify a guilty suspect but they are eventually convicted regardless. Exonerations present yet another potentially fruitful research approach. The presence and frequency of exonerations provide one of the few direct measures of False IDs. The challenges of generating statistical power with long-term outcomes lead us to expect significant data requirements, but with the frequency of exonerations in contemporary criminal justice it is possible that, with sufficiently rich data, researchers could identify whether implementation of reforms lead to a decrease in the 10-year (or 15-year?) rate of exoneration. Even if only a partial measure, the latter would be the most poignant measure of the potential benefits of reforms.

The other major area of opportunity for field research is in confirming and exploring the heterogeneous effects of eyewitness reforms that we provide limited evidence for in this paper. We posit that if there are moderating factors that make reforms more effective in some states than in others, then identifying these factors will allow policymakers to adjust to the specifics of their state or city and ensure that reforms have optimal impact. This area of research will likely benefit from the greater specificity of examining county or agency-level reforms rather than aggregated state-level reforms. Furthermore, it is the combination of the latter research agenda and other research ideas regarding False IDs that will allow researchers to finally answer which reform elements lead to high reductions in False IDs with minimal cost to Correct IDs, which is the root of the debate regarding sequential superiority.

CHAPTER 5

Conclusion

This dissertation presents three independent research projects, each with a unique contribution to research in the field of criminal justice.

Chapter 2 estimates the causal relationship between prison crowding and violent behavior. This study exploits exogenous variation in California prison populations, resulting from a Supreme Court mandate to reduce prison crowding, to estimate the effect on violence. Using both difference-in-differences and instrumental variables identification strategies, a significant positive relationship is identified that is robust to a variety of model specifications. These are the first empirical estimates showing a causal link between crowding and violence, suggesting that reducing prison crowding by 10 percentage points leads to a reduction in the rate of assault and battery of approximately 15%. In addition, differential reductions in the rates of violence between population types is presented as evidence of a compositional effect associated with shocks to prison crowding, which poses a threat to the validity of empirical estimates of the link between crowding and violence.

Chapter 3 synthesizes existing research on prison gangs into an explicit modeling framework that treats gangs as profit-maximizing suppliers and sources of informal governance in an illicit marketplace. The model offers broad policy implications for prison enforcement and highlights the futility of certain policy approaches that don't account for the profit motive underlying gang activity.

Chapter 4 tests for the presence of an identifiable impact on police clearance rates from the implementation of statewide reforms to eyewitness procedures. We find insufficient evidence to identify an average effect for all reform states, but do find evidence that this is the result of heterogeneous effects among the reform states. We are also able to statistically bound the possible effect on clearance rates to rule out concerns that reforms lead to large reductions in positive identifications.

In summary, I observe that the practice of criminal justice is a field with many overlapping policies and confounding factors. This leads to complex issues with answering even the most basic questions in a robust manner. The field deserves a great deal more research attention and careful development of rich data sources to better answer questions of great import to the fair and balanced administration of justice.

Bibliography

(1967). *Stovall v. Denno*, volume 388 U.S. 293.

(2005). *Wisconsin v. Dubose*, volume 272 Wis. 2d 856, 679 NW 2d 927.

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505.

Amendola, K. L. and Wixted, J. T. (2015a). Comparing the diagnostic accuracy of suspect identifications made by actual eyewitnesses from simultaneous and sequential lineups in a randomized field trial. *Journal of Experimental Criminology*, 11(2):263–284.

Amendola, K. L. and Wixted, J. T. (2015b). No possibility of a selection bias, but direct evidence of a simultaneous superiority effect: a reply to wells et al. *Journal of Experimental Criminology*, 11(2):291–294.

Anderson, T. L. and Hill, P. J. (2002). Cowboys and contracts. *The Journal of Legal Studies*, 31(S2):S489–S514.

Association, A. B. et al. (2004). American Bar Association statement of best practices for promoting the accuracy of eyewitness identification procedures. *Washington, DC: American Bar Association. Retrieved on*, 27.

Blevins, K. R., Listwan, S. J., Cullen, F. T., and Jonson, C. L. (2010). A general strain theory of prison violence and misconduct: An integrated model of inmate behavior. *Journal of Contemporary Criminal Justice*, 26(2):148–166.

Bottoms, A. E. (1999). Interpersonal violence and social order in prisons. *Crime and Justice*, 26:205–281.

Butler, M., Slade, G., and Dias, C. N. (2018). Self-governing prisons: Prison gangs in an international perspective. *Trends in Organized Crime*, pages 1–16.

Calandra, D. and Carey, J. (2005). Field guide for the California peace officers legal sourcebook. *Sacramento: California District Attorneys Association*.

Camp, S. D. and Gaes, G. G. (2005). Criminogenic effects of the prison environment on inmate behavior: Some experimental evidence. *Crime & Delinquency*, 51(3):425–442.

Camp, S. D., Gaes, G. G., Langan, N. P., and Saylor, W. G. (2003). The influence of prisons on inmate misconduct: A multilevel investigation. *Justice Quarterly*, 20(3):501–533.

Campana, P. (2011). Eavesdropping on the mob: The functional diversification of mafia activities across territories. *European Journal of Criminology*, 8(3):213–228.

Chen, M. K. and Shapiro, J. M. (2007). Do harsher prison conditions reduce recidivism? A discontinuity-based approach. *American Law and Economics Review*, 9(1):1–29.

Clark, S. E. (2008). The importance (necessity) of computational modelling for eyewitness identification research. *Applied Cognitive Psychology*, 22(6):803–813.

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science*, 7(3):238–259.

Clark, S. E., Erickson, M. A., and Breneman, J. (2011). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior*, 35(5):364.

Clay, K. (1997). Trade without law: Private-order institutions in Mexican California. *The Journal of Law, Economics, and Organization*, 13(1):202–231.

Cordingly, D. (2006). *Under the black flag: The romance and the reality of life among the pirates.* Random House Incorporated.

Cunningham, M. D. and Sorensen, J. R. (2007). Predictive factors for violent misconduct in close custody. *The Prison Journal*, 87(2):241–253.

DeLisi, M., Berg, M. T., and Hochstetler, A. (2004). Gang members, career criminals and prison violence: Further specification of the importation model of inmate behavior. *Criminal Justice Studies*, 17(4):369–383.

Devenport, J. L., Penrod, S. D., and Cutler, B. L. (1997). Eyewitness identification evidence: Evaluation commonsense evaluations. *Psychology, Public Policy, and Law*, 3(2-3):338.

DiIulio, J. J. (1990). *Governing Prisons*. Simon and Schuster.

Dolovich, S. (2011). Strategic segregation in the modern prison. *Am. Crim. L. Rev.*, 48:1.

Drury, A. J. and DeLisi, M. (2011). Gangkill: An exploratory empirical assessment of gang membership, homicide offending, and prison misconduct. *Crime & Delinquency*, 57(1):130–146.

Franklin, T. W., Franklin, C. A., and Pratt, T. C. (2006). Examining the empirical relationship between prison crowding and inmate misconduct: A meta-analysis of conflicting research results. *Journal of Criminal Justice*, 34(4):401–412.

Gaes, G. G. and McGuire, W. J. (1985). Prison violence: The contribution of crowding versus other determinants of prison assault rates. *Journal of Research in Crime and Delinquency*, 22(1):41–65.

Gaes, G. G., Wallace, S., Gilman, E., Klein-Saffran, J., and Suppa, S. (2002). The influence of prison gang affiliation on violence and other prison misconduct. *The Prison Journal*, 82(3):359–385.

Glaze, L. and Herberman, E. (2013). Correctional population in the United States. Technical report, U.S. Department of Justice, Bureau of Justice Statistics.

Goetting, A. and Howsen, R. M. (1986). Correlates of prisoner misconduct. *Journal of Quantitative Criminology*, 2(1):49–67.

Griffin, M. (2007). Prison gang policy and recidivism: Short-term management benefits, long-term consequences. *Criminology & Public Policy*, 6(2):223–230.

Griffin, M. L. and Hepburn, J. R. (2006). The effect of gang affiliation on violent misconduct among inmates during the early years of confinement. *Criminal Justice and Behavior*, 33(4):419–466.

Gross, S. R., Jacoby, K., Matheson, D. J., and Montgomery, N. (2004). Exonerations in the united states 1989 through 2003. *J. Crim. l. & CrimiNology*, 95:523.

Jiang, S. and Fisher-Giorlando, M. (2002). Inmate misconduct: A test of the deprivation, importation, and situational models. *The Prison Journal*, 82(3):335–358.

Kamphuis, J., Meerlo, P., Koolhaas, J. M., and Lancel, M. (2012). Poor sleep as a potential causal factor in aggression and violence. *Sleep Medicine*, 13(4):327–334.

Kearney, M. S., Harris, B. H., Jácome, E., and Parker, L. (2014). Ten economic facts about crime and incarceration in the united states. *The Hamilton Project, May. Accessed*, November 4, 2014.

Klobuchar, A., Steblay, N. K. M., and Caligiuri, H. L. (2006). Improving eyewitness identifications: Hennepin county's blind sequential lineup pilot project. *Cardozo Pub. L. Pol'y & Ethics J.*, 4:381.

Knox, G. W. (2005). The problem of gangs and security threat groups (stg's) in American prisons today: Recent research findings from the 2004 prison gang survey. *National Gang Crime Research Center*.

Kurzfeld, J. (2018). Prison crowding and violent misconduct: Evidence from California public safety realignment. *Forthcoming*.

Leeson, P. T. (2007a). An-arrgh-chy: The law and economics of pirate organization. *Journal of political economy*, 115(6):1049–1094.

Leeson, P. T. (2007b). Trading with bandits. *The Journal of Law and Economics*, 50(2):303–321.

Leeson, P. T. (2010). Pirational choice: the economics of infamous pirate practices. *Journal of Economic Behavior & Organization*, 76(3):497–510.

Leeson, P. T. (2014). Pirates, prisoners, and preliterates: anarchic context and the private enforcement of law. *European Journal of Law and Economics*, 37(3):365–379.

Leeson, P. T. and Skarbek, D. B. (2010). Criminal constitutions. *Global Crime*, 11(3):279–297.

Lindsay, R., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., and Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology*, 76(6):796.

Lindsay, R. and Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, 70(3):556.

Lofstrom, M. and Raphael, S. (2013). Impact of realignment on county jail populations. *Public Policy Institute of California. Available at www. ppic. org/main/publication. asp.*

Lofstrom, M. and Raphael, S. (2015). Realignment, incarceration, and crime trends in California. *Public Policy Institute of California.*

Lofstrom, M., Raphael, S., and Grattet, R. (2014). Is public safety realignment reducing recidivism in California. *Public Policy Institute of California.*

McQuiston-Surrett, D., Malpass, R. S., and Tredoux, C. G. (2006). Sequential vs. simultaneous lineups: A review of methods, data, and theory. *Psychology, Public Policy, and Law*, 12(2):137.

Mecklenburg, S. H., Bailey, P. J., and Larson, M. R. (2008). The illinois field study: A significant contribution to understanding real world eyewitness identification issues. *Law and Human Behavior*, 32(1):22–27.

Megargee, E. I. (1977). The association of population density, reduced space, and uncomfortable temperatures with misconduct in a prison community. *American Journal of Community Psychology*, 5(3):289–298.

Morris, R. G. and Worrall, J. L. (2014). Prison architecture and inmate misconduct: A multilevel assessment. *Crime & Delinquency*, 60(7):1083–1109.

Orrick, E. A. and Morris, R. G. (2015). Do parole technical violators pose a safety threat? an analysis of prison misconduct. *Crime & Delinquency*, 61(8):1027–1050.

Peden, J. R. (1977). Property rights in Celtic Irish law. *Journal of Libertarian Studies*, 1(2):81–95.

Petersilia, J. and Snyder, J. (2013). Looking past the hype: 10 questions everyone should ask about California's prison realignment.

Radford, R. A. (1945). The economic organisation of a pow camp. *Economica*, 12(48):189–201.

Rakoff, J. S., Daumier, H., and Case, A. C. (2014). Why innocent people plead guilty. *The New York Review of Books*, 20.

Reisig, M. D. (2002). Administrative control and inmate homicide. *Homicide studies*, 6(1):84–103.

Ruback, R. B. and Carr, T. S. (1993). Prison crowding over time: The relationship of density and changes in density to infraction rates. *Criminal Justice and Behavior*, 20(2):130–148.

Skaperdas, S. (2001). The political economy of organized crime: providing protection when the state does not. *Economics of Governance*, 2(3):173–202.

Skaperdas, S. and Syropoulos, C. (1995). Gangs as primitive states. *The Economics of Organised Crime*, pages 61–82.

Skarbek, D. (2010). Putting the "con" into constitutions: The economics of prison gangs. *The Journal of Law, Economics, & Organization*, pages 183–211.

Skarbek, D. (2011). Governance and prison gangs. *American Political Science Review*, 105(4):702–716.

Skarbek, D. (2012). Prison gangs, norms, and organizations. *Journal of Economic Behavior & Organization*, 82(1):96–109.

Skarbek, D. (2014). *The social order of the underworld: How prison gangs govern the American penal system*. Oxford University Press.

South, C. R. and Wood, J. (2006). Bullying in prisons: The importance of perceived social status, prisonization, and moral disengagement. *Aggressive Behavior*, 32(5):490–501.

Steblay, N., Dysart, J., Fulero, S., and Lindsay, R. C. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: a meta-analytic comparison. *Law and human behavior*, 25(5):459.

Steblay, N. K., Dysart, J. E., and Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law*, 17(1):99.

Steiner, B. and Wooldredge, J. (2013). Implications of different outcome measures for an understanding of inmate misconduct. *Crime & Delinquency*, 59(8):1234–1262.

Sundt, J., Salisbury, E. J., and Harmon, M. G. (2016). Is downsizing prisons dangerous? *Criminology & Public Policy*, 15(2):315–341.

Symkovych, A. (2017). The 'inmate code' in flux: A normative system and extralegal governance in a ukrainian prison. *Current Sociology*, page 0011392117744596.

Trulson, C. R. (2007). Determinants of disruption: Institutional misconduct among state-committed delinquents. *Youth Violence and Juvenile Justice*, 5(1):7–34.

Trulson, C. R., Marquart, J. W., and Kawucha, S. K. (2006). Gang suppression and institutional control. *Corrections Today*, 68(2):26.

Wagenaar, W. A. and Veefkind, N. (1992). Comparison of one-person and many-person lineups: A warning against unsafe practices. *Psychology and law: International perspectives*, pages 275–285.

Walters, G. D. and Crawford, G. (2014). Major mental illness and violence history as predictors of institutional misconduct and recidivism: Main and interaction effects. *Law and Human Behavior*, 38(3):238.

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology*, 14(2):89–103.

Wells, G. L., Dysart, J. E., and Steblay, N. K. (2015a). The flaw in amendola and wixted's conclusion on simultaneous versus sequential lineups. *Journal of Experimental Criminology*, 11(2):285–289.

Wells, G. L., Steblay, N. K., and Dysart, J. E. (2015b). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*, 39(1):1.

White, A. A. (2008). The concept of less eligibility and the social function of prison violence in class societies. *Buff. L. Rev.*, 56:737.

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., and Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences*, 113(2):304–309.

Wooldredge, J., Griffin, T., and Pratt, T. (2001). Considering hierarchical models for research on inmate behavior: Predicting misconduct with multilevel data. *Justice Quarterly*, 18(1):203–231.

Yarmey, A. D., Yarmey, M. J., and Yarmey, A. L. (1996). Accuracy of eyewitness identifications in showups and lineups. *Law and Human Behavior*, 20(4):459–477.

APPENDIX A

Chapter 1 Appendix

A.1 Theory Appendix

Return to Section 2.4.

The objective of this model is to capture the major channels by which policy interventions that increase or decrease the population (and thus crowding) may impact violent behavior in prisons. Of particular interest is the role of changes in the composition of the prison population, with regards to individuals' tendency to resort to violence, that are the result of the manner in which the population size is increased or decreased. To this end, it is not the intent of this model to explore the motives underlying individual choices. Nor is it intended that the model capture the determinants of all forms of violent behavior. Since premeditated assaults and strategic gang violence are unlikely to be responsive to the crowdedness of the prison, the model is oriented to violent behavior that may be seen as impulsive or extemporaneous.

Recall that the three mechanisms defined in Section 2.4 were a behavioral mechanism, a structural mechanism, and a compositional mechanism. The setting is distilled into a reduced form: an applied probability model where pairwise interactions occur between inmates. These interactions are assumed to be contentious in nature and each has a probability of resulting in violence. Since the interest of this model is the effect of changing the composition of a set of heterogeneous individuals, the underlying rational choice problem of each individual is suppressed in this presentation of the model.[1] Instead, each individual is assumed to have a baseline *propensity* for violence, a_i, that is distinct from their *probability* of resorting to violence in any given situation. Propensity is assumed to be a fixed individual characteristic, whereas the probability is a function of situational factors (e.g. crowding) and the individual's fixed propensity. However, it is the overall rate of violence for the prison population that is modeled here, so the implications of these individual propensities are aggregated to the prison level variables in the following model set up.

V	*Total violence per unit time.*
P	*Total prison population.*
K	*Design capacity of the prison.*
c	*Degree of crowding, defined as the population/capacity ratio (P/K).*
s	*Scale effect as a function of P, generally taken to exhibit constant returns to scale.*
n	*Number of potentially violent interactions per unit time for each individual, with $n'(c) \geq 0$.*
π	*Probability that a particular interaction will be violent, with $\pi'(c) \geq 0$.*
λ^n	*A shift parameter for policy 'n' with a positive relationship to crowding.*
a_i	*An index measuring individual i's baseline propensity for violence.*
$F(\cdot)$	*The distribution of all inmates' a_i, with support* $[0,1]$.[2]

Given this setting, the aggregate incidence of violence in the prison is defined by the identity in Equation A.1. The equation simply states that the total violence in a unit of time will equal the total number of pairwise interactions that occur per unit of time[3] multiplied by the average probability that a single interaction becomes violent. This very basic representation of the setting allows the structural crowding mechanism to be captured by $n(c)$. Meanwhile $\pi(c, \lambda^n)$ captures both

[1] Future work on this research will include treatment of the underlying choice setting for inmates, as well as the strategic interaction between inmates that leads to taunting and baiting their opponents. As it pertains to the broader idea of compositional change, the author does not believe those stages to be instrumental.

[2] $F(\cdot)$ is an approximation of the true distribution from which an inmate's "partner" in any pairwise interaction is randomly is randomly drawn. For sufficiently large populations, the distributions are effectively identical.

[3] Under CRS, $s(P) \cdot n(c)$ can be simplified to $P \cdot n(c)$, which is the total population times the number of interactions per individual per unit of time.

the behavioral and compositional mechanisms, respectively the direct effect of c and the marginal effect of the shift parameter λ^n.

$$V = [s(P)][n(c)][\pi(c, \lambda^n)]. \tag{A.1}$$

To derive the responsiveness of violence to crowding and incorporate the principle that changes in crowding are the result of some policy change, it is assumed that crowding (and thus total population) is a function of the policy parameter (λ^n) with an elasticity of one ($E_{c:\lambda^n} = 1$). Furthermore, we allow that the relationship between the shift parameter (λ^n) and probability of violence ($\pi(\cdot)$) can vary between policy options. That is, the partial derivatives $\frac{\partial \pi}{\partial \lambda^1} < 0$ and $\frac{\partial \pi}{\partial \lambda^2} > 0$ are explicitly permitted in this setting.

The elasticity of violence with respect to a population shock is derived as follows.

$$\begin{aligned}
V &= s(P)n(c)\pi(c, \lambda^n) \\
&= s(cK)n(c)\pi(c, \lambda^n) \\
\implies ln[V] &= ln[s(cK)] + ln[n(c)] + ln[\pi(c, \lambda^n)] \\
\implies \frac{1}{V}\frac{dV}{d\lambda^n} &= \frac{1}{s(P)}s'(P) \cdot K\frac{dc}{d\lambda^n} + \frac{1}{n(c)}n'(c)\frac{dc}{d\lambda^n} + \frac{1}{\pi(c, \lambda^n)}\Big[\frac{\partial \pi}{\partial c}\frac{dc}{d\lambda^n} + \frac{\partial \pi}{\partial \lambda^n}\Big].
\end{aligned}$$

Then multiply through by λ^n.

$$\begin{aligned}
\frac{\lambda^n}{V}\frac{dV}{d\lambda^n} &= \Big[\frac{1}{s(P)}s'(P) \cdot K + \frac{1}{n(c)}n'(c) + \frac{1}{\pi(c, \lambda^n)}\frac{\partial \pi}{\partial c}\Big]\frac{dc}{d\lambda^n}\frac{\lambda^n}{c}c + \frac{\lambda^n}{\pi(c, \lambda^n)}\frac{\partial \pi}{\partial \lambda^n} \\
\implies E_{V:\lambda^n} &= \Big[\frac{1}{s(P)}s'(P) \cdot K + \frac{1}{n(c)}n'(c) + \frac{1}{\pi(c, \lambda^n)}\frac{\partial \pi}{\partial c}\Big]c \cdot E_{c:\lambda^n} + E_{\pi:\lambda^n} \\
&= \Big[\frac{cK}{s(P)}s'(P) + \frac{c}{n(c)}n'(c) + \frac{c}{\pi(c, \lambda^n)}\frac{\partial \pi}{\partial c}\Big]E_{c:\lambda^n} + E_{\pi:\lambda^n} \\
&= \big[E_{s:P} + E_{n:c} + E_{\pi:c}\big]E_{c:\lambda^n} + E_{\pi:\lambda^n}.
\end{aligned}$$

Now recall that λ^n is defined such that $E_{c:\lambda^n} = 1$ and note that CRS[4] requires that $E_{s:P} = 1$. Then the total policy impact can be decomposed into the scale effect plus an elasticity that represents each of the three mechanisms, as shown in Equation A.2.

$$E_{V:\lambda^n} = 1 + E_{n:c} + E_{\pi:c} + E_{\pi:\lambda^n}. \tag{A.2}$$

The first terms on the right-hand side of Equation A.2 constitute the direct elasticity of violence with respect to crowding, $E_{V:c}$. Therefore Equation A.3, the same as presented in Section 2.4, is an equivalent representation of the relationship between violence and the policy parameter.

$$E_{V:\lambda^n} = E_{V:c} + E_{\pi:\lambda^n}. \tag{A.3}$$

There is much yet to be discovered about the motivations and determinants of violent behavior, in and out of prisons. What is offered here is the simple insight that there are several potential mechanisms at work that determine the net effect of changing prison crowding on violent behavior. Furthermore, with sufficiently rich data, it should be possible to perform a decomposition of the overall impact of a policy intervention and differentiate between these mechanisms. Lastly, in the absence of very granular data, estimates using any significant variation in crowding should be viewed as estimates of the relationship represented by $E_{V:\lambda^n}$ and not $E_{V:c}$.

There is an argument to be made that the latter point is not necessarily an issue. For example, in a study that is a straightforward impact evaluation following a new law or regulation, the effects of each mechanism are all part of the impact that the law had on violence and thus rightfully included in the analysis. The caution raised by this model is in proper interpretation. AB 109 and the results of this research provide a poignant example. Separate impact evaluations for reception centers and level 2 facilities would likely show no significant impact on violence in the former and a decrease in the rate of violence for the latter. Attributing these directly to crowding would imply that violent

[4] CRS in this setting implies that, in the absence of any crowding or compositional changes, doubling the population size will double the total amount of violence.

behavior in reception populations is not responsive to crowding. While the model presented here does not preclude that possibility, it does raise a plausible alternate explanation that simultaneously accounts for the difference in outcomes at level 2 facilities.

The framework is also helpful for conceptualizing the role that the design of an intervention plays. The impact of AB 109 on reception centers is a good illustration. In the initial stage, AB 109 dramatically reduces crowding in reception facilities and does so by eliminating the flow of incoming non-violent offenders. Since reception centers take custody of all other offenders upon arrival in the prison system, now mostly very serious offenders, it can be expected that there is a significant compositional change, in addition to the large decrease in crowding. These are expressed in Equation A.2 as $E_{n:c} > 0$ and/or $E_{\pi:c} > 0$ from the crowding mechanisms and $E_{\pi:\lambda^n} < 0$ from compositional change. But there is also the reception adjustment, which consolidates the remaining reception populations into fewer facilities. This diminishes the overall drop in crowding, which pertains to the first two elasticities, but maintains the compositional change associated with the very large reduction in total reception population size, which acts through the latter elasticity.

In summary, empirical work on the relationship between prison crowding and violence typically claims to estimate that relationship directly. The model presented here questions whether that is really the case, based on two assumptions: there is heterogeneity among inmates with regards to their propensity for violence and increasing or decreasing crowding is associated with altering the composition of the inmate population with respect to this heterogeneity. Where possible, a decomposition of the policy effects on violence should be performed to properly illuminate the distinct roles of crowding and composition. With less precise data, logic dictates that some policies will have a predictable pattern of selection by which they cause compositional change and this can be used to make inferences about the manner in which compositional change may bias estimates from the data.

Return to Section 2.4.

A.2 Empirical Appendix

This appendix provides figures and regression estimates that have been excluded from the main body of the paper. This additional information tests alternate specifications of the models and provides additional context to the California prison setting.

One piece of context for understanding violence in California prisons is the difference in baseline rates of violence between the subpopulations. Due to the same data limitations with units of observation that constrain the main identification,[5] exact averages by subpopulation are not possible. Instead, Table A.1 reports a basic OLS regression of the rate of assault on shares of each major subpopulation. The regression is estimated without a constant term, so the point estimates can be viewed as rough approximations of the average rate of assault for facilities with that subpopulation. The negative coefficient on the special needs population emphasizes the fact that there are confounding factors in these approximations. Nonetheless, the estimates follow an intuitive pattern that should be expected. The rates of violence are monotonically increasing in security classification and reception centers fall within bounds set by the security levels. Even a high rate of violence at reception centers relative to level 2 and 3 facilities aligns well with the idea that stability in reception centers is disrupted by the high rate of turnover.

Table A.2 presents the first-stage estimates of the IV specifications from Section 2.6. The level 2 and level 3 instruments are very consistent predictors of decreased crowding across all specifications. Most of the interesting variation between specifications occurs in the reception instrument ($Months * Rec$) and the months since implementation variable ($Months$). Adding the additional RA interaction with the reception instrument in column (2) dramatically increases the significance and the magnitude of the coefficient on $Months * Rec$. At the same time, column (2) is the only specification for which $Months$ has little correlation with decreased crowding. Furthermore, the coefficients on the two terms for the reception instrument are the inverse of each other, neatly off-

[5] The unit of observation in the data is a prison. Each prison has several facilities that typically house different types of subpopulation, therefore the rate of violence for each prison cannot be attributed to a single subpopulation.

Table A.1: OLS Regression of Population Shares on Assaults

Dependent Variable: Rate of Assault per 100 Inmates	
	(1)
VARIABLES	Assault
Security Level 1	0.244***
	(0.0273)
Security Level 2	0.311***
	(0.0179)
Security Level 3	0.623***
	(0.0266)
Security Level 4	1.148***
	(0.0370)
Reception Center	0.934***
	(0.0236)
Special Needs	-0.295***
	(0.0412)
Observations	1,440

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

setting one another once a reception adjustment occurs. Together, this suggests that the pre-AB 109 reception share is correlated with large reductions in crowding during the early months of the new law, but this correlation dissipates as reception adjustments begin to offset the crowding reductions. When the specifications do not allow for this mid-shock change in the marginal effect of the reception instrument, the correlation with initial reductions in crowding due to reception share are spread between the reception instrument and the $Months$ variable, which has a more consistent negative correlation.

Tables A.3 and A.4 are tests of the robustness of the DD estimation strategy. Table A.3s replicates the Table 2.5 specifications, but excludes the outlier that is shown in Figure 2.9 and excluded from Figure 2.6. The coefficients for the reception treatment are clearly not dependent upon this outlier. Table A.4 repeats the DD specification time fixed effects for alternate outcome variables, as Table 2.7 did with the IV strategy. Using incidents rather than disciplinaries to measure assaults undermines the estimates for this strategy. As alluded to earlier, there remain unanswered questions about the

Table A.2: First-stage Regressions from IV Estimates in Table 2.6

VARIABLES	(1) Base IV	(2) Interact RA	(3) Exclusion	(4) Drop 3mo.
Months	-0.222	-0.150	-0.222	-0.194
	(0.0958)	(0.0883)	(0.0958)	(0.1000)
Months_sq	0.146	0.0630	0.146	0.132
	(0.0793)	(0.0702)	(0.0793)	(0.0853)
Months*Lv2	-0.290	-0.309	-0.290	-0.302
	(0.0751)	(0.0761)	(0.0751)	(0.0760)
Months*Lv3	-0.324	-0.352	-0.324	-0.329
	(0.0743)	(0.0719)	(0.0743)	(0.0729)
Months*Rec	-0.158	-1.223	-0.158	-0.0733
	(0.0929)	(0.195)	(0.0929)	(0.0883)
Months*Rec*RA		1.210		
		(0.216)		
Observations	1,440	1,440	1,440	1,350
Number of ID	30	30	30	30
Controls	X	X	X	X

Robust standard errors in parentheses

selection process by which incidents are reported because they are far less frequent than disciplinaries. The significant coefficient for the reception treatment on drug possession is likely a function of the law disrupting the channels by which contraband is funneled into prisons.[6]

Tables A.5 and A.6 test the effects of changing the measure of violence on the IV strategy. Table A.5 includes murder and attempted murder with the original measures of assault and battery. This leads to slightly larger point estimates, but no substantive changes in outcomes. Table A.6 excludes assaults on staff members from the outcome variable and finds diminished statistical significance and point estimates in each specification. This suggests that variation in the rate of aggression towards staff is a dimension in which crowding impacts violence.

The effects of the reception adjustment on alternate subpopulations is illustrated in Figures A.1 through A.4. Although there is some time variation following AB 109 for each of these, none of them follow the distinct pattern shown for the security level 3 population.

[6] During an informal interview with an inmate in a California prison, the author was told that one way prison gangs funnel drugs into the prison was to have someone out on parole swallow baggies, then violate their parole to bring them inside the prison. By redirecting parole violators to county jails, AB 109 disrupted part of the supply chain.

Table A.3: DD Estimation: DVI (Outlier) Excluded from Observations.

Dependent Variable: Log Rate of Assaults

VARIABLES	(1) No Controls	(2) Main	(3) TimeFE	(4) 3mo.Gap
ShareLv2*Post ($\hat{\beta}_{12}$)	-0.537	-0.405	-0.406	-0.459
	(0.176)	(0.144)	(0.123)	(0.165)
ShareLv3*Post ($\hat{\beta}_{13}$)	-0.304	-0.217	-0.211	-0.307
	(0.278)	(0.267)	(0.145)	(0.307)
ShareRec*Post ($\hat{\beta}_{1R}$)	0.241	0.277	0.207	0.398
	(0.187)	(0.176)	(0.180)	(0.0993)
Observations	1,421	1,421	1,421	1,334
Controls	None	X	X	X
Trend/Gap	No	No	No	Yes

Robust standard errors in parentheses

Table A.4: DD Model Placebo Tests

Dep. Variable: Log Rate of the given form of misconduct.

VARIABLES	(1) Assaults	(2) Incidents	(3) Drugs	(4) Cellphone
ShareLv2*Post ($\hat{\beta}_{12}$)	-0.314	-0.175	-0.162	-0.210
	(0.156)	(0.143)	(0.178)	(0.242)
ShareLv3*Post ($\hat{\beta}_{13}$)	-0.233	-0.0917	0.00244	0.197
	(0.175)	(0.160)	(0.200)	(0.273)
ShareRec*Post ($\hat{\beta}_{1R}$)	0.0585	-0.0689	-0.550	-0.0891
	(0.204)	(0.187)	(0.233)	(0.320)
Observations	780	780	780	779
Controls	X	X	X	X
Trend/Gap	No	No	No	No

Standard errors in parentheses

Table A.5: IV Model: Murder and Attempted Murder Included in Outcome Variable

Dep. Variable: Log Rate of Violence per 100 inmates				
	(1)	(2)	(3)	(4)
VARIABLES	Base IV	Interact RA	Exclusion	Drop 3mo.
Crowding (P/K)	2.208***	1.656***	1.911**	2.338***
	(0.604)	(0.481)	(0.755)	(0.605)
Observations	1,440	1,440	1,440	1,350
Number of ID	30	30	30	30
Controls	X	X	X+Months	X
Exclusions	Base	Months*S*RA	Months*S	Base
F-test IVs	10.34	26.81	10.05	10.59

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.6: IV Model: Staff Assaults Excluded from Outcome Variable

Dep. Variable: Log Rate of Inmate-on-inmate Assault				
	(1)	(2)	(3)	(4)
VARIABLES	Base IV	Interact RA	Exclusion	Drop 3mo.
Crowding (P/K)	1.763**	1.123*	1.521*	1.958***
	(0.752)	(0.605)	(0.872)	(0.744)
Observations	1,440	1,440	1,440	1,350
Number of ID	30	30	30	30
Controls	X	X	X+Months	X
Exclusions	Base	Months*S*RA	Months*S	Base
F-test IVs	10.34	26.81	10.05	10.59

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

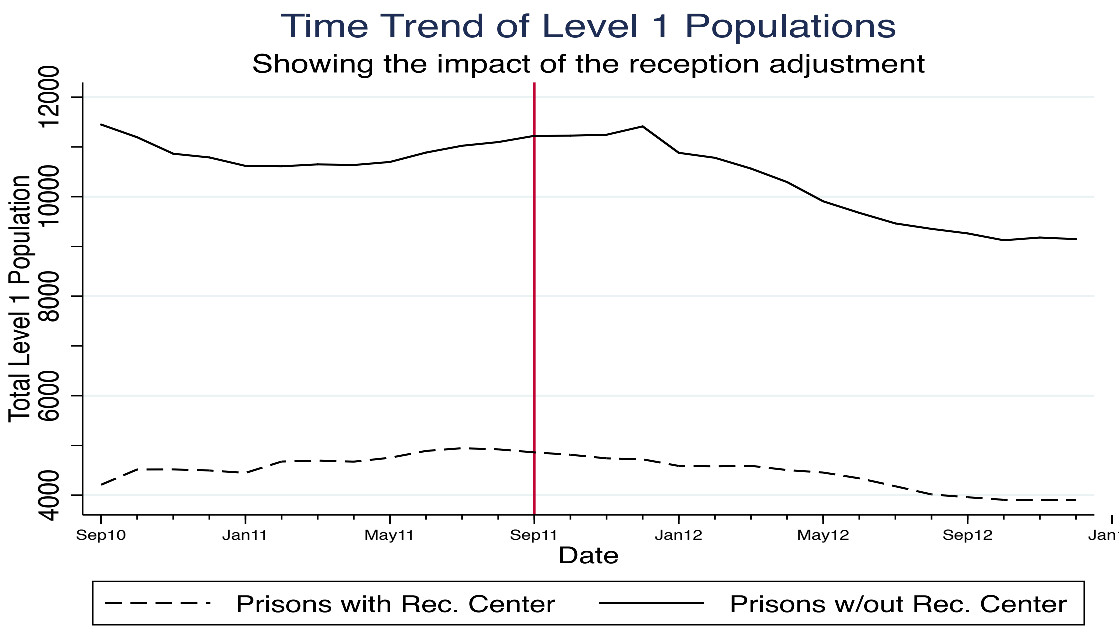


Figure A.1: *The impact of the reception adjustment on the level 1 subpopulation. The time trends are for sum of security level 1 populations parsed by whether the prison has a reception center facility or not. The vertical line denotes the last observation prior to implementation of AB 109.*

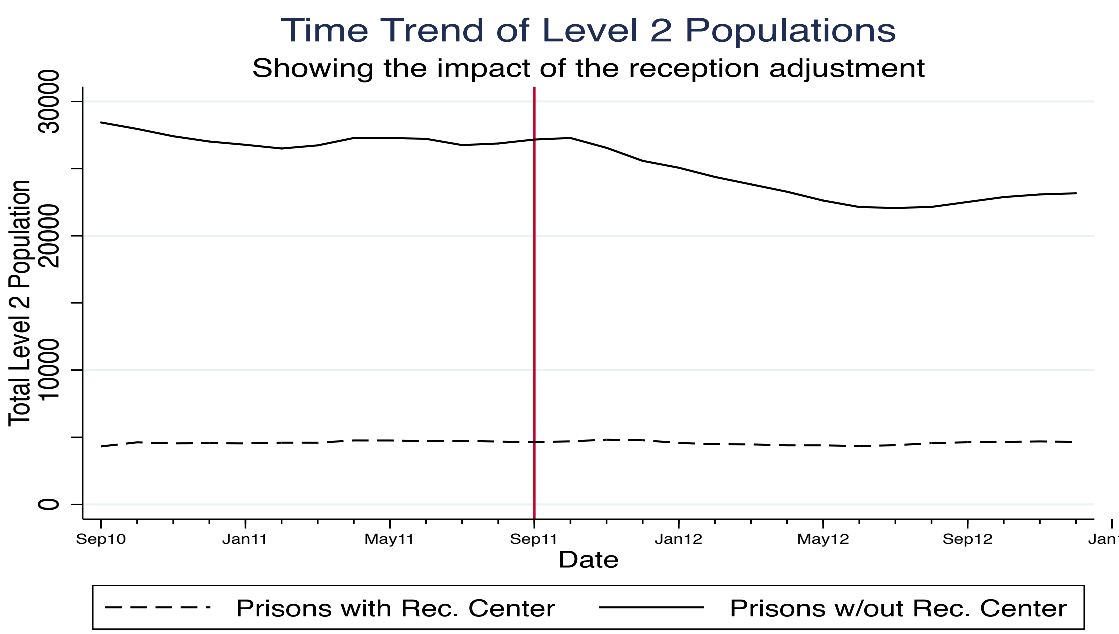


Figure A.2: *The impact of the reception adjustment on the level 2 subpopulation. The time trends are for sum of security level 2 populations parsed by whether the prison has a reception center facility or not. Note that the effect of RA is conflated with that of AB 109 for this subpopulation. The vertical line denotes the last observation prior to implementation of AB 109.*

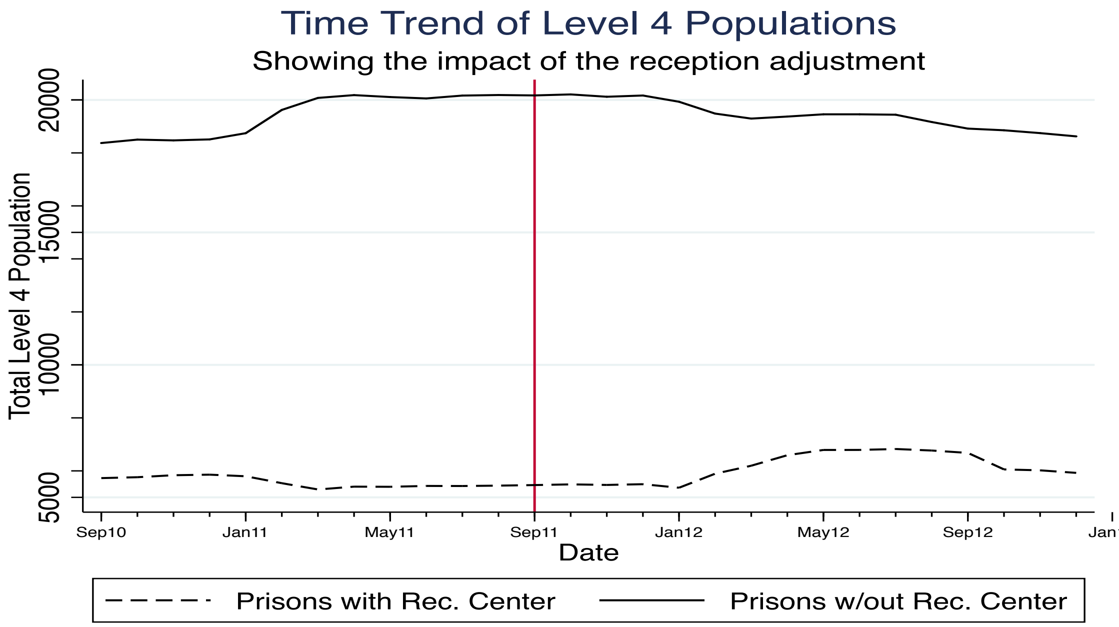


Figure A.3: *This figure shows the impact of the reception adjustment on the level 4 subpopulation. The time trends are for sum of security level 4 populations parsed by whether the prison has a reception center facility or not. The vertical line denotes the last observation prior to implementation of AB 109.*

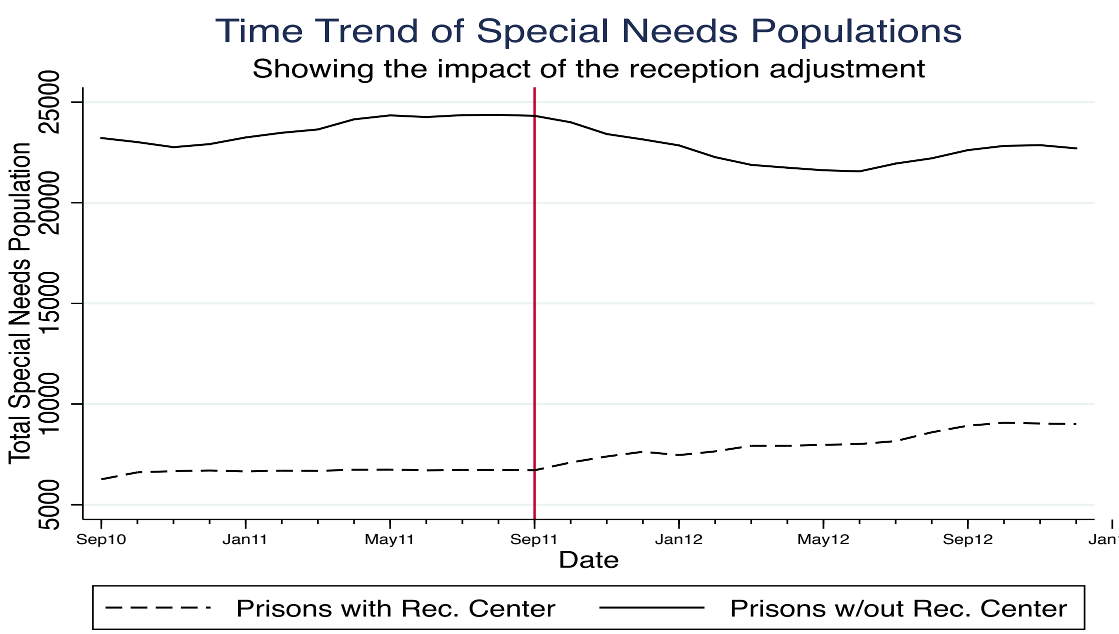


Figure A.4: *The impact of the reception adjustment on the special needs subpopulation. The time trends are for sum of special needs populations parsed by whether the prison has a reception center facility or not. The vertical line denotes the last observation prior to implementation of AB 109.*

Figure A.5 shows the change in several measures of misconduct, measured over the six months preceding implementation of AB 109 and the fourth through ninth months following it. This figure provides context for the decision to focus on misconduct measured via rates of assault. The other forms of misconduct are generally less stable and subject to more uncertain sources of variation.

Figure A.6 shows changes in the participation rate (per 100 inmates) in the programs included in the standard set of controls. For these, some increases are expected since population is decreasing and there is no reason to expect a decrease in program capacities. Most notable in this figure are the decrease in academic enrollment and the large increase in Substance Abuse Treatment Facility (SATF) participation at reception facilities. The change in academic enrollment, if significant at all, is likely due to some statewide institutional change (such as a reduction in education funding) since it is relatively uniform across the three types of prisons. The increase in SATF beds can be partially explained by the population decrease, but could also imply increased demand for substance abuse treatment since the waitlist for the program increases quite a bit as well.[7]

Return to Section 2.6

[7] The curious null effect for both SATF categories in the 'Others' group of prisons is due to the fact that all SATFs are in prisons with a reception center or high proportion of security level 2 inmates.

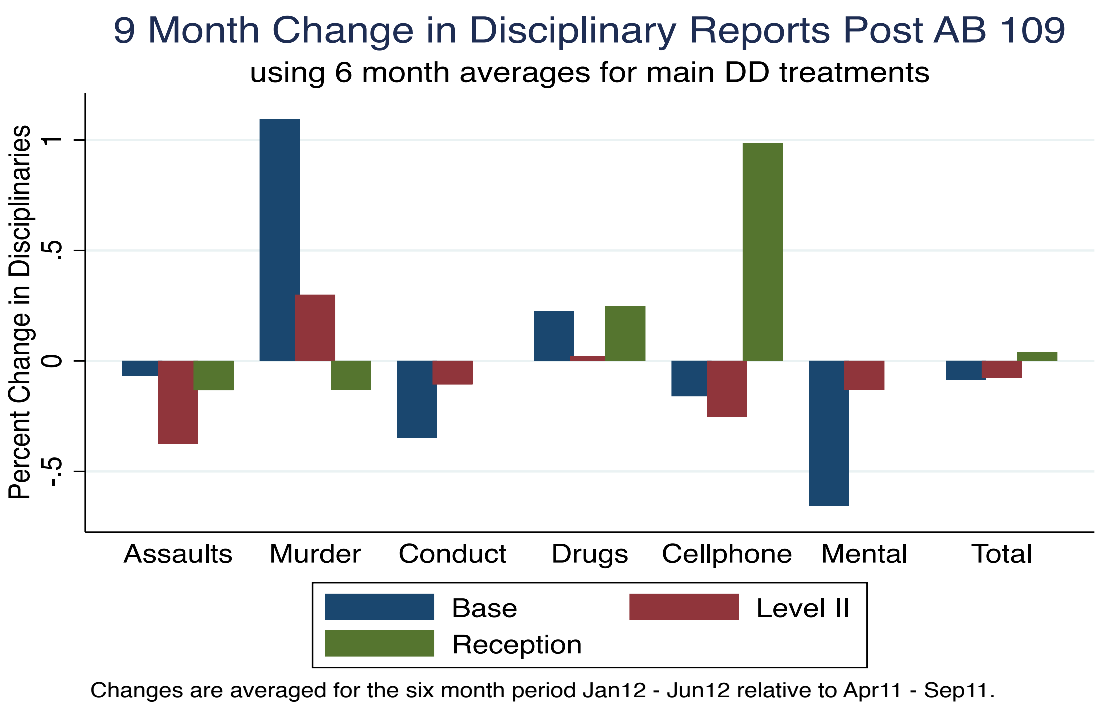


Figure A.5: *Changes in the rate (per 100 inmates) of several types of disciplinaries. The changes are in the six month average measured from January 2012 through June 2012, relative to the average just prior to implementation of AB 109, April 2011 through September 2011. Source: Generated from CDCR CompStat reporting data.*

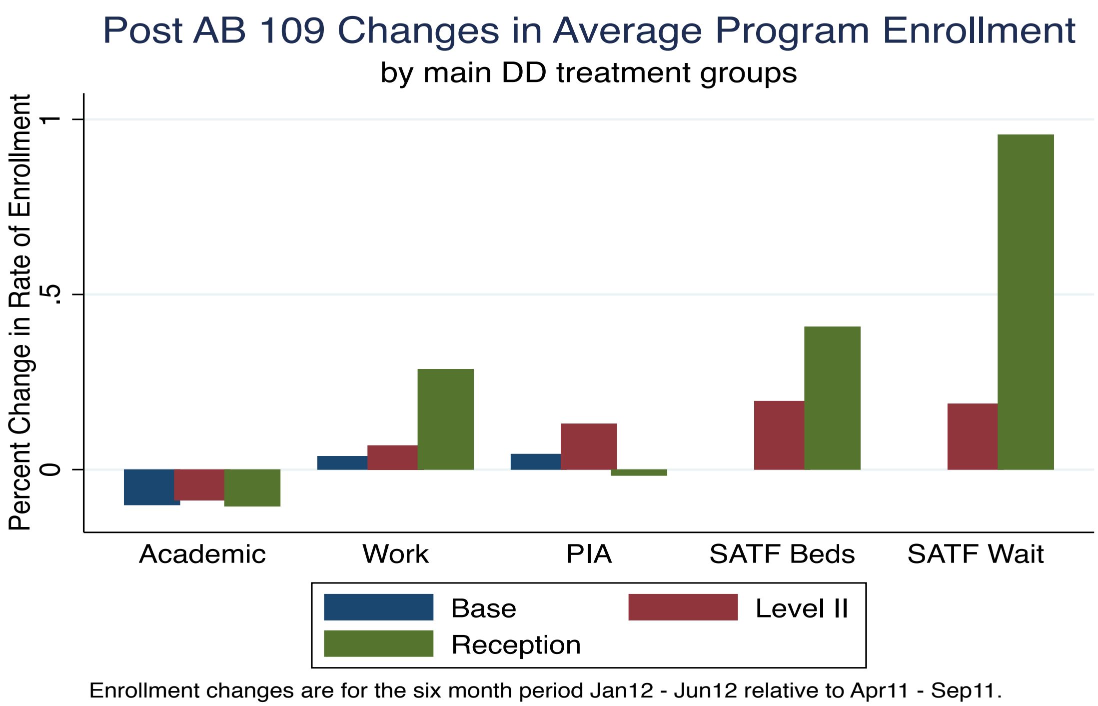


Figure A.6: *Changes in the rate (per 100 inmates) of several types of program enrollment. The changes are in the six month average measured from January 2012 through June 2012, relative to average just prior to implementation of AB 109, April 2011 through September 2011. Source: Generated from CDCR CompStat reporting data.*

APPENDIX B

Chapter 2 Appendix

B.1 Technical Appendix

Derivatives of $\phi(v_i, \bar{v}_{-i}; c_i, \theta, \rho)$.

Each derivative is fairly straightforward, the presence of implicit functions adding a moderate degree of complexity. Where they appear in the text, they are represented in more compact form.

From the text:

- $\phi(v_i, \bar{v}_{-i}; c_i, \theta, \rho) = (P_i - c_i)Q(P_i, \rho)\frac{\partial s_i}{\partial v_i} - \frac{\partial f(v_i, r_i, \theta)}{\partial v_i}$
- $\frac{\partial f(v_i, r_i, \theta)}{\partial v_i} = f_1(v_i, r_i, \theta) + f_2(v_i, r_i, \theta)\frac{\partial r_i}{\partial v_i}$
- $\frac{\partial^2 f(v_i, r_i, \theta)}{\partial v_i^2} = f_{11}(v_i, r_i, \theta) + f_{22}(v_i, r_i, \theta)(\frac{\partial r_i}{\partial v_i})^2 + 2(f_{21}(v_i, r_i, \theta)\frac{\partial r_i}{\partial v_i}) + f_2(v_i, r_i, \theta)\frac{\partial^2 r_i}{\partial v_i^2}$

Own violence:

$$\begin{aligned}
\frac{\partial \phi}{\partial v_i} &= (P_i^* - c_i)Q(P_i^*, \rho)\frac{\partial^2 s_i}{\partial v_i^2} - \frac{\partial^2 f(v_i, r_i, \theta)}{\partial v_i^2} \\
&= (P_i^* - c_i)Q(P_i^*, \rho)\frac{\partial^2 s_i}{\partial v_i^2} - f_{11}(v_i, r_i, \theta) + f_{22}(v_i, r_i, \theta)(\frac{\partial r_i}{\partial v_i})^2 + 2(f_{21}(v_i, r_i, \theta)\frac{\partial r_i}{\partial v_i}) + f_2(v_i, r_i, \theta)\frac{\partial^2 r_i}{\partial v_i^2} < 0
\end{aligned}$$

Each term is strictly less than zero except for $-2(f_{21}(v_i, r_i, \theta)\frac{\partial r_i}{\partial v_i})$ – so the derivative is negative as long as the cross partial of $f(\cdot)$ is sufficiently small, which is implied in the equilibrium conditions of the Hessian.

Others' violence:

$$\begin{aligned}
\frac{\partial \phi}{\partial \bar{v}_{-i}} &= (P_i^* - c_i)Q(P_i^*, \rho)\frac{\partial^2 s_i}{\partial v_i \partial \bar{v}_{-i}} - (f_{12}(v_i, r_i, \theta) + f_{22}(v_i, r_i, \theta)\frac{\partial r_i}{\partial v_i})\frac{\partial r_i}{\partial \bar{v}_{-i}} - f_2(v_i, r_i, \theta)\frac{\partial^2 r_i}{\partial v_i \partial \bar{v}_{-i}} \\
&= (P_i^* - c_i)Q(P_i^*, \rho)\frac{\partial^2 s_i}{\partial v_i \partial \bar{v}_{-i}} - (f_{12}(v_i, r_i, \theta) + f_{22}(v_i, r_i, \theta)\frac{\partial r_i}{\partial s}\frac{\partial s}{\partial v_i})\frac{\partial r_i}{\partial s}\frac{\partial s}{\partial \bar{v}_{-i}} \\
&\qquad - f_2(v_i, r_i, \theta)(\frac{\partial^2 r_i}{\partial s^2}\frac{\partial s}{\partial \bar{v}_{-i}} + \frac{\partial r_i}{\partial s}\frac{\partial^2 s}{\partial v_i \partial \bar{v}_{-i}}) >< 0
\end{aligned}$$

As discussed in the text, the sign of ϕ_2 is indeterminate due to the several factors. First, the sign of $\frac{\partial^2 s_i}{\partial v_i \partial \bar{v}_{-i}}$ is indeterminate, generally switching from negative to positive as v_i due to the constraint that all shares sum to one. Second, for any sign that $\frac{\partial^2 s_i}{\partial v_i \partial \bar{v}_{-i}}$ takes and given the restrictions made on the functional forms, one term of the derivative has the opposing sign to the other two. The second term, as it is presented above, is strictly positive. The first and last terms are positive and negative, respectively, when $\frac{\partial^2 s_i}{\partial v_i \partial \bar{v}_{-i}}$ is positive, and vice versa. Since $\frac{\partial^2 s_i}{\partial v_i \partial \bar{v}_{-i}} = 0$ for some value $\hat{v}_i$, there must exist a "Competitive" zone where $\phi_2 > 0$ implies that gangs respond to the violence of their rivals with violence of their own. In either direction from $\hat{v}_i$ the sign of ϕ_2 can become negative. Furthermore, as v_i goes to zero ϕ_2 necessarily becomes negative because $\frac{\partial^2 s_i}{\partial v_i \partial \bar{v}_{-i}} < 0$, $\frac{\partial r_i}{\partial s} \implies 0$, $f_2(v_i, r_i, \theta) \implies 0$, and the profit margin, $(P_i^* - c_i)$, remains strictly positive.

Marginal production costs:

$$\frac{\partial \phi}{\partial c_i} = [(P_i^* - c_i)\frac{\partial Q}{\partial P_i^*}\frac{\partial P_i^*}{\partial c_i} + (\frac{\partial P_i^*}{\partial c_i} - 1) \cdot Q(P_i^*, \rho)]\frac{\partial s_i}{\partial v_i} < 0$$

This sign is a straightforward result of the monopolistic first order conditions on P^*, which give $0 < \frac{\partial P_i^*}{\partial c_i} < 1$ and make both terms in the brackets strictly negative. Intuitively the market rents are decreasing in c_i, so the incentive to exert costly effort is decreasing as well.

Marginal cost of violence:

$$\frac{\partial \phi}{\partial \theta} = -\frac{\partial^2 f(v_i, r_i, \theta)}{\partial v_i \partial \theta} < 0$$

Market scale:

$$\frac{\partial \phi}{\partial \rho} = [\frac{dP_i^*}{d\rho} Q(P_i^*, \rho) + (P_i^* - c_i)(\frac{\partial Q}{\partial P_i^*}\frac{\partial P_i^*}{\partial \rho} + \frac{\partial Q}{\partial \rho}]\frac{\partial s_i}{\partial v_i} > 0$$

By definition, ρ scales up demand at all prices. Thus, $Q(P_i)$ increases for all P_i and by $FOC(P_i^*)$ we get $\frac{\partial Q}{\partial \rho} > \frac{\partial Q}{\partial P_i^*}\frac{\partial P_i^*}{\partial \rho}$.

Signing the Comparative Statics

The full system of equations:

$$\begin{bmatrix} \Phi_1^1 & \Phi_2^1 \\ \\ \Phi_1^2 & \Phi_2^2 \end{bmatrix} \begin{bmatrix} \partial v_1 \\ \\ \partial v_2 \end{bmatrix} + \begin{bmatrix} \Phi_3^1 \\ \\ \Phi_3^2 \end{bmatrix} \partial c_1 + \begin{bmatrix} \Phi_4^1 \\ \\ \Phi_4^2 \end{bmatrix} \partial c_2 + \begin{bmatrix} \Phi_5^1 \\ \\ \Phi_5^2 \end{bmatrix} \partial \theta + \begin{bmatrix} \Phi_6^1 \\ \\ \Phi_6^2 \end{bmatrix} \partial \rho = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix} \tag{B.1}$$

Comparative statics results:

$$\frac{\partial v_1^*}{\partial c_1} = \frac{1}{\Delta}\Big(-\phi_3^1 \cdot \phi_2^2 - (-\phi_3^2) \cdot \phi_2^1 \Big) < 0$$

$$\frac{\partial v_1^*}{\partial c_2} = \frac{1}{\Delta}\Big(-\phi_4^1 \cdot \phi_2^2 - (-\phi_4^2) \cdot \phi_2^1 \Big) \sim ??$$

$$\frac{\partial v_1^*}{\partial \theta} = \frac{1}{\Delta}\Big(-\phi_5^1 \cdot \phi_2^2 - (-\phi_5^2) \cdot \phi_2^1 \Big) <^* 0$$

$$\frac{\partial v_1^*}{\partial \rho} = \frac{1}{\Delta}\Big(-\phi_6^1 \cdot \phi_2^2 - (-\phi_6^2) \cdot \phi_2^1 \Big) >^* 0$$

Stability Condition:

$$\frac{1}{\Delta}\Big(\phi_1^1 \cdot \phi_2^2 - \phi_1^2 \cdot \phi_2^1 \Big) \geq 0$$

Now converting the previous results to the duopolist notation, we have:

- $\Phi_1^1 < 0$ and $\Phi_2^2 < 0$
- Φ_2^1 ? 0 and Φ_1^2 ? 0
- $\Phi_3^1 < 0$ and $\Phi_3^2 = 0$
- $\Phi_4^1 = 0$ and $\Phi_4^2 < 0$

- $\Phi_5^1 < 0$ and $\Phi_5^2 < 0$
- $\Phi_6^1 > 0$ and $\Phi_6^2 > 0$

Now when we consider the comparative statics from the text, the sign for the c_1 is a direct application of the signs above and the indeterminance of the sign on the rival's production cost, c_2, is because it is the inverse of the sign on Φ_1^2, which is also indeterminant. The signs of the last two comparative static conditions, for θ and ρ, are generally stable due to the stability condition that is also presented above. However, when we the difference between the stability condition and the comparative static condition, it is possible for the sign to flip for a range of values when the magnitude of $\frac{\Phi_6^1}{\Phi_6^2}$ is sufficiently smaller than that of $\frac{\Phi_1^1}{\Phi_1^2}$ (and analogously for Φ_5^1 and Φ_5^2).

APPENDIX C

Chapter 3 Appendix

C.1 Technical Appendix

This appendix supplies several figures and tables of estimates that support the findings and analysis in the body of the paper. The first three tables are robustness checks showing that the point estimates from table 4.4 are not sensitive to specific choices that were made in modeling and cleaning the data. The first table shows the DDD model estimates when early years, where we believe there is more measurement error, are excluded; the second table presents the estimates when HI, ME, and NY are excluded because of irregularities int their clearance data; the last table excludes the four reform states that did not include sequential lineups in their eyewitness reforms.

The first set of figures show our best representation of parallel trends for the DD specification. Figure C.1 is the most pertinent of them, since the other two are for the crime categories that are used as counterfactuals. Since we need a single threshold year to show parallel trends leading up to, the three early-adopting reform states were excluded from the figures. Thus the trend lines are the averages for the non-reform states and the 10 reform states that adopted after 2010.

The final set of figures show the clearance rate trends for each state individually. The trends for each of the crime categories used in this paper are shown. A red line indicates the year of eyewitness reforms, if the state had one at all. The raw data is shown in these figures, meaning that missing data has not been filled in with artificial observations.

DDD Model 2b - Period beginning 1997

	(1)	(2)	(3)
VARIABLES	Robbery	Auto Theft	Burglary
ReformPost (α_1)	-0.00930	-0.0137	0.00181
	(0.0163)	(0.0110)	(0.00570)
Observations	881	881	881
FE - State	Y	Y	Y
FE - Year	Y	Y	Y
$\hat{\beta}$		0.0044	-0.0111

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C.1: DDD Model Robustness – Early data. This table is a replication of table 4.4 from section 4.3.2, except the first six years of data have been excluded to ensure that the estimates are not driven by the higher rate of measurement error in the early years of data.

DDD Model 2c - Excluding More States

	(1)	(2)	(3)
VARIABLES	Robbery	Auto Theft	Burglary
ReformPost (α_1)	-0.0144	-0.00859	0.000801
	(0.0155)	(0.0131)	(0.00678)
Observations	1,146	1,146	1,146
FE - State	Y	Y	Y
FE - Year	Y	Y	Y
$\hat{\beta}$		-0.0058	-0.0152

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C.2: DDD Model Robustness – Unusual trends. This table is a replication of table 4.4 from section 4.3.2, except the states of HI, ME, and NY are excluded from the regressions (in addition to KS and IL) to ensure that the estimates are not driven by unusual variation in the time trends of these states.

DDD Model 2d - Excluding Non-Sequential States

VARIABLES	(1) Robbery	(2) Auto Theft	(3) Burglary
ReformPost (α_1)	-0.0123	-0.0165	-0.000799
	(0.0225)	(0.0160)	(0.00830)
Observations	1,123	1,123	1,123
FE - State	Y	Y	Y
FE - Year	Y	Y	Y
$\hat{\beta}$		0.0042	-0.0115

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C.3: DDD Model Robustness – Non-sequential reforms. This table is a replication of table 4.4 from section 4.3.2, except the states of FL, MT, OH, and RI are excluded from the regressions (in addition to KS and IL) as a robustness check since these states did not include sequential lineups in their reforms.

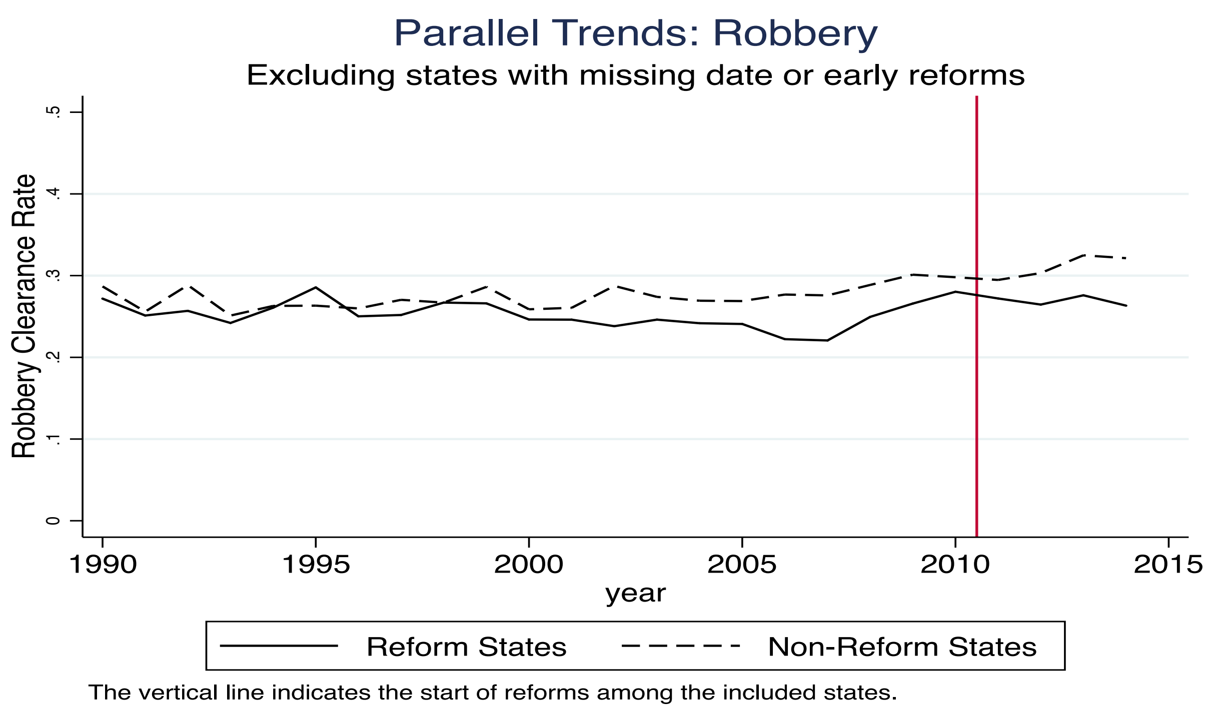


Figure C.1: *Parallel Trends for DD model – Robbery. Reform states that adopted reforms prior to 2011 excluded. Dependent variable: Robbery Clearance Rate.*

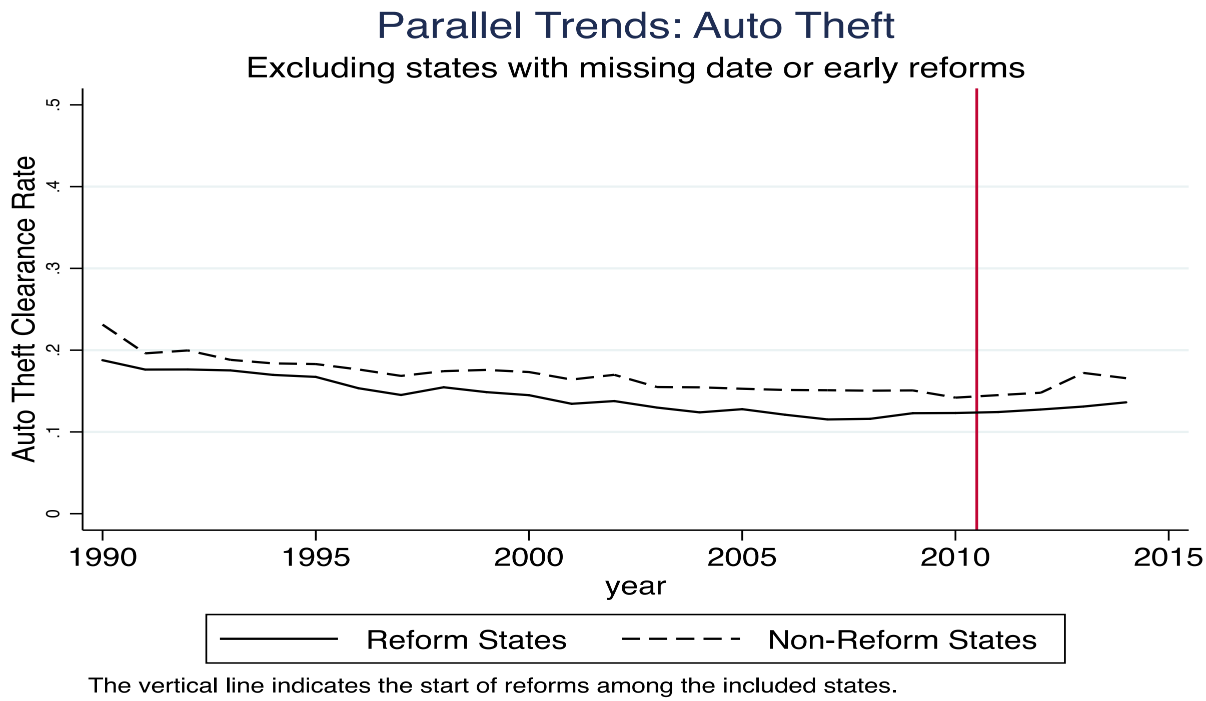


Figure C.2: *Parallel Trends for DD model – Auto theft. Reform states that adopted reforms prior to 2011 excluded. Dependent variable: Auto Theft Clearance Rate.*

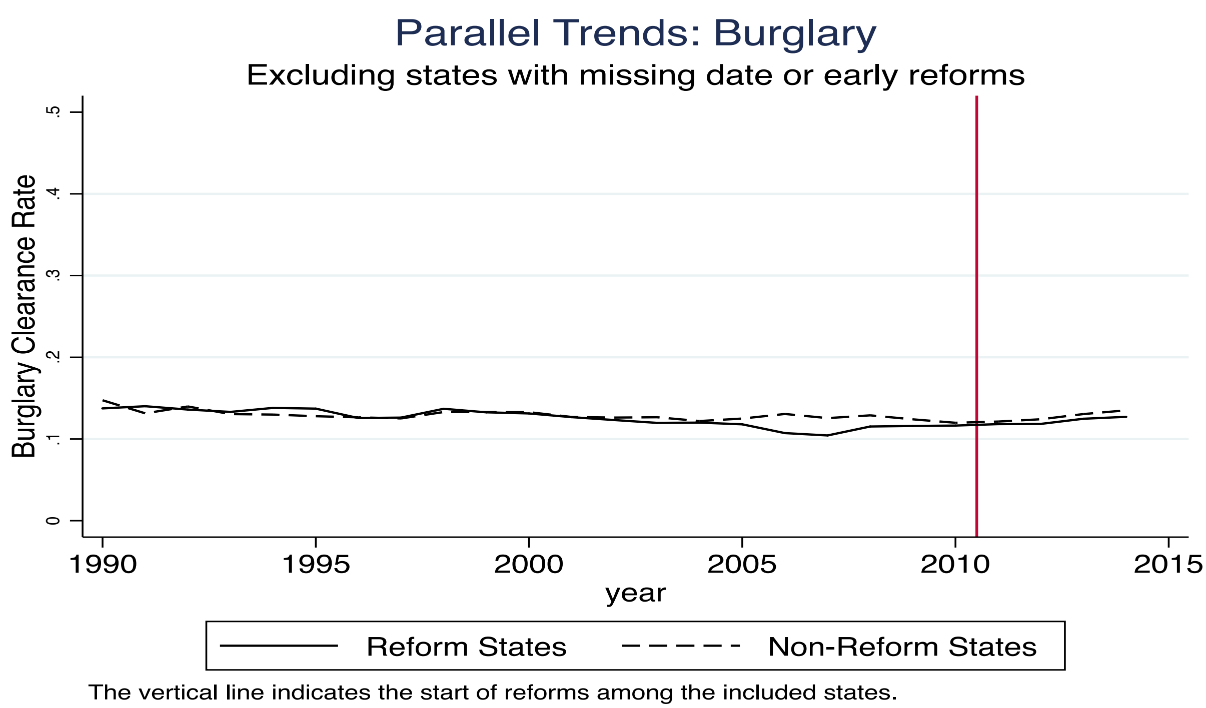
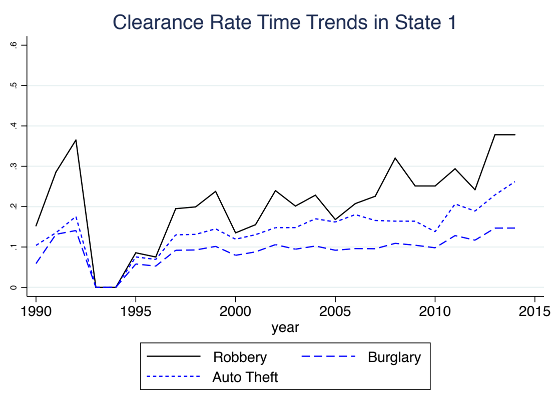
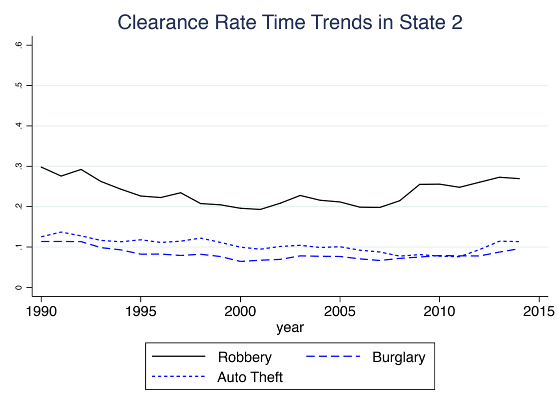


Figure C.3: *Parallel Trends for DD model – Burglary. Reform states that adopted reforms prior to 2011 excluded. Dependent variable: Burglary Clearance Rate.*

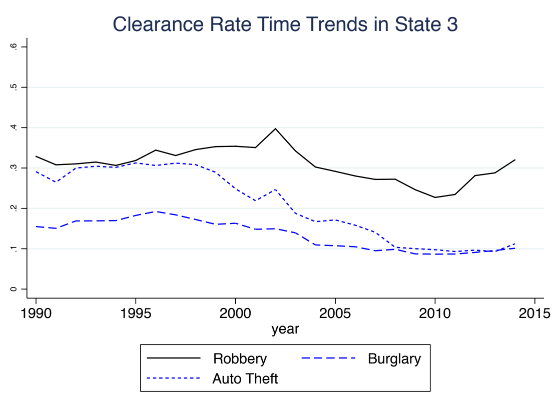
(a) Alabama



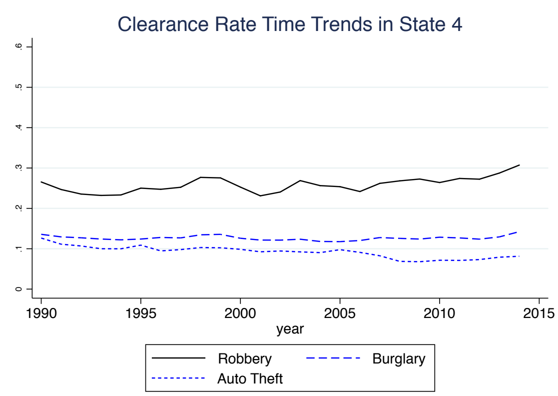
(b) Arizona



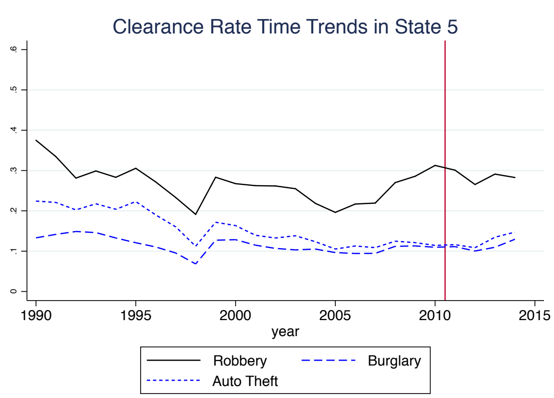
(c) Arkansas



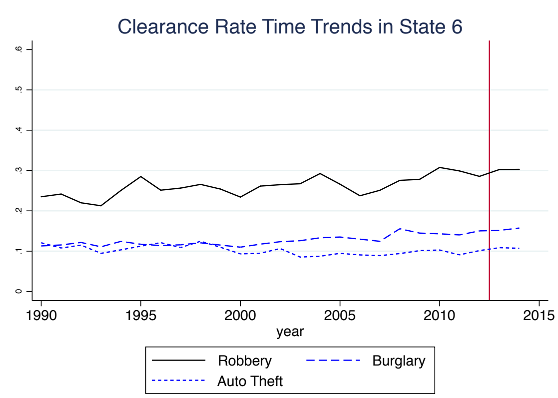
(d) California



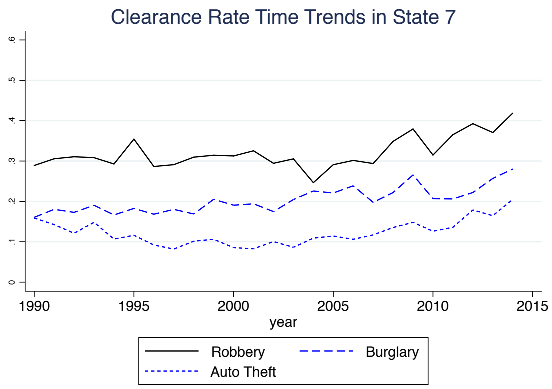
(e) Colorado



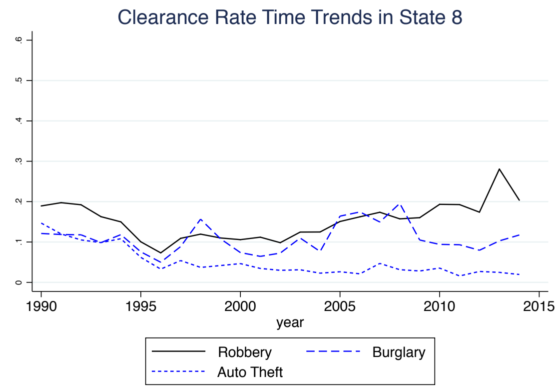
(f) Connecticut



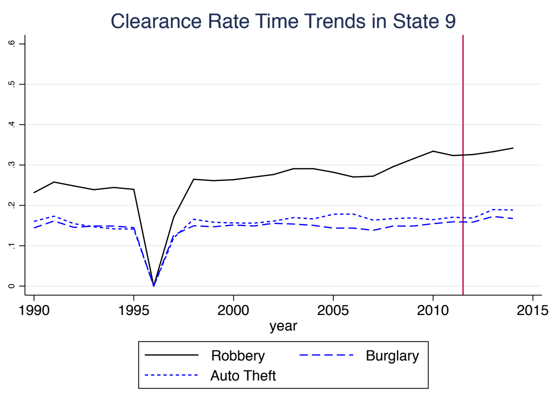
(a) Delaware



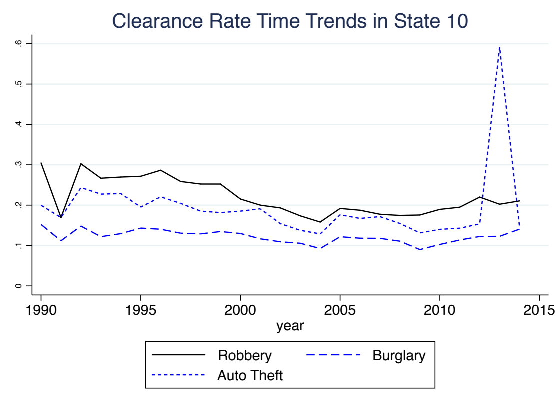
(b) District of Columbia



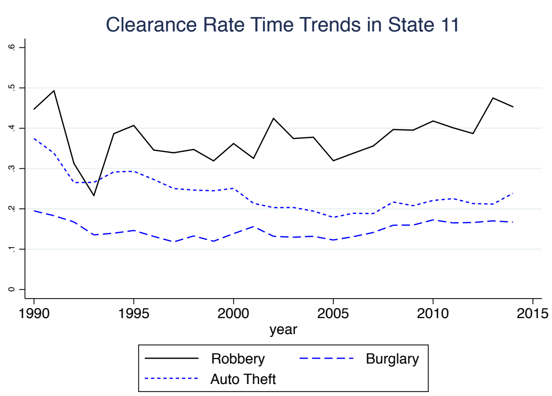
(c) Florida



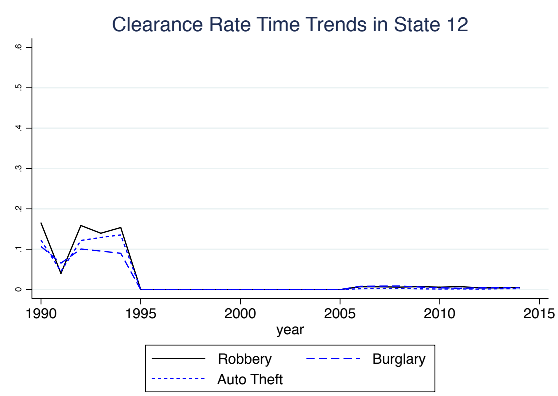
(d) Georgia



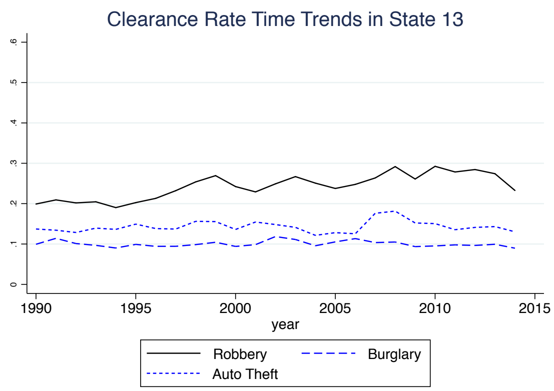
(e) Idaho



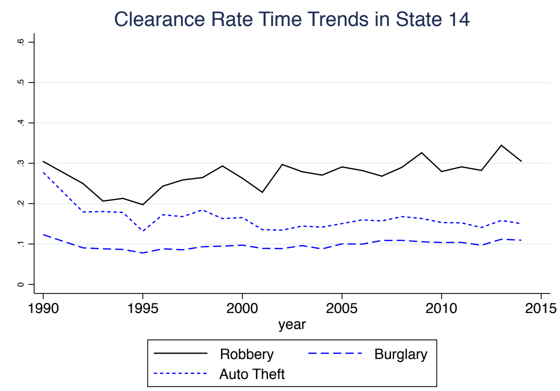
(f) Illinois



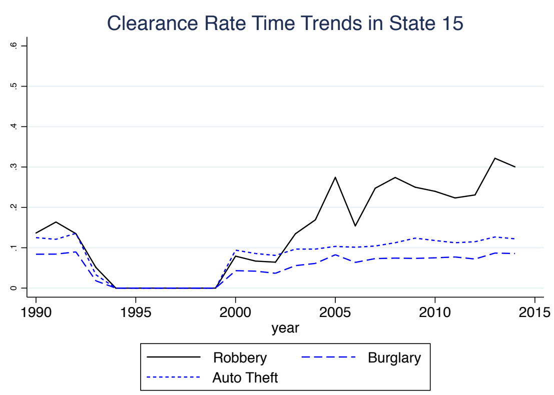
(a) Indiana



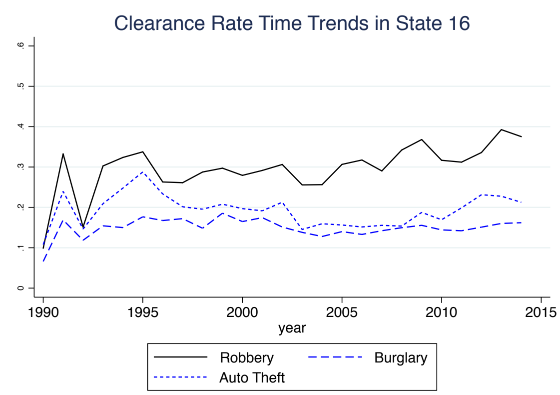
(b) Iowa



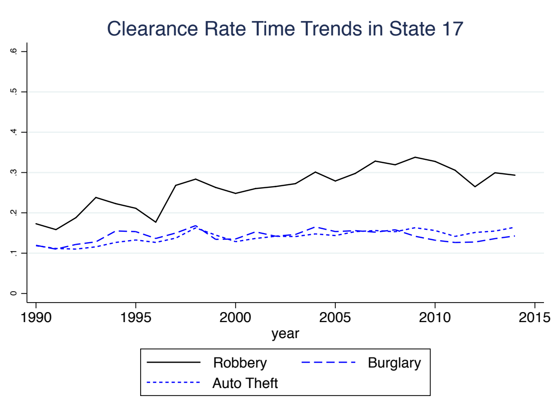
(c) Kansas



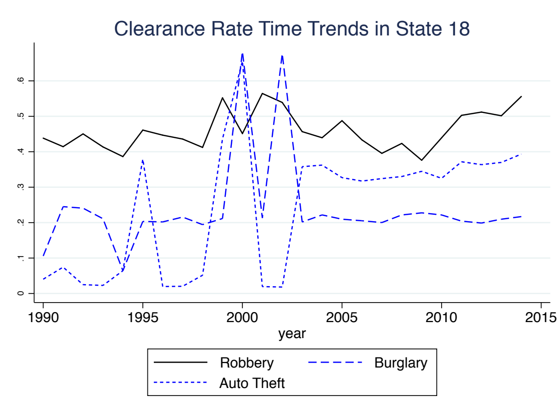
(d) Kentucky



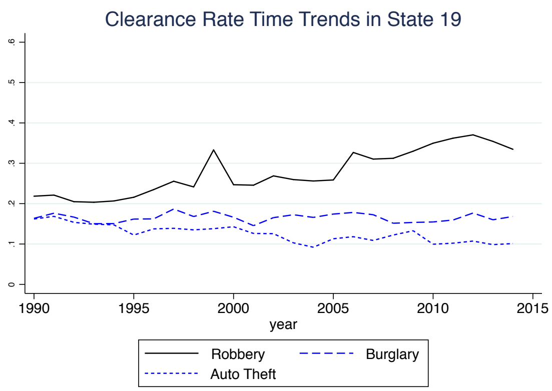
(e) Louisiana



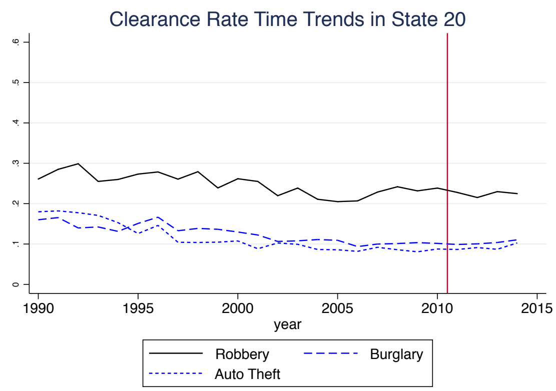
(f) Maine



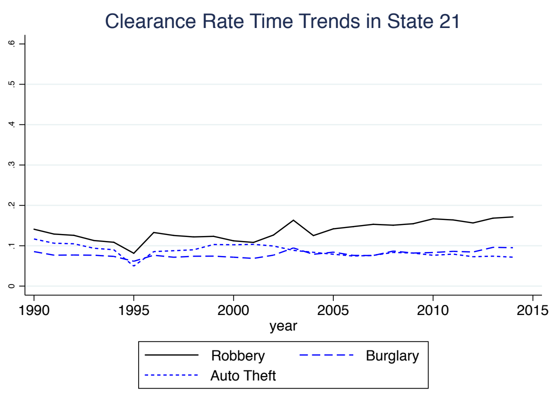
(a) Maryland



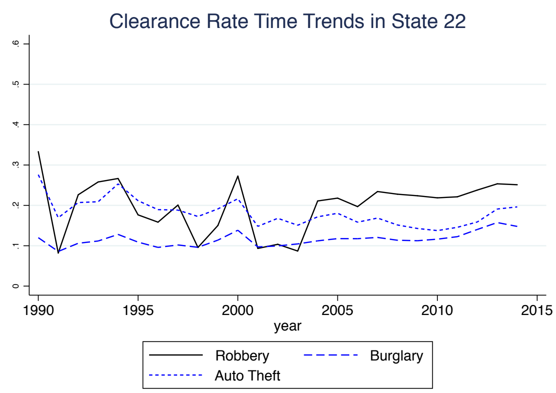
(b) Massachusetts



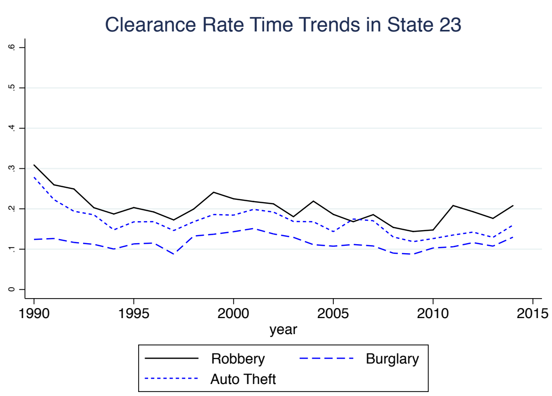
(c) Michigan



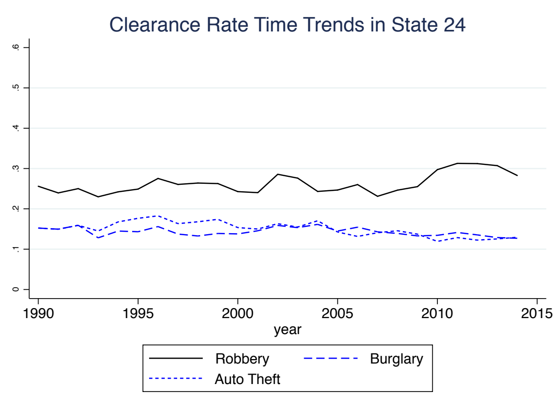
(d) Minnesota



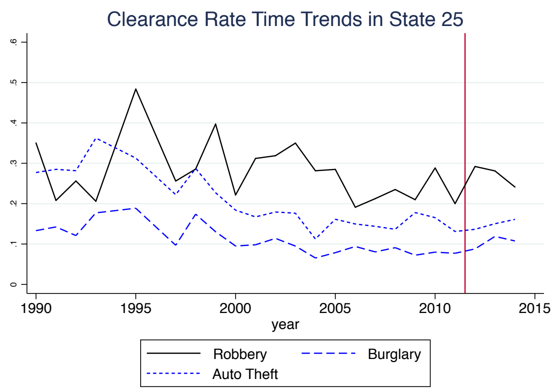
(e) Mississippi



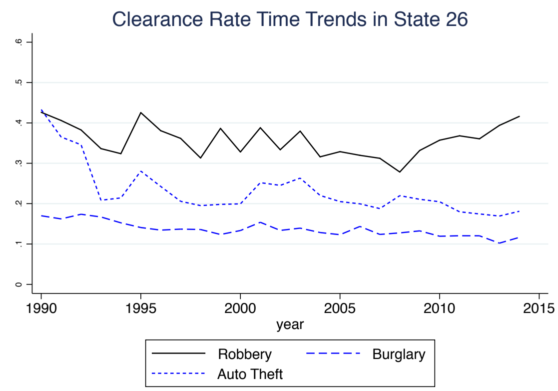
(f) Missouri



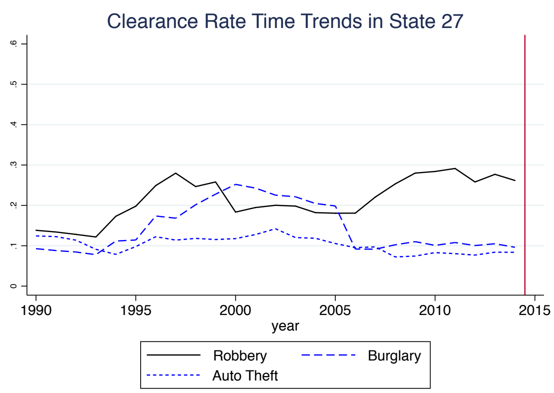
(a) Montana



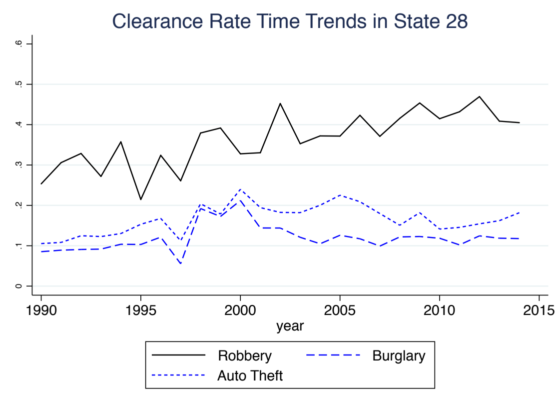
(b) Nebraska



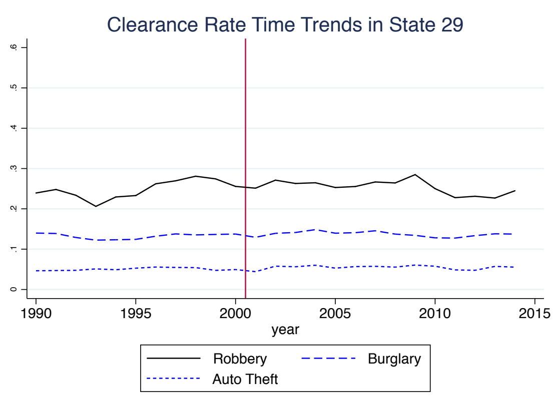
(c) Nevada



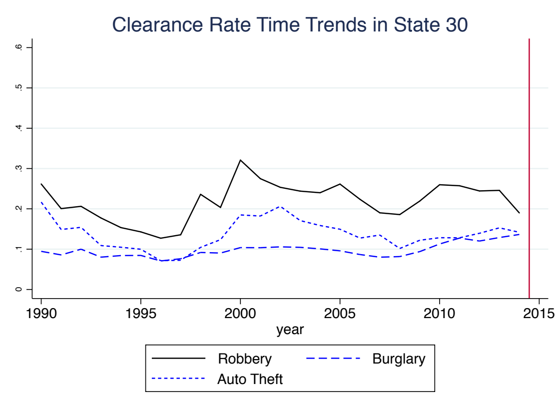
(d) New Hampshire



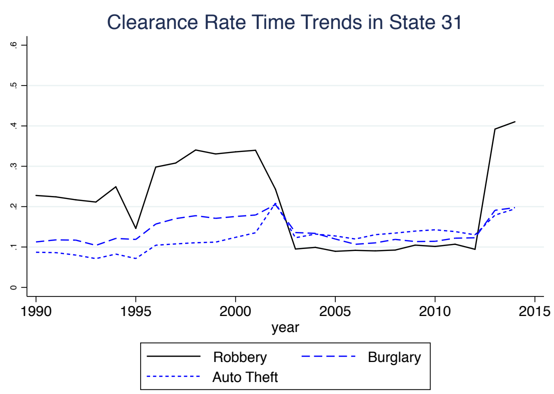
(e) New Jersey



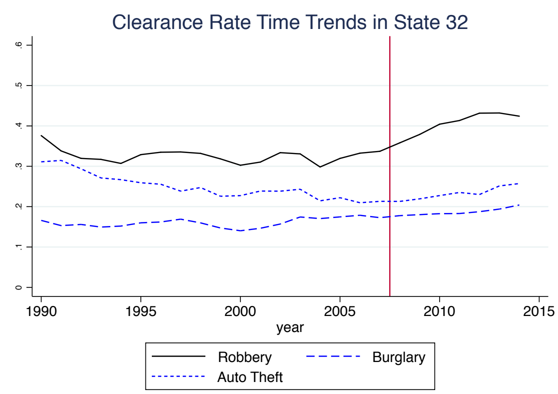
(f) New Mexico



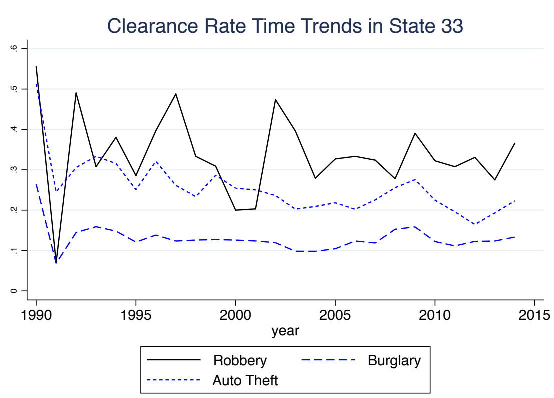
(a) New York



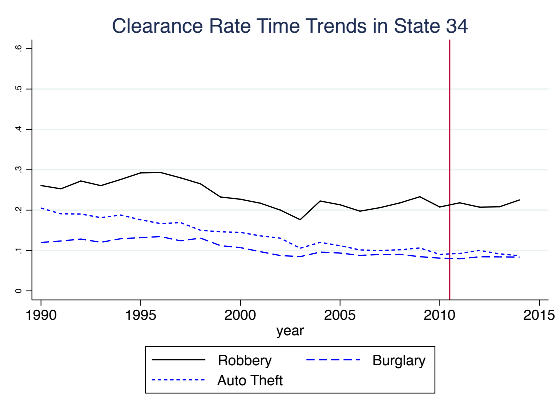
(b) North Carolina



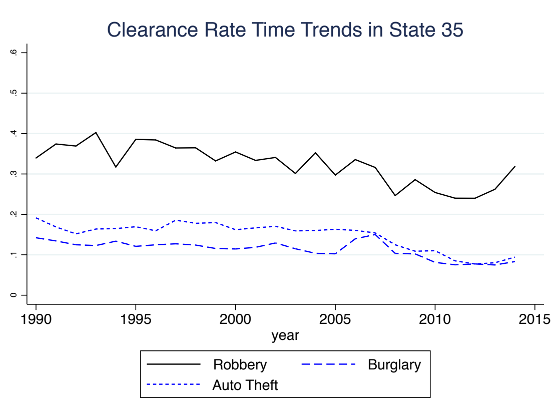
(c) North Dakota



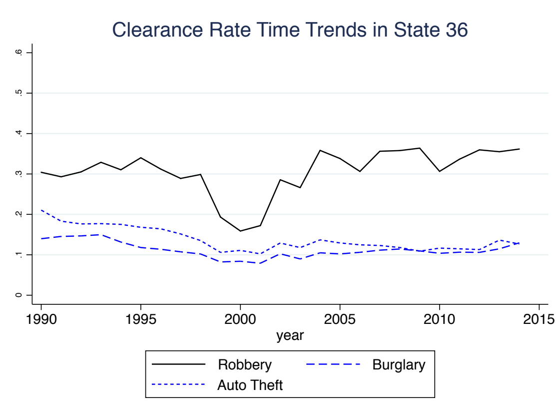
(d) Ohio



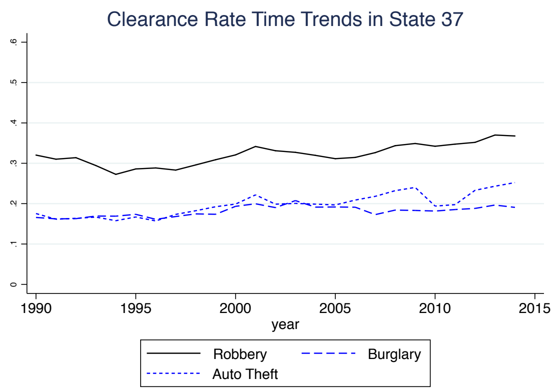
(e) Oklahoma



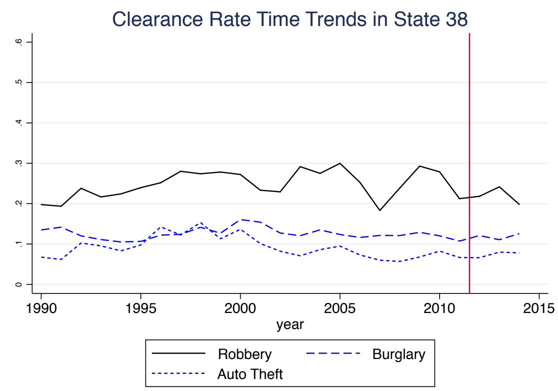
(f) Oregon



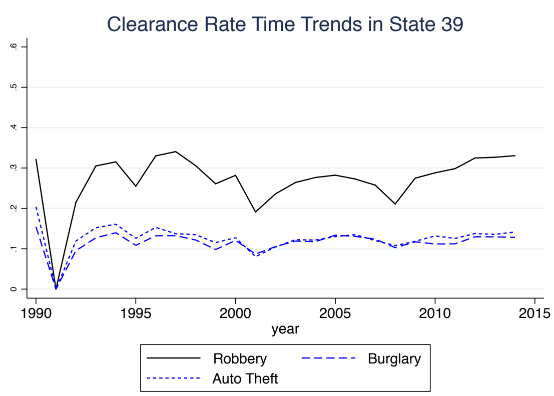
(a) Pennsylvania



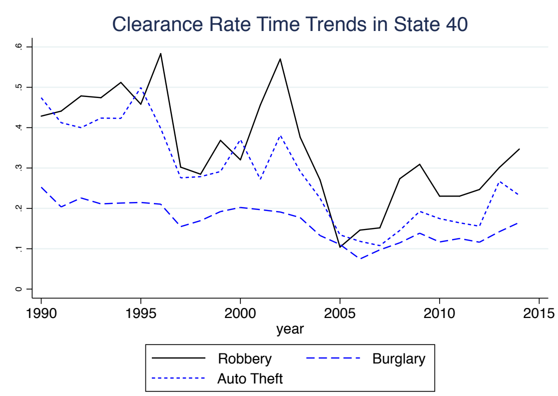
(b) Rhode Island



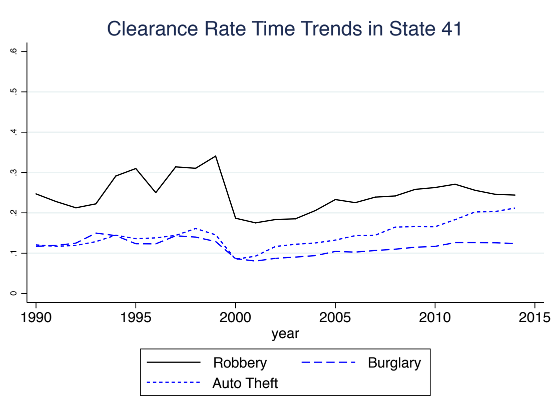
(c) South Carolina



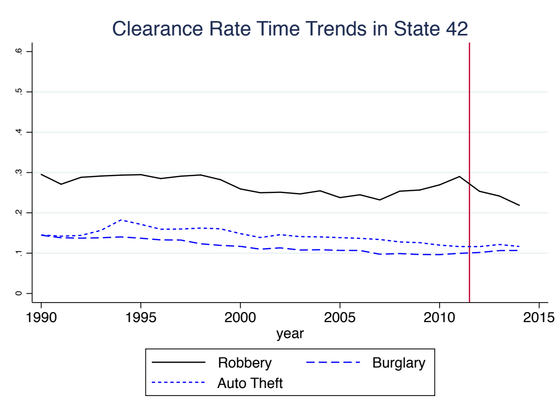
(d) South Dakota



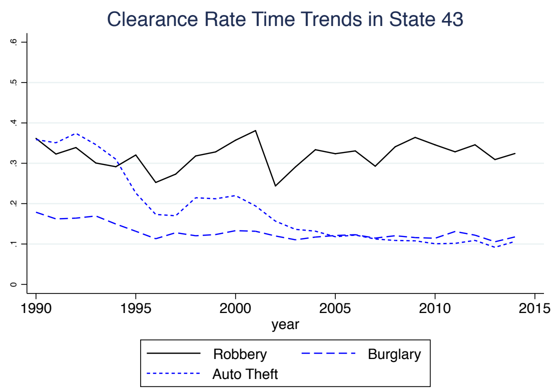
(e) Tennessee



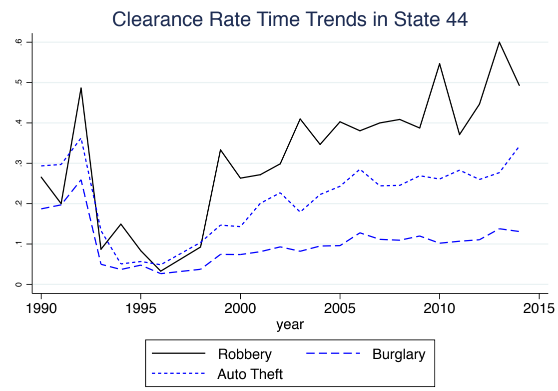
(f) Texas



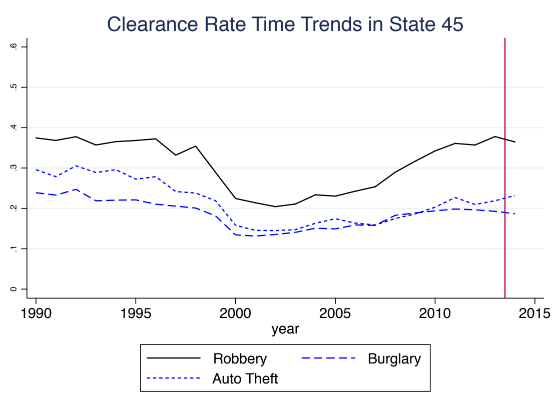
(a) Utah



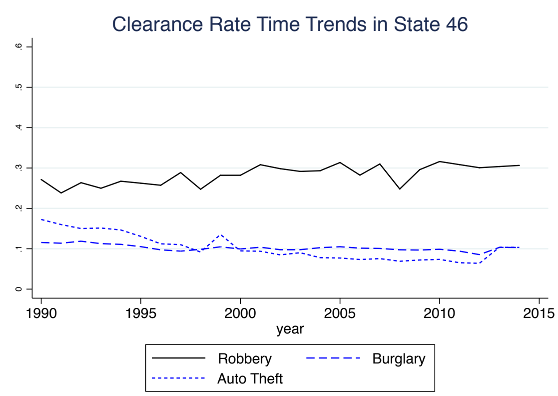
(b) Vermont



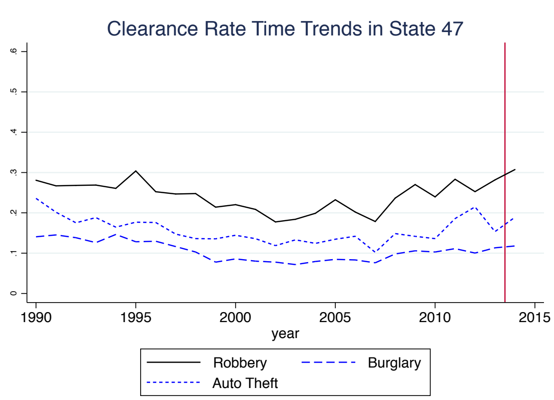
(c) Virginia



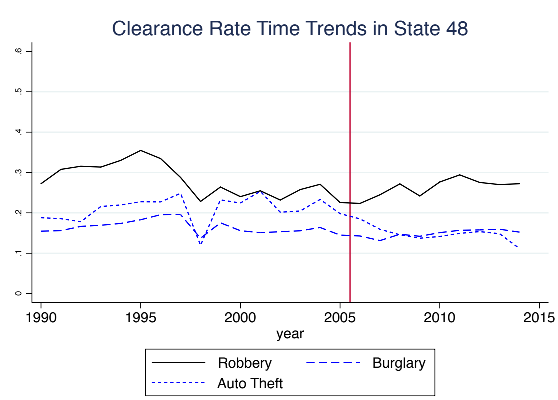
(d) Washington



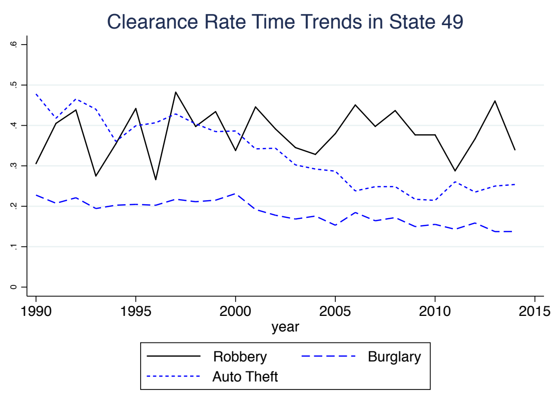
(e) West Virginia



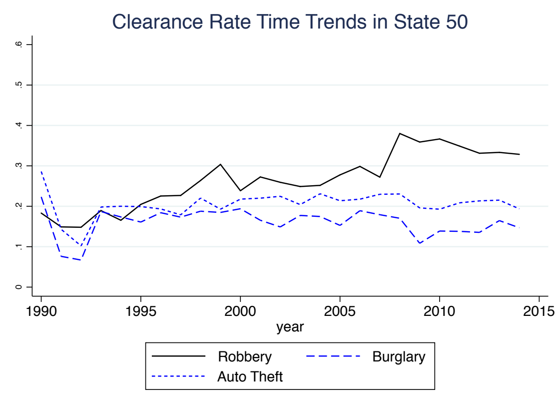
(f) Wisconsin



(a) Wyoming



(b) Alaska



(c) Hawaii