Who Gets a Second Chance? Effectiveness and Equity in Supervision of Criminal Offenders

Evan K. Rose*

November 24, 2019

[click here for most recent version]

Abstract

Most convicted criminals are sentenced to probation and allowed to return home. On probation, however, a technical rule violation such as not paying fees can result in incarceration. Rule violations account for more than 30% of all prison spells in many states and are significantly more common among black offenders. I test whether technical rules are effective tools for identifying likely reoffenders and deterring crime and examine their disparate racial impacts using administrative data from North Carolina. Analysis of a 2011 reform eliminating prison punishments for technical violations reveals that 40% of rule breakers would go on to commit crimes if their violations were ignored. The same reform also closed a 33% black-white gap in incarceration rates without substantially increasing the black-white reoffending gap. These effects combined imply that technical rules target riskier probationers overall, but disproportionately affect low-risk black offenders. To justify black probationers' higher violation rate on efficiency grounds, their crimes must be roughly twice as socially costly as that of white probationers. Exploiting the repeat-spell nature of the North Carolina data, I estimate a semi-parametric competing risks model that allows me to distinguish the effects of particular types of technical rules from unobserved probationer heterogeneity. The estimates reveal that the deterrent effects of harsh punishments for rule breaking are negligible. Rules related to the payment of fees and fines, which are common in many states, are ineffective in tagging likely reoffenders and drive differential impacts by race. These findings illustrate the potentially large influence of facially race-neutral policies on racial disparities in criminal justice outcomes.

*Ph.D. Candidate, U.C. Berkeley Department of Economics; ekrose@berkeley.edu. I thank Pat Kline, David Card, Danny Yagan, and Chris Walters for their help and encouragement. This paper has benefited tremendously from comments and suggestions from Alessandra Fenizia, Fred Finan, Ingrid Haegele, Jonathan Holmes, Hilary Hoynes, Peter Jones, Nick Li, Juliana Londoño-Vélez, Maxim Massenkoff, Justin McCrary, Conrad Miller, Steven Raphael, Emmanuel Saez, Jonathan Schellenberg, Yotam Shem-Tov, Francis Wong, and seminar participants at the University of California at Berkeley. I am grateful to the North Carolina Department of Public Safety and Administrative Office of the Courts for their help in securing and understanding the data, as well as to Ginny Hevener, Linda Mitterling, George Pettigrew, Alan Pistick, Cara Stevens, and the officers of the 14th Probation District for their feedback and patience.

1 Introduction

For many black men, encounters with police, courts, and prisons are as common as employment. Black high school dropouts, for example, are almost as likely to be incarcerated as to be holding a job. Recent research has studied racial disparities in decisions by police, bail judges, prosecutors, and juries (Fryer, 2019; Arnold, Dobbie and Yang, 2018; Rehavi and Starr, 2014; Anwar, Bayer and Hjalmarsson, 2012) and how arrest, conviction, and incarceration affect economic outcomes (Agan and Starr, 2018; Dobbie, Goldin and Yang, 2018; Harding et al., 2018; Chetty et al., 2018; Bayer and Charles, 2018; Mueller-Smith and Schnepel, 2019; Bhuller et al., 2019). However, less attention has been paid to the impact of probation, the most common way criminals are punished in the United States. Every year, more than 3.7 million probationers are sent home after conviction on the condition that they obey strict technical rules. Breaking these rules, which forbid alcohol and drugs, entail frequent meetings with a probation officer, and require timely payment of fees and fines levied by the court, can result in incarceration. Such "technical violations" account for more than 30% of prison spells in many states (CSG, 2019) and are significantly more common among black men. This "second chance" sentence is therefore a key driver of incarceration overall and of racial disparities in prison exposure.[1]

Incarcerating technical rule breakers is also expensive. Supervising a probationer at home costs just 5% as much as prison. It nevertheless may be effective to shift technical rule breakers from probation to prison if violations are a strong indicator of future criminal behavior and thereby help the state incapacitate risky offenders, or if the threat of imprisonment for rule breaking deters reoffending. The *effectiveness* of enforcing probation's technical rules thus depends on how well violations serve as predictors of future crime and on the behavioral responses to potential punishments. The *equity* implications depend on racial differences in the association between rule breaking and criminality (Kleinberg et al., 2017; Kleinberg, Mullainathan and Raghavan, 2017) and on differences in any behavioral responses to punishments.

This paper examines the effectiveness and equity of the probation system. I test whether technical rules target probationers who would otherwise commit crimes, measure their deterrence effects, and examine racial differences in targeting and deterrence. To do so, I focus on a major 2011 reform that eliminated the use of incarceration as a punishment for breaking rules related to cash fees and fines, drug and alcohol use, and other violations. As a result of this change, many probationers who would have been imprisoned for rule breaking prior to the reform were instead permitted to remain in their communities. Measuring the resultant increases in crime thus allows me to assess how effectively rule breaking "tagged" would-be reoffenders and measure any behavioral response. Analyzing the reform separately by race allows me to assess equity by examining differences in tagging and deterrence effects between black and white probationers.

To begin, I develop an empirical model focused on measuring how rule breaking predicts crim-

[1] These concerns became headline news in 2017 when the rapper Meek Mill was incarcerated for breaking the terms of a decade-old probation sentence over technical violations that included riding a dirt bike without a helmet and traveling to perform. Jay-Z, writing in the New York Times, argued "What's happening to Meek Mill is just one example of how our criminal justice system entraps and harasses hundreds of thousands of black people every day....Instead of a second chance, probation ends up being a land mine, with a random misstep bringing consequences greater than the crime" (Nov. 17, 2017).

inality. The simplest one-period version of this model (e.g., covering the first few months of a probation spell) describes the relationship between two binary events: whether the probationer is incarcerated for rule violations and whether the probationer commits a crime. Rule breakers' criminal activity is not observed because their outcomes are censored by incarceration. However, by comparing observed rule-breaking incarceration and crime in two regimes—one where rule breaking is punished by incarceration and one where it is not—I can recover the full joint distribution of the two events. This allows me to characterize the accuracy, type-I (i.e., false positive), and type-II (i.e., false negative) error rates associated with using technical rules as "tags" of criminality. I define "bias" in the rules as a difference in error rates or accuracy.

I then extend this model to a more realistic multi-period setting where rule breaking and crime are competing risks. That is, there is a latent risk of crime across a series of periods (e.g., months of a probation spell) and individuals can be incarcerated for a rule violation in any period. Again, by comparing regimes with and without rule-breaking incarceration, it is possible to estimate accuracy and error rates in this dynamic setting. Importantly, both models require that changes in rule-breaking punishments do not affect criminal activity, an assumption that has strong empirical support. Nevertheless, I show later that results change little when using a semi-parametric competing hazards model to relax this exclusion restriction (Cox, 1962; Tsiatis, 1975; Heckman and Honoré, 1989; Honoré, 1993; Abbring and Van Den Berg, 2003).

I begin my empirical analysis by estimating the simplified one-period model. I treat the first year of probation as a single period for measuring outcomes and focus on successive cohorts of probationers who started their probation spell within four years before or after the reform. To control for any time trends in crime, I compare probationers' outcomes to those of individuals convicted of similar offenses and placed on unsupervised probation, an alternative punishment regime. Unsupervised probationers provide a useful control group because they are not subject to most technical rules and thus were unaffected by the reform.

Difference-in-differences estimates reveal that prison punishments for technical rule violations in the first year of a spell declined by 5.5 percentage points (p.p.) as a result of the reform, a 33% drop relative to the pre-reform mean of 15.4%. Arrests increased by 2.0 p.p. These two effects combined imply that roughly 37% of individuals who escaped incarceration for rule breaking due to the reform were arrested instead. This figure is an estimate of the accuracy, i.e., the probability of offending conditional on breaking a rule, of the drug and administrative rules affected by the reform. While accuracy is roughly 10 p.p. higher than mean arrest rates, both type-I and type-II errors are large, at 6% and 94%, respectively. These high error rates imply that rules catch a meaningful fraction of non-reoffenders and few potential reoffenders.

Remarkably, the reform's impact on black offenders' incarceration for rule breaking was nearly twice as large as its impact on white offenders'. As a result, black-white gaps in prison punishments for technical rule violations were practically eliminated, and thousands more black probationers were allowed to remain in their community. Black probationers, however, saw only slightly larger increases in arrests after the reform than white probationers. Roughly 50% of white probationers spared incarceration were arrested, while among black probationers the arrest rate was only 30%. The implied accuracy of technical rules is therefore 66% higher in the white population. Moreover, while type-II error rates are similar in both groups, type-I error rates are three times higher in the

black population. In other words, substantially more black offenders who would not have offended in the first year of their spell were incarcerated due to technical rules.

My results suggest these race gaps reflect the disparate *impact* of facially race-neutral rules rather than disparate *treatment* by those who enforce them. For example, there is no disparity in how black and white probationers are punished conditional on breaking the same technical rule. Moreover, technical rules for which officers have wide enforcement discretion and those for which violations are detected automatically both exhibit large race gaps. There is also no evidence of caseworker-probationer race match effects. This setting thus highlights the potential importance of how rules and polices are designed rather then how they are applied by practitioners for generating racial disparities.[2] By contrast, the economics literature on taste-based and statistical discrimination (Becker, 1957; Phelps, 1972; Arrow, 1973) has largely focused on disparate treatment based on race throughout the criminal justice process (Abrams, Bertrand and Mullainathan, 2012; Rehavi and Starr, 2014; Fryer, 2019; Arnold, Dobbie and Yang, 2018).

I estimate the multi-period version of the model next. This model better reflects the reality that arrests and rule violations can occur throughout a probation spell and well beyond one year. The estimates show that black offenders are targeted more aggressively by technical rules regardless of their risk. There are especially large race gaps for the lowest risk probationers, who would only be re-arrested three years after starting probation or later, and possibly never. A decomposition of aggregate race gaps into differences in risk distributions and differences in targeting shows that targeting explains ∼90% of black probationers' higher violation rate. The conclusions of simplified model are thus also supported in this richer setting.

I use these results to conduct a partial cost-benefit analysis that compares the costs of incarcerating a technical rule breaker to the social costs of crime they would commit and any attendant punishments if allowed to remain free. The results show that for every $100 the state spends incarcerating technical rule breakers, it saves $30 in prison costs that it would have paid anyway. To justify the state's use of incarceration for technical rule breaking, the social costs of crime averted by incarcerating a rule breaker must fill the gap, implying a break-even valuation of roughly $40,000 per arrest. Because black probationers are targeted more aggressively, break-even valuations for black offenders are roughly twice as large as for white offenders. Using estimates from the existing literature, I find that the social cost of averted offenses falls near or below this benchmark. Importantly, however, these calculations also assign no value to the impact of the reform on racial disparities.

I conclude by re-examining the effectiveness and equity of technical rules using a semi-parametric model of competing hazards. The model is useful first because it allows me to explicitly test for behavioral responses to weaker punishments for breaking technical rules. These responses are not directly identified by the reform alone because increases in crime after it took effect could reflect weaker deterrence effects, mechanical reductions in rule-breaker incarceration, or a mixture of both. The model, however, fully characterizes the joint distribution of criminal arrest and rule-driven incarceration both before and after the reform. The multiple-spell nature of my data allows me to flexibly incorporate unobserved heterogeneity that is potentially correlated with observed covariates (Heckman and Honoré, 1989; Honoré, 1993; Abbring and Van Den Berg, 2003) and capture rich

[2] Termed "Type B" discretion by Bushway and Forst (2013).

differences in the relationship between rule breaking and criminality across groups. I can then test whether the increases in crime after the reform took effect reflect simple reductions in rule-breaker incarceration and no behavioral response, as assumed in the previous analysis.

The estimates show limited evidence of behavioral responses to the reform. The implied elasticity of one-year latent offending risk with respect to one-year rule-breaking incarceration rates is 0.05. Due to this response, observed arrest rates post-reform are roughly 0.5 p.p. higher than under no response. Probationers, police, and judges therefore do not appear responsive to weaker rule regimes. Moreover, estimated behavioral responses are similar across race groups, suggesting disparities in technical violations are not justified by larger deterrent effects among black offenders. The estimates can closely replicate difference-in-differences estimates of the reform's impacts, corroborating the model's functional form restrictions and the small estimated behavioral response.

The model also allows for a quantification of the overall equity and effectiveness of technical rules by simulating what would happen if policy makers *further* reduced incarceration punishments for technical rules after the 2011 reform. These estimates show that technical violations overall—not just those impacted by the reform—target black men more aggressively and that technical violation and arrest risk are more tightly correlated among white men. The rules left in place after the 2011 reform, however, are slightly better predictors of criminal risk and are less biased. Further reductions in technical rules would therefore generate slightly larger increases in crime than the reform, but come with smaller benefits in terms of reduced racial disparities.

A simple extension that examines the impacts of specific rule types (e.g., drug tests vs. cash fees and fines) helps explain these results. These estimates show that all rules tend to target black offenders more aggressively. However, rules related to cash fees and fines are particularly problematic. Not enforcing them would increase the share of future reoffenders who break technical rules and decrease the share of future non-reoffenders incarcerated for doing so. Hence, eliminating this type of rule provides a double social benefit by improving the effectiveness of the probation regime overall and reducing existing disparities. Since the 2011 reform directly addressed financial rules, it had large impacts on disparities within more limited impacts on crime. Other rule types, such as drug abuse and reporting rules, tend to perform better.

Taken together, my results show how facially race-neutral policies—in this case common sense rules designed to promote public safety—can generate large racial disparities not justified by the policies' ultimate goals. In some contexts, opting to give local decision makers more discretion instead of relying on uniform rules may increase policies' effectiveness and fairness by taking advantage of agents' superior information and encouraging effort (Aghion and Tirole, 1997; Kuziemko, 2013; Duflo et al., 2018). North Carolina's reform shows that holding discretion fixed, however, there is the potential to redesign rules themselves to improve outcomes. Poorly designed rules and policies are a potentially powerful explanation for many observed racial disparities in criminal justice, where the use of detailed guidelines to constrain decisions has become increasingly popular.

The remainder of this paper is structured as follows. I first describe the probation system both nationally and in North Carolina, explain the sources and content of my data, and estimate observational racial disparities in Section 2. Section 3 lays out the empirical model. Section 4 presents the main results that analyze the 2011 reform. Section 5 estimates a competing risk model for probation violations and crime. Section 6 concludes.

2 Setting and data

2.1 The probation system

Over the past several decades, the US probation system has grown in tandem with incarceration rates. The national probation population now stands at 3.67 million, a 230% increase over levels in 1980. Since probation spells can be quite short, this population turns over quickly—1.6 million individuals entered probation in 2016, and 1.9 million individuals exited (Kaeble, 2018). Many millions more US residents living today have thus likely served a probation sentence at some point in the past.[3]

The size of the probation system reflects the popularity of probation as a criminal sentence. In the 75 largest counties in the US, 51% of felony defendants receive probation as part of their sentences, with higher rates for non-violent property and drug offenders (Reaves, 2013). Misdemeanor defendants, who account for the bulk of cases processed in state courts, receive probation at even higher rates. While probation is common overall, it is used most often for young and first-time offenders facing their first serious criminal case. In North Carolina, for example, 78% of first-time felons are placed on probation, along with 70% of 16-25 year-old offenders.[4]

Probation spells typically last between one and three years (Reaves, 2013). Over this period, offenders must comply with a set of conditions imposed by the court as "reasonably necessary to ensure that the defendant will lead a law-abiding life or to assist him to do so" (NC General Statues §15A-1343). In North Carolina, these conditions include a set of standard "regular" rules: pay fees and fines ordered by the court, including a monthly fee for supervision itself and repayment for any indigent defense provided, remain within the jurisdiction of the court unless given permission to travel, report regularly to a probation officer, submit to drug and alcohol tests and warrantless searches, and remain gainfully employed. Occasionally, judges impose special conditions such as substance abuse treatment programs and electronic monitoring.[5] Of course, all probationers are also required to commit no new criminal offenses during their spell. As is clear from North Carolina's statute, public safety is a primary motivation for enforcing technical rules in probation. Interviews conducted with probation officials, probationers, judges, and attorneys across the country by the University of Minnesota's Robina Institute show that many other jurisdictions have a similar focus Robina Insitute (2016).

North Carolina, like many other states, operates two forms of probation: supervised and unsupervised. Supervised probationers are assigned a probation officer who is personally responsible for monitoring them. These officers oversee 60-80 offenders at a time, conducting regular interviews, drug tests, searches, and arrests. Most officers have four-year degrees in a criminal justice related field. Roughly 50% of officers are female and 40% are black. Unsupervised probationers are not

[3]Roughly 870,000 individuals are currently serving parole sentences in the US. Parole is qualitatively similar to probation, but typically follows an incarceration spell. Probationers, on the other hand, most often go directly back into the community upon conviction with no intervening prison spell. For much of the last 25 years, North Carolina has operated a very limited parole system, opting to release most incarcerated individuals with no supervision. I thus focus exclusively on the probation system in this analysis.

[4]Individuals granted deferred prosecution are also typically placed on probation. Unlike regular probationers, however, after successfully completing their spell their records may be cleared.

[5]The full set of regular and special probation conditions are listed in North Carolina's general statutes, available at: https://www.ncleg.net/EnactedLegislation/Statutes/PDF/ByArticle/Chapter_15A/Article_82.pdf.

assigned a probation officer. They are technically subject to the same rules as their supervised peers, except those related to supervision, such as reporting regularly to an officer. While in most cases judges have discretion to assign either supervised or unsupervised probation, unsupervised probation tends to be reserved for misdemeanants and individuals convicted of driving while intoxicated or with a revoked license. Due to the lack of monitoring, unsupervised probationers are rarely subject to technical rule violations and thus were largely unaffected by North Carolina's 2011 reform, making them a useful control group.

When a offender breaks a technical rule, they must report to a local judge for a violation hearing. Judges can respond by "revoking" probation and sending the individual to jail or prison for the duration of their original, suspended sentence. I call this type of punishment technical incarceration or revocation. Judges can also modify specific conditions, extend the supervision term, and issue verbal reprimands and warnings. In practice, judges closely follow probation officers' recommendations, agreeing to revoke in 85% of hearings where the officer favors doing so. Revocation is also very common. Over the 2000s, for example, probationers remanded to prison without a new criminal conviction accounted for ∼40% of new state prison spells.

2.2 2011 reform

In 2011, North Carolina made major changes to the state's criminal justice system by passing the Justice Reinvestment Act (JRA).[6] Among the most consequential changes was the introduction of strong limits on courts' authority to revoke probation. For all probation violations occurring on or after December 1, 2011, supervision could be revoked only for new criminal offenses, for fleeing supervision, or if the defendant had two or more violations in the past. Previously, judges could revoke for any technical violation, including non payment of fees and fines, not reporting, or failing drug and alcohol tests. As I will show below, this change dramatically reduced prison punishments for technical violations and provides an important source of variation I use throughout this study.

JRA also made several other changes to the probation and parole system. Probation officers received expanded authority to impose conditions such as additional community service in response to failures to comply with certain conditions. JRA also introduced a new violation response—Confinement in Response to Violation (CRV)—that imposes confinement for up to 90 days, although this appears to be used relatively infrequently, especially in the years just after the reform took effect. Finally, JRA also made several changes to other parts of the court system, including increasing the scope of post-release supervision (i.e., parole), adjusting some sentencing enhancements, and re-defining some conditions of supervision. Since my focus is on the probation system, most of these changes are beyond the scope of this study.[7]

[6] The law reflected several years of work by the Council of State Governments' Justice Center (CSG). After studying North Carolina's corrections system, the CSG concluded that technical incarceration of probationers was responsible for hundreds of millions in annual costs (CSG, 2011). Law makers passed the JRA in an effort to reduce these costs and lower projected correctional spending in the future.

[7] A useful feature of the JRA reforms is that changes to revocation authority applied to all *probation violations* after December 1, 2011. Other changes largely applied to sentences for *offenses committed* after December 1. This allows me to study the effects of the change to revocations while holding other factors constant by looking in a relatively narrow window around December 1, which I do in robustness checks. Appendix Table A11, for example, shows that similar results hold when examining effects on the reform within just one year after it took effect, when the vast majority of offenders were not subject to additional changes.

2.3 Data sources

This project primarily analyzes administrative datasets provided by the North Carolina Department of Public Safety (DPS). The core data consist of records for the universe of individuals serving supervised probation sentences that started between 2006 and 2018 (inclusive). These data detail individual demographics, the duration of the probation spell, the original convictions that resulted in the probation spell, and the probation officers assigned to the individual over the course of the spell. The data also record all violations (coded in dozens of unique categories), the probation officer's recommended response, and the ultimate disposition.

In addition to these records, I utilize data on all criminal court cases disposed from 2006 to the present provided by the North Carolina Administrative Office of the Courts (AOC). Because police officers are the charging agency in North Carolina, these records capture close to the universe of arrests.[8] I use the AOC data to measure new criminal offenses, the type and length of any incarceration sentences meted out as a result, and criminal histories. I also use the AOC data to identify my control group, individuals placed on unsupervised probation. I combine this data with additional records from the DPS that detail all sentences to supervised probation and incarceration from the 1970s to the present as an additional source of criminal history information.

Lastly, in some descriptive regressions I use scores on standardized, state-wide tests administered in math and reading at the end of grades three through eight. These data are housed at the North Carolina Education Research Data Center (NCERDC) and were linked to North Carolina criminal records for related work in Rose, Schellenberg and Shem-Tov (2019). Test scores are only available for about a third of the sample, since not all offenders were educated in the state at times covered by the NCERDC data.

All data are linked using a combination of personal and administrative identifiers. This includes full name and date of birth in all cases, but also partial social security numbers, driver's license numbers, and unique codes assigned to individuals by the State Bureau of Investigation, Federal Bureau of Investigation, and the DPS.

Throughout the analysis, I define technical incarceration as having probation revoked without an intervening arrest in AOC data. Although most probation violations for new criminal behavior are accompanied by a new criminal charge in court records, occasionally they are not. This definition thus avoids relying on violation codes themselves to define technical incarceration, which is attractive because violation coding may vary across groups or be affected by the reform. Alternative definitions of technical incarceration, such as revocation for violations consisting exclusively of non-criminal behaviors, yield similar results.

2.4 Descriptive statistics

Descriptive statistics for the treated and control samples are provided in Table 1. Both groups are young, with 50% of the sample 30 or under at the start of their spell, predominately male, and overrepresent minorities relative to North Carolina's population. Supervised probation spells last

[8] In Charlottle-Mecklenberg, where I have collected jail booking records directly from the Sheriff, 93.3% of arrests appear in the AOC data. The remaining 6.7% of Charlotte records reflect events unlikely to be captured in AOC data, such as federal prison transfers.

about 20 months on average and are the result of a relatively even mix of felony, misdemeanor, and driving while intoxicated or driving with a revoked license offenses. The treated sample has very limited criminal histories, with the median defendant having just one prior misdemeanor conviction and no prior sentences to supervised probation or incarceration. As expected, unsupervised probationers were convicted of less severe offenses and have more limited criminal histories. Despite these differences, I show below that control units' outcomes closely track those of treated units for many years leading up to the reform, supporting their use as a counterfactual.

Almost all probationers break at least one rule during their spell. As shown in Table 2, over 90% of probation spells include at least one violation, with citations for non-payment of fees and fines occurring in 80%. The next most common violation is for not reporting to a probation officer—for example by missing a weekly check-in at the local probation office. This violation occurs in 46% of spells. Drug violations and treatment program failures are also common, occurring in 30% and 25% of spells, respectively. New misdemeanor arrests are the fourth most common violation; new felony arrests are the 11th. Strikingly, probationers are twice as likely to be cited for moving or changing jobs without notifying their probation officer as for committing a new felony crime.

Rather than work with the full list of detailed violation types, I categorize them into four groups: Drug related, administrative, absconding, and new crime. The top violations in each category are reported in Appendix Table A1. Drug related violations are predominately for failing a drug test, dropping out of a substance abuse program, or admitting to drug use. Administrative violations are predominately for non-payment of fees, not reporting, moving without permission, breaking curfew, failing to secure employment, etc. Absconding is a special violation issued when probation officers can no longer locate the offender. Arrest warrants are issued for absconders, and they are typically caught soon after. After the JRA reforms, offenders could only be incarcerated for new crime or absconding violations. Beforehand, they could be revoked for any violation.

2.5 Racial disparities

Racial disparities are a pervasive feature of the US criminal justice system. As shown in Figure 1, for example, black men who did not complete high school are almost as likely to be incarcerated as at work and are employed half as frequently as similarly educated white men. Probation contributes to these patterns. Black offenders are more likely to face technical violations of virtually all types. These disparities are summarized in Figure 2, which reports the coefficients from regressions of a black indicator on an indicator for an event occurring within the probation spell using the North Carolina data. The blue bars report the coefficient when no additional controls are added, while the regressions underlying the red bars feature a battery of other controls, including covariates capturing demographics, geography, criminal history, and standardized math and reading test scores.[9] The first blue bar, for example, shows that black probationers are 17 p.p. more likely to face administrative violations, a 30% increase relative to the white mean. After including all controls, this difference drops to about 10 p.p. In all cases, however, the black coefficient remains

[9]Tables showing full regression results, including the effect of adding controls sequentially, are available starting with Appendix Table A2. Test scores available due to related work in North Carolina described in Rose, Schellenberg and Shem-Tov (2019).

large and statistically significant after including controls. Similar patterns have been documented in multiple other jurisdictions (Jannetta et al., 2014).

Because black offenders face more technical violations, they are also more likely to be incarcerated for breaking technical rules. The black effect for this outcome is roughly 10% of the white mean after including the full suite of control variables. However, the final two bars show that black offenders also more likely to be arrested. These effects are correlated across geographies, as shown in Figure 3. Each dot in this figure plots the black coefficient from a pair of regressions—one with any technical violation and one with any arrest as the outcome—estimated separately for each of the 30 probation districts in North Carolina. In parts of the state where black offenders are more likely to commit crime relative to their white peers, they are also more likely to face technical violations. This pattern suggests that at least part of the racial disparities in technical violations may in fact reflect that potential criminals are also very likely to break technical rules.

3 Defining effective rules and biased rules

In this section, I provide a framework for assessing the effectiveness and equity of rules when viewed as simple tools for predicting socially costly behavior. In my context, these rules—curfews, limitations on travel, and bans on drug and alcohol use, etc.—are intended to identify offenders who are not committed to rehabilitation and thus likely to commit socially costly crimes. The same ideas, however, apply to other contexts, including bail setting (Kleinberg et al., 2017), parole release (Kuziemko, 2013), classroom discipline, and background screening. I then show how with the use of an instrument one can construct a test for biases in accuracy and type-I and type-II error rates, as well as a method for quantifying the contribution of any bias to aggregate disparities in outcomes.

3.1 Illustration of approach

To build intuition, consider a simple one-period model. Individuals are either technically imprisoned due to technical rule violations or not. Individuals who are not imprisoned have the opportunity to commit crimes. Let Y_i^* be a latent binary outcome that equals 1 if individual i would offend if not incarcerated. Let R_i be a binary outcome that equals one if an individual is technically incarcerated. Throughout this section, I suppress an additional subscript s for "spell," treating each person-spell observation as a separate unit for simplicity.

Effectiveness depends on the shares of criminals and "innocents" technically imprisoned, or $\Gamma_1 = Pr(R_i = 1 \mid Y_i^* = 1)$ and $\Gamma_0 = Pr(R_i = 1 \mid Y_i^* = 0)$, respectively. In this one-period model, these parameters correspond to true positive (i.e., $1-$ type-II error) and false positive (i.e., type-I error) rates, respectively, and govern how useful technical rules are as tags for criminal risk. When Γ_1 is close to one, all individuals who would commit a crime also commit technical violations, making it easy to use rules to identify and imprison potential offenders. When Γ_0 is sufficiently high, however, technical rules may catch more innocents than criminals. Thus, for any level of total technical incarceration cost (i.e., $Pr(R_i = 1)$), more effective rules have higher Γ_1 (or equivalently lower Γ_0), implying they ensnare a greater share of criminals and let more innocents go free. In

other words, more effective rules are better classifiers of criminal risk.

My primary concept of equity depends on how Γ_1 and Γ_0 vary across groups. A high Γ_1 for black offenders but low Γ_1 for white offenders, for example, implies that rules target black criminals aggressively, while letting relatively more white offenders off the hook. Higher Γ_0 for one group, on the other hand, implies more non-reoffenders are imprisoned. Unbiased rules are those for which both Γ_1 and Γ_0 do not depend on race, which implies that differences in technical incarceration across groups arise solely because of differences in $Pr(Y_i^* = 1)$, the underlying targeted behavior.

Similar notions of equity have been explored recently in work on "algorithmic fairness" (Corbett-Davies et al., 2017; Kleinberg et al., 2017; Berk et al., 2018). A standard result in this literature holds that it is impossible to simultaneously equalize type-I and type-II errors rates and predictive accuracy (e.g., $Pr(Y_i^* = 1|R_i = 1)$) across groups when an algorithm either does not yield perfect predictions or rates of the predicted outcome differ across groups.[10] Although I will consider accuracy in what follows as well, I focus on type-I and type-II errors because they are most closely connected to the concept of "disparate impact" discrimination in employment law. Type-I errors are also particularly troubling in the criminal justice context, where the presumption of innocence is a core value.

How can Γ_1 and Γ_0 be estimated? Given data on technical incarceration and offending, we can observe $Pr(R_i = 1)$ and $Pr(Y_i^* = 1 \mid R_i = 0)$, but not $Pr(Y_i^* = 1 \mid R_i = 1)$, since these individuals are incapacitated and their criminal outcomes are therefore censored. Despite this, we can always construct an indicator for being *observed* offending, or $Y_i = Y_i^*(1 - R_i)$. Now suppose that we have a binary instrument Z_i that satisfies two assumptions:

1. $E[R_i|Z_i = 1] = 0$
2. $E[Y_i^*|Z_i] = E[Y_i^*]$

That is, the instrument eliminates the possibility of technical incarceration and is mean independent of Y_i^*. The latter assumption implies that when $Z_i = 1$ and technical violations are not punished with imprisonment, probationers do not respond by committing more crime. Such responses are potentially plausible. For example, offenders might use more drugs when failed drug tests are not punished with prison time, which could increase crime. I see no evidence of such behavior, however, as I discuss further below. Moreover, I relax this assumption later in the paper and show that any behavioral responses to weaker punishments for rule breaking are small.

With an instrument that satisfies these two assumptions, it is easy to see that

$$\begin{aligned} \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[Y_i|Z_i = 1]} &= \frac{Pr(Y_i^* = 1) - Pr(Y_i^* = 1, R_i = 0)}{Pr(Y_i^* = 1)} \\ &= \frac{Pr(Y_i^* = 1, R_i = 1)}{Pr(Y_i^* = 1)} \\ &= Pr(R_i = 1 \mid Y_i^* = 1) = \Gamma^1 \end{aligned} \tag{1}$$

[10] To see this, note that:

$$Pr(Y_i^* = 1|R_i = 1) = \Gamma_1 \frac{Pr(Y_i^* = 1)}{Pr(R_i = 1)} = \frac{\Gamma_1}{\Gamma_1 + \Gamma_0(Pr(Y_i^* = 1)^{-1} - 1)}$$

Hence unless $\Gamma_0 = 0$ for both groups or $Pr(Y_i^* = 1)$ is the same, accuracy will differ.

A simple rescaling of the reduced form effect of Z_i thus reveals Γ_1. Since $Pr(R_i = 1)$ is observed, we can also easily estimate Γ_0. The intuition is that because crime is uncensored when $Z_i = 1$, any increases in offending vs. when $Z_i = 0$ must come from individuals who would have offended but were incarcerated instead. Normalizing by uncensored arrest rates yields the fraction of would-be offenders thwarted by technical rules.

By estimating both objects in the black and white populations separately, one can readily test whether technical rules satisfy the notion of equity put forward above. With race specific estimates of Γ_1 and Γ_0, one can also decompose differences in $Pr(R_i = 1)$, or technical incarceration, into a share attributable to targeting and a share attributable to risk. Specifically, letting $B_i \in \{0, 1\}$ denote race, we have:

$$\underbrace{Pr(R_i = 1|B_i = 1) - Pr(R_i = 1|B_i = 0)}_{\text{difference in technical incarceration}} = \tag{2}$$

$$\sum_{k=0}^{1} \underbrace{Pr(Y_i^* = k|B_i = 0)}_{\text{white risk}} \underbrace{[Pr(R_i = 1|Y_i^* = k, B_i = 1) - Pr(R_i = 1|Y_i^* = k, B_i = 0)]}_{\text{difference in targeting}}$$

$$+ \underbrace{Pr(R_i = 1|Y_i^* = k, B_i = 1)}_{\text{black targeting}} \underbrace{[Pr(Y_i^* = k|B_i = 1) - Pr(Y_i^* = k|B_i = 0)]}_{\text{difference in risk}}$$

Thus the total difference is comprised of a component driven by differences in risk ($Pr(Y_i^* = 1)$ and $Pr(Y_i^* = 0)$) and a component driven by differences in targeting. As always with Oaxaca-style analyses, it is possible to construct alternative decompositions by adding and subtracting other composite terms (Oaxaca and Ransom, 1999). Here, I decompose the difference using the white risk distribution and the black targeting rates as the baseline.

The analysis below extends the one-period approach in two ways. First, I incorporate multiple periods. This requires allowing both R_i and Y_i^* to be integer-valued variables indicating how many days into a spell a probationer would be incarcerated for rule breaking or reoffend, rather than the simple binary measures used above. The logic remains the same, however—one simply rescales the difference in crime when $Z_i = 1$ vs. $Z_i = 0$ at each horizon, generating a measure of the share of offenders targeted at that point.

Second, I account for the fact that the reform does not completely eliminate technical rules. In the one period example, this implies that $E[R_i \mid Z_i = 1] > 0$. As a result, I need to introduce a notion of compliers for the reform. These are individuals who *could* be incarcerated for breaking rules if assigned $Z_i = 0$ but would not be if $Z_i = 1$. Because the reform affected only drug and administrative rules, these compliers are simply individuals at risk of breaking these rules alone.

3.2 Full model

Let $Y_i^* \in \{0, 1, 2, \ldots, \infty\}$ measure the time in days it would take individual i to be arrested for a new criminal offense from the start of her probation spell absent any intervention. An infinite duration implies the individual would never be arrested. $R_i^* \in \{0, 1, 2, \ldots, S_i\} \cup \{\infty\}$ measures days to technical incarceration. This event must occur between 0 and S_i, which is the length of the probation spell. Individuals are targeted by technical incarceration whenever $R_i^* < Y_i^*$, implying

they would be imprisoned before they get a chance to commit their crime. Unlike in the single period model, here both objects are latent.

The multi-period version of the bias definition introduced above implies that unbiased rules should target black and white potential criminals similarly at each value of Y_i^*.[11] That is:

Definition 1 *Racially unbiased technical rules satisfy:*

$$Pr(R_i^* < Y_i^* \mid Y_i^* = k, race_i) = Pr(R_i^* < Y_i^* \mid Y_i^* = k) \ \forall \ k$$

Relying on Y_i^* to define risk is akin to a single index restriction. That is, I assume that Y_i^* characterizes risk completely, including the frequency and severity of future offending. Similar assumptions are used in other recent work on racial bias in criminal justice, such as Dobbie, Goldin and Yang (2018), in the spirit of a Becker (1968) outcomes tests. An alternative interpretation of this assumption is that I focus on the extensive margin of *any* offending rather than cumulative measures, as is common in the literature.

Because Y_i^* is unobserved, it is difficult to test the assumption directly. However, I show below that it is not the case that black offenders targeted by technical incarceration (i.e., with $R_i^* < Y_i^*$) commit more severe or more frequent offenses. The increases in crime by crime type across race groups are highly similar after the reform, with black offenders in fact seeing slightly smaller increases in felonies. Moreover, estimated increases in the total cost of crime, where each offenses is assigned a social costs estimate taken from the literature, are statistically indistinguishable between the two groups.

3.3 Impacts of the reform

The reform shifts R_i^*. I model this by allowing each offender to have two *potential* times to technical incarceration: one pre-reform, where drug and administrative rules are enforced, and one post-reform, when they are not. I denote these $R_i^*(0)$ and $R_i^*(1)$, respectively. This setup is an example of the standard Neyman-Rubin potential outcomes model, where, for example, treatment status is indexed by a binary instrument. As usual, only one potential outcome is ever observed for each spell, so that in single-spell data $R_i^* = Z_i R_i^*(1) + (1 - Z_i) R_i^*(0)$.

I make three assumptions about the impacts of the reform. These assumptions are analogous to the standard monotonicity and independence / exclusion assumptions made in estimation of local average treatment effects, or LATEs (Angrist, Imbens and Rubin, 1996), but adapted to the duration context.

Assumption 1. (Monotonicity) $R_i^*(1) \geq R_i^*(0) \ \forall i$

Assumption 2. (Exogeneity) $R_i^*(0), R_i^*(1) \perp\!\!\!\perp Z_i$

Assumption 3. (Exclusion) $Y_i^* \perp\!\!\!\perp Z_i$

[11] This restriction is implied by a stronger definition of bias that requires $R_i^* \perp\!\!\!\perp race_i \mid Y_i^*$. This definition generates many other restrictions, such as that $Pr(R_i^* < l \mid Y_i^* = k, race_i) = Pr(R_i^* < l \mid Y_i^* = k) \ \forall \ l < k$. Since these restrictions are not testable given my variation, I focus on the weaker definition.

Assumption 1 implies that the reform does not *reduce* anyone's time to technical imprisonment. This assumption seems highly plausible in my setting, since the reform simply eliminated prison punishments for some technical rules without introducing additional ones. Assumption 1 does, however, rule out changes in probationers', caseworkers', or judges' behavior that would lead to offenders being technically imprisoned earlier in their spell (for example, by fleeing supervision). I find no empirical evidence that behaviors change in such a way.

Assumption 2 requires that potential technical incarceration durations are independent of exposure to the reform, Z_i. This assumption is supported by a battery of balance and validation checks grounding the claim that individuals placed on probation before the reform provide a good counterfactual for those serving sentences afterwards. There is no evidence of changes in the characteristics of offenders entering probation before and after the reform, no sharp changes in the quantity of offenders on probation, and no trends in technical violations' frequency or type in anticipation of the reform.

Assumption 3 requires that the reform has no direct effect on Y_i^* and was introduced in the one-period model above. It rules out offenders adjusting their criminal behavior because probation overall has become a more lenient punishment as a result of the reform. This implies that offenders also do not increase proscribed behaviors, such as drug use, that may have an indirect effect on crime. Doing so would require probationers to be forward looking. This idea finds little support in the data. The risk of breaking a rule (regardless of the ultimate punishment) does not change after the reform takes effect, for example, despite the fact that the incentives to break some rules (e.g., passing drug tests) changed substantially. Nor do arrest hazards.

Table 3 demonstrates this by estimating a post-reform effect in Cox proportional hazards models for arrests and rule breaking, as well as for drug violations and fees and fines violations specifically. When studying arrests, these regressions treat any technical rule violation as a source of censoring (in essence, pretending that all rule breakers are incarcerated both before and after the reform). If the reform's change to rule-breaking punishments did not affect rule breaking itself, then the population at risk at each duration should also be unchanged. If the change in punishments also did not affect criminality, then there should also be no change in arrest hazards. Analogous arguments apply when treating arrests as censoring and studying violation risk. The results show no change in any behaviors as a result of the reform. I show below that the observable characteristics of offenders also do not change after the reform, suggesting selection *into* probation does not change. Nevertheless, while I impose this assumption initially, in the final part of the paper I relax it and test for behavioral responses directly. I find very limited evidence of any response.

Because the reform did not completely eliminate technical imprisonment, it did not shift R_i^* for all individuals. Specifically, probationers who flee supervision can be still be incarcerated both before and after the reform. To account for this, it useful to introduce an indicator $D_i = 1$ for individuals who could be "caught" by the drug and administrative rules impacted by the reform. These individuals have $Y_i^* < R_i^*(1)$ and are the compliers alluded to above. Individuals with $D_i = 0$ have $R_i^*(1) < Y_i^*$ and thus would be caught by technical rules even after drug and administrative violations are no longer enforced. There is no information in the variation induced by the reform about their criminal outcomes.

3.4 Testing for equity

This framework allows me to use the same logic illustrated above to test whether drug and administrative rules target similar shares of black and white offenders. To do so, I estimate rescaled reduced form effects of Z_i on a composite outcome $Y_i^k = 1\{R_i^* \geq k\}1\{Y_i^* = k\}$, which is an indicator for having an *observed* offending time of k (and hence not being technically imprisoned beforehand). The result, Γ_k, can be interpreted as the multi-period version of Γ_1 studied in the one-period model above. It measures the share of time k offenders who are caught by technical rules. As such, it is also simply the percentage decrease in offenses at each horizon k as a result of imposing technical rules.[12]

Proposition 1 *Under Assumptions 1-3, the rescaled reduced form effect at each horizon k yields:*

$$\begin{aligned}\Gamma_k &= \frac{E[Y_i^k|Z_i = 1] - E[Y_i^k|Z_i = 0]}{E[Y_i^k|Z_i = 1]} \\ &= Pr(R_i^*(0) < Y_i^* \mid Y_i^* = k, D_i = 1)\end{aligned}$$

I leave the short proof of this result for Appendix A1. The intuition is that if offending at time k increases after the reform, it must be because individuals who counterfactually would be technically incarcerated before k now have the opportunity to commit crimes instead. Thus the increase in observed arrests at time k is the product of the probability of having arrest duration k and the conditional probability of breaking the drug and administrative rules impacted by the reform before k. Dividing by the post-reform mean of Y_i^k eliminates the first probability. The result conditions on $D_i = 1$ because the reform did not affect violations for fleeing supervision, so there is no information on individuals incarcerated due to these rules in the reform.

As shown above, race-specific estimates of Γ_k can also be used to measure the contributions of differences in criminal risk and differences in targeting to aggregate racial disparities. In the full model, however, individuals who would never be arrested have $Y_i^* = \infty$. Given a limited time window K over which outcomes are measured, I can at most observe whether $Y_i^* \geq K$. Hence in the full decomposition, the summation in Equation 2 runs from 1 to K and includes a residual component that captures the contributions of all individuals who would offend at time $K + 1$ or later (and possibly never).

4 Results

First, I analyze the effects of the 2011 JRA reform on technical incarceration and arrests over a one-year time horizon using a difference-in-differences estimator. This analysis implements the one-period model used to illustrate my approach in the previous subsection. This one-period analysis is also sufficient to conduct a simple cost-benefits analysis of the effectiveness of technical rules as tags for potential reoffenders and to compare the relative social return to enforcing rules across race groups. I then present estimates from the full model over a three-year time horizon, including tests for bias and a decomposition of aggregate racial disparities.

[12] Ignoring dynamics effects on repeat offending, of course.

4.1 Unadjusted time series

I analyze the 2011 JRA reform using two possible outcomes for each probation spell: 1) new criminal arrest; and 2) technical incarceration. These events are mutually exclusive—a offender cannot be technical revoked if they are arrested first by definition. For each probationer, I measure which event occurs first (if any) and the time to the event. I then calculate the share of probationers technically incarcerated and the share arrested over the course of their spell.

Figure 4 plots the raw data for these two outcomes in Panels A and B, respectively, for three-month cohorts of supervised probationers. These cohorts all start their spells within four years of the reform's effective date, which is marked with the black solid line. The leftmost line in Panel A, for example, plots the share of probationers starting their spells in the beginning of 2007 who were technically incarcerated over the next 365 days. By the end of that period, where the line ends, roughly 15% of the cohort was imprisoned for technical violations. Similar shares experience the same fate in each cohort for the next 12 quarters.

Cohorts beginning probation within a year of the reform, however, begin to see reductions in technical incarceration. These cohorts were affected because the reform's limitations on technical imprisonment applied by the violation date and not the probationer's start or offense date. Thus these cohorts spend a portion of their spell under the new policy regime and see reductions in technical incarceration as a result. The more time each cohorts spends under the new regime, the larger the reductions. Probationers who begin their spell after the reform are fully exposed to its changes. For these cohorts, technical incarceration reduces to 9%, a 33% drop relative to the pre-reform mean. Technical incarceration then stabilizes for the next several years.

The large decrease in technical incarceration means many more probationers had the opportunity to commit crimes instead of being imprisoned. Panel B plots the share who did so. After a slight decline over several years, offending is relatively flat in the 4 quarters before the reform. It then jumps up slightly for spells interrupted by the reform and remains 1-2 p.p. higher afterwards. Thus while the reform sharply reduced technical incarceration, these gains came at a cost. A meaningful share—roughly 30%—of probationers spared technical incarceration in the first year of their probation spells were arrested instead.

This simple interrupted time series analysis may be misleading if selection into probation changed as a result of the reform or if changes in aggregate crime coincided with its implementation. Figure 5 shows that the first threat is not a concern. Predicted offending rates formed using all available covariates are stable over the four years before and after the reform and I cannot reject the null the predicted 1-year crime rates are identical for spells starting in the year before vs. after the reform. Appendix Figure A2 shows that the quantity of offenders on supervised and unsupervised probation also did not change discretely around the reform, indicating that judges' sentencing behavior was unaffected. Thus, although probation overall became more lenient after the reform, there is no evidence that either judges changed their sentencing behavior or potential offenders changed their crime choices in response. Nevertheless, I return to this important point in the final section of the paper, where I estimate behavioral responses to the reform directly.

4.2 Difference-in-differences estimates

To account for potential time-varying confounders, I use a difference-in-differences approach that compares supervised probationers' outcomes to unsupervised probationers'. Panel C of Figure 4 plots the difference in these groups' one-year technical incarceration and arrest rates (i.e., the end-points of the lines in Panels A and B).[13] Specifically, it plots estimates of β_l^T from the linear regression:

$$Y_{is}^j = \alpha + \sum_{l=-16}^{16} 1\{S_{is} = l\}(\beta_l + \beta_l^T T_{is}) + e_i \tag{3}$$

where Y_{is}^j measures whether individual i in spell s experienced outcome j (either arrest or technical incarceration), S_{is} measures how many quarters before or after the reform's effective date i started probation, and T_{is} is an indicator for being on supervised probation. The β_l^T effects are normalized relative to the cohort starting four quarters before the reform, the last group to spend the entirety of their first year of probation under the old regime. The dotted red line marks the first cohort of probationers who start after the reform took effect.

Because unsupervised offenders are not assigned probation officers, less than 1% of them experience technical incarceration in the first year of their spell. As a result, the reform had virtually no impact on this group. The blue line in Panel C thus closely tracks the declines in Panel A—decreases of roughly 6 p.p. after a prolonged period of no substantial changes. Because unsupervised probationers saw no decline in technical incarceration, their arrest rates evolved smoothly over the reform. Beforehand, their outcomes tracked supervised probationers' closely for three plus years. The red line reflects this pattern, showing increases of 2 p.p. with no evidence of pre-trends.

To obtain point estimates of the reform's effects, I collapse Specification 3 to a simple difference-in-difference comparison using probation spells that begin 1-3 years before the reform and 0-2 years afterwards, thus using two years of pre/post data while omitting cohorts whose first year of probation was interrupted by the reform and were therefore only partially affected.[14] These results are presented in Panel A of Table 4. The estimated effect on revocation is 5.5 p.p and easily distinguishable from zero at conventional confidence levels. The increase in arrests is roughly 2 p.p. Thus, over this one-year horizon 30-40% of probationers spared technical incarceration find themselves arrested instead. For both outcomes, it makes little difference whether demographic and criminal history controls are included. Moreover, the small coefficients on the post indicators show that over this narrow window, results would be similar if only treated units were included.

Are these effects small or large? A simple benchmark for the reform's expected effects uses the share of probationers arrested pre-reform, which was 29%. If a similar share of probationers spared technical incarceration instead commit crimes, we would expect offending to go up by roughly 1.6%. The observed increase falls slightly above this simple benchmark, suggesting individuals targeted by technical incarceration are somewhat more risky than average. Since technical incarceration occurs over the course of a probation spell, however, this benchmark is potentially too high. For example, in the extreme case where all technical incarceration occurs on day 355 of the spell, the

[13] The raw rates for unsupervised probationers are presented in Appendix Figure A3.

[14] I use these partially affected cohorts in estimation of the structural model that follows.

reform would only give offenders *one* extra day to commit crimes in their first year, and finding any increase would be surprising. I return to this question in Section 5, where I estimate arrest and technical incarceration hazards directly and show that they are highly correlated across individuals.

In the last two rows of Panel A, I use these results to estimate false positive (Γ_0) and false negative rates ($1 - \Gamma_1$), treating the full first year of the spell as a single period.[15] Specifically, $Y_i^* = 1$ if an individual would commit a crime in the first year of probation and is zero otherwise. The estimated false negative rate shows that just 6.5% of potential criminals are caught by drug and administrative rules affected by JRA reforms. The estimated false positive rate shows that 5.8% of non-offenders (over the one-year horizon), however, violate the same rules. Of course, many of these individuals may offend later, a fact I account for in the full model estimates that follow. Nevertheless, in this simplified setting rules appear almost as likely to target non-reoffenders as reoffenders.

Remarkably, the reform's impact on black offenders' technical incarceration was nearly twice as large as its impact on white offenders'. As a result, the reform eliminated raw racial disparities in technical incarceration. Panel A of Figure 6 demonstrates this result by plotting technical incarceration rates in the sample used for difference-in-differences estimation separately by race. While black offenders were 30-40% more likely to face technical imprisonment over the first year of their spell before the reform, afterwards the race gap is reduced to less than 1%.

Because many more black offenders were spared technical incarceration, one might expect crime in the black population to increase more than in the white population after the reform. Panel B of Figure 6 shows that this did not happen. While more probationers in both groups were arrested after the reform, the racial gap does not change substantially. Race-specific difference-in-difference estimates in Panels B and C of Table 4 imply that the arrest rate among white offenders who, but for the reform, would have been imprisoned for technical violations is above 55%. However, the correspond figure among black offenders is only 30%.

Appendix Table A10 shows that the increase in crimes by crime type do not differ substantially across the two race groups. In fact, the absolute increase in felony offenses is *smaller* in the black population than in the white population, and a larger share of the total increase is accounted for by traffic related offenses. It therefore does not appear that black probationers targeted by technical violations pose lower average risk, but higher risk for more socially costly crimes such as felonies.

Estimates of false negative and false positive rates by race are reported the bottoms of Panels B and C. False negatives are similar by race—roughly 93%—indicating that similar shares of potential reoffenders in both groups are targeted by rules over a one-year period. False positive rates are three times higher for black offenders, however, implying that far more innocent black offenders are technically incarcerated relative to white offenders. In the one-period model, therefore, there is evidence of substantial bias.

Table 5 uses these results to conduct the simple Oaxaca decomposition exercise described in the previous section. This analysis measures the relative contributions of risk (i.e., $Pr(Y_i^* = 1)$) and targeting (i.e., $Pr(R_i = 1|Y_i^* = 1)$) to aggregate racial gaps in technical violations among the

[15] Appendix A2 shows how additive time effects can be incorporated into the model to justify using the difference-in-difference estimates to do so. Using this design introduces a negligible bias, which I estimate to be on the order of 10% of the main post-x-treat effect.

complier population for the reform.[16] As expected, the first two rows show that rates of technical incarceration and offending are both higher in the black population. The next two rows, however, show that in total risk explains a very small share of the aggregate gap. While black offenders' higher likelihood of offending contributes slightly, it is more than fully offset by harsh treatment of non-offenders. This implies that the bulk of differences in technical incarceration are in fact driven by differences in how non-offenders are targeted. The last row of the table confirms this, showing that differences in false positive rates explain 105% of the aggregate gap.

4.3 Cost-benefit analysis

When the state incarcerates an offender for technical violations, it must pay close to $100 a day to do so.[17] If the state instead opts to leave the offender in the community, she may then commit a crime and be sentenced to incarceration as a result. The social value of technically incarcerating individual i can thus be written as:

$$V_i = \underbrace{-J_i}_{\text{Cost of tech. incar.}} + \underbrace{Pr(Y_i^* = 1|R_i = 0)}_{\text{Pr(offend) if not incar.}} \left[\underbrace{E[U(Y_i^*)|R_i = 0, Y_i^* = 1]}_{\text{Cost of crime}} + \underbrace{J_i'}_{\text{Cost of new sent.}} \right] \tag{4}$$

where J_i is the cost of the technical jail/prison spells, R_i and Y_i^*, as before, are indicators for technical incarceration and offending, $U(Y_i^*)$ represents the social cost of this crime, and J_i' represents the total cost of incarceration as a result of the new crime, including any resulting revocation.

Enforcing technical violations for a group offenders is beneficial if $E[V_i] > 0$. I assess this criterion for offenders affected by the 2011 JRA reforms in two ways. First, I use changes in observed costs of incarceration and offending rates over a fixed horizon to back out a "break-even" $E[U(Y_i^*)|R_i = 0, Y_i^* = 1]$ that sets $E[V_i] = 0$ for this population. That is, I solve for:

$$E[U(Y_i^*)|R_i = 0, Y_i^* = 1] = \frac{\triangle E[-J_i \cdot R_i] - \triangle E[(1 - R_i)J_i']}{\triangle E[Y_i^*]} \tag{5}$$

This exercise asks what the *minimum* social cost of crime would be to justify the state's use of technical incarceration for the drug and administrative rules impacted by the reform. The numerator captures the change in net incarceration costs—spending on technical incarceration minus spending on incarceration due to crime. The denominator divides this gap by the increase in crime to arrive at break-even valuation for these marginal offenses.

In a second approach, I use existing estimates from the literature to benchmark crime costs and compare it to these break-even values. This analysis assigns a cost to each category of arrest ranging from $500 (for simple drug possession) to close to $20 million (for homicides) primarily sourced from Cohen et al. (2011).[18] I then compare the change in net incarceration costs due to the reform to estimated increases in costs of crime.

[16] Appendix Section A3 provides complete details on how the decomposition is calculated.

[17] 2018 average daily cost per inmate for the North Carolina Department of Public Safety (https://www.ncdps.gov/adult-corrections/cost-of-corrections). Supervision costs roughly $5 a day in 2018.

[18] See the appendix to Rose and Shem-Tov (2019) for a detailed list of crime costs and their sources. Each arrest is assigned a lower and upper bound for costs based on existing estimates and the categorization of the offense.

This analysis omits several other factors that might contribute to the aggregate costs and benefits of technical incarceration. In particular, the foregone earnings of incarcerated offenders, the utility costs of imprisonment, and the court costs associated with processing technical incarceration are excluded. The excluded potential benefits mainly relate to deterrence effects. As shown earlier, however, there is little evidence that the reform impacted the perceived punitiveness of probation enough to shift potential criminals' offending calculus. Nor is there any change in technical violation behavior after the reform, including for payment of fees or fines.[19] On net, therefore, I view this analysis as providing a lower bound on costs while capturing most potential benefits.

Importantly, these cost-benefit calculations also place no weight on racial equity. Since the reform dramatically reduced black-white gaps in technical incarceration, this is a potentially important factor. Indeed, the more policy makers value reducing black-white disparities, the more attractive the reform becomes regardless of its impact on crime. A full social welfare analysis of the reform—including putting a price on racial equity—is beyond the scope of this paper, however.

I consider costs and benefits of technical incarceration that begins and arrests that occur in the first year of a probation spell. Extending to longer windows tends to reduce the benefits of technical incarceration because many imprisoned individuals will be released and have the opportunity to reoffend. However, because the suspended sentences activated by technical incarceration are usually 3-4 months long, these results are highly similar to comparing the cumulative change in offending over the first year of a spell to the cumulative changes in incarceration costs over the same horizon.

The results are reported in Table 6. The first column reports the change in spending on technical incarceration spells activated in the first year of a probation spell after the reform took effect. This declined by $680 per probationer on average. The second column reports the increase in costs of incarceration attributable to new crimes committed in the first year of a spell. This is relatively close to zero because the majority of new crimes after the reform do not merit an actual prison sentence. The estimates thus imply that for every dollar the state spent on technical incarceration, it saved roughly 30 cents it would have spent on prison costs anyway.

Column 4 reports the implied break-even valuations discussed above. These average about $40k per offense. Although this may seem relatively low, consider that the modal offense committed by a probationer is a relatively minor misdemeanor. In fact, excluding all misdemeanor and traffic offenses raises the marginal valuation to $100k. Columns 5 and 6 report the estimated costs of new crimes generated by the reform. Unfortunately, due to the wide dispersion in reported costs of crime, these estimates are relatively noisy. The point estimates, however, suggest that costs may fall at or below break-even valuations.

The remainder of Table 6 repeats the same exercise for various sub-populations. The second and third rows, which compare black and white probationers, provide a concise summary of the degree to which drug and administrative violations target black offenders more aggressively. The decrease in spending on technical incarceration in the black population is roughly twice as large as in the white population, while increases in the costs of incarceration attributable to new crimes are only slightly larger. Combined with similar increases in reoffending rates for both groups shown earlier, the result is that implied break-even valuations for black offenders are 2-3 times larger than

[19] There is no data available on collection rates for court costs in North Carolina. Surveys in other districts have found overall repayment rates ranging from 50% to 9% in other states (Pepin, 2016).

for white offenders. Unfortunately, estimates in Columns 5 and 6 are too noisy to ask whether differences in costs of crime justify these disparities. However, racial gaps in break-even valuations are even larger when only felony offenses are considered in Column 4, suggesting that differences in the severity of crime committed are unlikely to justify the gap. The final two rows of Table 6 shows that similar but more extreme patterns hold when considering black and white men.

4.4 Full model estimates

The previous estimates abstracted from the durational nature of probation spells by treating the first-year as a single period. I now extend the results to incorporate multiple periods and a longer time horizon, thus accounting for any differences in the distribution of offending times across race groups. Rather then estimating the full model at the daily level, I construct estimates of Γ_k with k binned into 90 day intervals to gain precision. I thus test for bias conditioning on Y_i^* falling somewhere within this interval rather than at k exactly, although results are not sensitive to the exact bin size. I bin all k beyond three years into a final period capturing censored values of Y_i^*—that is, individuals who would reoffend more than three years after starting probation, or possibly never. I continue to include unsupervised probationers as controls to ensure that the results are robust to time-trends in offending.

If drug and administrative violations are unbiased, Γ_k should not vary by race for all horizons k. Figure 7 plots estimates Γ_k for k up to three years and for a final period indicating $Y_i^* >$ 3 years. Although at the shortest durations drug and administrative violations target black and white probationers similarly, large gaps appear later. For all k above six months except one, black probationers are more likely to be targeted. Thus we can clearly reject that Γ_k does not depend on race, and therefore that drug and administrative rules are unbiased.

How important is this bias for the raw racial differences in technical incarceration? As in the one-period example, two factors contribute to these race gaps—the distribution of risk Y_i^* and the conditional probability each risk level is targeted by technical incarceration. The latter factor is exactly Γ_k. Appendix A1 also shows that the distribution of risk among compliers can be calculated using $E[Y_i^k|Z_i = 1]$ for each k. Having estimates of both objects allows me to decompose racial differences in drug and administrative violations into the contributions of each factor.

The results of this exercise are reported in Table 7. The first two columns report the share of technical probationers targeted by drug and administrative violations and their risk distributions separately by race. The first row corresponds to the effect of the reform—i.e., the quantity of technical incarceration due to drug and administrative rules over the full course of the probation spell. The next four rows show the quantity of offenders targeted by such rules who have arrest durations less than 1 year, 2 years, 3 years, etc. For example, the last row says that 25% of white offenders targeted by drug and administrative rules would otherwise be arrested three years later or beyond (including never), while 42% of targeted black offenders would do the same.

The next columns reports the differences between black and white offenders in each row and a decomposition into the relative contributions of Γ_k and the distribution of risk types. This decomposition is akin to asking how many white offenders would be hit by technical imprisonment if they were targeted like black offenders and vice versa. Because black offenders are riskier on

average, differences in risk explain a non-zero portion of race gaps in technical imprisonment. However, differences targeting—the Γ_k estimated above—explain the majority of the differences. As shown in the first row, black technical imprisonment for drug and administrative violations would have been 90% lower if they were targeted like white offenders, but their risk left the same.

4.5 Behaviors or biased responses?

In general, racial disparities in technical violations could arise for two reasons. First, black offenders may be more likely to exhibit the proscribed behaviors. For example, black offenders may have more limited wealth and income and thus find it more difficult to pay fees and fines. Likewise, some populations may have less access to transport, making it more difficult to report to probation officers. In these cases, however, disparities reflect genuine differences in behavior across the populations, whatever their root cause. Alternatively, caseworkers and judges may respond more aggressively to identical behaviors when the offender is black instead of white.

Several pieces of evidence suggest that racial disparities are largely driven by differences in behaviors rather then responses to them. First, there is limited cross-officer variation in black offenders' likelihood of technical violations relative to whites. As shown in Appendix Table A9, controlling for assigned officer has no measurable impact on the black effect for technical violations and only slightly increases the R^2, despite adding hundreds of parameters. Relatedly, as Appendix Table A9 also shows, there is no consistent evidence of same-race effects—black officers are as likely to cite black offenders for administrative violations as white offenders.[20] Meaningful same-race effects have been found in other criminal justice contexts (e.g., West (2018)).

Second, racial disparities are large for technical violation categories where officers have relatively limited discretion as well as those where they have more. For example, relative to their mean incidence, black offenders are equally more likely to face violations for not reporting as for failing drug tests. While officers could fairly easily ignore a forgotten meeting, drug tests are initiated with an automated form produced by the Department of Public Safety's offender tracking computer system and thus harder to sweep under the rug.[21] Black effects divided by the white mean for all violation categories are presented in Appendix Figure A1. This is consistent with officers closely following detailed guidelines in the NC Department of Community Corrections' policy manual, which specify appropriate responses to different probationer behaviors.

Third, racial disparities in technical incarceration are entirely driven by how often offenders pick up violations, not how those violations are punished. Conditional on the violation type, probation officers are equally likely to recommend revocation for black and white offenders and judges are equally likely to grant it, as shown in Appendix Table A8. In fact, simple fixed effects capturing violation types explains 40% of the variation in revocations, implying limited discretion overall in

[20] For drug violations, black officers treat black offenders slightly *more* harshly on average. There is no same-race effect in revocations overall, however, and small negative same race effect for technical revocations.

[21] I shadowed probation officers at work in Durham, N.C. for several days during the summer of 2018. Officers rely heavily on their forms and computer systems. They are primarily incentivized to ensure that all appopriate policies and procedures are followed in each case. Many interactions with offenders consist of probation officers clicking through automated forms on their desktop computers while the probationer answers a standard set of questions. Most officers described their responsibilities as ensuring that their caseload respects all conditions imposed in their sentences, not helping to identify and incapacitate the riskiest offenders.

technical incarceration punishments.

5 Beyond the reform

The previous results demonstrate that the technical rules affected by North Carolina's 2011 reform proxy for latent criminality, but target black offenders substantially more aggressively. As a result, eliminating them increased crime but sharply reduced racial gaps in incarceration. How would these results change if probationers responded to weaker rule regimes by increasing criminal activity? Are these results unique to the rules affected by the reform, or would effects be similar if North Carolina further reduced technical incarceration? And if policy makers opt to keep some rules, which types are the most effective and fair? Answering these questions is not possible with the reduced form evidence alone. In this section, I introduce a semi-parametric model of competing hazards for technical violations and criminal offending that allows me to address them. The model directly characterizes the effectiveness and equity of rules overall and allows me to measure any behavioral responses in crime due to changes in technical rules. In addition, the model allows me to disaggregate among rule types and study their relative effectiveness.

5.1 Model set-up

I work with a multivariate mixed proportional hazards (MMPH) model, where the cause-specific hazard for each outcome at duration t is the product of a baseline hazard and unobserved heterogeneity.[22] To reduce notation and adhere more closely to convention in the duration analysis literature, I relabel Y_{is}^* as T_{is}^0 and R_{is} as T_{is}^1 and reintroduce the s subscript for spells to emphasize the important role repeated spells play in identification of this model. The specification for each hazard j is therefore:

$$Pr(T_{is}^j = t \mid T_{is}^j \geq t, V_i^j, Z_{ist}) = \theta_j(t|V_i^j, Z_{ist}) \tag{6}$$

$$= \theta_j^0(t)\exp(V_i^j + Z_{ist}\tilde{V}_i^j) \tag{7}$$

$\theta_j^0(t)$ is a baseline hazard that is common to all individuals. No restrictions are placed on the shape of $\theta_j^0(t)$. In practice, I estimate a high degree polynomial in duration, although results are similar if dummies for fixed intervals are used instead. The term V_i^j captures multiplicative shifts in this baseline hazard as a function of unobserved heterogeneity or "frailty." The hazards for different causes j are correlated due to this term. Z_{ist} is a binary indicator for whether period t of spell s falls after the reform. I allow the distribution of unobserved heterogeneity to be directly affected by the reform to capture decreases in technical incarceration after it took effect. Note that I do not restrict Z_{ist} to have no impact on criminal offending hazards, an assumption I imposed in the previous analysis. Here, I instead directly measure any behavioral response to the reform.

Identification of this model in both single cause and competing risks settings has been studied extensively. Elbers and Ridder (1982) showed that unobserved heterogeneity is identified in the

[22] I focus on cause-specific hazards as opposed to the sub-distribution hazard of Fine and Gray (1999) due to its clear connection to the potential outcomes notation introduced above.

single-cause setting with the use of covariates. Results from Cox (1962) and Tsiatis (1975) originally showed that generally competing risks are not empirically distinguishable from independent risks. However, Heckman and Honoré (1989) proved that when covariates are included as an additional multiplicative shifter of the baseline hazard (usually as $\exp(X_i'\beta_j)$), the MMPH model is identified with sufficient variation in X_i and under some restrictions on V_i^j. When the data contain multiple spell observations per person, these conditions can be relaxed substantially and no covariates are needed. Restrictions on the distribution of V_j and its relationship with X_i are also unnecessary (see Honoré (1993) and Proposition 3 of Abbring and Van Den Berg (2003)). Since many probationers frequently reoffend, I rely on this source of identifying variation to estimate the model.[23] I omit covariates, allowing all heterogeneity (including observed) to flow through V_i^j, and estimate the model completely separately by race and gender.

Intuitively, identification with multiple-spell data comes from the joint distribution of survival times across spells. If there is no unobserved heterogeneity within a covariate group (e.g., white males), then the joint distribution should factor into the product of marginal survival time distributions for each spell. If, on the other hand, individuals who survive for longer in their first spell are also likely to survive for longer in their second, that is evidence of an unobserved component in the hazard common to both spells. The same logic applies to the joint distribution of survival times across different, competing causes.

The reform itself also generate additional identifying variation and is especially useful for pinning down behavioral responses in crime. Since the reform reduced the risk of technical violations, it exposes the latent arrest propensities for individuals who counterfactually would have been censored by this competing risk. If crime increases by *more* than what would be predicted by the decrease in censoring alone, that implies some behavioral response to the reform is necessary to rationalize the data. As I will show, however, observed increases in crime are highly consistent with very small behavioral responses.

In Appendix Section A5, I discuss technical issues in identification and estimation explicitly when failure times are discrete as opposed to continuous, as is my case, and when the distribution of unobserved heterogeneity is approximated using a finite number of mass points, as I do below following Heckman and Singer (1984). Interestingly, the MMPH model's core assumption of a multiplicatively separable relationship between the baseline hazard and V_i^j can be relaxed by allowing $V_i^j = V_i^j(t)$ vary stepwise over duration. This allows unobserved heterogeneity to play a bigger or smaller role at different points in the spell. In the limit, when $V_i^j(t) = V_{it}^j$ is fully flexible, the discrete-time, discrete-types model is still set identified. In practice, however, these modifications have little impact in my setting. Instead, I find that the baseline MMPH model provides a good fit to the data and I therefore use it for the primary results.

5.2 Estimation

The MMPH model describes continuous time durations, whereas failure times in my data are observed at the daily level. It is straightforward to show that the MMPH discrete time hazard from

[23] As shown in Table 1, there are 1.33 spells per person in the treated sample.

periods t to t' is given by:[24]

$$\theta_j^d(t, t', V_i^j, Z_{ist}) = 1 - \exp(-\exp(\tilde{\theta}_j^0(t, t') + V_j^i + Z_{ist}\tilde{V}_i^j)) \tag{8}$$

where d indicates "discrete", $\tilde{\theta}_j^0(t, t')$ is the log of the integrated baseline hazard from t to t'. Given that hazards must fall between zero and one, this discrete hazard is equivalent to a log-log binary choice model for failure in time t to t' conditional on survival up to t. For p close to zero, the log-log link $ln(-ln(1-p))$ is extremely close to the logit transform $ln(p/(1-p))$. Using a logit instead of a log-log binary choice model is attractive because it constraints probabilities to fall between zero and one, so I follow Efron (1988) and use this link function in estimation. To reduce computational burden, I discretize to the weekly level and censor spells at three years. In practice, therefore, I estimate a logit model for failure in the weekly duration panel, stacking each cause. The likelihood is written in Appendix A4. Estimation is conducted in Python using the Boyd-Fletcher-Goldfarb-Shanno algorithm and the analytic gradient.

I estimate the model separately by race (black vs. white) and gender (male vs. female), generating four total groups. In the baseline specification, I include a fifth order polynomial in weekly duration. Rather than incorporating untreated probationers to account for time variation in offending, I include simple time trends in the intercept of the duration polynomial, although results are not sensitive to this choice. Following Heckman and Singer (1984), I discretize the unobserved heterogeneity V_i^j and $\tilde{V}_i^j$. For the J competing causes, each type w has a set of J unobserved components, one for each cause, i.e., $V_w = \{V_w^j\}_{j=1}^J$. While I normalize types so that the first has the lowest unobserved criminal offending risk, I make no restrictions on the relative risk of all other causes across types. In practice, I find that $\tilde{V}_w^0$—i.e., the effects of the reform offending hazards—are very small. To gain precision, I impose that $\tilde{V}_w^0 = \tilde{V}^0 \; \forall \; w$, so that the reform's effects on offending hazards are captured by a single behavioral response parameter.

5.3 Results

I begin by estimating the MMPH model for the same two causes considered in the reduced form analysis presented above: arrest and technical incarceration. Doing so allows me to test the fit of the MMPH model, to examine behavioral responses to the reform, and to simulate further reductions in technical incarceration. I allow for four distinct types in the core specifications. Results change little if more types are allowed, however.

Estimated parameters and their associated standard errors are presented in Table 8. Given that I use the logit link function in estimation, these estimates can be interpreted as partial effects on the log-odds of the weekly hazard for the relevant outcome. Estimates of baseline hazards show negative duration dependence in arrest risks and positive duration dependence in technical violation. Since these coefficients are difficult to interpret on their own, Figure 8 plots average cause-specific hazards for black and white men over the first year of a spell. As expected, black men have both higher arrest and technical violation hazards. The degree of duration dependence in arrest hazards for both groups is relatively minor, decreasing roughly 0.3 percentage points over

[24] This follows from the fact that the survivor function $S_j(t|V_i^j, Z_{ist}) = exp\left(-\int_0^t \theta_j(u|V_i^j, Z_{ist})du\right)$.

the first year before flattening out. Technical violation risk, however, peaks mid-way through the first year of a spell.

Estimates of type effects and their associated probabilities show that unobserved heterogeneity is an important feature of the data. Among black men, for example, the lowest criminal risk type comprises 41% of the population and has a 3 log point lower weekly odds of offending then the highest risk type, which makes up 3.7% of the population. White men show similar degrees of unobserved heterogeneity, although as shown in Figure 8 their average arrest risk is lower. Black and white women also show wide variation and qualitatively similar patterns in arrest risk. I focus on men in what follows since they make up the bulk of offenders and capture the cross-race patterns well.

The estimated type effects on technical violation show large degrees of unobserved heterogeneity and a strong correlation with arrest risk. The highest criminal risk black males, for example, have 2.3 log point higher weekly odds of facing technical violations than the lowest risk types. Low-risk white men have even lower risk of technical violations, with 4% of the population belonging to Type 1, which is relatively low risk and virtually never subject to technical violations. Risks are not perfectly correlated, however. Both black and white men show evidence of imperfect correlation between technical violation and arrest risks, indicating that not all variation in technical violations is driven by criminal propensities. That is, both dimensions of heterogeneity cannot be collapsed into single factor with separate loadings.

Comparing the model's cause specific hazards to Kaplan-Meier (KM) (Kaplan and Meier, 1958) estimates of the same objects, which are presented in Appendix Figure A4, further illustrates the impact of unobserved heterogeneity in this setting. The KM estimator is simply the weekly probability of failure for each cause conditional on not failing due to *any* cause previously. Hence, KM only accurately estimates hazards when the risks are independent. In this case, the strong positive correlation in risk both depresses the KM hazard estimates overall for each cause and exacerbates observed negative duration dependence, as is expected (Van Den Berg, 2001). KM estimates of arrest hazards, for example, suggest declines in risk of close to 66% for black men over the first year of a spell.

Estimates of behavioral effects of the reform on arrest hazards in Table 8 are captured in the rows labeled "Type w x post." As noted above, since type-specific effects are all small and noisy, I impose equal coefficients for all types to increase precision and produce a single behavioral response parameter. These estimates show positive but economically small responses to the reform. Neither coefficient are statistically significant, however, at the 1% level. In Appendix Figure A5, I plot the implied effect of these responses on one-year failure rates for men. Black and white male arrests rates increase by 0.9 p.p. and 0.3 p.p., respectively. Moreover, these responses diminish as more flexible controls for calendar time are included in the model or more types are added. Thus the model shows very weak evidence for real behavioral responses to the reform, suggesting our previous assumption of zero response was correct.

To test the model's fit to the data, Figure 9 compares the model's predicted increases in arrests as a result of the reform to difference-in-difference estimates of the reform's effects, an exercise similar in spirit to testing the fit of control function-based reproductions of non-parametric estimates of treatment effects (Kline and Walters, 2016; Rose and Shem-Tov, 2019). For each race-by-gender

group, I estimate the increase in observed offending at 90, 180, 270, and 360 days in the spell using the same specification as in the preceding difference-in-differences analysis, yielding a total of 16 points. I then simulate increases in offending in the model at each horizon and for each race-by-gender group using the model's estimates of offending and technical violation hazards and the effects of the reform on the latter. The model-based predictions fall within the difference-in-difference estimates' confidence intervals for all points.

I next examine racial bias by plotting the Γ_k (i.e., technical rule targeting), as studied in the reduced form analysis, implied by the model estimates. To do so, I simulate arrest and technical violation failure times separately by race using the pre-reform coefficients and plot $Pr(T^r_{is} < k | T^c_{is} = k)$. Figure 10 show results for k up to 1080, with $k > 1080$ shown as a single final point at the rightmost extreme of the figure. Unlike in the earlier reduced-form analysis, the Γ_k defined by the model here captures the impact of all technical violations, not just those impacted by the 2011 reform. The pattern remains the same, however: Black men are more likely to be targeted by technical violations regardless of their offending risk.

How important is variation from the reform to this conclusion? Appendix Figure A6 explores this by plotting estimates of Γ_k from the model when estimated on pre-reform data only. In this case, all of the information about the correlation between arrest and technical violation risks and differences across race groups comes from repeated spells. This figure is highly similar to Figure 10 and continues to show that black men are more likely to be targeted at all risk levels. In particular, the rightmost dots can be interpreted as type-I error rates for arrests over the three year period after probation begins. While roughly 14% of white non-offenders are hit with technical rules, roughly 24% of black offenders meet the same fate.

Estimated unobserved heterogeneity components can speak to the source of this bias. These coefficients show, for example, that bias appears at all risk levels. The highest criminal risk black male offenders, for example, have 0.3 log point higher weekly arrest risk than the highest risk whites yet have 0.6 log point higher technical violation risk. Low risk black men, however, appear particularly likely to pick up technical violations. For example, Type 2, which has the second lowest criminal risk, has the highest technical violation risk and faces a higher risk than any white male type. Hence, a simple regression of type effects for technical violations on those for arrests yields a slope roughly 50% larger among white men than among black men.

In principle, the patterns in Figure 10 could arise if the unobserved types were the same in black and white populations but the shares differed. However, these estimates imply that rules are also *ex-ante* biased against black probationers. That is, subpopulations of black and white probationers with the same likelihood of reoffending face different likelihood of technical violations. To make this point more clearly, Table A12 re-estimates the model constraining all parameters except unobserved technical violation risk and type shares to be the same between black and white offenders (by gender). The estimates thus directly compare the technical violation risk of black and white probationers with equal *ex-ante* reoffending risk. The estimates show that black offenders are targeted by technical rules more aggressively, especially among low-risk probationers.

5.4 Eliminating all technical violations

What would happen if policy makers further reduced technical incarceration? The answer depends on whether rules unaffected by the reform behave similarly to the drug and administrative rules it relaxed. Using the model estimates of technical incarceration hazards pre vs. post reform, it is straightforward to compare rules overall, those affected by the reform, and those left in place afterwards.

In particular, the model implies that switching to the post-reform regime generates increases in three-year arrest rates of 2.4 p.p. and decreases in technical incarceration of 6 p.p., implying the three-year accuracy rate of impacted rules is 44%. Moving from the post-reform regime to *zero* rules, on the other hand, would increasing crime a further 7 p.p. and drop technical incarceration by 12 p.p., implying a 56% accuracy rate. Thus remaining rules are slightly better predictors of crime than those affected by the reform, but not dramatically so. Cost benefit analyses of reducing technical rules further are thus likely to yield similar results to the previous analysis of the impacts of the 2011 reform itself.

As shown previously, however, after the reform black-white gaps in technical incarceration were effectively eliminated, implying limited scope for any remaining bias. Appendix Figure A7, which plots estimates of Γ_k for technical rules after the reform, shows that the model estimates agree with this assessment: Racial gaps in targeting are significantly narrower post-reform, especially for $k <$ three years. Thus while reducing technical rules further poses a similar tradeoff in terms of increased crime, it would bring smaller benefits in terms of reduced inequity. This result implies that not all technical rules have similar effectiveness and equity and that the drug and administrative rules impacted by the reform may be particularly biased. To more precisely characterize the impact of specific rule types, I next extend the model to accommodate additional risks for different types of violations.

5.5 Disaggregating violation types

I examine the impact of specific rule types by adding separate hazards for each violation type to the model so that there are now five completing causes: 1) criminal arrests, 2) failure to pay fees and fines, 3) drug or alcohol use, 4) not reporting, and 5) all other violations.[25] In this section, I code a simple rule violation as a failure regardless of whether the probationer was punished with prison time. Doing so allows me to examine how the violation behavior itself, as opposed to technical incarceration overall, correlates with criminal behavior.

Appendix Figure A8 plots the estimated average hazards for each outcome from the model estimated with five types, as before. Encouragingly, estimated arrest hazards are almost identical to the two-cause model, suggesting that both models capture similar degrees of unobserved heterogeneity in criminality. Other hazards have the expected shapes, with reporting and drug / alcohol violations peaking halfway through the first year of a spell. Fees and fines violations are

[25] There is a natural hierarchy to some violation types. For example, offenders who stop reporting almost always have unpaid fees. Hence I code a violation as "failure to pay" if the offender did not pay fees or fines and also did not have any drug/alcohol related or reporting violations at the same time. "Reporting violations" are for not reporting / absconding without simultaneous drug violations, while violations for "drug or alcohol" are coded as such whenever substance use / possession is included in the violation.

concentrated towards the end of a first year, when many spells are coming to a close and financial obligations are due.

Figure 11 uses the model to examine how enforcing subsets of these rule types impacts the equity and effectiveness of technical rules. The x-axis plots the share of probationers who would reoffend over the first three years of a spell, but break technical rules first—in other words, the share of would-be reoffenders caught by technical rules, or the true positive rate. The y-axis plots the share of non-reoffenders over the same period who do not violate any rules. Each point comes from enforcing a technical rule "regime" using the rules indicated in the labels: "F" for fees / fines violations, "D" for drug / alcohol violations, "R" for reporting violations, and "O" for all other rules.[26] Rules' effectiveness improves moving to the top right corner of the graph, indicating that the rules catch more would-be offenders and imprison fewer non-offenders. Regimes reflecting the rules enforced with incarceration punishments before vs. after North Carolina's 2011 reform are highlighted with red arrows.

Figure 11 illustrates several interesting features of technical rules. First, using rules related to fees and fines is almost always dominated by not doing so for both race groups. For black offenders, for example, regimes that use fees and fines lie below and to the left of regimes that do not. Switching from a regime that only enforces fees to one that only enforces drug violations would result in 1-2 p.p. more would-be reoffenders being caught and 18 p.p. fewer non-reoffenders being imprisoned. Eliminating fees and fines violations thus offers a clear improvement over the current status quo.[27] North Carolina's reform achieved some of this impact by addressing this violation category.

Second, most regimes for black men are interior to those of white men, indicating that all rule types generally have a tougher time discriminating between black offenders and innocents. Some rules, however, appear to be particularly unfair to black offenders. While fees and fines, for example, reduce the effectiveness of all almost all regimes for white offenders, the decreases in true negative and true positive rates when using them in combination with other rules are smaller than for black offenders. Hence, dropping fees and fines rules thus not only improves effectiveness but also reduces disparities, as in North Carolina's 2011 reform. Indeed, the post-reform regime for black men appears to reside on a frontier to the exterior of the pre-reform regime.

Third, drug and reporting rules both appear to perform similarly. Using them in combination tends to simply increase the aggressiveness of the regime overall, trading off increases in the share of would-be reoffenders incarcerated for increases in the share of non-reoffenders locked up. The regimes that tend to produce the most similar results for black and white offenders, however, include simply using drug violations or reporting violations alone. The optimal technical rule regime depends on how policy makers assign benefits to catching would-be offenders and costs to incarcerating innocents. If the former is assigned more weight than the later, combinations of drug, reporting, and all other rules will be preferred. If the latter is assigned more weight, on the other hand, relying on smaller subsets of rules will be optimal.

Consider, for example, a counterfactual in which only reporting rules are enforced. These rules can reduce three-year reoffending rates by 9.4 p.p. and 6.8 p.p. among black and white male

[26] Other rules include violations rarely charged, such as failing to pursue vocational training are contacting a victim.
[27] Ignoring impacts on collection, as discussed above.

probationers, respectively. Black men would be 5 p.p. more likely to be subject to technical violations (18% vs. 13%), but targeting bias (as measured by the Oaxaca-Blinder decomposition employed previously) explains the smallest share of this gap among all regimes considered. Adding fees and fines violations reduces crime by an additional 6 p.p. and 4 p.p. among black and white men, but at substantial cost: the share of offenders subject to technical violations nearly doubles, rising to 35% and 24% among each group, respectively. Adding drug use rules, on the other hand, reduces crime by slightly *more* while subjecting 30% and 21% of probationers to technical violations.

Many states continue to use technical violations extensively today, as shown in Figure 12. This figure lists the top 20 US states ranked by the share of state prison admissions due to technical violations of probation and parole from data collected recently by the Council of State Governments Justice Center (CSG, 2019). In Kentucky, South Dakota, Kansas, Missouri, Utah, and Wyoming, technical violations among probationers and parolees account for more than 40% of all new prison spells. Many other states sit at well over 25%, including New York, Ohio, Mississippi, and South Carolina. Most of these states—those with blue bars—have no statutory limitations on which technical violations can lead to prison time. Those that do—the grey bars—have simple "hardship" exceptions for fees and fines violations. Reduced reliance on fees and fines in probation is therefore likely to be an attractive reform for many jurisdictions. Indeed, related reforms have become increasingly popular in other areas of the criminal justice system, such as California's recent efforts to eliminate cash bail for pre-trial detention.

6 Conclusion

This paper studies the probation system. Probation is the primary way the US criminal justice system gives convicted offenders a second chance to avoid prison and get back to work. Probationers return home, but are subject to technical rules that forbid drugs and alcohol, require payment of fees and fines, and limit travel, among other constraints. Rule violators can be sent to prison, making probation an important driver of incarceration. Since black men are significantly more likely to break rules, probation also drives racial disparities in prison exposure.

I use a 2011 reform in North Carolina that reduced prison punishments for technical rules to study whether rule violations are strong predictors of future crime and deter reoffending, to examine how their predictive power and deterrence effects differ across racial groups. I find that while rule violations are correlated with criminality overall, they are significantly less predictive of future offending among black probationers. As a result, the reform closed black-white gaps in imprisonment for breaking technical rules without affecting black-white gaps in crime. Using a semi-parametric model of competing risks, I find that rules related to fees and fines are particularly poor predictors of future crime and drive racial disparities. I also find harsh punishments for rule violations have negligible deterrence effects that do not differ by race.

More broadly, my results show how ostensibly race-neutral policies—in this case the imposition of common sense rules designed to encourage desistance from crime and promote public safety—can generate large racial disparities not justified by the policies' ultimate goals. Poorly designed rules and policies are a potentially powerful explanation for many observed racial disparities in criminal justice and beyond. Fortunately, correcting bias due to disparate impact may be easier

than changing biased decision makers' behavior—be they cops, judges, or prosecutors—since doing so is a matter of simply changing the rules themselves. The findings presented here provide clear evidence that such changes are both feasible and can have large, persistent impacts on racial disparities.

References

Abbring, Jaap H., and Gerard J. Van Den Berg. 2003. "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica*, 71(5): 1491–1517.

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2012. "Do Judges Vary in Their Treatment of Race." *The Journal of Legal Studies*, 41(2): 1239–1283.

Agan, Amanda, and Sonja Starr. 2018. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment." *The Quarterly Journal of Economics*, 133(1): 191–235.

Aghion, Philippe, and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy*, 105(1): 1–29.

Angrist, Joshua D., and Alan B. Krueger. 1995. "Split-Sample Instrumental Variables Estimates of the Return to Schooling." *Journal of Business & Economic Statistics*, 13(2): 225–235.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, 91(434): 444–455.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials*." *The Quarterly Journal of Economics*, 127(2): 1017–1055.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*, 133(4): 1885–1932.

Arrow, Kenneth. 1973. "Higher education as a filter." *Journal of Public Economics*, 2(3): 193–216.

Bayer, Patrick, and Kerwin Kofi Charles. 2018. "Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940." *The Quarterly Journal of Economics*, 133(3): 1459–1501.

Becker, Gary S. 1957. *The Economics of Discrimination.* University of Chicago Press.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy*, 76(2): 169–217.

Berk, Richard, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. 2018. "Fairness in Criminal Justice Risk Assessments: The State of the Art." *Sociological Methods & Research*, 0049124118782533.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. 2019. "Incarceration, Recidivism, and Employment." *Journal of Political Economy*, Forthcoming.

Bushway, Shawn D., and Brian Forst. 2013. "Studying Discretion in the Processes that Generate Criminal Justice Sanctions." *Justice Quarterly*, 30(2): 199–222.

Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2018. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." *NBER Working Paper No. 24441.*

Cohen, Mark A., Roland T. Rust, Sara Steen, and Simon T. Tidd. 2011. "Willingness-To-Pay For Crime Control Programs." *Criminology*, 42(1): 89110.

Corbett-Davies, Sam, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. "Algorithmic Decision Making and the Cost of Fairness." *KDD '17*, 797–806. New York, NY, USA:ACM.

Cox, David R. 1962. *Renewal Theory.* Methuen.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review*, 108(2): 201–40.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2018. "The Value of Regulatory Discretion: Estimates From Environmental Inspections in India." *Econometrica*, 86(6): 2123–2160.

Efron, Bradley. 1988. "Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve." *Journal of the American Statistical Association*, 83(402): 414–425.

Elbers, Chris, and Geert Ridder. 1982. "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model." *The Review of Economic Studies*, 49(3): 403–409.

Fine, Jason P., and Robert J. Gray. 1999. "A Proportional Hazards Model for the Subdistribution of a Competing Risk." *Journal of the American Statistical Association*, 94(446): 496–509.

Fryer, Roland G. 2019. "An Empirical Analysis of Racial Differences in Police Use of Force." *Journal of Political Economy*, 127(3): 1210–1261.

Harding, David J., Jeffrey D. Morenoff, Anh P. Nguyen, and Shawn D. Bushway. 2018. "Imprisonment and Labor Market Outcomes: Evidence from a Natural Experiment." *American Journal of Sociology*, 124(1): 49–110.

Heckman, James, and Burton Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica*, 52(2): 271–320.

Heckman, James J., and Bo E. Honoré. 1989. "The Identifiability of the Competing Risks Model." *Biometrika*, 76(2): 325–330.

Honoré, Bo E. 1993. "Identification Results for Duration Models with Multiple Spells." *Review of Economic Studies*, 60(1): 241–46.

Jannetta, Jesse, Justin Breaux, Hellen Ho, and Jeremey Porter. 2014. "Examing Racial and Ethnic Disparities in Probation Revocation." Urban Institute Washington, D.C.

Kaeble, Danielle. 2018. "Probation and Parole in the United States, 2016." Bureau of Justice Statistics BJC Bulletin NCJ 251148.

Kaplan, E. L., and Paul Meier. 1958. "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association*, 53(282): 457–481.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. "Human Decisions and Machine Predictions*." *The Quarterly Journal of Economics*, 133(1): 237–293.

Kleinberg, Jon, Snedhil Mullainathan, and Manish Raghavan. 2017. "Inherent Trade-Offs in the Fair Determination of Risk Scores." New York, NY, USA.

Kline, Patrick, and Christopher R. Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start*." *The Quarterly Journal of Economics*, 131(4): 1795–1848.

Kuziemko, Ilyana. 2013. "How Should Inmates Be Released From Prison? An Assessment of Parole Versus Fixed Sentence Regimes." *Quarterly Journal of Economics*, 128(1): 371–424.

Mueller-Smith, Michael, and Kevin T. Schnepel. 2019. "Diversion in the Criminal Justice System." Working Paper.

Oaxaca, Ronald L., and Michael R. Ransom. 1999. "Identification in Detailed Wage Decompositions." *The Review of Economics and Statistics*, 81(1): 154–157.

Pepin, Arthur W. 2016. "The End of Debtors' Prisons: Effective Court Policies for Successful Compliance with Legal Financial Obligations." Conference of State Court Administrators Policy Paper.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *The American Economic Review*, 62(4): 659–661.

Reaves, Brian A. 2013. "Felony Defendants in Large Urban Counties, 2009 - Statistical Tables." Bureau of Justice Statistics State Court Processing Statistics NCJ 243777.

Rehavi, M. Marit, and Sonja B. Starr. 2014. "Racial Disparity in Federal Criminal Sentences." *Journal of Political Economy*, 122(6): 1320–1354.

Robina Insitute. 2016. "Probation Revocation And Its Causes: Profiles of State and Local Jurisdictions." University of Minnesota Policy Paper.

Rose, Evan K., and Yotam Shem-Tov. 2019. "Does Incarceration Increase Crime?" *Working Paper.*

Rose, Evan K., Jonathan Schellenberg, and Yotam Shem-Tov. 2019. "The Effects of Teacher Quality on Criminal Behavior." *Working Paper.*

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2019. "PUMS USA: Version 9.0 [dataset]." *Minneapolis, MN: IPUMS.*

The Council of State Governments Justice Center. 2011. "Justice Reinvestment in North Carolina."

The Council of State Governments Justice Center. 2019. "Confined and Costly: How Supervision Violations are Filling Prisons and Burdening Budgets."

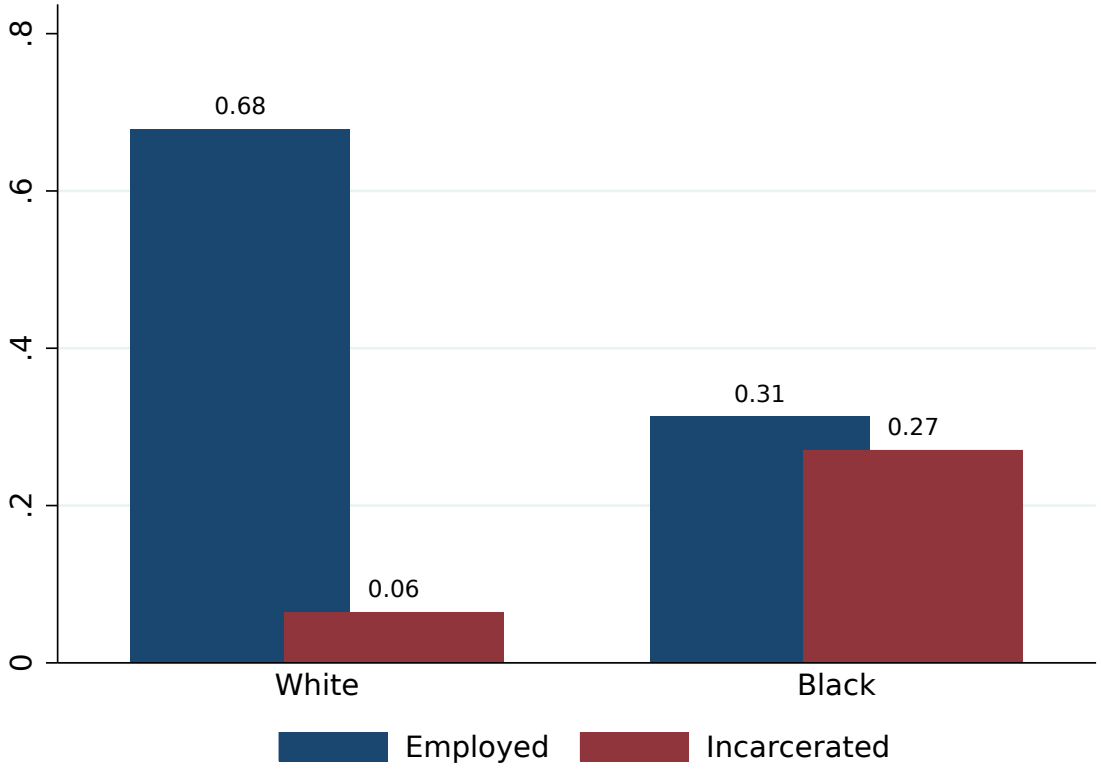
Tsiatis, Anastasios. 1975. "A Non-identifiability Aspect of the Problem of Competing Risks." *Proceedings of the National Academy of Sciences*, 72(1): 20–22.

Van Den Berg, Gerard J. 2001. "Duration models: specification, identification and multiple durations." In *Handbook of Econometrics*. Vol. 5 of *Handbook of Econometrics*, , ed. J.J. Heckman and E.E. Leamer, Chapter 55, 3381–3460. Elsevier.

West, Jeremy. 2018. "Racial Bias in Police Investigations." *Working Paper.*

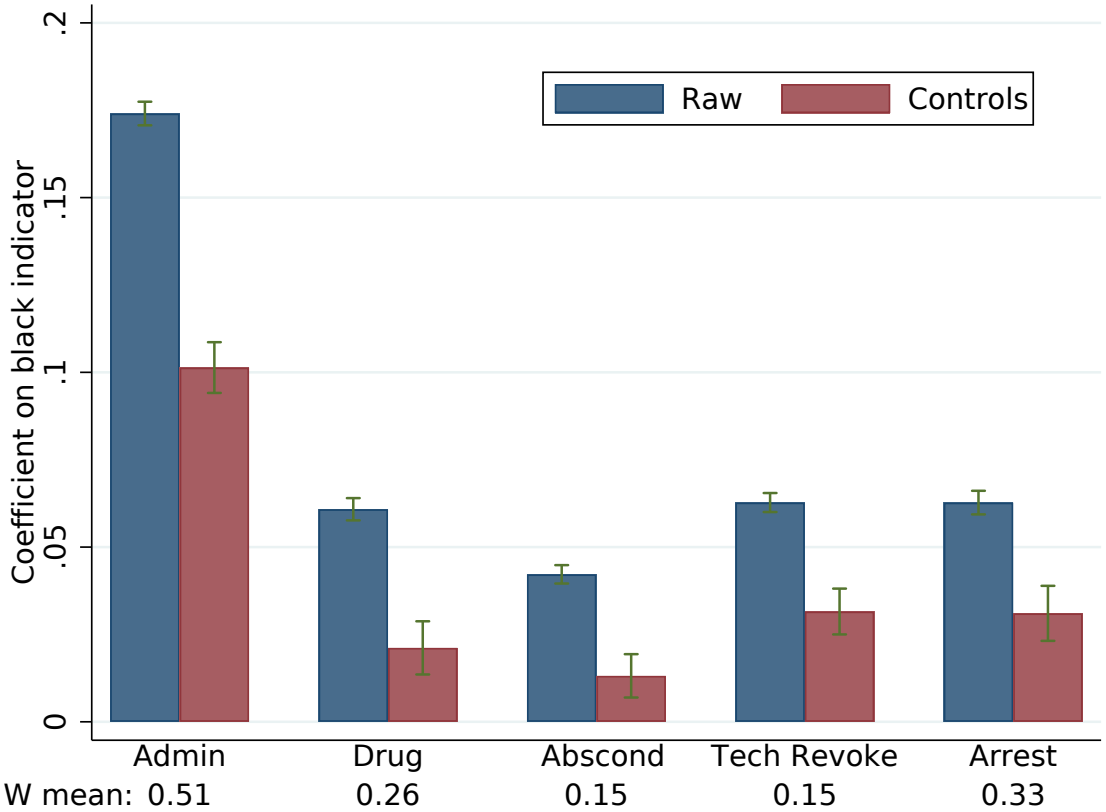
Figures

Figure 1: Male High School Dropouts: Employment and Incarceration



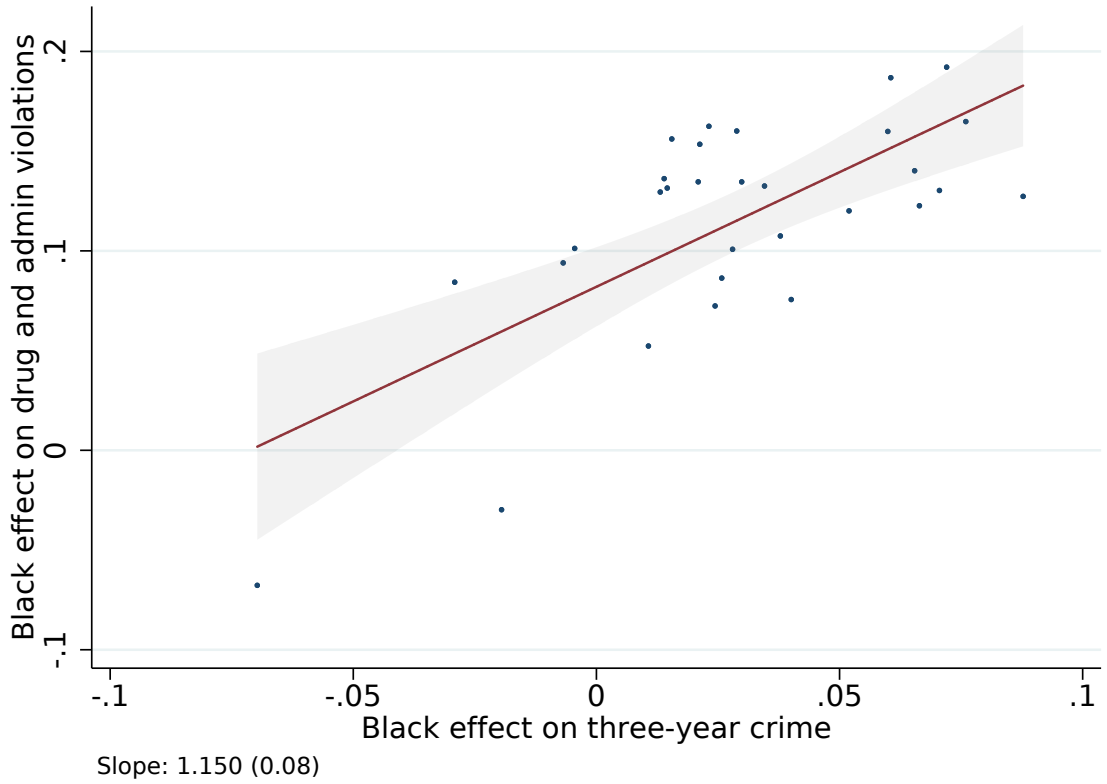
Notes: Figure constructed using the 2013-2017 5-year public use American Community Survey data (Ruggles et al., 2019). Includes White and African-American men aged 20-40 with less than 12 years of education. All estimates constructed using IPUMS person weights. Blue bars are means of an indicator for being at work at the time of enumeration. Red bars are means of an indicator for being enumerated in institutional group quarters, which includes adult correctional facilities, mental institutions, and homes for the elderly, handicapped, and poor. Breakouts for correctional facilities alone are not available in public use data, but adult correctional facilities account for 95% of the total institutional group quarters population for men 18-54 in the 2013-2017 ACS, according to Census Bureau tabulations.

Figure 2: Racial Disparities in Probation Outcomes



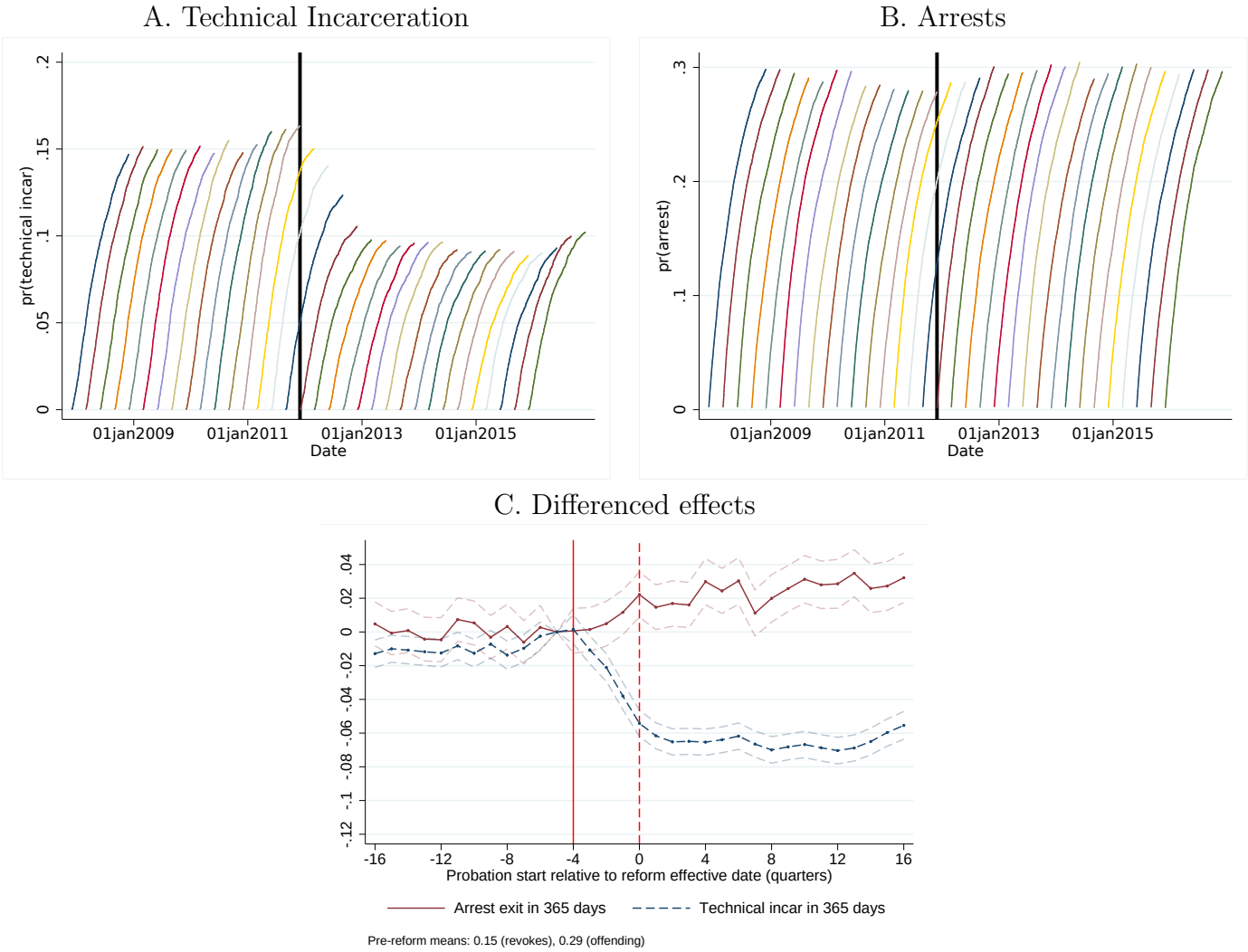
Notes: Regressions include all supervised probationers starting spells in 2006-2010. W mean refers to the white mean of the dependent variable, which is an indicator for the relevant outcome occurring at any point in the spell. Admin includes violations such as non-payment of fees and fines. Drug includes drug-related violations. Absconding is fleeing supervision. Technical revocations are incarceration for rule breaking without a preceding arrest. Adjusted estimate is from an OLS regression with controls for gender, 20 quantiles of age effects, district fixed effects, fixed effects for the offense class of their focal conviction, a linear control for the length of the supervision spell, fixed effects for prior convictions and revokes, a linear control for previous incarceration duration, and the most recent math and reading standardized test scores (normalized to have mean 0 and standard deviation 1 in the full test-taker population) observed between grades 3 and 8.

Figure 3: Relationship Between Black Effects on Technical Violations and Crime



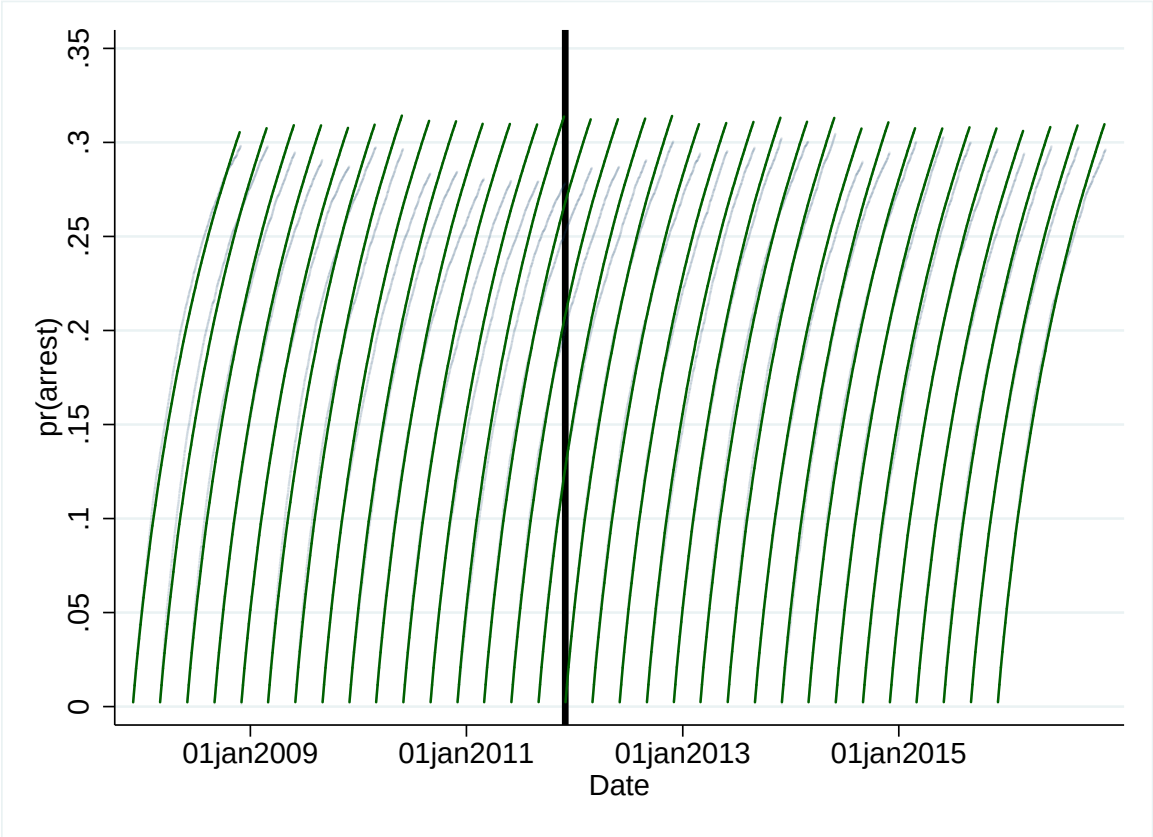
Notes: Regressions include all spells starting in 2006-2010. Each dot plots the coefficient on a black indicator from two regressions estimated separately in each of the 30 probation districts in the state. The outcome in the first regression is an indicator for any criminal arrest within three years of starting probation. The outcome in the second regression is an indicator for any drug or administrative violation in the spell. All regressions include the demographic, sentencing, and criminal history controls used in Figure 2. To avoid mechanical relationships, I randomly split the sample in half and run regressions for each outcome in separate samples, as in a split-sample IV estimate (Angrist and Krueger, 1995). The positive slope indicates that racial gaps in technical violations and racial gaps in criminal risk are positively correlated across the state, as would be expected if criminally riskier probationers incur more technical violations.

Figure 4: Effect of Reform on Technical Incarceration and Crime



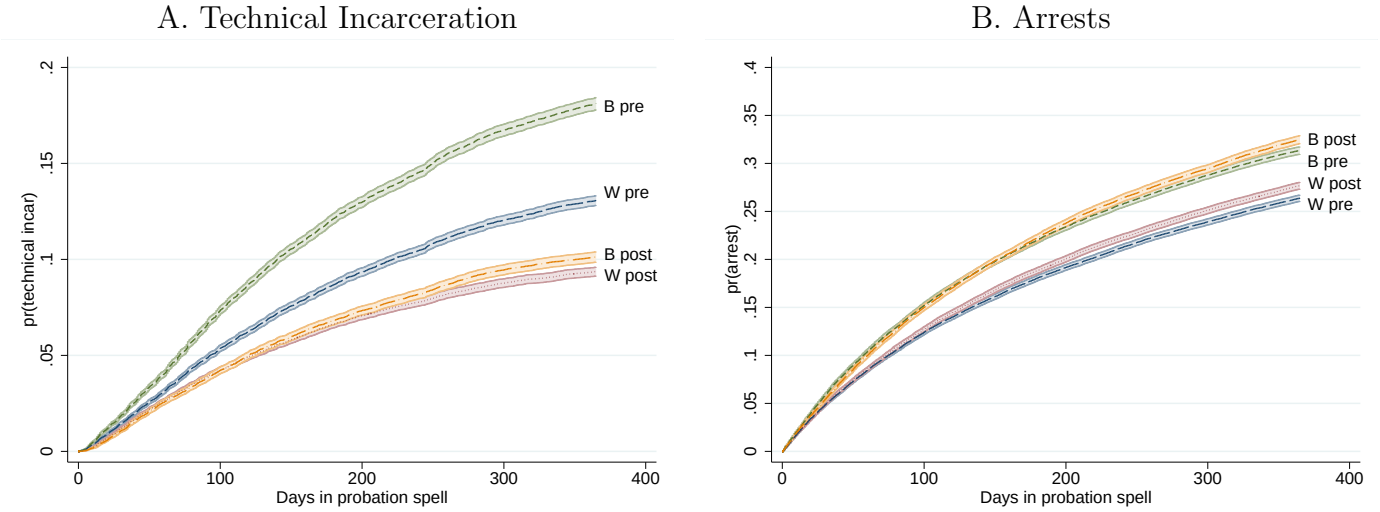
Notes: Panels A and B include all supervised probationers starting their spells within four years of the reform. Each line represents a three-month cohort of probationers who start their spells where the line intersects the x-axis. The y-axis measures the share of this cohort experiencing the relevant outcome over the following year. Technical incar is an indicator for having probation revoked for rule violations with no intervening criminal arrest. Arrest is an indicator for a criminal arrest before incarceration for any rule violations. Events are therefore mutually exclusive. Panel C plots mean one-year technical incarceration and arrest rates for supervised probationers minus the same measure for unsupervised probationers. The same cohort definitions are used. Effects are normalized relative to the cohort starting 4 quarters before the reform, indicated by the solid red line. The dotted red line indicates the first cohort whose first year of probation falls completely post-reform.

Figure 5: Predicted Offending Around Implementation of Reform



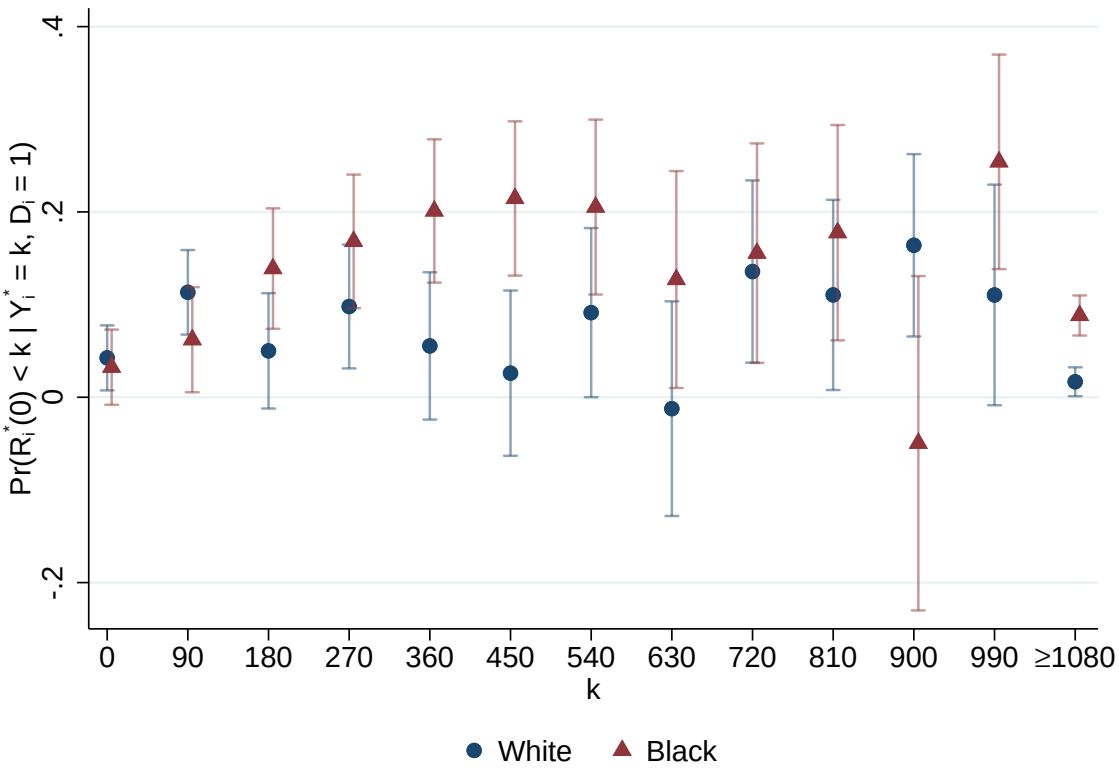
Notes: Includes all supervised probationers starting their spells within four years of the reform. Each line represents a three-month cohort of probationers who start their spells where the line intersects the x-axis. The y-axis measures the predicted share of this cohort arrested over time formed using linear regressions of arrest within t days on 5-year age bins interacted with race and gender, indicators for criminal history, and indicators for arrest offense. The regression is estimated for all $t \leq 365$ in the unsupervised (i.e., control group) probation population starting spells within 4 years of the reform. Treated (i.e., supervised) probationers' actual outcomes are reproduced in the light grey lines in the background.

Figure 6: Effects of Reform by Race



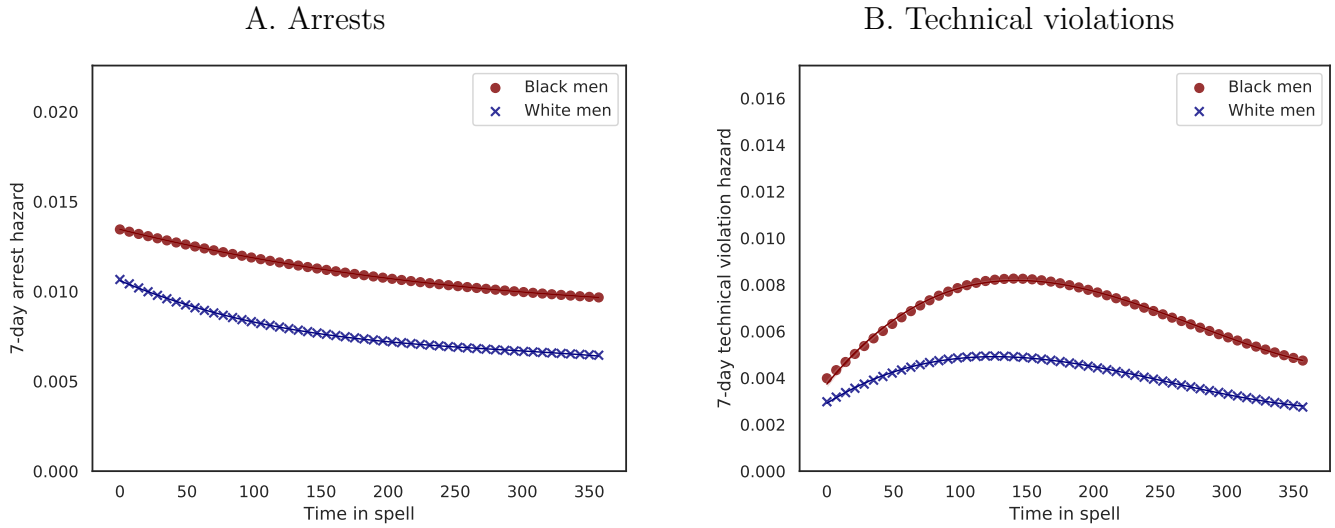
Notes: Includes all supervised probationers starting their spells either 1-3 years before (pre) or 0-2 years after the reform (post). The y-axis measures the share of each group experiencing the relevant outcome over the first year of their probation spell. Technical incar is an indicator for having probation revoked for rule violations with no intervening criminal arrest. Arrest is an indicator for a criminal arrest before incarceration for any rule violations.

Figure 7: Estimates of Targeting Bias in Drug and Administrative Violations



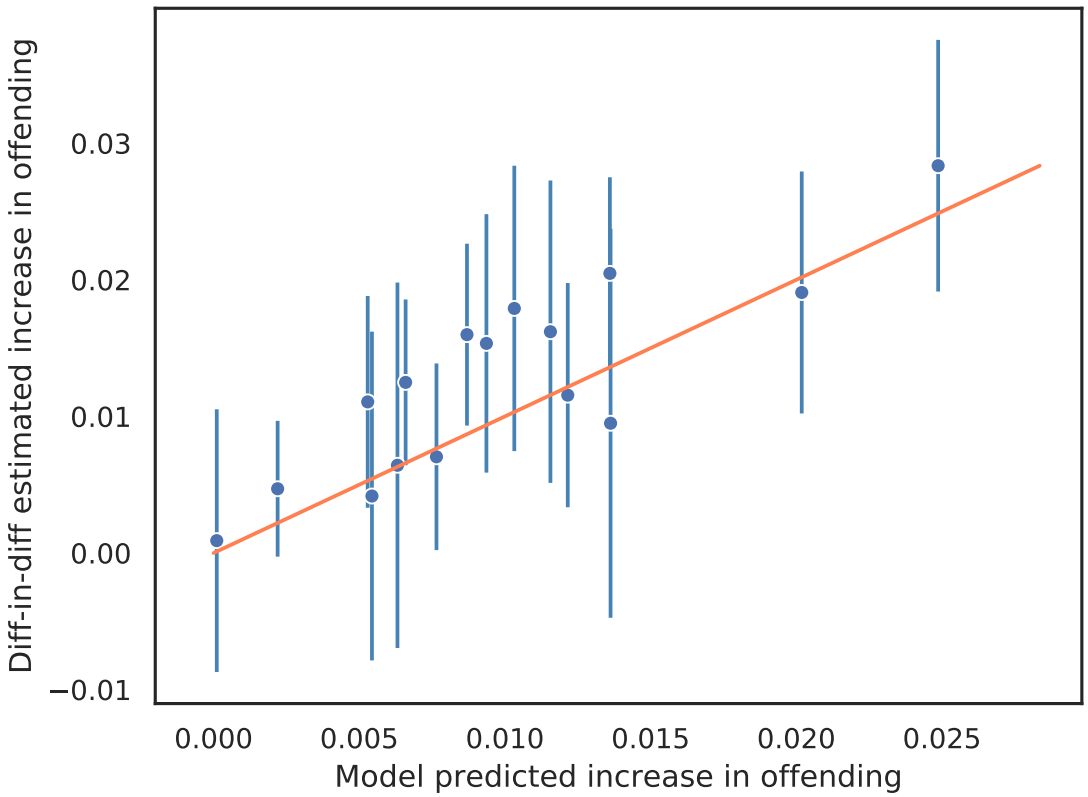
Notes: Figure plots estimates and 95% confidence intervals for Γ_k by race using the core diff-in-diff sample. Γ_k estimates the fraction of potential reoffenders at each horizon k who are incarcerated for technical rule violations before k. Higher values for black probationers indicate that among probationers who would otherwise be rearrested at the same time, technical rules target black probationers more aggressively. Γ_k is estimated using the ratio of coefficients from the core diff-in-diff specification in Table 4. The outcome for each k is Y_i^k, an indicator for being rearrested within $k, k+89$ days of probation start without any intervening technical incarceration. The numerator is the coefficient on post-x-treat. The denominator is the sum of coefficients on post-x-treat, treat, and the constant. The final estimate for $k \geq 1080$ is computed using 1 - an indicator for being rearrested within 1080 days as the outcome. Spells starting pre-reform with sentenced lengths that imply finishing post reform are dropped, since these spells are only partially affected.

Figure 8: Average Hazards for Arrest and Technical Incarceration



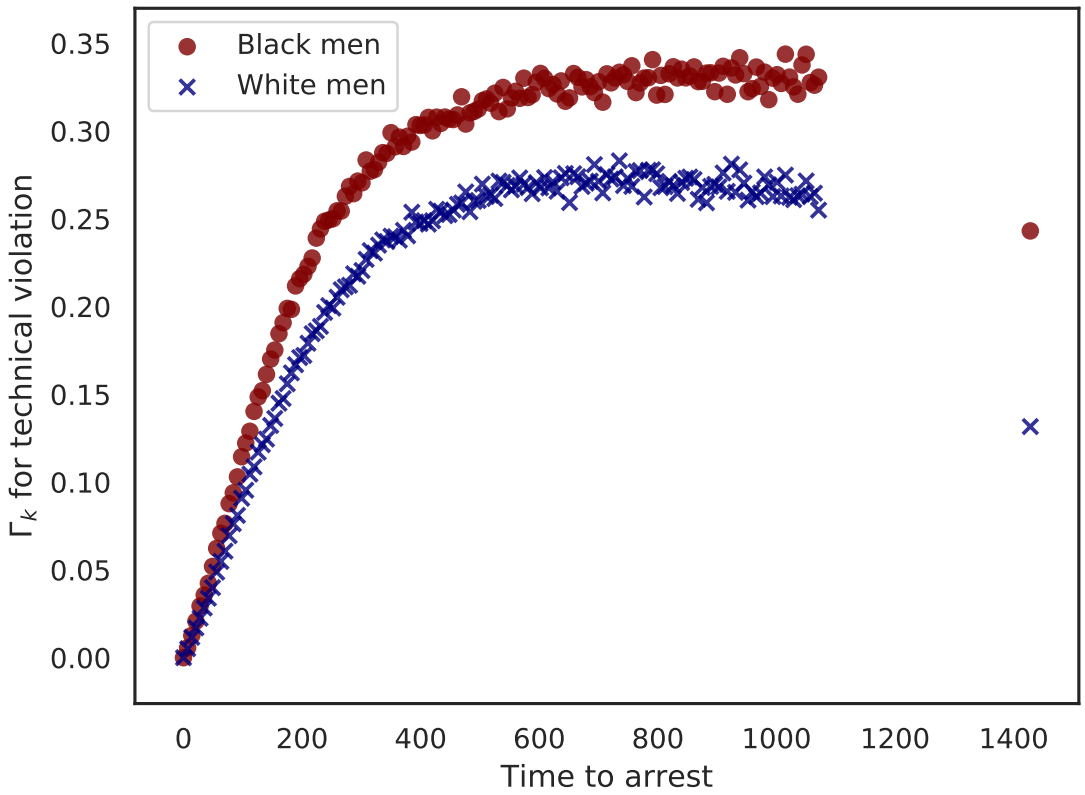
Notes: Figure plots mean latent cause-specific weekly hazard rates for each risk implied by estimates of the MMPH model. See text for details on sample and specification of unobserved heterogeneity used in estimation. Mean weakly hazards are similar but not identical to the baseline hazard, since the partial effects of unobserved heterogeneity on the hazard depend on baseline levels in the logit formulation.

Figure 9: Model-based Replication of Difference-in-Difference Estimates



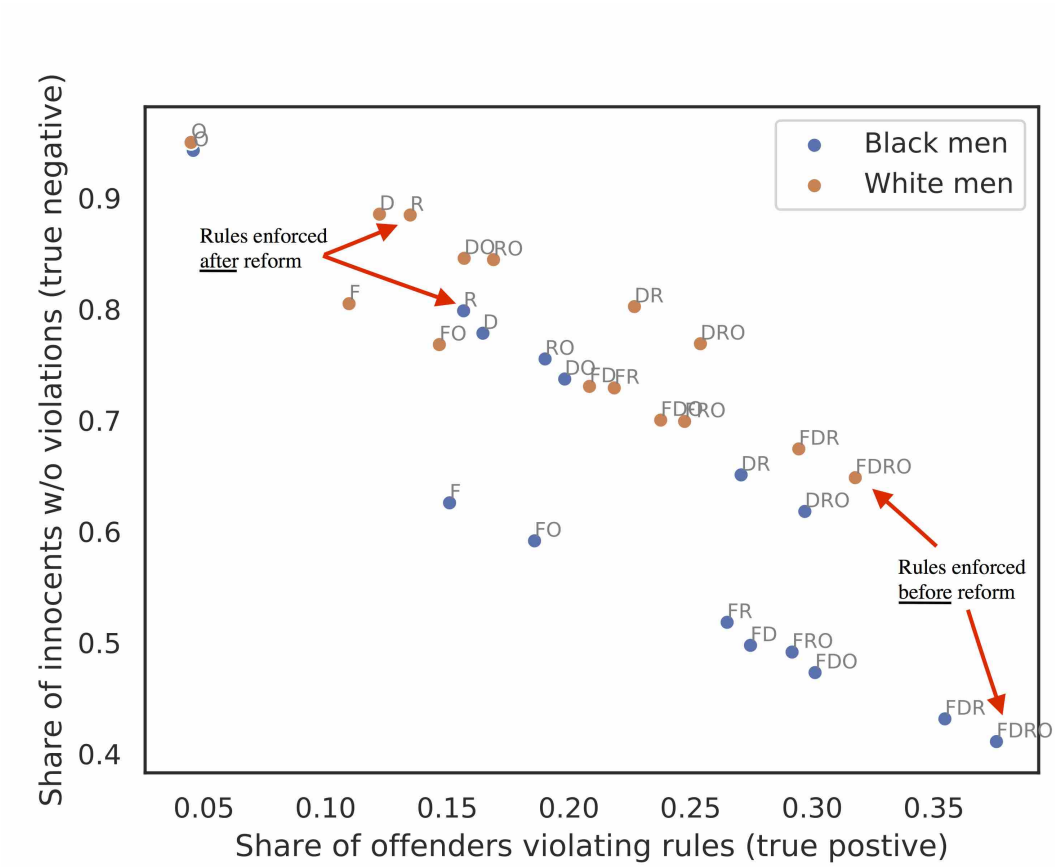
Notes: Figure compares difference-in-difference estimates of increase in observed arrests at 90, 180, 270, and 360 days for each race-by-gender group to the MMPH model's prediction of the same object. Vertical lines reflect 95% confidence intervals for the diff-in-dif estimates, while the orange line lies on a 45 degree angle. The diff-in-diff estimates are constructed using the sample sample and specification as in the reduced-form analysis and with no covariates included. Model predictions come from simulating observed arrests at each horizon with and without the "post" coefficients (Z_{ist}) on technical violation risk turned on.

Figure 10: Targeting Bias in the MMPH Model



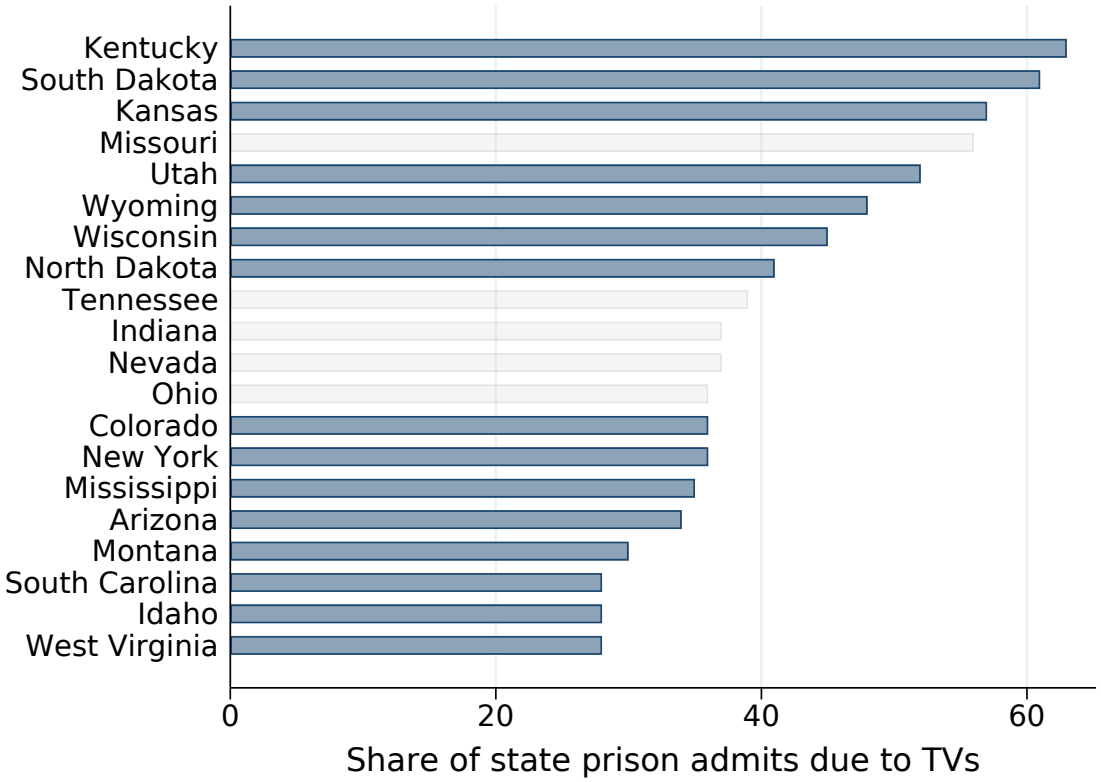
Notes: Figure plots estimates of Γ_k from simulating arrest and technical violation failure types 25 times using pre-reform coefficients from the MMPH model. Γ_k is the share of observations who have arrest failure times equal to k but technical violation times $< k$. Higher values for black probationers indicate that among probationers who would otherwise be rearrested at the same time, technical rules target black probationers more aggressively. The final dots at the right of the graph plot the probability of technical violation failure times ≤ 1080 conditional on having arrest failure times > 1080 (and possibly never).

Figure 11: Efficiency and Equity of Technical Violation Rule Types



Notes: Figure plots estimates of the share of potential reoffenders over a three year period who break technical rules before they reoffend (x-axis) against the share of non-reoffenders who do not break technical rules. Estimates come from simulating the model estimated in Section 5.5 using a different set of rules. Each point is labeled with a combination of "F" for fees / fines violations, "D" for drug / alcohol violations, "R" for reporting violations, and "O" for all other, reflecting the sets of rules enforced in the simulation.

Figure 12: Top States by Share of Prison Admissions Due to Technical Violations



Notes: Figure plots the share of state prison admissions due to technical violations of probation and parole using data from the Council of State Governments Justice Center (CSG, 2019) for the 20 states with the highest shares. States with blue bars have no statutory limits on which technical violations can result in prison time, while states with grey bars restrict incarceration for failure to pay fees and fines when the defendant can demonstrate a financial "hardship."

Tables

Table 1: Descriptive Statistics

	Supervised (treated)			Unsupervised (control)		
	Mean	Sd.	p50	Mean	Sd.	p50
Demographics:						
Age at start	32.059	10.85	29.83	32.707	10.77	30.29
Male	0.738	0.44	1.00	0.732	0.44	1.00
Black	0.435	0.50	0.00	0.355	0.48	0.00
White	0.490	0.50	0.00	0.522	0.50	1.00
Other race	0.074	0.26	0.00	0.124	0.33	0.00
Sentence:						
Sup. length (m)	19.449	9.58	18.17	14.841	8.77	12.00
Felon	0.429	0.49	0.00	0.032	0.18	0.00
Misd.	0.318	0.47	0.00	0.502	0.50	1.00
DWI / DWLR	0.208	0.41	0.00	0.457	0.50	0.00
Criminal history:						
Crim. hist. score	2.059	2.97	1.00	0.988	1.76	0.00
Prior sentences	1.917	3.28	0.00	1.251	2.69	0.00
Prior inc. spells	0.860	2.22	0.00	0.497	1.74	0.00
N	708623			895090		
Individuals	531099			661103		

Notes: Treated and control samples include all supervised and unsupervised probation spells beginning between 2006 and 2018, respectively. Felon, misdemeanor, and DWI / DWLR measure the most serious offense that resulted in the spell, with DWL / DWLR referring to driving while intoxicated and driving with license revoked. A small share of spells result from offenses with no classification. Criminal history score is a weighted sum of prior convictions used by North Carolina's sentencing guidelines. A prior misdemeanor conviction is typically worth 1 point, while a prior felony is worth two or more. Prior sentences refer to previous sentences to supervised probation or incarceration. Prior incarceration spells refers to previous incarceration in state prison.

Table 2: Frequency of Top 20 Probation Violations

	Violation	**Share of violations**	**Share of spells**
	Any violation	1.000	0.908
1	Not paying fees	0.282	0.802
2	Not reporting	0.134	0.462
3	Positive drug test	0.097	0.298
4	New misdemeanor charge	0.071	0.224
5	Treatment / program failure	0.060	0.252
6	Fleeing supervision	0.057	0.264
7	Moving / job change without notifying	0.038	0.136
8	Breaking curfew	0.036	0.105
9	Not completing community service	0.031	0.165
10	No employment	0.028	0.095
11	New felony charge	0.026	0.065
12	Admitting drug use	0.013	0.037
13	Possessing drugs	0.009	0.020
14	No education / training	0.008	0.030
15	Traveling without permission	0.007	0.022
16	Electronic monitoring failure	0.006	0.016
17	Contacting drug users	0.004	0.008
18	Refuse drug test	0.004	0.012
19	Possessing weapons	0.004	0.009
20	Disobeying curfew	0.004	0.012
	All others	0.081	0.285

Notes: Includes all treated observations starting probation in 2006-2010. Share of violations measures share of all violation recorded over this period. Share of spells measures the share of probation spells with any violation of the listed type.

Table 3: Behavioral Responses to Reform

	Arrest		Any violation		Drug use		Fees and fines	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post reform	-0.000972	0.00133	-0.0230*	-0.0180	0.0163	0.0225	-0.0000153	0.00582
	(0.0117)	(0.0117)	(0.0101)	(0.0101)	(0.0176)	(0.0176)	(0.0118)	(0.0118)
N	152734	152734	152734	152734	152734	152734	152734	152734
Controls		Yes		Yes		Yes		Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Table reports estimates of Cox proportional hazard regressions using all supervised probation spells starting within one year of the reform. "Post reform" is a time-varying indicator for dates after Dec. 1, 2011. Each pair of columns considers the listed behavior as failure and the other behaviors as a source of independent censoring. If rule breaking and arrests are unaffected by the reform's decrease in punishments for rule violations, then the populations at risk at each duration and measured hazards should also be unaffected. Controls include demographic and criminal history covariates where indicated. All spells are censored at 365 days.

Table 4: Difference-in-Differences Estimates of Reform Impacts

	Technical incarceration		Arrest	
	(1)	(2)	(3)	(4)
A. All offenders				
Post-reform	-0.00172***	-0.00203***	-0.00787***	-0.00699***
	(0.000274)	(0.000290)	(0.00167)	(0.00159)
Treated	0.147***	0.136***	0.0306***	-0.0156***
	(0.00105)	(0.00102)	(0.00166)	(0.00164)
Post-x-treat	-0.0546***	-0.0546***	0.0199***	0.0198***
	(0.00137)	(0.00136)	(0.00242)	(0.00233)
N	546006	546006	546006	546006
Pre-reform treated mean	.154	.154	.286	.286
Accuracy			.365 (.044)	.365 (.042)
False negative rate ($1 - \Gamma_1$)			.935 (.008)	.935 (.007)
False positive rate (Γ_0)			.058 (.004)	.058 (.004)
B. Non-black offenders				
Post-reform	-0.000522	-0.000867**	-0.00688***	-0.00661***
	(0.000317)	(0.000336)	(0.00199)	(0.00190)
Treated	0.126***	0.114***	0.0442***	-0.000306
	(0.00131)	(0.00127)	(0.00208)	(0.00207)
Post-x-treat	-0.0366***	-0.0371***	0.0201***	0.0182***
	(0.00175)	(0.00174)	(0.00304)	(0.00295)
N	328784	328784	328784	328784
Pre-reform treated mean	.131	.131	.264	.264
Accuracy			.549 (.083)	.543 (.079)
False negative rate ($1 - \Gamma_1$)			.929 (.01)	.93 (.01)
False positive (Γ_0)			.027 (.005)	.027 (.005)
C. Black offenders				
Post-reform	-0.00389***	-0.00411***	-0.0117***	-0.0111***
	(0.000509)	(0.000538)	(0.00295)	(0.00281)
Treated	0.172***	0.164***	-0.00603*	-0.0467***
	(0.00168)	(0.00168)	(0.00274)	(0.00268)
Post-x-treat	-0.0760***	-0.0756***	0.0232***	0.0237***
	(0.00217)	(0.00216)	(0.00399)	(0.00383)
N	217222	217222	217222	217222
Pre-reform treated mean	.181	.181	.314	.314
Accuracy			.305 (.052)	.306 (.049)
False negative rate ($1 - \Gamma_1$)			.931 (.011)	.931 (.011)
False positive rate (Γ_0)			.095 (.007)	.094 (.007)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Includes all treated and untreated probation spells beginning 1-3 years before the reform and 0-2 years afterwards. Post is indicator for starting probation after Dec. 1, 2011, the date JRA reforms took effect. Demographic controls include five-year age bins and indicators for race and gender. Criminal history controls include fixed effects criminal history points and prior sentences to supervised probation or incarceration.

Table 5: Decomposition of Racial Gaps in Technical Violations Using One-Period Model

	Overall rates		Decomposition	
	White	Black	Difference	Share of gap
Probability of T.V.				
$Pr(R_i = 1 \mid D_i = 1)$	0.040	0.085	0.045	100.0%
Distribution of risk				
$Pr(Y_i^* = 1 \mid D_i = 1)$	0.314	0.377	0.063	9.7%
$Pr(Y_i^* = 0 \mid D_i = 1)$	0.686	0.623	-0.063	-13.4%
Targeting				
$Pr(R_i = 1 \mid Y_i^* = 1, D_i = 1)$	0.071	0.069	-0.002	-1.3%
$Pr(R_i = 1 \mid Y_i^* = 0, D_i = 1)$	0.027	0.095	0.068	105.0%

Notes: Table decomposes the difference in technical violation risk between black and white probationers into the contributions of differences in arrest risk and differences in the likelihood of violation conditional on arrest risk. Estimates are based on core difference-in-differences results without controls from Table 4. The decomposition calculates the contribution of differences in risk using black targeting rates as baseline, and differences in targeting using white risk as baseline. The first row is -1 times the race-specific post-x-treat effect for technical violations. The second row is the sum of the constant, treat, and post-x-treat effects from difference-in-differences estimates for arrests. Both rows are re-scaled by 1 minus the sum of the constant, treat, and post-x-treat effects for technical violations, since this measures the size of the complier population. The final two rows are calculated as described in the text. Appendix Section A3 provides complete details on how the decomposition is calculated.

Table 6: Cost-Benefit Analysis of Reform

	(1)	(2)	(3)	(4)	(5)	(6)
	$\triangle$ in rev. \$	$\triangle$ indir. \$	Break-even	Break-even fel.	Cost lb	Cost ub
All	-676***	246*	39,813***	100,863**	23,512	195,295
	(26)	(118)	(10,079)	(31,183)	(36,126)	(109,304)
Non-black	-450***	213	24,991*	50,576*	2,114	47,363
	(34)	(128)	(10,343)	(22,161)	(39,639)	(120,331)
Black	-957***	296	50,037**	188,899	36,439	339,574
	(40)	(224)	(17,379)	(107,553)	(62,285)	(189,895)
Non-black men	-533***	197	31,863*	55,798*	-13,146	39,561
	(43)	(164)	(13,243)	(23,950)	(43,565)	(136,574)
Black men	-1,085***	376	44,156*	149,230	38,920	340,983
	(50)	(297)	(17,615)	(87,676)	(68,152)	(206,603)

Notes: Table calculates the minimum mean social costs of arrests necessary for the state to "break-even" on changes in incarceration costs and arrest rates induced by the reform. Includes all treated and untreated probation spells beginning 1-3 years before the reform and 0-2 years afterwards. Includes same controls as in Table 4.

Table 7: Decomposition of Racial Differences in Drug and Administrative Technical Incarceration

	Overall rates		Decomposition	
	White	Black	Difference	Share of gap
Probability of T.V.				
$Pr(R_i^* < Y_i^* \mid D_i = 1)$	0.045	0.100	0.056	100.0%
Distribution of risk				
$Pr(Y_i^* < 360 \mid D_i = 1)$	0.313	0.363	0.05	6.5%
$Pr(Y_i^* < 720 \mid D_i = 1)$	0.426	0.488	0.061	10.1%
$Pr(Y_i^* < 1080 \mid D_i = 1)$	0.498	0.558	0.060	11.0%
$Pr(Y_i^* \geq 1080 \mid D_i = 1)$	0.502	0.442	-0.060	-9.6%
Total contribution				1.5%
Targeting				
$Pr(R_i^* < Y_i^* \mid Y_i^* < 360, D_i = 1)$	0.070	0.077	0.007	4.5%
$Pr(R_i^* < Y_i^* \mid Y_i^* < 720, D_i = 1)$	0.063	0.106	0.043	34.6%
$Pr(R_i^* < Y_i^* \mid Y_i^* < 1080, D_i = 1)$	0.073	0.110	0.037	34.3%
$Pr(R_i^* < Y_i^* \mid Y_i^* \geq 1080, D_i = 1)$	0.017	0.088	0.072	64.3%
Total contribution				98.5%

Notes: Table decomposes the difference in technical violation risk between black and white probationers into the contributions of differences in arrest risk and differences in the likelihood of violation conditional on arrest risk using the multi-period model described in Section 3. The first row reports the share of white and black compliers caught by the drug and administrative rules affected by the reform and the black rate minus the white rate. The remainder of the table decomposes this differences into the share explained by targeting (differences in Γ_k) and risk (differences in $Pr(Y_i^* = k)$. The rows under "Distribution of Risk" show the share of compliers by race with Y_i^* falling in certain ranges, the black-white gap, and the contribution of this gap to the total disparity. The rows under "Targeting" show mean values of Γ_k for compliers with Y_i^* in certain ranges (weighted by the distribution of Y_i^*), the gap, and the contribution of this gap to the total disparity. Since crime is measured up to a max of a 3 year horizon, risk distributions are not observed beyond this point. Y_i^* is therefore binned in 90-day intervals up to 3 years with a final bin reflecting 3 years or later. Additional details are available in Section A3.

Table 8: MMPH Model Parameter Estimates

	Black men				White men			
	Arrest		Tech. Violation		Arrest		Tech. Viol.	
	β	se	β	se	β	se	β	se
Duration	-0.44	(0.09)	3.95	(0.15)	-1.08	(0.09)	3.05	(0.15)
Duration^2	0.28	(0.60)	-22.44	(1.06)	3.34	(0.59)	-19.33	(1.14)
Duration^3	0.74	(1.55)	44.97	(2.96)	-5.72	(1.54)	40.46	(3.25)
Duration^4	-1.32	(1.71)	-41.99	(3.49)	4.93	(1.71)	-39.17	(3.87)
Duration^5	0.57	(0.68)	14.82	(1.45)	-1.68	(0.68)	14.22	(1.62)
Calendar time	-0.10	(0.01)	-0.28	(0.02)	-0.03	(0.01)	-0.10	(0.02)
Calendar time^2	0.01	(0.01)	-0.16	(0.01)	0.04	(0.01)	-0.09	(0.01)
Type 1	-6.30	(0.06)	-7.07	(0.11)	-6.80	(0.07)	-8.12	(0.09)
Type 2	-5.27	(0.24)	-4.71	(0.11)	-4.83	(0.06)	-6.66	(0.11)
Type 3	-4.41	(0.06)	-6.20	(0.08)	-4.83	(0.06)	-4.94	(0.08)
Type 4	-3.31	(0.06)	-4.80	(0.11)	-3.59	(0.07)	-5.44	(0.08)
Type 1 x post			-0.98	(0.10)			-0.63	(0.12)
Type 2 x post	↕ 0.041	↕ (0.02)	-0.34	(0.17)	↕ 0.016	↕ (0.01)	-0.85	(0.12)
Type 3 x post			-0.67	(0.09)			-0.01	(0.09)
Type 4 x post			-0.04	(0.09)			-0.22	(0.13)
Pr(type 1)	0.406		(0.02)		0.449		(0.02)	
Pr(type 2)	0.097		(0.01)		0.356		(0.02)	
Pr(type 3)	0.369		(0.01)		0.095		(0.01)	
Pr(type 4)	0.128		(0.01)		0.1		(0.01)	
	Black women				White women			
Duration	-0.67	(0.19)	3.88	(0.34)	-1.06	(0.16)	2.74	(0.26)
Duration^2	1.34	(1.28)	-21.87	(2.46)	2.25	(1.03)	-20.61	(2.00)
Duration^3	-0.72	(3.26)	42.48	(6.85)	-2.13	(2.62)	46.59	(5.77)
Duration^4	-0.73	(3.56)	-38.88	(8.09)	0.56	(2.87)	-47.16	(6.94)
Duration^5	0.63	(1.39)	13.58	(3.38)	0.16	(1.13)	17.49	(2.94)
Calendar time	-0.08	(0.02)	-0.21	(0.05)	-0.03	(0.02)	-0.01	(0.03)
Calendar time^2	0.01	(0.01)	-0.10	(0.03)	0.02	(0.01)	0.02	(0.02)
Type 1	-6.51	(0.07)	-4.56	(0.19)	-6.54	(0.07)	-5.36	(0.43)
Type 2	-6.51	(0.07)	-7.68	(0.08)	-6.54	(0.07)	-8.13	(0.22)
Type 3	-4.49	(0.07)	-6.40	(0.08)	-4.53	(0.07)	-6.49	(0.08)
Type 4	-3.22	(0.14)	-4.37	(0.26)	-3.75	(0.18)	-4.48	(0.17)
Type 1 x post			-0.43	(0.25)			-0.41	(0.33)
Type 2 x post	↕ 0.011	↕ (0.03)	-0.95	(0.16)	↕ 0.044	↕ (0.02)	-0.69	(0.20)
Type 3 x post			-0.52	(0.12)			-0.39	(0.09)
Type 4 x post			-0.17	(0.26)			0.03	(0.14)
Pr(type 1)	0.036		(0.01)		0.041		(0.02)	
Pr(type 2)	0.601		(0.02)		0.565		(0.02)	
Pr(type 3)	0.327		(0.02)		0.352		(0.02)	
Pr(type 4)	0.037		(0.01)		0.042		(0.01)	

Notes: Table reports estimates of the MMPH model described in Section 5. Duration and calendar time are standardized (mean 0 and s.d. 1) in estimation. Standard errors are the robust "sandwich form" clustered by individual.

Appendix

A1 Proof of bias test derivation

The numerator of Γ_k is derived as follows:

$$E[1\{R_i^* \geq k\}1\{Y_i^* = k\}|Z_i = 1] - E[1\{R_i^* \geq k\}1\{Y_i^* = k\}|Z_i = 0] \tag{9}$$
$$= E[1\{R_i^*(1) \geq k\}1\{Y_i^* = k\} - 1\{R_i^*(0) \geq k\}1\{Y_i^* = k\}] \tag{10}$$
$$= Pr(Y_i^* = k, R_i^*(0) < k \leq R_i^*(1)) \tag{11}$$

The denominator is:

$$E[1\{R_i^* \geq k\}1\{Y_i^* = k\}|Z_i = 1] = Pr(Y_i^* = k, k \leq R_i^*(1)) \tag{12}$$

Taking the ratio of these two objections coverts the joint probability to the desired conditional probability.

$$\frac{Pr(Y_i^* = k, R_i^*(0) < k \leq R_i^*(1))}{Pr(Y_i^* = k, k \leq R_i^*(1))} = Pr(R_i^*(0) < k|Y_i^* = k, D_i = 1) \tag{13}$$

The notation for $D_i = 1$ is equivalent to writing $Pr(R_i^*(0) < k|Y_i^* = k, R_i^*(1) \geq k)$.

A2 Additive time effects

Since the instrument Z_i is a simple indicator for beginning probation pre/post reform, time effects are a form of violation of the exclusions restriction, which requires that $Y_i^* \perp\!\!\!\perp Z_i$. This violation can be accounted for if the control group provides a good measure of the effect of Z_i on Y_i^* in the compliers group, so that it can be differenced off.

Let S_i be a binary indicator for whether the individual is on supervised vs. unsupervised probation and thus is in the treated vs. control group, respectively. Let the population shares with offending durations k be given by:

$$Pr(Y_i^* = k|Z_i, S_i) = \alpha_k + \beta_k^1 Z_i + \beta_k^2 S_i \tag{14}$$
$$\tag{15}$$

Observed offending rates in the post period (i.e., the Y_i^k used in estimation of Γ_k) be given by:

$$Pr(Y_i^* = k, R_i^*(1) \geq k|Z_i, S_i) = Pr(Y_i^* = k|R_i^*(1) \geq k, Z_i, S_i)Pr(R_i^*(1) \geq k|Z_i, S_i) \tag{16}$$
$$= \left(\alpha_k + \beta_k^1 Z_i + \beta_k^2 S_i\right) Pr(R_i^*(1) \geq k|Z_i, S_i) \tag{17}$$

Observed offending rates in the pre period are given by:

$$Pr(Y_i^* = k, R_i^*(0) \geq k|Z_i = 0, S_i) = \left(\alpha_k + \beta_k^2 S_i\right) Pr(R_i^*(0) \geq k|Z_i, S_i) \tag{18}$$

Because the control group is virtually never subject to technical incarceration, both $Pr(R_i^*(1) \geq k|Z_i, S_i)$ and $Pr(R_i^*(0) \geq k|Z_i, S_i)$ are 1 when $S_i = 0$. Taking the difference-in-difference between these two probabilities and across S_i thus yields:

$$Pr(Y_i^* = k, R_i^*(1) \geq k|Z_i = 1, S_i = 1) - Pr(Y_i^* = k, R_i^*(1) \geq k|Z_i = 1, S_i = 0) \tag{19}$$

$$- Pr(Y_i^* = k, R_i^*(0) \geq k|Z_i = 0, S_i = 1) - Pr(Y_i^* = k, R_i^*(0) \geq k|Z_i = 0, S_i = 0) \tag{20}$$

$$= \left(\alpha_k + \beta_k^1 + \beta_k^2\right) Pr(R_i^*(1) \geq k|S_i = 1) - \alpha_k - \beta_k^1 \tag{21}$$

$$- \left(\alpha_k + \beta_k^2\right) Pr(R_i^*(0) \geq k|S_i = 1) + \alpha_k \tag{22}$$

$$= \left(\alpha_k + \beta_k^2\right) (Pr(R_i^*(1) \geq k|S_i = 1) - Pr(R_i^*(0) \geq k|S_i = 1)) + \beta_k^1(Pr(R_i^*(1) \geq k|S_i = 1) - 1) \tag{23}$$

$$= \left(\alpha_k + \beta_k^2\right) Pr(R_i^*(0) < k \leq R_i^*(1)|S_i = 1) + \beta_k^1(Pr(R_i^*(1) \geq k|S_i = 1) - 1) \tag{24}$$

$$= Pr(Y_i^* = k, R_i^*(0) < k \leq R_i^*(1)|S_i = 1) + \beta_k^1(Pr(R_i^*(1) \geq k|S_i = 1) - 1) \tag{25}$$

Thus the difference-in-differences estimator yields the correct probability plus a bias term. This term reflects the fact that although Z has the same effect on the $Pr(Y_i^* = k)$ for both treatment and control units, the effect is partially muted in the treatment group by the fact that $Pr(R_i^*(1) \geq k|S_i = 1) < 1$, so that only a portion of the effect of Z is revealed, whereas the full effect is revealed in the control group. This bias term is decreasing in $Pr(R_i^*(1) \geq k|S_i = 1)$. Empirically, this value is roughly 0.9 at one-year horizons. Thus practically speaking this size of any bias is roughly 10% of the estimated post-effect, which is very small as well.

A3 Calculation of Oaxaca decomposition

I use the primary results from Table 4 to construct the one-period Oaxaca decomposition. The first row, which reports $Pr(R_i = 1|D_i = 1)$ by race is -1 times the coefficient on post-x-treat, which is an estimate of $Pr(R_i(0) = 1, D_i = 1)$, rescaled by the probability of being a complier, or $Pr(D_i = 1)$. This probability is easily estimated as one minus the share of individuals incarcerated for technical violations in the first year of their spell in the post period. That is, the sum of the constant, the treated indicator, and the post-x-treat indicator from Column 1.

The second row reports estimates of $Pr(R_i = 1|D_i = 1)$. This object is estimated as the probability of offending within the first year of a probation spell after the reform, or the sum of the constant, the treated indicator, and the post-x-treat indicator from Column 3, again re-scaled by the estimate of $Pr(D_i = 1)$. The third row is 1 minus the second row.

The fourth row is simply the re-scaled reduced form discussed in Section 3. It is the coefficient on treat-x-post from Column 3 divided by the sum of the coefficients on post-x-treat, treat, and the constant from Column 3.

The fifth row is estimated by first subtracting the coefficient on post-x-treat in Column 3 from -1 times the coefficient on post-x-treat from Column 1. This object reflects $Pr(R_i(0) = 1, Y_i^* = 0, D_i = 1)$. Rescaling by complier probability converts to $Pr(R_i(0) = 1, Y_i^* = 0|D_i = 1)$. I then divide by 1 minus the sum of coefficients on post-x-treat, treat, and the constant from Column 3 divided by the complier probability. This estimates $Pr(Y_i^* = 0|D_i = 1)$. The ratio gives the desired

object, $Pr(R_i = 1|Y_i^* = 0, D_i = 1)$.

The Oaxaca decomposes the differences in $Pr(R_i = 1|D_i = 1)$ as described in Section 3, but with all objects conditioning on $D_i = 1$.

Calculation of the multi-period Oaxaca is analogous. The estimate of $Pr(R_i^* < Y_i^*|D_i = 1)$ is the post-x-treat effect on ever being imprisoned for technical violations. The complier probability is 1 minus the probability of any imprisonment for technical violations in the post period. Risk distributions are given by diff-in-diff estimates of increases in offending in each 90-day time bin, rescaled by complier probabilities. Targeting is estimated as discussed on Section 3.

Since outcomes are only observed for 3 years, share of compliers with $Y_i^* \geq 1080$ is simply 1 minus the sum of complier shares with $Y_i^* < 1080$. Targeting for this population is calculated as in the one-period version, but treating the first three years of a spell as single period.

Table 7 performs the same Oaxaca decomposition, but summing over all k (instead of the binary indicator). The targeting parameters reports are averages over the relevant time bins, weighted by estimated distributions of risk.

A4 MMPH model likelihood

The failure hazard for cause j at duration t for individual i in spell s and with type w is given by:

$$Pr(T_{is}^j = t|T_{is}^j \geq t \mid type_i = w) = \frac{1}{1 + exp(-\psi_j(t) - v_w^j)} \tag{26}$$

The hazards for each cause j are independent conditional on type, so this is also the hazard conditional on survival due to *all* causes: $Pr(T_{is}^j = t|T_i^k \geq t \ \forall \ k|type_i = w)$.

Let T_{is} denote the failure time in spell s and $C_{is} \in \{1, \ldots, J\}$ indicate the cause. Then the type-conditional failure of at time t due to cause j is given by:

$$\begin{aligned} Pr(T_{is} = t, C_{is} = j|type_i = w) = & \frac{1}{1 + exp(-\psi_l(t) - -v_w^j)} \\ & \prod_{l \neq j} \left(1 - \frac{1}{1 + exp(-\psi_l(t) - v_w^l)}\right) \\ & \prod_{k=0}^{T_s - 1} \prod_{l=1}^{J} \left(1 - \frac{1}{1 + exp(-\psi_l(k) - v_w^l)}\right) \end{aligned} \tag{27}$$

which is simply the probability of survival across all causes up to $t - 1$, then failure due to cause j. Ties are possible in theory, but do not occur in practice. The likelihood for censored spells contains only the product in the third line, since failure times are not observed.

The conditional individual likelihood is the product of the spell-specific failure probabilities across spells. Let $T_i = \{T_{i1}, \ldots, T_{iS}\}$ and $C_i = \{C_{i1}, \ldots, C_{iS}\}$ collect the vector of failure times and causes across spells. Then:

$$Pr(T_i, C_i \mid type_i = w) = \prod_{s=1}^{S} Pr(T_{is}, C_{is}|type_i = w) \tag{28}$$

Finally, the full, unconditional likelihood simply sums over the unobserved types w:

$$Pr(T_i, C_i) = \sum_w \pi_w Pr(T_{is}, C_{is} \mid type_i = w) \tag{29}$$

where π_w is the population share of type w.

A5 Identification and Estimation of the MMPH

Consider the single-cause MPH model given by:

$$Pr(T_{is} = t | T_{is} \geq t, v_i) = \theta(t) v_i \tag{30}$$

Using the relationship between survival and the integrated hazard, in this case the joint distribution of survival times across two spells is given by:

$$\begin{aligned} S(s_1, s_2) &= \int exp\left(-\int_0^{s_1} \theta(u) v du - \int_0^{s_2} \theta(u) v du\right) dG(v) \\ &= \int exp\left(-v(Z(s_1) + Z(s_2)\right) dG(v) \end{aligned} \tag{31}$$

where $G(v)$ denotes the distribution of unobserved heterogeneity v_i in the population of interest and $Z(t)$ is the integrated baseline hazard from zero to t. Identification results for this model are neatly summarized in Van Den Berg (2001). Honoré (1993) provides an intuitive proof by showing that the ratio of derivatives with respect the first and second argument of the survivor function yields the baseline hazards (up to a normalizing constant). With the hazards in hand, cumulative hazards can be found via integration, which yields G due to the uniqueness of the Laplace transform.

In practice, I estimate the multiple-cause, discrete-time version of Equation 30 and approximate G with discrete mass points. With discrete failure time data, it is not possible to estimate all features of Equation 30. For example, if failures are coded at the daily level, it is clearly not possible to identify the within-day variation in baseline hazards using the arguments from Honoré (1993).

To see what can be estimated in the double-discrete setting, consider the same model where $t \in \{1, 2, 3, \dots\}$ (e.g., days) and $v_i \in \{v_1, \dots, v_K\}$ (i.e., is one one of K discrete types). In this case, the joint distribution of survival times across spells is given by:

$$S(s_1, s_2) = \sum_{k=1}^{K} \pi_k \prod_{t=1}^{s_1} (1 - \theta(t) v_k) \prod_{u=1}^{s_2} (1 - \theta(u) v_k) \tag{32}$$

where π_k is the population share of type k. In this case, it is straightforward to show that the the baseline discrete-time hazards $\theta(t)$ can be recovered (up to normalization) using the relation:

$$\frac{S(s_1, s_2 - 1) - S(s_1 - 1, s_2 - 1)}{S(s_1 - 1, s_2) - S(s_1 - 1, s_2 - 1)} = \frac{\theta(s_1)}{\theta(s_2)} \tag{33}$$

which follows from the fact that $S(s_1, s_2 - 1) = \sum_{k=1}^{K} \pi_k (1 - \theta(s_1) v_k) \prod_{t=1}^{s_1 - 1} (1 - \theta(t) v_k) \prod_{u=1}^{s_2 - 1} (1 - \theta(u) v_k)$.

Note that this also suggests one test of the MPH model, since it must be that for any s_1, s_1'

$$\frac{(S(s_1, s_2 - 1) - S(s_1 - 1, s_2 - 1)) \,/\, (S(s_1 - 1, s_2) - S(s_1 - 1, s_2 - 1))}{(S(s_1', s_2 - 1) - S(s_1' - 1, s_2 - 1)) \,/\, (S(s_1' - 1, s_2) - S(s_1' - 1, s_2 - 1))} = \frac{\theta(s_1)}{\theta(s_1')} \tag{34}$$

does not depend on s_2. As I show below, however, by allowing unobserved heterogeneity v_k to vary across durations this restriction can be relaxed.

Baseline hazards must rationalize observed survival times across spells given in Equation 38 (and implicitly marginal distributions as well). These survival times form a set of polynomial equations that pin down v_k and π_k. The quantity of periods over which survival times are observed thus also determines the estimable number of mass points and type shares. For K types, for example, there are $2K - 1$ parameters to estimate in addition to the baseline hazards. Hence observing just a single period, for example, would not be sufficient to estimate a model with three types, since there would be just two equations. Indeed, the maximum degree in these polynomial equations is $S_1 + S_2$, where S_1 and S_2 are the periods after which spells are censored in each of the two repeated spells.

This model can be relaxed slightly by allowing $v_k(t) = v_{kl(t)}$ to be piecewise linear in intervals of t instead of constant across all duration. In this case, the complete distribution of survival times can be broken up into the conditional survival time distributions within each interval. By the same argument just made, clearly both baseline hazards and the interval-specific mass points v_{kl} are identified whenever each interval contains a sufficient number of periods for the given number of types.

More specifically, divide durations into L intervals segmented by $\{t_1, t_2, \ldots, t_L\}$. That is, let the lth interval span $\{t_{l-1}, \ldots, t_l - 1\}$ with $t_0 = 0$. The joint distribution of survival time is now given by:

$$S(s_1, s_2) = \sum_{k=1}^{K} \pi_k \prod_{t=1}^{s_1} (1 - \theta(t) v_{kl(t)}) \prod_{u=1}^{s_2} (1 - \theta(u) v_{kl(u)}) \tag{35}$$

Using identical arguments to the above for any s_1, s_2 in the same interval, baseline hazards can be identified for each interval up to normalization. As before, survival times yield a system of polynomial equations. Now, however, there are $(L + 1)K - 1$ type parameters to estimate. These parameters are identified whenever the non-linear system allows it.

In the limit, when $L = S_1 = S_2$, so that unobserved heterogeneity varies arbitrarily across each discrete period, the model is still set identified. To see this by way of example, consider failure times s_1, s_2 across two spells with just one period observed:

$$S(1, 0) = S(0, 1) = \sum_{k=1}^{K} \pi_k (1 - \theta(1) v_{k1}) \tag{36}$$

$$S(1, 1) = \sum_{k=1}^{K} \pi_k (1 - \theta(1) v_{k1})^2 \tag{37}$$

This system is underdetermined. For example, when $K = 2$ there are three unknowns π_1, v_{11}, and v_{21} ($\theta(1)$ can be normalized to 1) but only two linearly independent equations. Nevertheless,

bounds on the type-specific hazards are easily derived by considering extremes and assigning an ordering without loss of generality to types, i.e., $v_{11} \geq v_{21}$. When $v_{21} = 0$, for example, it must be that $v_{11} = 1 - \frac{S(1,1)}{S(1,0)}$, which is therefore an upper bound for v_{11}, and $\pi_1 = \frac{S(1,0)^2}{S(1,1)}$. Analogous bounds can be found by setting $v_{11} = 1$. These type-specific discrete-time hazards are thus set-identified (and possibly weakly).

A5.1 Competing risks

The preceding discussion focused on the single-cause discrete-time MPH model. Abbring and Van Den Berg (2003) extends the results on multi-spell identification from Honoré (1993) to the competing risk setting. Highly similar arguments allow for estimation of both baseline hazards and unobserved heterogeneity components in the discrete MMPH model as well. For two competing causes, for example, note that:

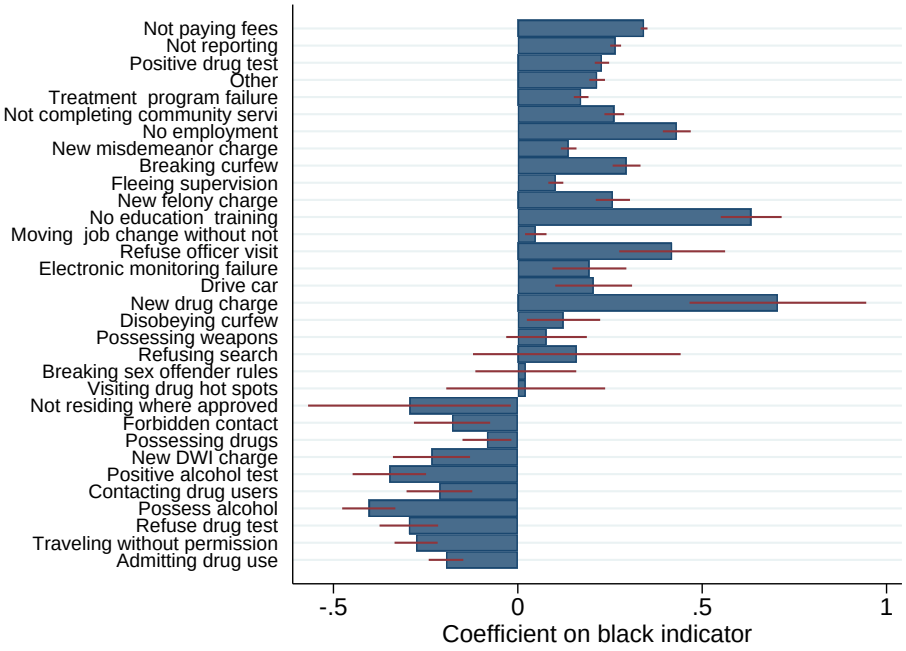
$$S(s_1^1, s_1^2, s_2^1, s_2^2) = \sum_{k=1}^{K} \pi_k \prod_{t=1}^{s_1^1} (1 - \theta^1(t) v_k^1) \prod_{u=1}^{s_1^2} (1 - \theta^2(u) v_k^2) \prod_{j=1}^{s_2^1} (1 - \theta^1(j) v_k) \prod_{m=1}^{s_2^2} (1 - \theta^2(m) v_k^2) \tag{38}$$

where s_t^j denotes survival times for cause j in spell t, $\theta^j(t)$ is the cause j-specific baseline discrete hazard, and v_k^j is type k's heterogeneity component for cause j. It can be shown just as before that:

$$\frac{S(s_1^1, s_1^2 - 1, s_2^1 - 1, s_2^2 - 1) - S(s_1^1 - 1, s_1^2 - 1, s_2^1 - 1, s_2^2 - 1)}{S(s_1^1 - 1, s_1^2 - 1, s_2^1, s_2^2 - 1) - S(s_1^1 - 1, s_1^2 - 1, s_2^1 - 1, s_2^2 - 1)} = \frac{\theta^1(s_1)}{\theta^1(s_2)} \tag{39}$$

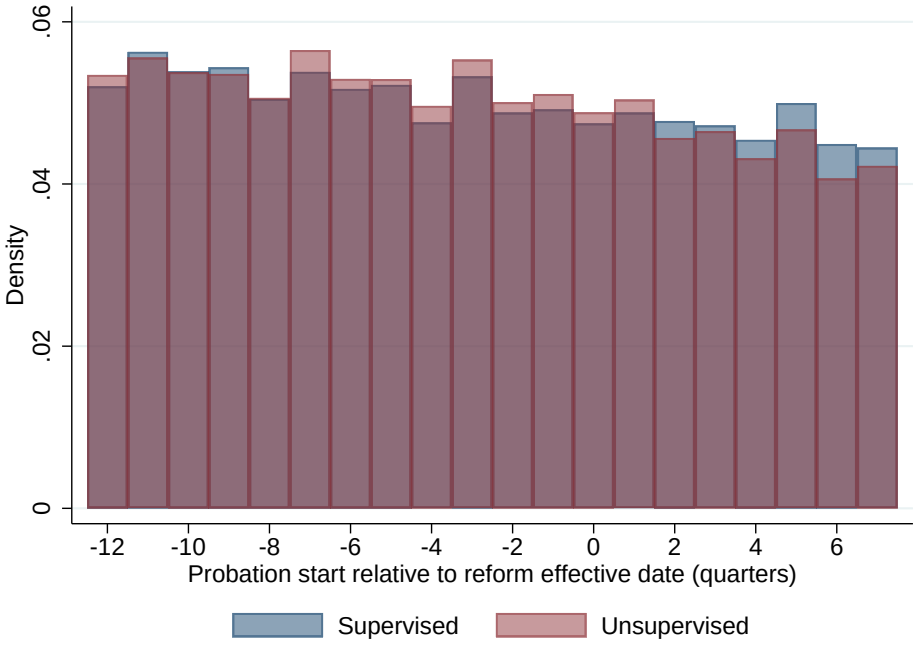
Hence baseline hazards and heterogeneity components can be estimated exactly as before. Note, however, that more periods are required to accommodate the additional heterogeneity terms.

Figure A1: Black Effects by Detailed Violation Type



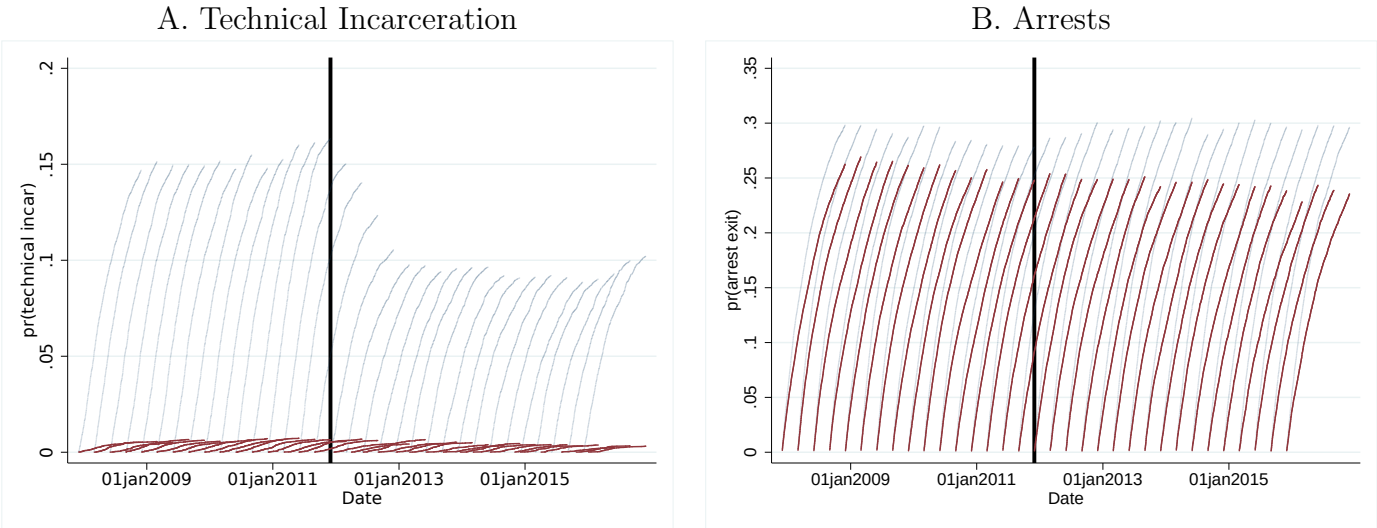
Notes: Sample and specification are the same as in Column 5 of Table A2, except the black coefficient is divided by the white mean of the dependent variable.

Figure A2: Sample Densities Around Reform



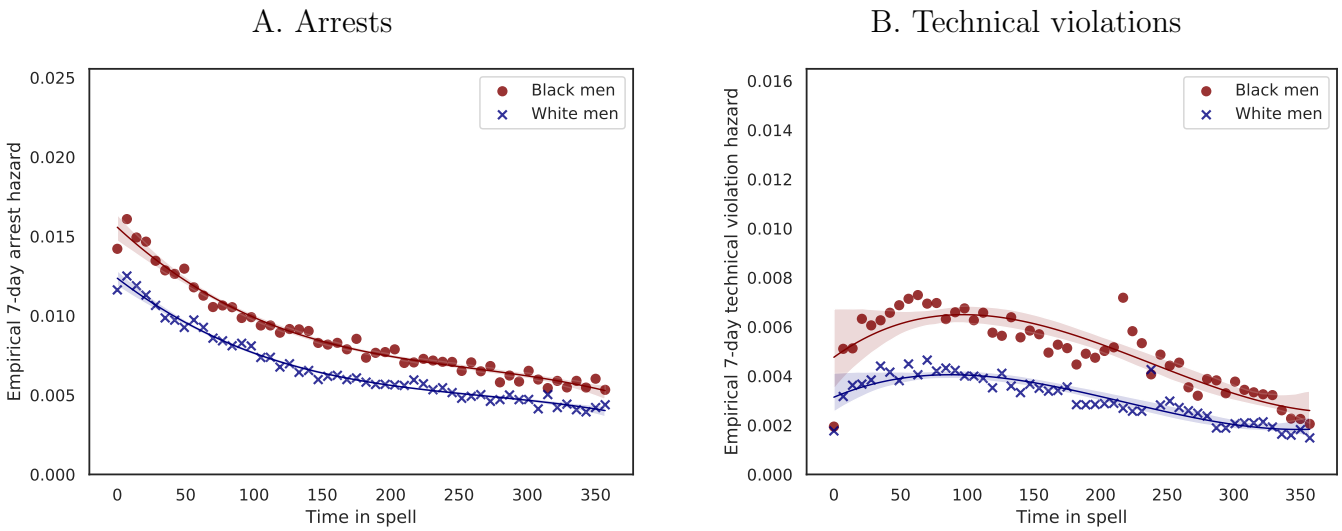
Notes: Figure plots the share of treated and untreated units in each quarter before and after the 2011 reforms for the core difference-in-differences estimates.

Figure A3: Effect of Reform on Unsupervised Probationers' Technical Incarceration and Crime



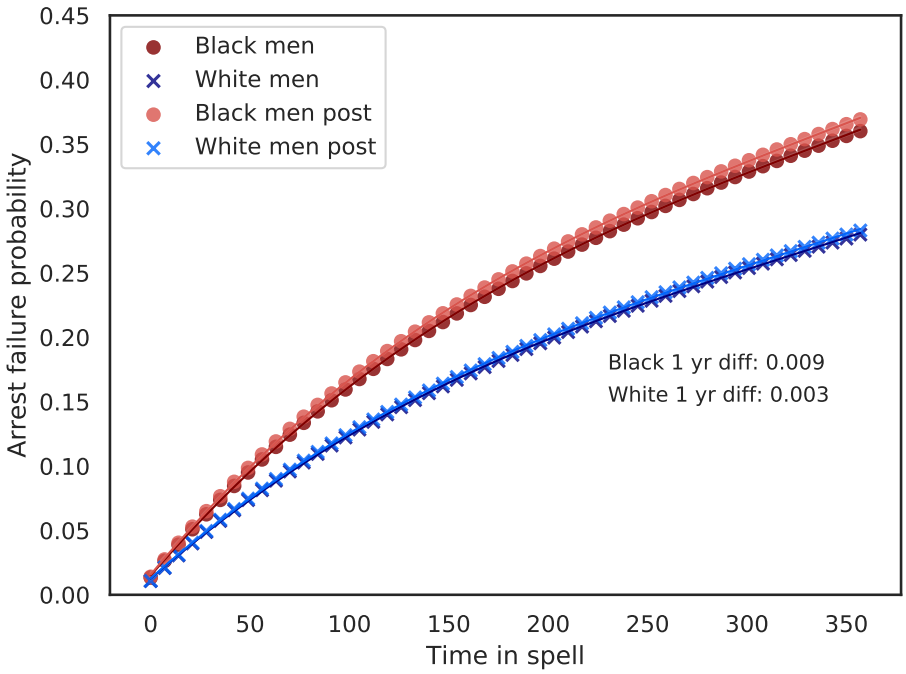
Notes: Includes all unsupervised probationers starting their spells over the included time window. Each line represents a three-month cohort of probationers starting their spells where the line intersects the x-axis. The y-axis measures the share of this cohort experiencing the relevant outcome over the following year. Technical incar is an indicator for having probation revoked with no intervening arrest. Arrest is an indicator for being arrested before being revoked. Treated (i.e., supervised) probationers' outcomes are reproduced in the light grey lines in the background.

Figure A4: Kaplan-Meier Empirical Hazards for Arrest and Technical Incarceration



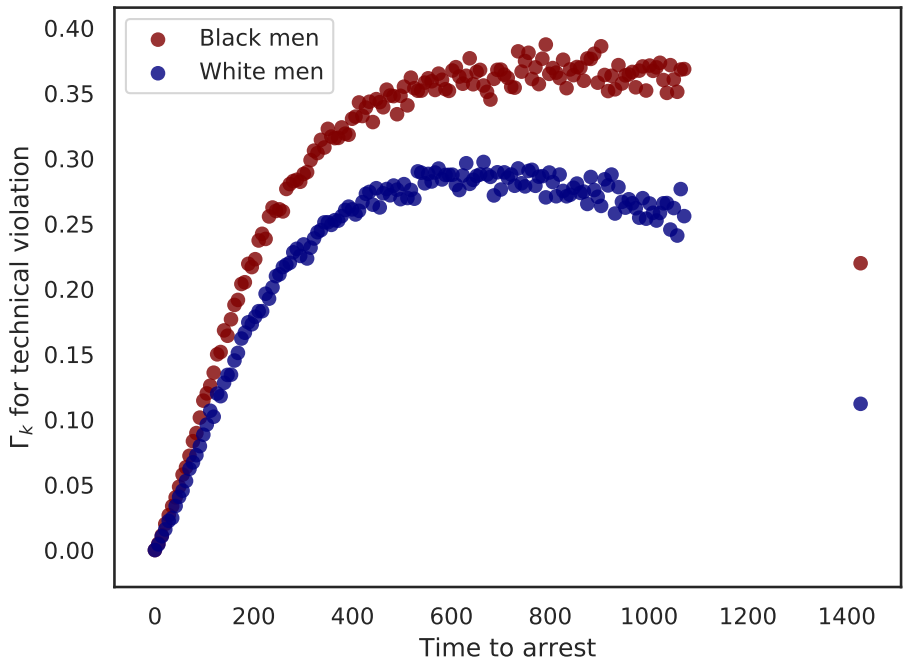
Notes: Figure plots Kaplan-Meier estimates of the cause-specific hazard using pre-reform data only. The Kaplan-Meier estimator in this context is simply the weekly probability of failing due to a given cause conditional on not failing due to *any* cause previously.

Figure A5: Model Estimates of Arrest Failure Functions Before / After Reform



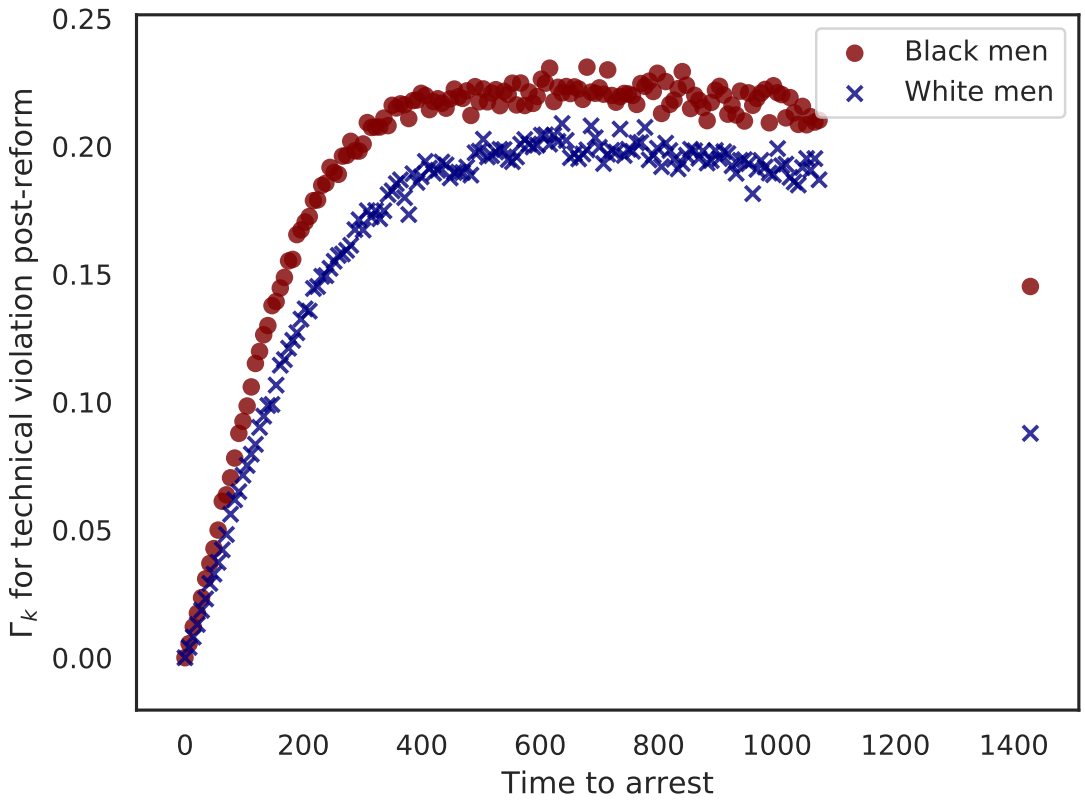
Notes: Figure plots mean (uncensored) arrest failure rates implied by estimates of the MMPH model. See text for details on sample and specification of unobserved heterogeneity used in estimation.

Figure A6: Targeting Bias in the MMPH Model Estimated Pre-Reform Only



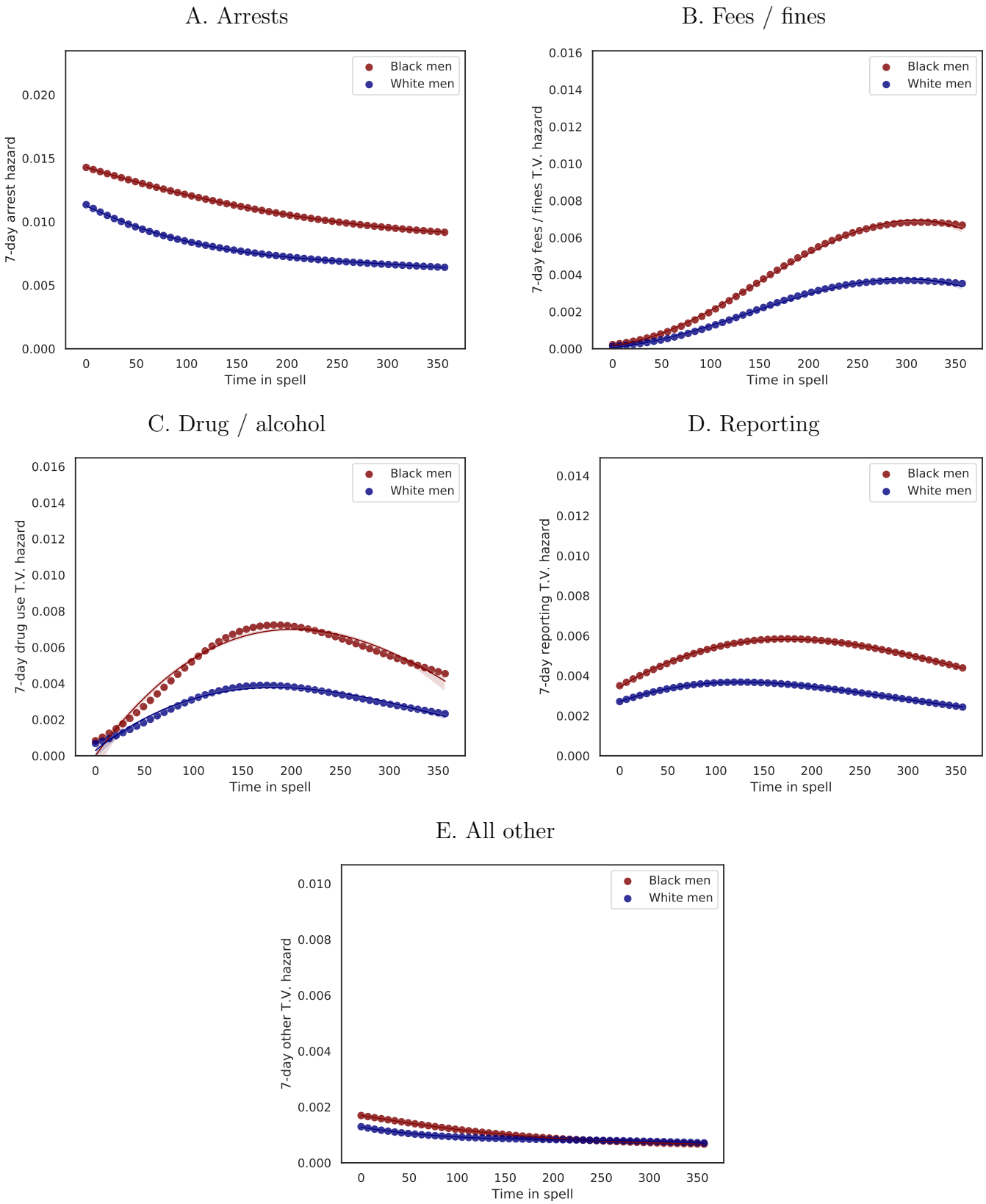
Notes: Figure plots estimates of Γ_k from the simulating arrest and technical violation failure times 10 times using coefficients from the model estimated using pre-reform data only. Γ_k is the share of observations who have arrest failure times equal to k but technical violation times $< k$. The final dots at the right of the graph plot the probability of technical violation failure times ≤ 1080 conditional on having arrest failure times > 1080 (and possibly never).

Figure A7: Targeting Bias in the MMPH Model for Post-Reform TVs



Notes: Figure plots estimates of Γ_k from simulating arrest and technical violation failure times 10 times using post-reform coefficients. Γ_k is the share of observations who have arrest failure times equal to k but technical violation times $< k$. The final dots at the right of the graph plot the probability of technical violation failure times ≤ 1080 conditional on having arrest failure times > 1080 (and possibly never).

Figure A8: Average Hazards for Arrest and Technical Violations



Notes: Figure plots mean cause-specific weekly hazard rates for each risk implied by estimates of the MMPH model. See text for details on sample and specification of unobserved heterogeneity used in estimation. Mean weakly hazards are similar but not identical to the baseline hazard, since the partial effects of unobserved heterogeneity on the hazard depend on baseline levels in the logit formulation.

Table A1: Violation Categorization

Violation type	Violation	Share of category
Absconding	-	1
Drug related	Positive drug test	0.526
	Treatment / program failure	0.295
	Admitting drug use	0.071
	Possessing drugs	0.036
	Contacting drug users	0.022
New criminal offense	New misdemeanor charge	0.716
	New felony charge	0.263
	New DWI charge	0.013
	New drug charge	0.007
Technical	Not paying fees	0.427
	Not reporting	0.202
	Other	0.099
	Moving / job change without notifying	0.058
	Breaking curfew	0.055
	Not completing community service	0.047
	No employment	0.043
	No education / training	0.012
	Traveling without permission	0.011

Notes: Includes all treated observations starting probation in 2006-2010.

Table A2: Effect of Race on Administrative Violations

	Outcome: Administrative violation in spell					
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.174***	0.190***	0.177***	0.145***	0.137***	0.101***
	(0.00172)	(0.00184)	(0.00185)	(0.00183)	(0.00195)	(0.00371)
N	315167	315167	315167	315167	315167	89122
R-squared	0.0309	0.0473	0.0697	0.114	0.128	0.107
Dep. var white mean	0.512	0.512	0.512	0.512	0.512	0.512
Demographic controls		Yes	Yes	Yes	Yes	Yes
Sentence controls			Yes	Yes	Yes	Yes
Criminal history controls				Yes	Yes	Yes
Zip code FE					Yes	Yes
Test score controls						Yes
Logit coefficient	0.734	0.817	0.779	0.665		
Logit AME	0.172	0.188	0.175	0.142		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Regressions include all spells beginning in 2006-2010. Demographic controls include gender, 20 quantiles of age, and probation district fixed effects. Sentence controls include fixed effects for the offense class of the focal conviction and a linear control for the length of their supervision spell. Criminal history controls include fixed effects for criminal history points and previous sentences to supervised probation or incarceration. Zip code FE are fixed effects for zip code at the time of initial arrest. Test score controls include the latest math and reading standardized test scores (normalized to have mean 0 and standard deviation 1 in the full population) observed from grades 3 to 8. Logit coefficient and AME are the coefficient and average marginal effects from logit estimations of the same specification. These are omitted for the last two columns where the number of fixed effects is high.

Table A3: Effect of Race on Drug Violations

	Outcome: Drug violation in spell					
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.0608***	0.0677***	0.0653***	0.0448***	0.0423***	0.0212***
	(0.00162)	(0.00171)	(0.00173)	(0.00173)	(0.00184)	(0.00388)
N	315167	315167	315167	315167	315167	89122
R-squared	0.00450	0.0241	0.0396	0.0614	0.0723	0.0695
Dep. var white mean	0.257	0.257	0.257	0.257	0.257	0.257
Demographic controls		Yes	Yes	Yes	Yes	Yes
Sentence controls			Yes	Yes	Yes	Yes
Criminal history controls				Yes	Yes	Yes
Zip code FE					Yes	Yes
Test score controls						Yes
Logit coefficient	0.298	0.340	0.331	0.233		
Logit AME	0.0603	0.0675	0.0646	0.0444		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: See notes to Table A2.

Table A4: Effect of Race on Absconding Violations

	Outcome: Absconded in spell					
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.0422***	0.0503***	0.0427***	0.0232***	0.0151***	0.0132***
	(0.00135)	(0.00143)	(0.00144)	(0.00144)	(0.00153)	(0.00317)
N	315167	315167	315167	315167	315167	89122
R-squared	0.00318	0.0176	0.0279	0.0555	0.0683	0.0725
Dep. var white mean	0.147	0.147	0.147	0.147	0.147	0.147
Demographic controls		Yes	Yes	Yes	Yes	Yes
Sentence controls			Yes	Yes	Yes	Yes
Criminal history controls				Yes	Yes	Yes
Zip code FE					Yes	Yes
Test score controls						Yes
Logit coefficient	0.303	0.367	0.310	0.181		
Logit AME	0.0418	0.0498	0.0417	0.0235		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: See notes to Table A2

Table A5: Effect of Race on Revocations

	Outcome: Revoked					
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.104***	0.118***	0.105***	0.0672***	0.0599***	0.0518***
	(0.00170)	(0.00179)	(0.00181)	(0.00177)	(0.00188)	(0.00390)
N	315167	315167	315167	315167	315167	89122
R-squared	0.0118	0.0397	0.0595	0.121	0.133	0.127
Dep. var white mean	0.296	0.296	0.296	0.296	0.296	0.296
Demographic controls		Yes	Yes	Yes	Yes	Yes
Sentence controls			Yes	Yes	Yes	Yes
Criminal history controls				Yes	Yes	Yes
Zip code FE					Yes	Yes
Test score controls						Yes
Logit coefficient	0.459	0.543	0.488	0.339		
Logit AME	0.102	0.117	0.103	0.0669		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: See notes to Table A2

Table A6: Effect of Race on Technical Revocations

	Outcome: Technical revocation in spell					
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.0627***	0.0710***	0.0649***	0.0485***	0.0418***	0.0316***
	(0.00139)	(0.00147)	(0.00150)	(0.00150)	(0.00159)	(0.00334)
N	315167	315167	315167	315167	315167	89122
R-squared	0.00664	0.0153	0.0219	0.0404	0.0503	0.0484
Dep. var white mean	0.150	0.150	0.150	0.150	0.150	0.150
Demographic controls		Yes	Yes	Yes	Yes	Yes
Sentence controls			Yes	Yes	Yes	Yes
Criminal history controls				Yes	Yes	Yes
Zip code FE					Yes	Yes
Test score controls						Yes
Logit coefficient	0.426	0.488	0.448	0.345		
Logit AME	0.0619	0.0704	0.0641	0.0485		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: See notes to Table A2

Table A7: Effect of Race on Criminal Arrests

	Outcome: Arrested in spell					
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.0627***	0.0690***	0.0562***	0.0284***	0.0300***	0.0310***
	(0.00172)	(0.00182)	(0.00184)	(0.00183)	(0.00194)	(0.00402)
N	315167	315167	315167	315167	315167	89122
R-squared	0.00423	0.0284	0.0453	0.0788	0.0893	0.0742
Dep. var white mean	0.330	0.330	0.330	0.330	0.330	0.330
Demographic controls		Yes	Yes	Yes	Yes	Yes
Sentence controls			Yes	Yes	Yes	Yes
Criminal history controls				Yes	Yes	Yes
Zip code FE					Yes	Yes
Test score controls						Yes
Logit coefficient	0.272	0.308	0.253	0.133		
Logit AME	0.0623	0.0688	0.0555	0.0282		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: See notes to Table A2

Table A8: Effect of Race on Revocation Conditional on Violation

	Outcome: Revoked (conditional on violation)				
	(1)	(2)	(3)	(4)	(5)
Black	-0.00444*	0.00829***	0.00304	-0.0112***	0.00241
	(0.00180)	(0.00193)	(0.00195)	(0.00193)	(0.00208)
N	296369	296369	296369	296369	296369
R-squared	0.0000205	0.0225	0.0308	0.0562	0.406
Dep. var white mean	0.401	0.401	0.401	0.401	0.401
Demographic controls		Yes	Yes	Yes	Yes
Sentence controls			Yes	Yes	Yes
Criminal history controls				Yes	Yes
Violations FE					Yes
Logit coefficient	-0.0185	0.0358	0.0139	-0.0479	
Logit AME	-0.00444	0.00838	0.00323	-0.0108	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Includes all violation hearings for spells beginning in 2006-2010. Controls are as defined in Table A2, except for violations FE, which are fixed effects for the unique violations categories disposed at the hearing. Logit coefficient and AME are the coefficient and average marginal effects from logit estimations of the same specification. These are omitted for specifications where the number of fixed effects is high.

Table A9: Officer-Offender Race Match Effect in Violations

	Outcome: Any outcome in spell							
	(1) Adm	(2) Adm	(3) Drug	(4) Drug	(5) Rev.	(6) Rev.	(7) Tech rev.	(8) Tech rev.
Black	0.092***	0.091***	0.026***	0.024***	0.040***	0.041***	0.031***	0.033***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Black x black off		0.0028		0.0075*		-0.0041		-0.0044
		(0.004)		(0.003)		(0.003)		(0.003)
N	306418	306418	306418	306418	306418	306418	306418	306418
W mean	0.37	0.37	0.18	0.18	0.21	0.21	0.12	0.12
Demo	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sent	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Crim hist	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Off FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Includes all spells starting in 2006-2010. Officer race is coded using the race of the first officer assigned in the spell. Controls are as defined in Table A2. Outcomes are an indicator for the listed event happening within the first year of a spell.

Table A10: Effect of Reform by Crime Type

	Black			Not-black		
	(1) Any	(2) Misd/fel	(3) Fel	(4) Any	(5) Misd/fel	(6) Fel
Post-reform	-0.0111***	-0.00923***	0.00207	-0.00668***	-0.00193	0.00325***
	(0.00281)	(0.00274)	(0.00168)	(0.00190)	(0.00178)	(0.000963)
Treated	-0.0468***	-0.0412***	-0.00292	-0.000288	0.00168	0.00745***
	(0.00268)	(0.00262)	(0.00163)	(0.00207)	(0.00195)	(0.00110)
Post-x-treat	0.0238***	0.0214***	0.00559*	0.0179***	0.0177***	0.00900***
	(0.00383)	(0.00374)	(0.00237)	(0.00295)	(0.00279)	(0.00163)
N	217214	217214	217214	328900	328900	328900
Pre-reform treated mean	.314	.29	.092	.264	.226	.062
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Criminal history FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Includes all treated and untreated probation spells beginning 1-3 years before the reform and 0-2 years afterwards. Post is indicator for starting probation after December 1, 2011, the date JRA reforms took effect. Demographic controls include five-year age bins and indicators for race and gender. Criminal history controls include fixed effects criminal history points and prior sentences to supervised probation or incarceration.

Table A11: Impact of Data Window for Measuring Effects of Reform

	White			Black		
	Technical incarceration					
	(1)	(2)	(3)	(4)	(5)	(6)
	1yr	2yr	3yr	1yr	2yr	3yr
Post-reform	-0.0013**	-0.00087**	-0.00064*	-0.0048***	-0.0041***	-0.0040***
	(0.00048)	(0.00034)	(0.00028)	(0.00077)	(0.00054)	(0.00044)
Treated	0.12***	0.11***	0.11***	0.16***	0.16***	0.16***
	(0.0018)	(0.0013)	(0.0010)	(0.0024)	(0.0017)	(0.0014)
Post-x-treat	-0.042***	-0.037***	-0.036***	-0.070***	-0.076***	-0.079***
	(0.0025)	(0.0017)	(0.0014)	(0.0031)	(0.0022)	(0.0018)
N	165936	328784	488779	109764	217222	319596
R-squared	0.081	0.079	0.078	0.090	0.091	0.092
Pre-reform treated mean	.136	.131	.128	.181	.181	.182
	Arrest					
Post-reform	-0.0036	-0.0066***	-0.0081***	-0.0036	-0.011***	-0.019***
	(0.0026)	(0.0019)	(0.0016)	(0.0039)	(0.0028)	(0.0024)
Treated	-0.0041	-0.00031	0.0019	-0.044***	-0.047***	-0.049***
	(0.0029)	(0.0021)	(0.0017)	(0.0038)	(0.0027)	(0.0022)
Post-x-treat	0.021***	0.018***	0.018***	0.016**	0.024***	0.029***
	(0.0041)	(0.0029)	(0.0024)	(0.0054)	(0.0038)	(0.0032)
N	165936	328784	488779	109764	217222	319596
R-squared	0.072	0.073	0.072	0.083	0.080	0.079
Pre-reform treated mean	.257	.264	.268	.31	.314	.317
Accuracy	.517	.543	.58	.205	.306	.363
False negative rate	.923	.93	.929	.956	.931	.917
False positive rate	.032	.027	.024	.099	.094	.091
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Criminal history FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Includes all treated and untreated probation spells beginning within 1, 2, and 3 years before the reform and within 0, 1, and 2 afterwards, as indicated in the column header. Post is indicator for starting probation after December 1, 2011, the date JRA reforms took effect. Demographic controls include five-year age bins and indicators for race and gender. Criminal history controls include fixed effects criminal history points and prior sentences to supervised probation or incarceration.

Table A12: MMPH Estimates With Constrained Arrest Types

	Arrest		Technical violation	
	β	(se)	β	(se)
Duration	-0.78	(0.06)	3.46	(0.10)
Duration^2	1.92	(0.42)	-20.82	(0.77)
Duration^3	-2.68	(1.09)	42.72	(2.19)
Duration^4	1.94	(1.21)	-40.65	(2.61)
Duration^5	-0.59	(0.48)	14.56	(1.09)
Calendar time	-0.06	(0.01)	-0.19	(0.02)
Calendar time^2	0.03	(0.01)	-0.13	(0.01)
Type 1	-6.71	(0.05)		
Type 2	-4.75	(0.04)		
Type 3	-4.75	(0.04)		
Type 4	-3.45	(0.04)		
White men				
Type 1			-8.08	(0.07)
Type 2			-6.55	(0.09)
Type 3			-4.86	(0.08)
Type 4			-5.29	(0.07)
Black men				
Type 1			-6.93	(0.04)
Type 2			-6.53	(0.11)
Type 3			-4.73	(0.07)
Type 4			-5.30	(0.05)
	Black men		White men	
Pr(type 1)	0.32	(0.01)	0.48	(0.01)
Pr(type 2)	0.39	(0.02)	0.35	(0.02)
Pr(type 3)	0.13	(0.02)	0.09	(0.01)
Pr(type 4)	0.15	(0.01)	0.09	(0.01)

Notes: Table reports estimates of the MMPH model described in Section 5 but with all coefficients except V_i^j for technical violation risk and type probabilities constrained to be equal across race within gender. Interactions between V_i^j are the reform Z_{ist} are omitted for brevity. Duration and calendar time are standardized (mean 0 and s.d. 1) in estimation. Standard errors are the robust "sandwich form" clustered by individual.