

Should the Punishment Fit the Crime? Discretion and Deterrence in Law Enforcement*

Felipe Goncalves[†] Steven Mello[‡]

September 11, 2022

Abstract

We study the implications of police discretion for public safety. Relying on variation across highway patrol officers in their propensity to issue harsh fines, we show that higher fines reduce future traffic offending. Motorists most likely to face harsh sanctions are least deterred by fines, inconsistent with an allocation of sanctions that maximizes public safety, and most likely to reoffend, suggesting an alternative model of officer behavior. Counterfactual punishment allocations can reduce the aggregate reoffending rate by as much as seven percent, highlighting efficiency costs associated with current officer practices, but require that the lowest risk drivers face harsh punishments.

JEL Codes: K42, D73, J45

*We are grateful to Leila Agha, David Arnold, Eric Chyn, Will Dobbie, Ben Hansen, Peter Hull, Elisa Jacome, Louis Kaplow, Ilyana Kuziemko, Emily Leslie, Adriana Lleras-Muney, Erzo Luttmer, Alex Mas, Emily Nix, Sam Norris, Rodrigo Pinto, Bruce Sacerdote, Yotam Shem-Tov, Chris Snyder, Doug Staiger, Megan Stevenson, and Emily Weisburst, as well as seminar participants at Princeton, UCLA, Dartmouth, Case Western, UBC, Claremont McKenna, SDSU, the SEA Annual Meeting, and NBER Summer Institute for helpful comments. We thank Jeffrey Bissainthe, Kiara Guzzo, Wilton Johnson, Timothy Kutta, Stacy Lehmann, Brenda Paige, and especially Beth Allman for assistance with acquiring and using the data from various Florida agencies. Jacklyn Pi provided excellent research assistance. An earlier version of this paper was circulated as “Does the Punishment Fit the Crime? Speeding Fines and Recidivism.” Any errors are our own.

[†]University of California, Los Angeles and NBER; fgoncalves@ucla.edu

[‡]Dartmouth College and NBER; steve.mello@dartmouth.edu.

1 Introduction

Ensuring public safety is a central function of the state. To that end, policy regimes specify sanctions for socially undesirable behaviors (Becker, 1968). However, beyond the set of official policies, the state also relies on bureaucrats tasked with allocating sanctions in practice. As in much of the public sector, these actors typically wield significant discretion (e.g., Lipsky 1980; Wilson 1989). For example, patrolling police officers can overlook minor crimes, prosecutors can reduce criminal charges, and judges can issue a wide range of sentences. Permitting this type of discretion can have important benefits if agents internalize the goals of the state and can effectively discern the allocation of sanctions that maximizes public safety (Banfield, 1975). On the other hand, these decision-makers may have poor information, attend to alternative objectives such as personal gain, or hold competing notions of the optimal allocation when making decisions (e.g., Bendor et al. 2001; Prendergast 2007).

A critical question raised by the pervasive discretion practiced by criminal justice agents, then, is whether these agents pursue public safety when making decisions. We study this question in the context of traffic enforcement, focusing on highway patrol officers choosing to impose harsh or lenient fines on apprehended speeders. First, we show that sanction choices have important deterrence effects: harsh fines reduce the likelihood of future traffic offending. Next, we examine how officers choose which motorists to punish harshly. We find that motorists most likely to face harsh fines are the least deterred by punishments, inconsistent with public safety maximization by officers, but reoffend at the highest rates, suggesting an alternative model of officer behavior. We conclude by considering potential explanations for this officer behavior, discussing policy implications, and quantifying the efficiency costs, in the form of foregone deterrence, associated with current officer practices.

Speeding enforcement is a high stakes setting in terms of public safety. In 2020, there were nearly twice as many traffic fatalities ($\sim 39,000$) as homicides ($\sim 22,000$) in the United States. Economic costs associated with motor vehicle accidents have been estimated at nearly \$250 billion per year, higher than annual costs of crime victimization (Blincoe et al. 2015; Chalfin 2016). Standard estimates suggest that at least one third of fatal crashes are caused by speeding and existing studies have found strong associations between average driving speeds and traffic fatalities (NHTSA 2014; Ashenfelter & Greenstone 2004).

While speeding sanctions are statutorily based only on a driver's speed relative to the posted limit, officers can manipulate fines by writing down a slower speed than was observed on the actual citation, resulting in a discounted fine (Anbarci & Lee 2014; Goncalves & Mello 2021). In Florida, the setting of our study, nearly one third of all speeding citations are issued for exactly nine miles per hour (MPH) over the limit, just below a \$75 increase in the statutory fine amount. Less than one percent are issued for either eight or ten MPH. Officers patrolling the same beat-shifts vary considerably in the degree of bunching in their charged distributions, highlighting that officer discretion, rather than driver behavior, explains the

bunching in cited speeds (Goncalves & Mello, 2021).

Our empirical approach leverages this variation across officers in the propensity to bunch motorists below fine increases. We use an instrumental variables (IV) framework where the treatment and outcome of interest are whether a driver receives a harsh (versus discounted) fine and whether a cited driver commits a new traffic offense in the following year. Our instrument is the citing officer’s propensity *not to bunch* other drivers, which we call stringency. Our design mirrors a growing literature leveraging randomly assigned judges for identification (e.g., Kling 2006; Maestas et al. 2013; Dahl et al. 2014; Dobbie & Song 2015), with the caveat that, in our setting, citing officers are not randomly assigned to drivers. An important concern for our approach, then, is whether an officer’s bunching propensity is correlated with the characteristics of her sample of cited drivers. We show that, conditional on beat-shift fixed effects, our stringency measure is uncorrelated with an officer’s ticketing frequency, driver characteristics that predict reoffending, and past traffic offending.

First, we use this IV framework to study the deterrence effects of sanctions. Specifically, we estimate the local average treatment effect (LATE) of harsher fines on the future driving behavior of cited drivers, instrumenting harsh fines with officer stringency. We find that a 125 dollar increase in fine amounts reduces the likelihood of any new traffic offense in the following year by 1.6 percentage points ($\epsilon = -0.07$). We document stronger effects on speeding offenses ($\epsilon = -0.13$) and statistically significant, but less precise, effects on the likelihood that a driver is involved in a traffic accident in the next year ($\epsilon = -0.04$).

In this setting, motorists receiving harsh and lenient fines face identical sanctions for future offenses. Hence, the ex-post response we document represents a specific deterrence effect, or the impact of the experience of punishment on offending (Nagin, 2013). While not our main contribution, our fine elasticity estimates advance a small literature focused on isolating specific deterrence effects (Dusek & Traxler 2021; Gehrsitz 2017; Hansen 2015) and add to a larger literature on the deterrence effects of traffic enforcement efforts (e.g., Makowsky & Stratmann 2011; DeAngelo & Hansen 2014; Luca 2014; Traxler et al. 2018).

Next, we turn to the question of how officers decide whom to punish harshly. Relying on the same IV framework, we use a marginal treatment effects (MTE) approach to describe heterogeneity in reoffending behavior for motorists with varying propensities to face harsh fines (Bjorklund & Moffitt 1987; Heckman & Vytlacil 2007). Building on a Roy (1951) selection model, the idea of our MTE approach is to characterize motorists who are shifted into harsh fines at different values of the stringency instrument. The offenders on the margin at the lowest values of stringency are those who would be punished harshly by all but the most lenient officers, or those that face harsh sanctions most often.¹ Marginal offenders at

¹Our baseline MTE approach relies on a strong monotonicity assumption (e.g., Imbens & Angrist 1994; Vytlacil 2002). In section 5.4, we show that our conclusions are unaltered when taking several approaches to relax this monotonicity assumption, including an alternative MTE estimator leveraging only the variation in propensity scores near zero and one.

the highest levels of stringency face harsh fines least often.

Holding fixed an officer’s willingness to issue harsh fines, an efficient (or safety-maximizing) allocation of sanctions would levy harsh fines only on the most responsive offenders. In other words, public safety maximization implies *positive selection on gains*: the motorists most likely to face harsh fines should exhibit the largest treatment effects. The logic of this hypothesis mirrors a recent literature connecting welfare analysis and treatment effect heterogeneity in settings with self-selection into public programs (e.g., [Deshpande & Li 2019](#); [Finkelstein & Notowidigdo 2019](#); [Ito et al. 2022](#)).

Instead, our MTE estimates show that officers’ punishment choices generate *reverse selection on gains*: the motorists facing harsh fines most often are the least responsive in terms of driving behavior, while the motorists treated least often are the most deterred by harsh fines. This difference in deterrence effects is sizable and statistically significant. We find, if anything, positive impacts of harsh sanctions on reoffending for the group of motorists facing harsh fines most often. Motorists least likely to face harsh sanctions, on the other hand, exhibit treatment effects that are about three times the size of our estimated LATE.

To explore potential explanations for this departure from efficient behavior by officers, we examine sorting on other dimensions. We document stark *selection on levels*: when issued a lenient fine, motorists most likely to face harsh sanctions reoffend about 30 percent more often than those least likely to see harsh sanctions. Hence, selection patterns imply that the “worst” drivers, as captured by recidivism risk, are the most likely to be fined harshly by officers. Interestingly, these selection patterns imply that the motorists most deterred by harsh fines and the motorists at the highest risk of reoffending are different groups, highlighting that, for a given number of harsh fines, officers face a tradeoff between maximizing public safety and punishing the “worst” drivers.

Although motorist characteristics that are easily observable to officers, such as race, gender, and offending history, are strongly predictive of harsh fines, the pattern of selection on levels and reverse selection on gains persists when re-estimating within detailed motorist covariate cells and when focusing on a subsample of first-time offenders. We find that these characteristics can only explain about 30 percent of the overall sorting patterns. Leveraging the fact that true driving speed is observed for the subset of officers that never bunch drivers, we also show that faster drivers are more likely to be punished harshly by estimating the average driving speeds among compliers at different values of stringency. Sanction decisions based on speed, however, can only explain a small share, about 12 percent, of the overall sorting. Hence, even conditional on offending history, demographics, and offense severity, we find that officers sort drivers into sanctions in a way that generates selection on levels and reverse selection on gains.

On the one hand, this behavior could arise from information problems. Officers may strive to maximize public safety but hold incorrect views about how to achieve this goal. The observed punishment decisions are consistent with, for example, officers seeking to maximize

the deterrence generated by their citations but mistakenly believing that reoffending levels and treatment effects are positively correlated. In this case, an information intervention that notifies officers of the safety gains associated with issuing harsh sanctions to the motorists typically granted lenience could be an effective policy for changing officer behavior and, accordingly, reducing the reoffending rate.

Alternatively, selection patterns could reflect officer preferences. Officers may explicitly strive to issue harsh fines to the the motorists most likely to reoffend, or to motorists with other unobservable characteristics strongly correlated with recidivism risk, despite their unresponsiveness to harsh fines. The desire to punish the “worst” drivers harshly captures well a notion of fairness, and it is plausible that officers could have strong preferences for fairness in this setting. Under the assumption that selection patterns are generated by officer preferences, we characterize those preferences by specifying and estimating a model of sanction choices. In the model, officers observe noisy signals of a motorist’s potential outcomes and choose whom to sanction harshly based on a weighted average of expected safety gains and recidivism risk. The weight captures the the rate at which officers are willing to trade off deterrence in pursuit of punishing the “worst” offenders, or alternatively punishing based on the characteristics that generate selection on levels. Our model estimates, which are set-identified based on the selection patterns in the data, imply that officers place at least as much weight on levels as gains when issuing sanctions.

Regardless of the underlying reasons for this officer behavior, the reverse selection on gains that we document implies that a lower aggregate recidivism rate could be achieved by reallocating sanctions, holding fixed the population of cited drivers and the share of harsh fines. We quantify these efficiency costs, in the form of foregone deterrence, by computing reoffending rates in two counterfactual reallocations of harsh fines that correspond closely with the two theories of officer behavior we posit above. First, we consider a counterfactual where officers are forced to sort motorists in reverse order of their current practice. And second, we consider the implications of forcing officers to consider only deterrence in our model of sanction choices. We estimate that the aggregate reoffending rate falls by between four and seven percent in these counterfactual scenarios. These safety gains are achieved by reassigning harsh sanctions to the most deterrable, but lowest risk, offenders.

Our central contribution is to a broad literature on the allocation choices of economic agents. A common practice in this literature is to examine selection patterns in settings where efficient allocations should exhibit selection on gains (e.g., [Carneiro et al. 2011](#); [Abaluck et al. 2016](#); [Van Dijk 2019](#); [Chandra & Staiger 2020](#)). Several such studies have nonetheless found sorting based on levels, such as parents choosing school districts for their children ([Abdulkadiroglu et al., 2020](#)) and hospitals opting into a Medicare reform ([Einav et al., 2022](#)). We document a similar departure from efficient decision-making in a new but high-stakes setting, law enforcement.

Our characterization of officer sanction choices advances a rapidly growing literature on

discretion in the criminal justice system (e.g., Weisburst 2017; Ba et al. 2021; Chalfin & Goncalves 2021; Goncalves & Mello 2021; Abrams et al. 2021; Feigenberg & Miller 2022; Norris 2022) and more broadly contributes to a literature on the implications of bureaucratic preferences for state effectiveness (e.g., Prendergast 2007; Best et al. 2017; Akhtari et al. 2022). We document a misalignment between police behavior and the public safety goals of the state and compute the associated efficiency costs. And finally, our paper speaks to a largely theoretical literature on fairness-efficiency tradeoffs in the design of legal institutions (e.g, Polinsky & Shavell 2000; Kaplow & Shavell 2006; O’Flaherty & Sethi 2019; Moore 2019) by documenting the empirical relevance of a tradeoff between efficiency and other potential law enforcement objectives, such as the targeting sanctions to the “worst” offenders.

The rest of our paper proceeds as follows. Section 2 describes our data and setting. We lay out our empirical framework in section 3 and estimate the causal effect of sanctions in section 4. Section 5 characterizes how officers allocate harsh sanctions and section 6 considers policy implications and computes efficiency costs. We conclude in section 7.

2 Data and setting

The Florida Clerks and Comptrollers provided administrative records of the universe of traffic citations issued in Florida for the years 2005–2018 from Florida’s Uniform Traffic Citation (UTC) database. These records include the date and county of the citation as well as information on the cited violation. When the violation is speeding, this information includes the charged speed and posted speed limit (e.g., 74 MPH in a 65 MPH zone). The UTC data also include all information provided on a stopped motorist’s driver license (DL): name, date of birth, address, race, gender, as well as the driver license state and number. Using the driver license number, we are able to link drivers across citations and construct our primary outcome measures of past and future traffic offending.

We augment the driver information in the UTC data with four auxiliary data sources. First, we match drivers on zip code of residence to estimated per-capita income at the zip code level from the IRS Statistics of Income files. Second, the make and year of the stopped automobile is provided for about 75 percent of citations. We use this information to construct an estimated vehicle value based on a database of online vehicle resale prices. Third, we recode a motorist’s race as Hispanic if, based on census records, their surname is associated with Hispanic status for more than 80 percent of individuals.² Finally, we link drivers on full name and date of birth to prison spell records from the Florida Department of Corrections to construct a measure of prior incarceration.

²As discussed in Goncalves & Mello (2021), there are clear inconsistencies in the recording of Hispanic status in the UTC data. Officers frequently write down race = *H* (for Hispanic). But in Miami-Dade county, where the population is over 60 percent Hispanic, less than one percent of citations are coded as being issued to a Hispanic motorist.

In the citations data, the ticketing officer is identified by name. We construct a consistent officer identifier by linking the officer name with data on Florida Highway Patrol (FHP) employment spells provided by the Florida Department of Law Enforcement. We focus on tickets issued by the FHP both because we can more consistently identify the citing officer and because speeding enforcement is a central duty of FHP officers. However, we measure past and future offending using all citations, not just FHP-issued citations.

2.1 Other data sources

We obtained administrative crash reports covering the universe of automobile accidents known to police over the period 2006–2018 from the Florida Department of Transportation (FDOT). These data are collected during a police response or investigation and include the date and county of the incident as well as information on injuries and property damage for a subset of crashes. The data also include the driver license numbers of involved drivers, which we use to link drivers with the citations data.

The Florida Clerks and Comptrollers also provided records from the Traffic Citation Accounting Transition System (TCATS) database, which includes information on the traffic court disposition associated with about 80 percent of the citations in our sample. We use these records to construct a measure of whether a citation was contested in traffic court and, based on the traffic court disposition, to construct measures of accrued, rather than statutory, sanctions. For additional details, see appendix D.

2.2 Sample construction

To construct our sample of focal citations, we first restrict attention to tickets written by the Florida Highway Patrol over 2007–2016 where the citing officer is identified.³ We further restrict the sample to include tickets where speeding is the only violation, no crash is indicated, and the charged speed is between nine and twenty-nine miles per hour over the posted speed limit. We choose twenty-nine as our baseline upper limit because (i) the available evidence suggests that motorists are still bunched with positive probability when their true speed is as high as twenty-nine MPH over the limit (see figure 1) and (ii) thirty MPH over the limit is the threshold for a misdemeanor speeding offense.

We also restrict to drivers with a valid Florida driver license number, so that we can reliably measure past and future offending, and require that officers have at least fifty citations meeting the above criteria to compute our instrument. Ultimately, our focal sample is comprised of 1,693,457 speeding citations issued by 1,960 FHP officers. There are 1.4M unique drivers in the sample. Table 1 presents summary statistics for our analysis sample.

³We focus on 2007–2016 so that we can measure other offending (including crash activity) for least one full year prior and one full year after the focal citation. Over this period, the ticketing officer is identifiable for 85 percent of FHP-issued speeding tickets.

At a few points in our analysis, we also rely on a sample constructed in an identical way using only tickets issued from 2010–2016 (the “late” sample; 1,152,791 citations issued by 1,625 officers). This alternative sample allows us to observe longer driving histories for all in-sample motorists and allows us to compute an out-of-sample measure of officer experience based on the number of speeding tickets issued over the period 2005–2009.

Again, reoffending and past offending are measured using *all* citations issued in the state rather than just the citations that comprise our sample of focal FHP tickets. Worth noting here is the fact that our main outcome measure will capture whether a motorist is caught and ticketed for a new traffic offense, which itself could be subject to officer discretion. If anything, we expect that officer discretion at the recidivism stage will bias our specific deterrence effect estimates towards zero. We compare reoffending rates for individuals (randomly) receiving harsh and lenient fines and find that those receiving harsh fines differentially reduce their offending rates. If officers are more likely to let drivers with less severe offending histories off with formal or informal warnings, that would bias our estimates towards zero by inflating the reoffending rates of those who are sanctioned harshly or deflating the reoffending rates of those who are issued lenient sanctions.

2.3 Florida highway patrol

State-level patrols are the primary enforcers of traffic laws on interstates and many highways, especially those in unincorporated areas. On patrol, officers are given an assigned zone over which they can combine roving patrol and parked observation patrol. Florida Highway Patrol (FHP) officers are divided into one of nine assigned troops, almost all of which patrol six to eight counties each. Officer assignments operate on eight-hour shifts and cover an assignment region that roughly corresponds to a county, though the size of a “beat” can vary based on an area’s population density. In practice, we use counties to proxy for assignment regions.

The FHP is comprised of approximately 1,500 full-time officers. Speeding enforcement is a primary duty of FHP officers and the FHP collectively issues between 150,000 and 200,000 speeding citations each year. Other responsibilities include enforcing a wide array of other traffic laws, investigating crashes, and responding to and assisting with highway emergencies. The FHP officer handbook reads “*Members should take the enforcement action they deem necessary to ensure the safety of the motoring public, reduce the number and severity of traffic crashes, and reduce the number of criminal acts committed on highways of this state,*” highlighting that officers are explicitly given discretion over enforcement decisions.

In Florida, speeding sanctions are based on an offender’s speed relative to the posted speed limit. Speeding 1-5 MPH over the limit carries a statutory warning but no sanctions, while speeding 30 or more MPH over the limit is a misdemeanor offense requiring the offender to appear in court. Between 6 and 29 MPH over the limit, the statutory fine is a step function, plotted as a red dotted line in figure 1.

Speeding offenses are also associated with “points” on an offender’s driver license (DL). Point assessments are also based on speed; speeding 6-15 MPH over the limit is associated with 3 points while speeding 16+ MPH over the limit is associated with 4 points. Points are used by car insurers to adjust premiums and offenders that collect a sufficient number of points (12 points in 12 months; 18 points in 18 months; 24 points in 36 months) have their license suspended for 30 days (6 months; 1 year).

After a citation has been issued, a driver can either submit payment to the county clerk or request a court date to contest the ticket. If the ticket goes to court, a judge or hearing officer typically decides either to uphold the original charge, reduce the charge, or dismiss the citation. At the time of payment, a subset of drivers can elect to attend an optional traffic school, completion of which combined with on-time payment will remove the citation from a driver’s record and prevent the accrual of the associated DL points.

During a traffic stop, the citing officer typically scans the offender’s DL and electronically receives a report on the motorist’s past traffic offending. To the best of our knowledge, officers do not receive any information on reoffending associated with their citations.

2.4 Discretion over sanctions

Panel (a) of figure 1 shows the speeding fine schedule in Florida and a histogram of charged speeds on FHP-issued speeding citations. Over one third of all citations are issued for exactly 9 MPH over the posted limit, just below a \$75 increase in the associated fine. Less than one percent of all citations are issued for eight or ten MPH over the limit. The dramatic bunching in the speed distribution suggests systematic manipulation by officers. Specifically, the distribution implies the practice of speed discounting, where officers observe drivers traveling at higher speeds but write down nine MPH on the citation as a form of lenience (Anbarci & Lee 2014; Goncalves & Mello 2021). An officer’s decision whether to bunch a driver, resulting in either a discounted or full fine, is the focus of our study.

We rely on several pieces of evidence to demonstrate that bunching in the speed distribution is generated by the behavior of officers rather than drivers (e.g., Traxler et al. 2018). First, following Goncalves & Mello (2021), figure 1 shows that all bunching is attributable to a subset of *lenient* officers.⁴ About 25 percent of officers, whom we term the non-lenient officers, almost never write tickets for nine MPH.

Moving beyond a binary split of officers, figure A-1 illustrates significant variation across officers in the propensity to bunch drivers. Panel (a) demonstrates full support across officers in bunching propensity, while panel (b) shows that this variation persists after netting out

⁴See appendix D-2 for details on the classification of officers as lenient versus non-lenient, which is based on the manipulation test from Frandsen (2017). To ensure that the pattern in figure 1 is not mechanical and to avoid the reflection problem in IV estimates, we randomly partition an officer’s stops into two groups, classify each officer \times partition as lenient versus not, and then use the officer’s classification in the *other* partition.

location and time fixed effects. Such variation is inconsistent with bunching due to driver behavior; if drivers systematically bunch below fine increases, then officers patrolling the same beat-shift should have similar degrees of bunching in their speed distributions.

However, this across-officer variation could alternatively be due to noise or estimation error. To confirm that the across-officer variation in bunching propensity is “true” variation (in a statistical sense), we estimate the following regression:

$$\mathbf{1}[\text{bunch}_{ijs}] = \gamma X_i + \psi_s + \alpha_j + u_{ijs}$$

where i indexes citations, j indexes officers, and s indexes beat-shifts; X_i is a vector of driver covariates, ψ_s is a beat-shift fixed effect, and α_j is an officer fixed effect.⁵ This regression has an $R^2 = 0.55$, with 0.32 (58 percent) attributable to the officer effects, 0.22 (41 percent) attributable to the beat-shift effects, and less than one percent attributable to the driver X ’s. In other words, the identity of the citing officer is significantly more predictive of a bunched citation than the beat-shift of the stop or the full set of driver characteristics. Moreover, there is significant variation in the estimated $\hat{\alpha}_j$ ’s ($\sigma^2 = 0.076$). Applying Empirical Bayes shrinkage (Morris, 1983) to adjust for estimation error has minimal impact on the dispersion of the estimated officer effects ($\sigma^2 = 0.71$). See panel (c) of figure A-1 for further details.

Finally, we show in figure A-2 that an officer’s bunching propensity is highly correlated across space and time. First, we randomly partition an officer’s citations into two location (county) groups and regress an officer’s bunching propensity, adjusted for beat-shift fixed effects, in one set of locations on the same officer’s adjusted bunching propensity in the other set of locations. This regression yields $\hat{\beta} = 0.68$ ($se = 0.02$). Next, we split an officer’s citations in half temporally and perform the same exercise, which gives $\hat{\beta} = 0.85$ ($se = 0.01$).

2.5 Why do officers bunch drivers?

Fine revenue is routed to the county government where the citation was issued. Hence, neither the officers themselves, nor the FHP or state government more broadly, have any financial stake in fine amounts. Officers do, however, potentially have a promotion incentive to write a certain number of tickets, as the number of tickets they write appears on their performance evaluations. We believe these set of institutional factors contribute to an environment in which officers are encouraged to write tickets but also have the freedom to write reduced charges, which is ideal for our research design (Goncalves & Mello, 2021).

Based on the available evidence, our view is that distaste for traffic court best explains officer lenience in this context. After receiving a traffic ticket, the cited driver has the option to contest the citation in traffic court. The citing officer is expected to attend the associated court hearing. Using the same identification strategy that we exploit to assess the causal

⁵The ψ ’s are the same fixed effects we use in our main analysis, described in section 3. They are at the level of county \times $\mathbf{1}[\text{highway}] \times$ year \times month \times $\mathbf{1}[\text{weekend}] \times$ shift.

effect of sanctions on offending, we find that a 125 dollar increase in fine (causally) increases the likelihood that a driver contests a ticket in court by about 40 percent (see table 3). Hence, distaste for appearing in traffic court generates an incentive to bunch drivers and heterogeneity in distaste for traffic court could explain the observed variation in lenience across officers.

3 Empirical framework

Our empirical approach leverages the variation across officers in the propensity to bunch drivers within an instrumental variables framework. The outcome of interest, Y_i , is whether cited driver i commits a new traffic offense in the following year. The treatment of interest is whether the driver i receives a harsh fine (as opposed to a lenient one), which we denote by $D_i = \mathbf{1}[\text{speed}_i \geq 10]$. The instrument, which we call officer stringency, is computed as:

$$Z_{ij} = 1 - \left(\frac{1}{N_j - 1} \sum_{k \neq i} \mathbf{1}[\text{speed}_{kj} = 9] \right) \equiv \text{stringency}$$

where i indexes motorists and j indexes officers. In words, Z_{ij} is the fraction of officer j 's citations issued to all *other* drivers that are for speeds of 10 MPH or more over the limit; or in other words, the fraction of citations that are not bunched.

To adjust for differential exposure of officers to groups of motorists based on patrol shift assignments, we condition on detailed beat-shift fixed effects, denoted by ψ_s , in all our analyses. These beat-shift effects are at the level of the county $\times \mathbf{1}[\text{highway}] \times \text{year} \times \text{month} \times \mathbf{1}[\text{weekend}] \times \text{shift}$. A county is approximately a patrol area for each officer. Officers work the same shift (day of week and time of day) for one month and then rotate.

Our empirical framework requires that the stringency instrument satisfy the standard local average treatment effect (LATE) assumptions (e.g., [Imbens & Angrist 1994](#)):

1. *Relevance.* $D(Z)$ is a nontrivial function of Z .
2. *Exogeneity.* $\{Y_{i1}, Y_{i0}, D_i(Z)\} \perp Z \mid \psi$
3. *Exclusion.* $Y_i(D, Z) = Y_i(D)$
4. *Monotonicity.* $\forall w, j \in J$, either $D_i(w) \geq D_i(h) \forall i$ or $D_i(w) \leq D_i(h) \forall i$

where J denotes the set of officers and $\{Y_{i1}, Y_{i0}\}$ are the potential outcomes of driver i when sanctioned harshly ($D = 1$) and leniently ($D = 0$).

The relevance assumption requires a relationship between stringency and harsh fines, which is empirically testable. Figure 3 plots the probability of harsh fines against officer stringency, conditional on beat-shift fixed effects, laid over a histogram of stringency, net of beat-shift effects. The figure documents a linear and statistically precise relationship, with

an estimated first stage coefficient of $\hat{\beta} = 0.944$ ($se = 0.006$) and associated $F \approx 22,000$. In figure A-4, we show the first stage estimates for other sanction measures. In terms of fine amounts, shown in panel (a), the estimated first stage is $\hat{\beta} = \$122$. We further discuss the exogeneity, exclusion, and monotonicity assumptions in turn below.

3.1 Exogeneity

Existing studies using examiner designs (e.g., Kling 2006, Dobbie & Song 2015, Maestas et al. 2013, Bhuller et al. 2020) have appealed to the institutional quasi-random assignment of examiners (e.g., bail judges) to satisfy the exogeneity assumption. Citing officers in our setting are, of course, not randomly assigned to drivers. Instead, officers can select their own samples by choosing (i) whom to pull over versus whom to let pass and (ii) whom to cite versus whom to let go with a formal or informal warning. We cannot observe formal or informal warnings in our data and cannot observe the full population of drivers passing by an officer during a given beat-shift.

An especially salient threat to our empirical design would be a correlation between stringency on the citing margin (whom to cite versus not) and the charging margin (whom to bunch versus not). To help illustrate this point, suppose there were two officers, $j \in \{1, 2\}$, with $j = 1$ an officer who bunches most drivers and $j = 2$ an officer who bunches very few drivers. Suppose that $j = 1$ is also very lenient on the citing margin; that is, she lets most motorists pass with no citation, while $j = 2$ is very stringent on the ticketing margin, citing most drivers. If $j = 1$ restricts her sample by only citing drivers with a higher expected Y_{i0} , then $E(Y_{i0} | j = 1) > E(Y_{i0} | j = 2)$, violating exogeneity.

There are two testable implications of the hypothesis that lenience on the intensive (bunch versus not) and extensive (ticket versus not) margins are correlated. First, our instrument Z should be correlated with an officer’s citation frequency. Holding constant the supply of offenders, officers with higher ticketing thresholds should have “missing” tickets relative to officers with lower ticketing thresholds. Second, Z should be correlated with driver characteristics that predict reoffending. We test both these predictions in figure 2. Panel (a) plots the relationship between officer stringency and an officer’s average monthly citations, both adjusted for beat-shift fixed effects. For both all citations and speeding citations, regression coefficients are quantitatively small and statistically indistinguishable from zero. Panel (b) illustrates that there is no relationship between stringency and either past offending or predicted reoffending based on driver covariates.

Table 2 presents the relationship between the full set of driver characteristics and recidivism, charged fines, and our stringency instrument Z . As shown in columns 1-2, driver covariates have substantial joint predictive power over reoffending ($F = 1734$) and are also quite predictive of reduced charges ($F = 29$). In contrast, motorist characteristics have considerably less ability to predict officer stringency ($F = 2.7$). While our test rejects the

null hypothesis of no statistical relationship between observables and officer stringency, a joint significance statistic of $F = 2.7$ is quite small in a setting with no institutional random assignment and with $N \approx 1.7\text{M}$.

Taken together, the evidence suggests that exogeneity violations generated by sample selection are unlikely. Officer stringency is uncorrelated with ticketing frequency and predicted offending and nearly uncorrelated with the full set of driver characteristics. Nonetheless, we take sample selection concerns seriously and subject our treatment effect estimates to a battery of associated robustness checks, described further in section 4.

3.2 Exclusion

The exclusion restriction requires that officer stringency affects future offending only through sanctions. Note that our strategy allows other (non-sanction) officer behaviors to affect drivers as long as those behaviors are uncorrelated with our stringency measure (Frandsen et al., 2019). On the other hand, features of the officer-driver interaction other than the sanction that cause a driver to change behavior would violate exclusion if those features are correlated with stringency.

Another plausible source of exclusion violations is downstream involvement in the traffic court system. As previously mentioned, stringency increases the likelihood that a driver contests a ticket in court and might influence traffic school elections. If anything about the court experience changes driver behavior, that could be considered an exclusion violation. However, whether traffic court involvement constitutes a violation of exclusion or simply a mechanism for fine effects is subject to interpretation. When viewed from the officer’s perspective, downstream events that are (i) caused by harsher sanctions and (ii) reduce reoffending still could be interpreted as a causal effect of sanctions themselves.⁶

Finally, the choice to bunch a driver indirectly affects the statutory “points” a driver receives on their license. In Florida, speeding offenses between 6 and 15 MPH over the limit carry 3 DL points, while speeding 16-29 MPH over the limit carries 4 DL points. Points can increase car insurance premiums and drivers that accrue sufficient points can face DL suspensions. As shown in figure A-4, officer stringency affects statutory points ($\beta_{FS} = 0.7$). However, drivers can mitigate their point exposure through the court system, and we find that, taking into account those downstream behaviors, there is almost no relationship between stringency and points, again shown in figure A-4. Hence, the burden of accrued license points cannot explain the effects we observe.

⁶Moreover, the evidence is largely inconsistent with the court system playing an important role in generating the treatment effects. As shown in figure B-4, treatment effects are very similar for local and non-local drivers. Because drivers need to travel to the citation county to attend court, local drivers are more likely to contest citations.

3.3 Monotonicity

Monotonicity violations are a natural concern in our setting given evidence of racial bias in officer leniency decisions (Anbarci & Lee 2014; Goncalves & Mello 2021). Importantly, Frandsen et al. (2019) show that IV estimates from examiner designs still recover the appropriate local average treatment effect under a weaker average monotonicity condition, which requires only that counterfactual reassignment to a more stringent officer increases the probability of harsh sanctions in expectation. In table A-2, we perform a standard check of the average monotonicity assumption by estimating our first stage regression for subgroups of motorists. We find that the first stage is statistically strong and that estimated magnitudes are remarkably consistent across subsamples.

However, our baseline marginal treatment effects approach in section 5 relies on a strict monotonicity assumption. In figure A-3, we present an ad-hoc version of the Frandsen et al. (2019) test of the joint assumptions of exclusion and strict monotonicity.⁷ Our version of the test is qualitatively useful but somewhat challenging to interpret because it does not account for estimation error in officer stringency and thus systematically over-rejects the joint null of exclusion and monotonicity. On the one hand, the test statistic is reassuringly small ($F = 2.3$); on the other hand, the (biased) p -value suggests that deviations from exclusion and strict monotonicity may be a salient concern.

We take several approaches to address monotonicity concerns, none of which alter our conclusions. First, we repeat our analyses computing the stringency instrument within cells based on driver characteristics, allowing for violations of strict monotonicity across motorist groups. Second, we develop an alternative marginal treatment effect estimator relying on a much weaker monotonicity assumption, described further in section 5. Lastly, we re-estimate our treatment effect and marginal treatment effect parameters using a subsample of officers ($N = 1,725$) satisfying the Frandsen et al. (2019) test (see figure A-3 for further details).

4 Deterrence effects

Given our interest in how officers allocate harsh fines, characterizing the average causal effect of sanctions on motorist behavior is an important first step in our analysis. Our approach follows directly from the empirical framework discussed above. We estimate:

$$Y_{ijs} = \beta D_{ij} + \psi_s + u_{ijs}$$

⁷We were unable to use the code provided by Frandsen et al. (2019) due to computing constraints. Our ad-hoc version replicates the fit component of their test by flexibly fitting reoffending rates to the stringency instrument (conditional on beat-shift effects), computing residuals, and then testing the ability of officer effects to explain the residuals. However, our test yields a p -value which is biased *towards zero*, meaning that we will over-reject the null of monotonicity and exclusion, because we do not account for estimation error in the construction of the residuals.

by two-stage least squares, instrumenting D_i with officer stringency Z_{ij} .

Note that our stringency instrument solves an important identification challenge arising from the nonrandom assignment of punishments. Not only do statutory sanctions increase with offense severity, as shown in figure 1, but officers further manipulate fines, as discussed in section 2.4. Naive OLS estimates, presented in table B-1, illustrate both dimensions of the identification challenge well. A regression of one-year reoffending on the charged fine (in \$100’s) and beat-shift fixed effects gives $\hat{\beta} = 0.043$ ($se = 0.002$), suggesting that harsher fines increase reoffending. Adding officer fixed effects increases the estimate to $\hat{\beta} = 0.055$ ($se = 0.002$), highlighting the nonrandom sorting of motorists into sanctions by officers.

Given the assumptions discussed in section 3, our deterrence IV estimates will recover a local average treatment effect (LATE) for the subgroup of marginal drivers (Imbens & Angrist, 1994). In our setting, compliers are motorists who would be punished harshly by some officers but given a break by others. Alternatively, with a continuous instrument, it is useful to think of compliers as those motorists who are neither always-takers nor never-takers, where these are the groups of drivers that would be fined harshly by any officer or no officer, respectively.

An interesting feature of our setting is the full support of our stringency instrument. Specifically, our sample includes subsets of officers that *always* bunch drivers and *never* bunch drivers. Under the LATE assumptions, the presence of officers that always (never) issue harsh fines implies that no motorist is a never-taker (always-taker). Because all drivers are compliers for some value of the stringency instrument, differences between naive OLS estimates and IV estimates should be interpreted as attributable to selection bias, rather than due to characteristic differences in the complier population. Moreover, our estimated LATE will be quite close to the average treatment effect.⁸

4.1 Results

In figure 4, we show the dynamic relationship between officer stringency and traffic offending. Specifically, we plot estimated coefficients (and 95 percent confidence intervals) from regressions of the form:

$$Y_{ijst} = \beta_{\tau} Z_{ij} + \psi_s + u_{ijst}$$

where Y_{ijst} is an indicator for whether driver i receives a traffic citation in quarter τ , which are quarters *relative* to the focal FHP citation. In the figure, $\tau = 0$ corresponds to the exact date of the focal FHP citation and $\tau = k$ corresponds to k quarters before or after the focal citation. The figure illustrates that the stringency of the citing officer at $\tau = 0$ has no ability

⁸Worth mentioning briefly here is the fact that, while all motorists are compliers for some value of the instrument, the characteristics of individuals may differ by which value of the instrument induces them into treatment. In section 5, we estimate these “marginal complier” characteristics using our marginal treatment effects framework.

to predict offending over the previous eight quarters but predicts a stark decline offending immediately after the focal citation. Impacts persist over the first four quarters and fade out considerably thereafter. Over the year following the focal citation, the reduced form estimate is $\hat{\beta} = -0.017$ ($se = 0.0053$).

Encouragingly, the dynamic pattern in figure 4 speaks to the validity of our research design. In order for the observed patterns to be generated by differential sample selection, it would have to be the case that more stringent officers differentially stop drivers with comparable offending histories but who are just about to reduce their offending rates.

Table 3 presents IV estimates for the full set of one-year offending outcomes.⁹ Column 1 reports the lenient officer mean of the outcome variable. Columns 2 and 3 report IV estimates excluding and including controls for driver characteristics. To help interpret magnitudes, column 4 reports the implied fine elasticity, which is computed by regressing the outcome on the (continuous) fine amount, driver controls, and beat-shift fixed effects, instrumenting the fine amount with stringency, and then scaling the IV estimate by the ratio of the average fine and average reoffending rates for lenient officers.

We find that harsh fines reduce the likelihood of a new traffic offense in the following year by about 1.6 percentage points ($\epsilon = -0.07$). The majority of this effect is attributable to reductions in speeding offenses; a harsh fine reduces the likelihood of a new speeding offense in the next year by about 1.4 percentage points. The IV estimate is precisely estimated, with a 95 confidence interval of $(-0.017, -0.012)$. Our point estimate for speeding offenses represents an 8.5 percent decline relative to the lenient officer mean and implies a fine elasticity of -0.13 . In other words, our estimate implies that a doubling of the fine amount would reduce the likelihood of speeding recidivism by 13 percent.

Estimated impacts of harsh fines on non-speeding offenses are also statistically significant but less pronounced ($\epsilon = -0.06$). The finding that speeding sanctions reduce other traffic offenses is consistent with Gehrsitz (2017), who finds specific deterrence effects of short-term license suspensions imposed on speeders in Germany on all forms of traffic offending.

Consistent with reductions in traffic offending implying a true behavioral response on the part of drivers, we also find that harsh fines reduces the likelihood of crash involvement over the following year by between 0.2 and 0.3 percentage points ($\epsilon = -0.04$). While less precisely estimated than the effects on traffic offenses, the IV estimates for crash involvement are statistically significant at the 10 percent level.

Finally, following our discussion in section 2.5, the last row of table 3 reports IV estimates of the impact of harsh fines on the likelihood that a driver contests a ticket in court. Relative to a lenient officer mean of 0.26, we find that a harsh fine increases the likelihood of a contested citation by about 11 percentage points, or about 42 percent, consistent with our

⁹In the appendix, we present graphical versions of the reduced form estimates (figure B-1), dynamic versions of the reduced form for other outcomes (figure B-2), and the full set of first stage and reduced form estimates with and without controls (table B-2).

hypothesis that court aversion motivates officer lenience.

4.2 Robustness

In the appendix, we present results from a battery of robustness checks. Table B-3 shows that estimated deterrence impacts are not sensitive to the computation of the stringency instrument. In particular, estimates are larger when using the binary (buncher versus not) version of the instrument which passes a conventional randomization test (see table 2) and satisfies monotonicity by construction because stringent officers never bunch motorists. Estimates are also quite similar when recomputing the instrument within cells based on driver demographics, as shown in the bottom panel of table B-3. Figure C-4 shows similar deterrence estimates when restricting to our Frandsen et al. (2019) subsample.

In figure B-3, we examine whether our deterrence estimates can be explained by selection of motorists. First, following Feigenberg & Miller (2022) we show that our IV estimate is robust to dropping officers with selected samples based on driver observables. Next, we show results from a Heckman (1979) selection correction based on officers' ticketing frequency. Finally, we show that results are similar when further interacting our beat-shift effects with stretch-of-road fixed effects, constructed by mapping the subset of geocoded tickets ($N = 244,858$) to Florida roads.

4.3 Interpretation

As highlighted in section 2, motorists issued harsh and lenient fines face the same sanctions for future offenses. Hence, our estimates capture a specific deterrence effect (e.g., Nagin 2013), or a behavioral responses to the experience of punishment, rather than the effects of statutorily higher sanctions for future offenses.

In figure B-4, we show that treatment effects are similar for local residents and out-of-county drivers, suggesting a minimal role for the traffic court system in explaining treatment effects, since drivers need to travel to the citation county to attend traffic court. Figure B-4 also illustrates that offending responses are nearly identical for motorists with higher and lower incomes, proxied by the zip code of residence. A purely financial mechanism, whereby the deleterious effects of fines on financial situations cause individuals to stop driving, is therefore unlikely based on Mello (2021), who finds that negative effects of fines on financial situations are concentrated among low-income drivers.

A mechanism that seems particularly consistent with with the dynamic patterns in figure 4 is drivers updating their beliefs about sanctions (Dusek & Traxler, 2021). In figure B-5, we show that offending responses are stronger in the county of the focal ticket and larger for motorists that have been exposed to stringent officers in the past, both of which are consistent with a learning hypothesis.

5 How are sanctions allocated?

Having established that sanction decisions have important deterrence effects, we turn to our analysis of how officers choose whom to fine harshly, with a particular interest in whether officers make decisions to maximize public safety.

To study officers’ sanction choices, we extend our instrumental variables framework with a marginal treatment effects model to explore the reoffending potential outcomes of motorists with varying propensities to face harsh fines. The motorists shifted from lenient to harsh fines at the lowest values of our stringency instrument, or those on the margin of treatment at the lowest stringency levels, are the motorists most likely to face harsh fines. Or alternatively, we could think of these motorists as those “most prioritized” for harsh sanctions by officers. Motorists on the margin of receiving harsh fines at the highest stringency values are the least likely to face harsh fines, or the motorists “least prioritized” for harsh sanctions.

Given a pool of cited drivers and a willingness to issue harsh fines, the allocation of sanctions that minimizes the reoffending rate is the one that levies harsh fines on only the most responsive offenders. Hence, public safety maximization by officers implies *positive selection on gains*: the motorists most likely to face harsh sanctions should exhibit the largest (most negative) treatment effects. The logic of optimization tests based on marginal treatment effects originates with Roy (1951) and has been applied in a variety of settings with treatment effect heterogeneity (e.g., Cornelissen et al. 2016; Mogstad & Torgovitsky 2018; Deshpande & Li 2019; Finkelstein & Notowidigdo 2019; Ito et al. 2022).

While our first-order concern is whether officer behavior is consistent with safety maximization, we also aim to provide a more complete picture of how officers sort drivers into sanctions. To that end, we estimate both how treatment responses and counterfactual reoffending rates vary with the propensity to face harsh fines. We focus on offering a positive characterization of officer sanction choices here and then discuss potential reasons for officer behavior, and associated policy implications, in section 6.

5.1 Estimating marginal treatment responses

Each individual has a pair of reoffending potential outcomes Y_D that depend on treatment status, $D \in \{0, 1\}$, where $D = 1$ denotes that the individual has received the harsh fine. The realized outcome can be written with the switching regression $Y_i = Y_1 D_i + Y_0(1 - D_i)$. We specify the potential outcomes to have the form $Y_j = X\beta_j + U_j$, where j indexes treatment status, β_j is a counterfactual-specific vector of coefficients on motorist and stop characteristics X , and U_j is a random variable with $E(U_j|X) = 0$.

Treatment status follows the threshold crossing rule, $D = \mathbf{1}[\mu_D(Z) > U_D]$, which depends on characteristics Z , including both X and our stringency instrument, and unobservable U_D . Without loss of generality, we impose that U_D has a uniform marginal distribution so that $\mu_D(Z)$ can be interpreted as a propensity score, which we denote with $P(Z)$. Each individual

has a fixed value of U_D , which we call their *resistance to treatment*. The higher one’s value of U_D , the greater their realization of $P(Z)$ must be for that individual to take up treatment.

Our aim is to identify the expected value of the counterfactual reoffending outcomes for individuals at each resistance to treatment:

$$E(Y_j|X, U_D) = X\beta_j + E(U_j|U_D, X)$$

Following Mogstad et al. (2018), we label these the marginal treatment response (MTR) functions. The difference in MTR functions is the marginal treatment effect (MTE):

$$MTE(X, U_D) \equiv E(Y_1 - Y_0|X, U_D) = X(\beta_1 - \beta_0) + E(U_1 - U_0|U_D, X)$$

In our estimation, we make the simplifying assumptions that the unobserved components of the MTRs are linear in U_D and independent of X . This second assumption imposes that all differences in counterfactual outcomes across observables are captured by the observable components $X\beta_j$.

We describe our approach to estimating the MTE and MTR functions briefly here and provide a more detailed explanation in appendix E-1. We take the “separate” approach of Heckman & Vytlacil (2007) and Brinch et al. (2017) to estimating the MTR functions for Y_0 and Y_1 . We first estimate the propensity score p_i for each individual by regressing D_i on the officer stringency instrument and beat-shift fixed effects and constructing a predicted treatment $\hat{D}_i \equiv p_i$. We then regress Y_i on beat-shift effects and the estimated propensity score, *restricting the sample to either untreated or treated drivers*, $D_i = 0, 1$. The coefficients on the fixed effects provide the level of the MTR, and the coefficient on the propensity score provides its linear slope term. We calculate the MTE as the difference between the two MTR’s. We present the estimated MTR and MTE functions for the average values of the fixed effects over all drivers in the sample.

At baseline, we are interested in whether officers sort drivers to maximize safety (or equivalently, minimize the reoffending rate). Hence, we start by estimating marginal treatment responses conditioning only on beat-shift fixed effects. We then consider the sensitivity of our estimates to including driver covariates in X , which yields MTR’s that reflect sorting within demographic groups. Throughout our presentation of MTR results, we show 95 percent confidence bands and report estimated slope standard errors obtained from a bootstrap clustered at the officer-level.

5.2 Baseline MTR results

Panel (a) of figure 5 presents our baseline estimates of the marginal treatment effects, $E(Y_1 - Y_0|U_D)$. We find that marginal treatment effects become more negative as resistance to treatment increases, indicating that the motorists *least likely* to face harsh fines are the *most deterred* by harsh fines. Deterrence impacts for the motorists most likely to be sanctioned

harshly ($U_D = 0$) are, if anything, marginally positive. The estimated MTE slope ($\Delta = -0.07$) is statistically significant at conventional levels. This pattern of reverse selection on gains is exactly the *opposite* of what we would expect if officers were sorting drivers to minimize the reoffending rate.¹⁰

The second panel of figure 5 illustrates our baseline estimates of the marginal treatment response functions, $E(Y_j|U_D)$ for $j \in \{0, 1\}$. Motorists with the lowest values of U_D , or those most likely to face harsh fines, exhibit the highest reoffending rates in either treatment state. In other words, the sorting of drivers into sanctions generates selection on levels, with officers most likely to issue harsh tickets to drivers with the highest reoffending rates.

The selection patterns we document in figure 5 suggest a few important takeaways. First, as explained above, reverse selection on gains is inconsistent with officers sorting drivers into sanctions to minimize the reoffending rate. Second, the opposite slopes of the MTE and MTR functions suggest that the most deterrable offenders are the motorists least likely to reoffend in the future, independent of their treatment status. And finally, the figures suggest that officers prioritize punishing drivers with the highest reoffending rates, at the expense of targeting sanctions to the most deterrable motorists, when allocating sanctions.¹¹

There are several potential explanations for the selection patterns we observe. Officers may sort drivers based on salient characteristics such as offending history or race (e.g., Goncalves & Mello 2021) that are incidentally correlated with reoffending levels and negatively correlated with deterrability. Moreover, officers may prioritize sanctioning drivers based on salient observables but then order motorists based on deterrability within those groups. Officers may also sanction drivers based on offense severity or, in this case, a driver’s speed relative to the posted limit. Alternatively, officers may sort drivers to optimize a different objective than total offending. We discuss these potential explanations below.

5.3 Sorting on other dimensions

5.3.1 Driver characteristics

Evidence that officers consider driver characteristics when allocating harsh fines can be seen in column 2 of table 2, which illustrates that our full set of driver covariates are jointly

¹⁰In Goncalves & Mello (2021), we found suggestive evidence that officers do not sort drivers to maximize deterrence in a robustness test for whether statistical discrimination could explain racial disparities in the likelihood of receiving lenient fines. Our current analysis not only further establishes this fact by estimating the MTE using different approaches, but also explores reasons for this officer behavior by exploring selection on levels and sorting on other dimensions.

¹¹These patterns are also summarized in table C-2, which shows that the Average Treatment Effect on the Treated (ATT) is a small and statistically insignificant -0.004 , compared to a statistically significant -0.04 Average Treatment Effect on the Un-Treated (ATUT). The average untreated level of reoffending (Y_0) for harshly fined motorists is significantly higher than for leniently fined motorists (0.379 v. 0.330).

predictive of a harsh charge ($F = 28.8$) after conditioning on beat-shift effects. We can further demonstrate the role that covariates play by estimating the characteristics of marginal compliers, or motorists on the margin of receiving harsh fines at each level of resistance to treatment. Our approach to estimating marginal complier characteristics, which follows directly from our strategy for estimating marginal treatment responses, is described in appendix E-3. By observing how complier characteristics change with resistance to treatment, we can ascertain the characteristics of motorists prioritized for harsh sanctions.

Panel (a) of figure 6 illustrates the share of marginal compliers with a traffic citation in the past year and the predicted reoffending rate, based on driver covariates, of marginal compliers. Both are sharply negatively correlated with resistance to treatment. The motorists most likely to face harsh fines ($U_D = 0$) are about 60 percent more likely to have committed a traffic offense in the past year than those least likely to face harsh fines ($U_D = 1$) and exhibit about 25 percent higher predicted reoffending rates. In figure C-10, we show identical figures for other driver demographics; drivers prioritized for harsh sanctions are more likely to be young, male, Black or Hispanic, and live in slightly lower-income neighborhoods.

However, panels (b) and (c) of figure 6 show that sorting based on driver characteristics cannot wholly explain the selection patterns we found in figure 5. These figures recompute our MTE and MTR estimates after conditioning on driver characteristics, which we parameterize with 32 fixed effects at the level of the driver’s gender \times race \times $\mathbf{1}[\text{age} \geq 35]$ \times $\mathbf{1}[\text{any citation in the past year}]$.¹² Relative to our baseline estimates, the within-covariate MTE and MTR functions are flatter but exhibit the same selection patterns. The estimated within-covariate MTE slope ($\Delta = -0.05$) is about 25 percent smaller than our baseline estimate but remains negative and statistically significant. The within-covariate untreated MTR slope ($\Delta = -0.06$) is about 40 percent smaller than our baseline estimate but also remains negative and statistically significant. In other words, the pattern of selection on levels and reverse selection on gains persists when estimating MTR’s within driver covariates.

Given the apparent salience of driving history for officer decisions, in figure 7, we examine sorting patterns when focusing only on first-time offenders. For this analysis, we restrict the sample to tickets issued from 2010–2016 (the “late” sample), which allows us to observe a full five years of past offending for all motorists in the sample. We consider a driver to be a first-time offender if they have no citations in the previous five years.

In panel (a) of figure 7, we show the estimated MTE for the full late sample and for the subset of first-time offenders. Restricting to the late sample has no impact on the baseline MTE (or MTR) estimates, relative to our main sample. Focusing on first-time offenders yields a flatter, but still downward-sloping MTE curve ($\Delta = -0.025$). We attempt to further purge these estimates of demographic biases by officers by restricting the sample to white first-time offenders ($\Delta = -0.012$) and white, first-time offenders ticketed by only

¹²In figure C-2, we show that the estimated MTE and MTR functions are nearly identical when using an expanded set of driver covariates.

white officers ($\Delta = -0.002$). The final restriction gives an MTE which is essentially flat.

Panel (b) reports the same set of estimates for the untreated MTR function, $E(Y_0|U_D)$. Because first-time offenders are less likely to reoffend, restricting to first-time offenders shifts down the MTR curve dramatically. The slope, however, is essentially unchanged when examining only first-time offenders ($\Delta = -0.11$ versus $\Delta = -0.099$). Our additional restrictions based on driver and officer race have almost no impact on the slope estimate.

The result that the extent of selection on levels that we see at baseline persists through these sample restrictions is sufficiently interesting to warrant some emphasis. The MTR curves in figure 7 suggest that, even among motorists with no offending history, officer sanctions decisions, either implicitly or explicitly, differentiate motorists on the basis of future recidivism risk.

5.3.2 Retribution

Another possible explanation for the sorting patterns we observe is that officers sort drivers based primarily on offense severity, or, in this case, the observed speed (i.e., the true rather than manipulated speed). If driving speed were positively correlated with the likelihood of reoffending and negatively correlated with treatment effects, sorting based on speed could explain the selection patterns that we document.

The fact that our sample includes a subset of officers that never bunch drivers allows us to estimate the “true” speeds of marginal compliers, or those motorists induced into treatment at varying levels of resistance to treatment, using the approach detailed in section E-3. Unsurprisingly, we find evidence that officers consider speed when sorting drivers, as shown in panel (a) of figure 8. We estimate that the motorists most prioritized for harsh sanctions ($U_D = 0$) were observed driving about 1.5 MPH faster than those drivers least prioritized for harsh sanctions ($U_D = 1$). We interpret this as evidence that, at least to some extent, officers exhibit retributivist preferences (e.g., Kaplow & Shavell 2006). In other words, officers prefer to allocate harsher sanctions to more serious offenders.

To quantify the contribution of this sorting based on “guilt” to the overall selection patterns we observe, we first simulate a counterfactual marginal compliers curve under the assumption that officers consider only speed when sorting drivers, which we also present in panel (a) of figure 8. This predicted marginal compliers curve ($\Delta \approx -13$) is dramatically more pronounced than the one we estimate in the data ($\Delta \approx -1.5$). Hence, while we find that officers do consider speed when choosing which drivers to punish harshly, speed alone can explain only a relatively small share of the overall sorting (about 12 percent based on a comparison of the two slopes).

Next, we ask whether this sorting based on offense severity can explain the selection on levels we observe at baseline. In panel (b) of figure 8, the maroon dashed line is our baseline estimate of the treated MTR curve, replicated from figure 5. The blue dashed line is a predicted version of the same curve based on (i) the extent of speed sorting and (ii)

the estimated relationship between true speed and reoffending.¹³ Comparing the two slopes suggests that speed-based sorting can explain about 18 percent of the selection on levels that we document in figure 5.

5.3.3 Other objectives

We also consider the possibility that officers sort drivers into sanctions to maximize another objective. For example, our measure of any traffic reoffending may not accord with the officers’ notion of maximizing public safety. In figure C-6, we present MTR and MTE estimates replacing the outcome of interest with whether a driver commits a new speeding offense and whether a driver is involved in a crash (both measured over the following year). In both cases, we find downward sloping MTE curves, inconsistent with officers allocating sanctions to maximize these alternative notions of safety.

In figure C-7, we ask whether officers allocate sanctions to minimize their time spent in traffic court. Replacing the outcome of interest with our measure of whether a driver contests their citation in court, we estimate an MTE which is inconsistent with officers optimizing on this margin. Motorists facing harsh sanctions most often are, in fact, the most responsive to harsh fines in terms of contesting. Interestingly, unlike motorist reoffending, officers receive direct feedback on contested tickets. Hence, we might particularly expect to see a “correctly” signed MTE curve here if minimizing court time were the true objective.

One might also ask whether officers allocate sanctions to maximize *general*, rather than *specific*, deterrence effects. Our analysis of marginal treatment responses focus on the question how officers choose which drivers receive harsh sanctions, holding constant the share of harsh citations. If drivers only consider the expected fine associated with speeding, there are no general deterrence effects associated with changing *which* drivers are punished harshly. There could be general deterrence effects associated with officers’ sorting choices if drivers perceive that their characteristics affect the likelihood of facing harsh fines, but this would entail a complex calculation by potential offenders, made more difficult by that fact that officer inference based on unobservables plays an important role in their decisions. Moreover, the fact that drivers learning about fines appears to be a relevant mechanism for the specific deterrence effects (see section 4.3) suggests that drivers may be, ex ante, unaware of average stringency levels, allowing minimal scope for general deterrence effects.

5.4 Robustness

There are two important caveats associated with our marginal treatment response estimates. The first is that our baseline approach imposes a restrictive linear parametric structure on

¹³Note that we can only simulate the treated MTR curve because true speed is observed only for treated (not bunched) drivers. See section E-4 for further details on the estimation of the predicted MTR curve in figure 8.

potential outcomes. In figure C-9, we present marginal treatment effect estimates using both a quartic functional form and two non-parametric approaches. While some curvature in the MTE is suggested by the more flexible fits, the polynomial and non-parametric MTE’s are consistent with our linear estimate and rule out a globally increasing marginal treatment effect, as would be predicted by deterrence maximization.

The second concern is that the empirical model underlying our MTR estimates implies the strict monotonicity of Imbens & Angrist (1994), since all individuals who take up treatment at a given value of $P(Z)$ would also take up treatment at greater values (Vytlacil, 2002). Intuitively, strict monotonicity, which implies that all officers share a common “ranking” of drivers for harsh sanctions, allows us to learn about a *within-officer* object using *across-officer* variation. In our setting, this is a very strong assumption that warrants caution.

Our first approach to relaxing the monotonicity assumption is to re-estimate the MTR functions using a stringency instrument that is re-computed within our baseline driver covariate cells, which allows for violations of strict monotonicity between, but not within, groups of motorists. We also estimate MTR functions that are allowed to vary by officer characteristics, defined as cells at the level of gender \times $\mathbf{1}[\text{nonwhite}] \times \mathbf{1}[\text{any college}] \times \mathbf{1}[\text{experienced}]$. Here, we use the “late” sample and define an officer as experienced if they issued an above median number of speeding tickets over 2005–2009. When we combine these two approaches, the weakened monotonicity assumption applies only within the interaction of each officer \times motorist group. As shown in figure C-3, the re-estimated untreated MTR and MTE functions are almost identical to our baseline within-covariate estimates.

We also leverage a unique feature of our setting, the full support of our stringency instrument (see figure C-1), to estimate a version of the MTE that satisfies monotonicity by construction. The idea of our method, which we call the *tails* approach, is to estimate the impact of being counterfactually reassigned from an officer with $Z \approx 0$ to an officer with a slightly positive Z . This yields an estimate of the marginal treatment effect near $Z = 0$ which satisfies monotonicity because $E(D|Z = 0) = 0$; in other words, there are no defiers. We can then estimate the marginal treatment effect at $Z \approx 1$ following the same logic and estimate the slope of the MTE curve using these two points. As illustrated in figure C-3, our estimated within-covariate MTE using the *tails* approach is almost identical to our benchmark linear estimate ($\Delta = -0.38$ and $\Delta = -0.04$), although the tails slope is less precisely estimated. We provide a more detailed discussion of estimation for the *tails* MTE in section E-2 and probe the robustness of the tails estimates in table C-1.

In figure C-4, we show estimated within-covariate MTE and MTR functions using the Frandsen et al. (2019) subsample of officers, which is a subgroup of 1,725 officers (88 percent of our original sample) that easily pass the Frandsen et al. (2019) joint test of monotonicity and exclusion. Sorting patterns are unchanged when examining this subset of officers. If anything, selection on levels is more pronounced in this subsample, while reverse selection on gains is slightly attenuated. The estimated MTE slope ($\Delta = -0.033$), however remains

negative and statistically significant at the 10 percent level.

Finally, we replicate the approach of Feigenberg & Miller (2022), detailed in section E-6, which requires only that monotonicity hold within patrol locations. Figure C-8 plots the non-parametric relationship between reoffending and stringency based on their approach. The curvature in this relationship is consistent with our baseline MTE estimate; the safety “return” to harsh sanctions is smallest at lower values of stringency and largest at higher values. Figure C-9 plots the non-parametric MTE implied by the Feigenberg & Miller (2022) approach, which is remarkably consistent with our linear estimate.

6 Policy discussion

The reverse selection on gains that we document implies that a lower aggregate recidivism rate could be achieved by reallocating harsh fines, holding fixed the population of cited drivers and the share of harsh tickets. In other words, there are efficiency costs, in the form of foregone deterrence, associated with current officer practices.

These efficiency costs, which we quantify below, are generated by the sanction choices of officers and exist regardless of the rationale for current officer behavior. However, the correct policy for mitigating efficiency losses may depend on the underlying reasons for officers’ sanction choices. In terms of characterizing officer behavior, the stark and robust selection on levels that we document is a particularly salient takeaway from our analysis of sorting patterns. Having ruled out that selection patterns can be explained by a targeting of sanctions based on easily observable characteristics in section 5.3, there are two key theories that can rationalize the simultaneous selection on levels and reverse selection on gains that we find. We discuss each in turn below.

6.1 Information problems

The officer behavior we document could result from information problems. Officers may seek to allocate sanctions in a way that maximizes public safety but have poor information on how to achieve that goal. The selection patterns we document, for example, are consistent with officers striving to minimize the reoffending rate but having an incorrect model for predicting treatment effects. In particular, officers may be able to discern recidivism risk but mistakenly believe that recidivism risk and treatment effects are positively correlated, or may rely on a model that yields treatment effect predictions which are more closely related to recidivism risk than to actual treatment effects.

Importantly, we should note that poor signals of deterrability alone cannot rationalize the estimated marginal treatment effects under the null hypothesis of deterrence maximization. Noisy signals of driver deterrability would attenuate the slope of the marginal treatment effects curve but would still generate (weakly) larger treatment effects for the motorists

most likely to face harsh fines. Our estimated MTE slopes, however, are wrong-signed and statistically significant.

Officers may also be unaware of treatment effect heterogeneity. With homogeneous treatment effects, any allocation of punishments generates the same aggregate recidivism rate, and prioritizing the motorists with the highest recidivism risk for harsh fines would enhance safety by reducing the total number of future offenses.¹⁴

If officers would like to allocate sanctions to maximize safety but information problems prevent them from doing so, a simple information intervention could be an effective policy for changing officer sanction choices and, as a result, reducing the reoffending rate. For example, the FHP could notify officers of the safety gains associated with harshly punishing the motorists that are typically issued lenient fines.

Officers may respond to this information by marginally increasing their stringency to include motorists whom they currently cite leniently. To consider the potential impacts of such a response, we conduct a simple policy simulation where all propensity scores are increased by 10 percent (up to a value of 1) relative to the estimated levels in the data. The treatment effect for the marginally impacted offenders is a commonly known form of the “policy-relevant treatment effect” (Heckman & Vytlacil 2007; Cornelissen et al. 2016). These estimates are presented in table C-2. Individuals induced into harsh punishments by this policy shift have a deterrence effect that, while marginally significant (-0.009 , $se = 0.005$), is double the impact for the motorists who are already fined harshly (-0.004 , $se = 0.008$).

Alternatively, officers may respond to an information intervention by holding fixed their average stringency levels but switching who receives harsh fines. We consider the impacts of such a change in officer behavior below in section 6.3.

6.2 Officer preferences

Alternatively, the selection patterns we document could reflect officer preferences. All the evidence we present is consistent with the view that officers prioritize drivers for harsh fines based on recidivism risk. In other words, officers may hold a different notion of the optimal allocation of sanctions than safety maximization and explicitly strive to ensure that the “worst” drivers, as captured by their likelihood of reoffending, face harsh fines. Such a preference would accord well with a notion of fairness or justice, and it seems quite plausible that officers desire a “fair” allocation of sanctions in this setting. In this case, an important goal may not be to change officer behavior, which is already optimal given their preferences, but instead to understand the efficiency costs associated with a preference for fairness.

Officers may also make sanction decisions based on other unobservable characteristics

¹⁴In figure C-11, we show that the pattern of selection on levels and reverse selection on gains holds when replacing the outcome of interest with a driver’s number of traffic offenses over the following year in our MTE and MTR estimates.

which are incidentally positively correlated with recidivism risk and negatively correlated with treatment effects. For example, officers may prefer to issue harsh sanctions only to drivers who behave confrontationally during traffic stops, and this behavior happens to predict higher reoffending rates and lower levels of deterrability.

Both of these theories are consistent with our MTE and MTR analyses. We treat them as conceptually equivalent as well, because in either case, officers hold a different notion of the optimal allocation of sanctions than safety maximization and pursue that allocation, independent of deterrence effects. Because the motorists with the highest recidivism risk are the least deterrable, officers face an explicit tradeoff between pursuing these other goals and reducing the reoffending rate.

Under the assumption that selection patterns are generated by officer preferences, we characterize those preferences by specifying and estimating a stylized model of sanction choices. The model allows officers to value both deterring crime and punishing individuals with high reoffending rates, which they balance against a private cost of issuing harsh sanctions. The parameter of interest in the model is the weight that officers place on recidivism risk, as opposed to deterrability, when choosing between harsh or lenient sanctions. Based on our analysis of sorting patterns, this model may be a reasonable representation of officer preferences and the weight captures the rate at which officers are willing to trade off deterrence for punishing the “worst” drivers, or alternatively, for attending to the underlying preference that generates selection on levels. Importantly, this model still allows for information problems, but makes a conceptual distinction from our discussion above in that we allow officers to depart from holding only safety maximization goals.

6.2.1 Officer decision model

Officer j is randomly matched to driver i . The driver is defined by a pair of binary potential outcomes, $(Y_{i0}, Y_{i1}) \in \{0, 1\}^2$, which indicate whether the driver will reoffend after the stop if receiving the lenient or harsh punishment, respectively. The officer does not necessarily see the driver’s true values for Y_{i0}, Y_{i1} but instead sees signals of each outcome, $\hat{Y}_{i0}, \hat{Y}_{i1}$, which are drawn from a cumulative distribution function $F(\hat{Y}_{i0}, \hat{Y}_{i1})$. While we allow officers to have only noisy signals of driver offending behavior, we impose that they reflect the expected value of each outcome: $E(Y_{ik}|\hat{Y}_{ik}) = \hat{Y}_{ik}$, $k = 0, 1$.¹⁵ This assumption is the key distinction

¹⁵This formulation of the information structure is general, and it can represent more specific models where the officer observes a signal S of $Y_i(0), Y_i(1)$ and constructs posteriors $E[Y_i(0)|S]$, $E[Y_i(1)|S]$. For example, an alternative model is that officers observe noisy signals, $\tilde{Y}_0 = Y_0 + \epsilon_{i0}$ and $\tilde{Y}_1 = Y_1 + \epsilon_{i1}$, where the error terms are mean zero and jointly normal, with variances σ_0^2 and σ_1^2 and correlation coefficient ρ . The officer would then take these noisy signals and, with the baseline rates of each pair of potential outcomes, infer values for the true potential outcomes, $E(Y_0|\tilde{Y}_0, \tilde{Y}_1) \equiv \hat{Y}_0$ and $E(Y_1|\tilde{Y}_0, \tilde{Y}_1) \equiv \hat{Y}_1$. This model generates a joint CDF of the signals $F(\hat{Y}_0, \hat{Y}_1)$ for a given set of model parameters. However, only a *subset* of functions $F(\hat{Y}_0, \hat{Y}_1)$ can be represented by this signal structure. We therefore allow for any distribution of posteriors and do not explicitly model the

from the previous section, where information problems may lead officers to have the “wrong” model of the world.

Officers receive some utility from issuing the harsh fine, which is increasing in the individual’s expected offending rate \hat{Y}_{i0} and in the (negative of the) treatment effect of the harsh fine:

$$u = \lambda \hat{Y}_{i0} - (1 - \lambda)(\hat{Y}_{i1} - \hat{Y}_{i0}) - c_j$$

The variable c_j is an officer-specific cost of issuing a harsh fine, which generates differences across officers in average stringency. The parameter λ reflects the relative weight an officer places on punishing individuals likely to reoffend versus punishing deterrable individuals. If officers have $\lambda = 1$, they care only about the motorist’s recidivism risk; $\lambda = 0$ indicates that they care only about deterrence. This weight is our main object of interest.

The utility of issuing the lenient fine is normalized to zero, which leads to the punishment rule $D = \mathbb{I}(\lambda \hat{Y}_{i0} - (1 - \lambda)(\hat{Y}_{i1} - \hat{Y}_{i0}) \geq c_j)$ for a given driver. Similarly, an officer’s probability of harsh punishment is given by $\theta_j = Pr(\lambda \hat{Y}_{i0} - (1 - \lambda)(\hat{Y}_{i1} - \hat{Y}_{i0}) \geq c_j)$. We suppose that all variation across officers in their behavior is due to differences in c_j , so that all officers have the same skill in identifying a driver’s potential outcomes (i.e., face the same posterior distribution $F(\hat{Y}_{i0}, \hat{Y}_{i1})$) and face the same trade-off between targeting levels and differences in reoffending. This assumption means that the values of θ and c are one-to-one, and we can write one as an invertible function of the other, $\theta = g(c)$, $c = g^{-1}(\theta)$.

6.2.2 Identification and estimation

Our goal is to map this model onto our estimated marginal treatment response functions to identify the values of λ that are consistent with the data. The model outputs corresponding to these estimates are the values of treated and untreated offending rates for individuals who are at the margin of punishment for officers at a given propensity to treat:

$$h_k(\theta) \equiv E[Y_k \mid \lambda \hat{Y}_1 - (1 - \lambda)(\hat{Y}_1 - \hat{Y}_0) = g^{-1}(\theta)], \quad k \in \{0, 1\}$$

Because the estimated MTR functions are linear, they are each characterized by two moments. Our model is therefore constrained to match four moments. The model contains the weight parameter λ and the distribution of signals $F(\hat{Y}_{i0}, \hat{Y}_{i1})$. Unless we place substantial restrictions by parameterizing the distribution of signals with three or fewer parameters, the model parameters are not point identified from the marginal treatment responses. However, they may provide informative bounds on their true values. We focus, in particular, on estimating the identified region for λ . We do so by solving optimization problems to find the smallest and largest values of λ such that, for some corresponding distribution of signals $F(\hat{Y}_{i0}, \hat{Y}_{i1})$, $h_k(\theta)$ falls within the confidence intervals of the four empirical MTR moments.

signals and officer inference based on those signals.

Details of the model estimation are provided in appendix E-5. For simplicity, we estimate a single value of λ across all officers in the sample, which we interpret as the average λ . In figure C-5, we estimate MTE and MTR curves that vary across 16 groups of officers based on gender, race, education, and experience. We cannot reject the null hypothesis that the MTE or MTR slopes are equal across officer groups, which we take as supporting evidence that our single parameter estimate provides a reasonable approximation to the data.

6.2.3 Model estimates

We estimate that λ lies in the interval $[0.61, 1]$. At the lowest value, officers place slightly more weight on recidivism risk than deterrability, but value both objectives. Because the MTR functions show that drivers most likely to be punished have the highest values of Y_0 , the data are also consistent with $\lambda = 1$, meaning that officers place no weight on the deterrability of motorists. Importantly, our estimates do rule out the possibility that officers care only about deterrence, as $\lambda = 0$ is inconsistent with the fact that deterrability is highest among motorists less likely to be sanctioned harshly.

6.3 Quantifying efficiency costs

To quantify efficiency costs associated with current officer behavior, we examine how the aggregate reoffending rate changes under counterfactual reallocations of harsh fines.¹⁶ We compute two related counterfactuals which correspond closely to the two theories for officer behavior we discuss above. First, we consider a counterfactual where officers sort drivers in reverse order of their current practice, so that a driver’s new (counterfactual) resistance to treatment is $\tilde{u} = 1 - u$. For example, a driver with $u = 0.1$ is induced into treatment by a propensity score of $p > 0.1$, and we now suppose they have $\tilde{u} = 0.9$ and are only induced into treatment by a propensity score of $p > 0.9$. This reversal leads to a mirroring of the marginal treatment responses, so that drivers least resistant to treatment now are more deterred and less likely to reoffend with either treatment status.

We should highlight that this counterfactual is closely related to a feasible, at least in theory if not in practice, policy change. The reverse-sorting approach generates a similar allocation of sanctions as a hypothetical mechanism that forces officers to make the opposite of their preferred sanction decision during each traffic stop. Hence, this counterfactual also highlights the important point that no additional information about motorists is necessary to achieve safety gains because current sorting practices successfully differentiate drivers based

¹⁶A different counterfactual of interest might consider the effects of removing officer discretion altogether. Such a policy change would induce both specific deterrence effects (by increasing fines on cited drivers) and general deterrence effects (by increasing the expected fine associated with speeding). Because there is no quasi-experimental variation available to estimate general deterrence effects in our setting, we do not consider this counterfactual policy.

on their deterrability but allocate punishments inefficiently.

The second counterfactual is based on our estimated officer decision model, setting $\lambda = 0$. We do so for every set of parameter values in our partially identified range, meaning that our estimates yield a range of counterfactual outcomes. In both calculations, we hold the overall probability of receiving a harsh fine fixed at the observed probability. These counterfactuals speak directly to the efficiency cost associated with current officer practices by quantifying the additional number of traffic offenses that occur relative to a scenario where sorting generates selection on gains. Our counterfactuals differ from the typical calculations based on marginal treatment effects, which consider the impact of changes to the distribution of treatment probabilities but hold fixed the resistance to treatment of each individual (Cornelissen et al. 2016; Mogstad & Torgovitsky 2018). Instead, our counterfactuals focus on changing how officers sort drivers into treatment.

The counterfactual estimates are presented in table 4. The first row presents the baseline reoffending rates and responses of individuals treated and untreated with the harsher punishment, where we use the estimates from the MTR functions. Again, treated drivers have higher ex-post offending rates than the untreated (0.376 v. 0.327) and are less deterrable (-0.001 v. -0.038). The second row shows the range of corresponding values when drivers are now sorted using their reversed resistance to treatment. Treated drivers are now more deterrable than the untreated (-0.027 v. 0.01). Unsurprisingly, this improved targeting leads to a lower overall reoffending rate of 0.341, relative to 0.358 observed in the data, a 1.7 percentage point (4.7 percent) reduction.

The third row shows the range of corresponding values from the officer decision model when λ is set to 0. The same tradeoff as in the previous counterfactual is apparent. The offending rate of treated individuals falls to between 0.335 and 0.338, and the untreated individuals are now worse offenders, with reoffending rates in the range of 0.362 to 0.363. As expected, the deterrability of treated drivers increases substantially in magnitude, from -0.001 to a range of -0.032 to -0.026 . In contrast, the deterrability of the untreated motorists decreases from -0.038 to a range of -0.010 to -0.005 .¹⁷

The improved targeting of harsh sanctions towards more deterrable drivers translates into fewer traffic offenses, with the overall reoffending rate declining from 0.358 to a range between 0.332 and 0.337. Hence, the weight that officers place on recidivism risk carries meaningful efficiency costs. Current officer practices forgo a 2.1 to 2.6 percentage point (5.8 to 7.3 percent) decline in the reoffending rate.

¹⁷In section C-1, we also consider the implications of counterfactual sanction allocations for racial disparities. Within beat-shifts, Black and Hispanic drivers are more likely to face harsh fines. These racial gaps are *flipped* in the reverse-sorting counterfactual.

7 Conclusion

In this paper, we study the public safety implications of police discretion over sanctions for speeding offenses. First, relying on variation across officers in the propensity to issue harsh fines, we show that sanctions decisions have important deterrence effects. Comparing motorists cited in the same beat-shifts by officers of varying stringency, we find that higher fines reduce the likelihood of a new traffic offense ($\epsilon = -0.07$), a new speeding offense ($\epsilon = -0.13$), and crash involvement ($\epsilon = -0.04$) in the following year.

Given that patrol officers are bureaucrats tasked by the state with producing safety, we then ask whether officers allocate sanctions to maximize deterrence effects. We use a marginal treatment effects approach to characterize the selection patterns generated by officers' sanction choices. Based on an underlying Roy (1951) framework, this exercise reveals which motorists are prioritized by officers for harsh fines. Current officer sorting practices generate *reverse* selection on gains, inconsistent with deterrence maximization by officers. The motorists most likely to face harsh fines are the least deterred by those fines.

To further explore why fines are inefficiently allocated, we also examine sorting on other dimensions. In particular, we document stark, positive selection on levels. Motorists most likely to be sanctioned harshly exhibit the highest probability of reoffending whether treated or untreated. Hence, officer sorting behavior results in a targeting harsh sanctions to the “worst” drivers, or those with the highest recidivism risk, at the expense of punishing the most deterrable offenders. We can rule out that these selection patterns are explained by officers sorting drivers based on observable characteristics or offense severity.

The selection patterns we document suggest either that officers strive to maximize deterrence but act on incorrect information about treatment effect heterogeneity or that officers allocate sanctions in pursuit of some other objective. In the case that information problems alone explain officer behavior, we discuss potential policy responses. Under the assumption that selection patterns reflect officer preferences, we specify and estimate a model of sanctions choices intended to quantify the rate at which officers are willing to trade off deterrence to accomplish other punishment goals.

Nonetheless, the reverse selection on gains that we document implies that a lower aggregate reoffending rate could be achieved with a reallocation of harsh sanctions, holding fixed the rate of harsh ticketing and the information available to officers. We quantify these potential safety gains by estimating counterfactual reoffending rates under different allocation rules. Our estimates indicate that the reoffending rate would decline by between four and seven percent were officers to allocate sanctions in a way that generates selection on gains, highlighting the efficiency costs associated with current officer practices.

References

- Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American statistical Association*, *97*(457), 284–292.
- Abaluck, J., Agha, L., Kabrhel, C., Raja, A., & Venkatesh, A. (2016). The determinants of productivity in medial testing: Intensity and allocation of care. *American Economic Review*, *106*(2), 3730–64.
- Abdulkadiroglu, A., Pathak, P., & Schellenberg, J. (2020). Do parents value school effectiveness? *American Economic Review*, *110*(5), 1502–1539.
- Abrams, D., Goonetilleke, P., & Fang, H. (2021). Do cops know who to stop? assessing optimizing models of police behavior with a natural experiment. *Unpublished manuscript*.
- Akhtari, M., Moreira, D., & Trucco, L. (2022). Political turnover, bureaucratic turnover, and the quality of public services. *American Economic Review*, *112*(2), 442–93.
- Anbarci, N. & Lee, J. (2014). Detecting racial bias in speed discounting: Evidence from speeding tickets in Boston. *International Review of Law and Economics*, *38*, 11–24.
- Ashenfelter, O. & Greenstone, M. (2004). Using mandated speed limits to measure the value of a statistical life. *Journal of Political Economy*, *112*(S1), S226–S267.
- Ba, B., Knox, D., Mummolo, J., & Rivera, R. (2021). Diversity in policing: The role of officer race and gender in police-civilian interactions in Chicago. *Science*, *371*(6530), 696–702.
- Banfield, E. (1975). Corruption as a feature of governmental organization. *Journal of Law and Economics*, *18*(3), 587–605.
- Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, *76*(2), 169–217.
- Bendor, J., Glazer, A., & T, H. (2001). Theories of delegation. *Annual Review of Political Science*, *4*, 235–268.
- Best, M. C., Hjort, J., & Szakonyi, D. (2017). Individuals and organizations as sources of state effectiveness. Technical report, National Bureau of Economic Research.
- Bhuller, M., Dahl, G., Loken, K., & Mogstad, M. (2020). Incarceration, recidivism, and employment. *Journal of Political Economy*, *128*(4), 1269–1324.
- Bjorklund, A. & Moffitt, R. (1987). The estimation of wage gains and welfare gains in self-selection models. *Review of Economics and Statistics*, *69*(1), 42–49.
- Blincoe, L., Miller, T., Zaloshnja, E., & Lawrence, B. (2015). The economic and societal impact of motor vehicle crashes, 2010. *National Highway Safety Administration Technical Report*.

- Brinch, C. N., Mogstad, M., & Wiswall, M. (2017). Beyond late with a discrete instrument. *Journal of Political Economy*, 125(4), 985–1039.
- Carneiro, P., Heckman, J., & Vytlacil, E. (2011). Estimating marginal returns to education. *American Economic Review*, 101(6), 2754–81.
- Cattaneo, M., Crump, R., Farrell, M., & Feng, Y. (2021). On Binscatter. *Unpublished Manuscript*.
- Chalfin, A. (2016). The economic cost of crime. *The Encyclopedia of Crime and Punishment*.
- Chalfin, A. & Goncalves, F. (2021). The professional motivations of police officers. *Unpublished manuscript*.
- Chandra, A. & Staiger, D. (2020). Identifying sources of inefficiency in healthcare. *Quarterly Journal of Economics*, 135(2), 785–843.
- Cornelissen, T., Dustmann, C., Raute, A., & Schönberg, U. (2016). From late to mte: Alternative methods for the evaluation of policy interventions. *Labour Economics*, 41, 47–60.
- Dahl, G., Kostol, A., & Mogstad, M. (2014). Family welfare cultures. *Quarterly Journal of Economics*, 129(4), 1711–1752.
- DeAngelo, G. & Hansen, B. (2014). Life and death in the fast lane: Police enforcement and traffic fatalities. *American Economic Journal: Economic Policy*, 6(2), 231–257.
- Deshpande, M. & Li, Y. (2019). Who is screened out? application costs and the targeting of disability programs. *American Economic Journal: Economic Policy*, 11(4), 213–248.
- Dobbie, W., Goldin, J., & Yang, C. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108(2), 201–40.
- Dobbie, W. & Song, J. (2015). Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection. *The American Economic Review*, 105(3), 1272–1311.
- Dusek, L. & Traxler, C. (2021). Learning from law enforcement. *Journal of the European Economic Association*, forthcoming.
- Einav, L., Finkelstein, A., Ji, Y., & Mahoney, N. (2022). Voluntary regulation: Evidence from medicare payment reform. *Quarterly Journal of Economics*, 137(1), 565–618.
- Feigenberg, B. & Miller, C. (2022). Would eliminating racial disparities in motor vehicle searches have efficiency costs? *Quarterly Journal of Economics*, 137(1), 49–113.
- Finkelstein, A. & Notowidigdo, M. (2019). Take-up and targeting: Experimental evidence from snap. *Quarterly Journal of Economics*, 134(3), 1505–1556.

- Frandsen, B., Lefgren, L., & Leslie, E. (2019). Judging judge fixed effects. *NBER Working Paper 25528*.
- Frandsen, B. R. (2017). Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. In *Regression Discontinuity Designs: Theory and Applications* (pp. 281–315). Emerald Publishing Limited.
- Gehrsitz, M. (2017). Speeding, punishment, and recidivism: Evidence from a regression discontinuity design. *The Journal of Law and Economics*, *60*(3), 497–528.
- Goncalves, F. & Mello, S. (2021). A few bad apples? racial bias in policing. *American Economic Review*, *111*(5), 1406–1441.
- Hansen, B. (2015). Punishment and deterrence: Evidence from drunk driving. *American Economic Review*, *105*(4), 1581–1617.
- Heckman, J. (1979). Sample selection bias as specification error. *Econometrica*, *71*(1), 53–161.
- Heckman, J. J. & Vytlacil, E. J. (2007). Chapter 71 econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. volume 6 of *Handbook of Econometrics* (pp. 4875–5143). Elsevier.
- Imbens, G. W. & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.
- Ito, K., Ida, T., & Tanaka, M. (2022). Selection on welfare gains: Experimental evidence from electricity plan choice. *American Economic Review*, *forthcoming*.
- Kaplow, L. & Shavell, S. (2006). *Fairness versus welfare*. Harvard University Press.
- Kling, J. R. (2006). Incarceration length, employment, and earnings. *The American economic review*, *96*(3), 863–876.
- Lipsky, M. (1980). Street-level bureaucracy: dilemmas of the individual in public services.
- Luca, D. (2014). Do traffic tickets reduce motor vehicle accidents? Evidence from a natural experiment. *Journal of Policy Analysis and Management*, *34*(1), 85–106.
- Maestas, N., Mullen, K., & Strand, A. (2013). Does disability insurance discourage work? using examiner assignment to estimate causal effects of ssdi receipt. *The American Economic Review*, *103*(5), 1797–1829.
- Makowsky, M. & Stratmann, T. (2011). More tickets, fewer accidents: How cash-strapped towns make for safer roads. *The Journal of Law and Economics*, *54*(4), 863–888.
- Mello, S. (2021). Fines and financial wellbeing. *Working paper*.

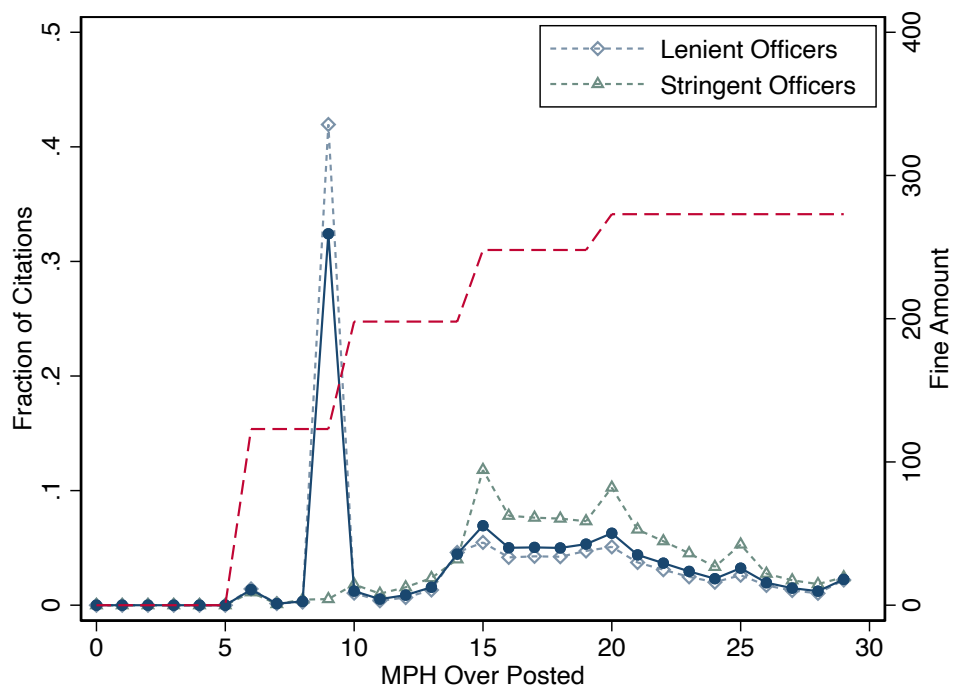
- Mogstad, M., Santos, A., & Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, *86*(5), 1589–1619.
- Mogstad, M. & Torgovitsky, A. (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics*, *10*, 577–613.
- Moore, M. S. (2019). The moral worth of retribution. In *Retribution* (pp. 337–388). Routledge.
- Morris, C. (1983). Parametric empirical bayes inference: Theory and applications. *Journal of the American Statistical Association*, *78*(381), 47–55.
- Nagin, D. (2013). Deterrence: A review of the evidence by a criminologist for economists. *Annual Review of Economics*, *5*, 83–105.
- NHTSA (2014). Traffic safety facts: 2012 data. *National Highway Safety Administration Technical Report*.
- Norris, S. (2022). Examiner inconsistency: Evidence from refugee decisions. *Unpublished manuscript*.
- O’Flaherty, B. & Sethi, R. (2019). *Shadows of doubt: Stereotypes, crime, and the pursuit of justice*. Harvard University Press.
- Polinsky, A. M. & Shavell, S. (2000). The fairness of sanctions: Some implications for optimal enforcement policy. *American Law and Economics Review*, *2*(2), 223–237.
- Prendergast, C. (2007). The motivation and bias of bureaucrats. *American Economic Review*, *97*(1), 180–196.
- Rose, E. (2021). Who gets a second chance? Effectiveness and equity in supervision of criminal offenders. *Quarterly Journal of Economics*, *136*(2), 1199–1253.
- Roy, A. (1951). Some thought on the distribution of earnings. *Oxford Economics Papers*, *3*(2), 135–46.
- Traxler, C., Westermeier, F., & Wohlschlegel, A. (2018). Bunching on the Autobahn? Speeding responses to a notched penalty scheme. *Journal of Public Economics*, *157*, 78–94.
- Van Dijk, W. (2019). The socioeconomic consequences of housing assistance. *Unpublished manuscript*.
- Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, *70*(1), 331–341.
- Weisburst, E. (2017). Whose help is on the way? the importance of individual police officers in law enforcement outcomes. *Unpublished manuscript*.
- Wilson, J. (1989). *Bureaucracy: What government agencies do and why they do it*.

Table 1: Summary Statistics

	(1) All	By Fines	
		(2) Discounted	(3) Harsh
<u>Panel A: Demographics</u>			
Female	0.384	0.415	0.368
Age	36.47	36.98	36.20
Age Missing	0.0002	0.0002	0.0002
Race = White	0.474	0.525	0.448
Race = Black	0.154	0.157	0.152
Race = Hispanic	0.187	0.144	0.209
Race = Other	0.041	0.034	0.044
Race = Unknown	0.144	0.140	0.147
<u>Panel B: Socioeconomic Status</u>			
Zip Income	57962	56459	58745
Zip Income Missing	0.101	0.107	0.097
Vehicle Value	17807	17297	18073
Vehicle Info Missing	0.143	0.139	0.145
<u>Panel C: Offending History</u>			
Prior Prison Spell	0.009	0.009	0.009
Citation Past Year	0.350	0.317	0.368
Speeding Past Year	0.179	0.158	0.189
Other Past Year	0.253	0.226	0.268
Crash Past Year	0.071	0.067	0.074
<u>Panel D: Offense Characteristics</u>			
MPH Over Posted	15.62	9.00	19.06
Fine Amount	207.70	123.00	251.79
Contest in Court	0.289	0.222	0.323
Observations	1,693,457	579,760	1,113,697

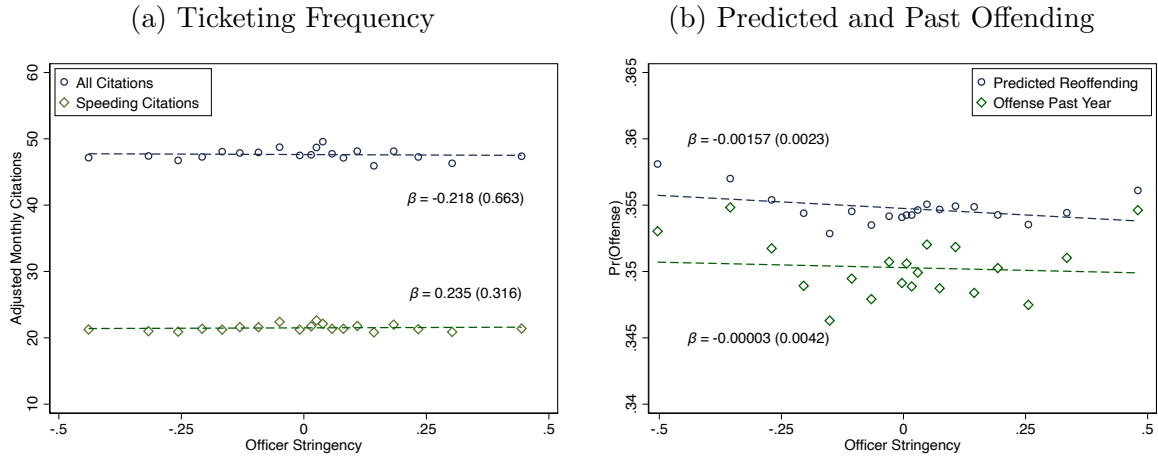
Notes: This table reports means for the analysis sample. See table [A-1](#) for officer characteristics.

Figure 1: Fine Schedule and Charged Speed Distribution



Notes: This figure plots the distribution of charged speeds on FHP-issued speeding citations in Florida. Dashed red line shows the fine schedule (right axis). Solid line and circles shows the aggregate distribution. Dashed lines with hollow diamonds and triangles plot the distribution for lenient and stringent officers, respectively, using the method described in section D.

Figure 2: Instrument Validity



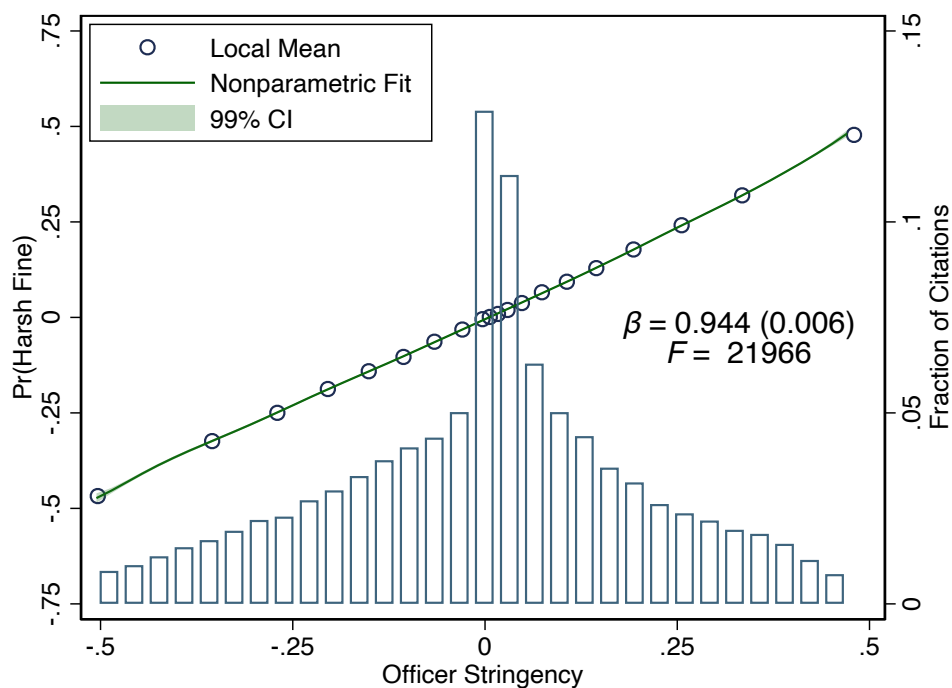
Notes: Panel (a) reports the relationship between officer stringency, residualized of beat-shift effects, and an officer’s average monthly number of citations, adjusted for beat-shift effects. Blue circles report the relationship for all citations and green diamonds report the relationship for only speeding citations. Panel (b) reports the relationship between officer stringency, residualized of beat-shift effects, and predicted reoffending based on covariates and past offending, both residualized of beat-shift effects. Predicted reoffending is computed in opposite sample partitions using only tickets issued by non-bunching officers (see section D-3 for further details).

Table 2: Randomization Test

	(1)	(2)	(3)	(4)
	Reoffend	Harsh Fine	Stringency	1 [Stringent]
Female	-0.0645 (0.000840)	-0.0233 (0.00173)	-0.00119 (0.000928)	-0.00161 (0.00130)
Age	-0.00631 (0.000151)	-0.00229 (0.000273)	0.000280 (0.000172)	0.000540 (0.000257)
Age Squared	0.0000190 (0.00000168)	0.0000151 (0.00000268)	-0.00000303 (0.00000166)	-0.00000586 (0.00000259)
Race = Black	0.0710 (0.00127)	0.0202 (0.00270)	-0.000326 (0.00176)	-0.00407 (0.00242)
Race = Hispanic	0.0324 (0.00117)	0.0345 (0.00290)	0.00633 (0.00212)	-0.000643 (0.00295)
Race = Other	0.000888 (0.00202)	0.0348 (0.00274)	0.00581 (0.00176)	-0.00101 (0.00250)
Race = Unknown	0.0120 (0.00299)	0.00505 (0.00550)	0.00377 (0.00266)	-0.00131 (0.00374)
Prior Prison Spell	0.139 (0.00393)	0.0107 (0.00359)	-0.00315 (0.00207)	-0.000259 (0.00275)
County Resident	0.00583 (0.00127)	-0.0186 (0.00309)	-0.00586 (0.00265)	0.00310 (0.00441)
Log Zip Income	-0.0161 (0.000992)	0.0103 (0.00211)	0.00486 (0.00164)	0.00180 (0.00194)
Log Vehicle Price	-0.00846 (0.000992)	0.0199 (0.00160)	0.00487 (0.00122)	0.00159 (0.00191)
Speeding Past Year	0.127 (0.00103)	0.0247 (0.00157)	0.000948 (0.000653)	-0.000595 (0.00110)
Other Past Year	0.154 (0.00103)	0.0149 (0.00115)	-0.000758 (0.000727)	-0.00201 (0.00118)
Crash Past Year	0.0393 (0.00148)	0.00736 (0.00132)	0.000990 (0.000769)	-0.000721 (0.00113)
Mean	.172	.658	.658	.763
F-Stat	5883.84	28.8	2.7	.92
F-test	<.0001	<.0001	.0002	.5484
Beat-Shift FE	Yes	Yes	Yes	Yes
Officers	1960	1960	1960	1960
Observations	1693457	1693457	1693457	1693457

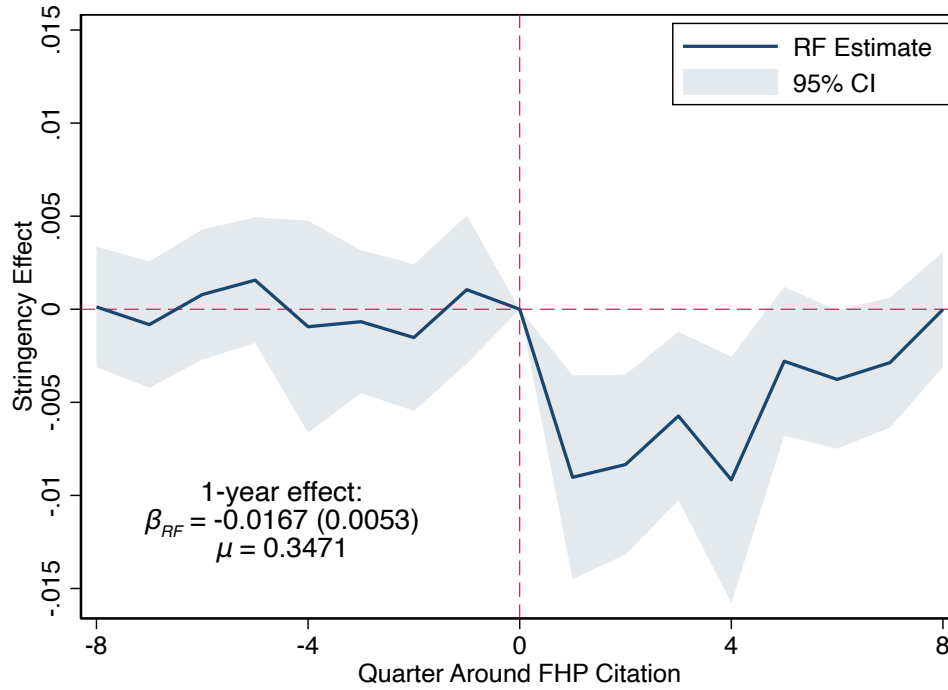
Notes: Standard errors clustered at the officer-level in parentheses. Regressions also include indicators for missing age (<1%), missing zip code income ($\approx 10\%$), and missing vehicle information ($\approx 14\%$); joint significance tests include these variables.

Figure 3: Distribution of Instrument and First Stage



Notes: Figure shows the first stage relationship between stringency and the probability of a harsh fine, both residualized of beat-shit fixed effects (left axis). Local binscatter means are denoted by blue circles and the green line shows a non-parametric fit, with a 99 percent confidence interval indicated by the shaded region. Figure also illustrates a histogram of the officer stringency instrument, residualized of beat-shit fixed effects (right axis). Figure reports the linear first stage estimate and associated F -statistic.

Figure 4: Reduced Form Over Time



Notes: This figure reports coefficients on officer stringency from regressions where the outcome of interest is an indicator for whether the driver received a traffic citation in each quarter relative to the date of their focal FHP citation. $\tau = 0$ denotes the exact date of the focal FHP citation (one day only, where all motorists receive a citation so the effect of stringency is zero by construction). Regressions also include beat-shift fixed effects. Shaded region denotes 95 percent confidence intervals, constructed from standard errors clustered at the officer-level. Identical figures for other outcomes are shown in figure B-2. Figure reports the reduced form coefficient for one-year reoffending as well as the mean one-year reoffending rate for lenient officers.

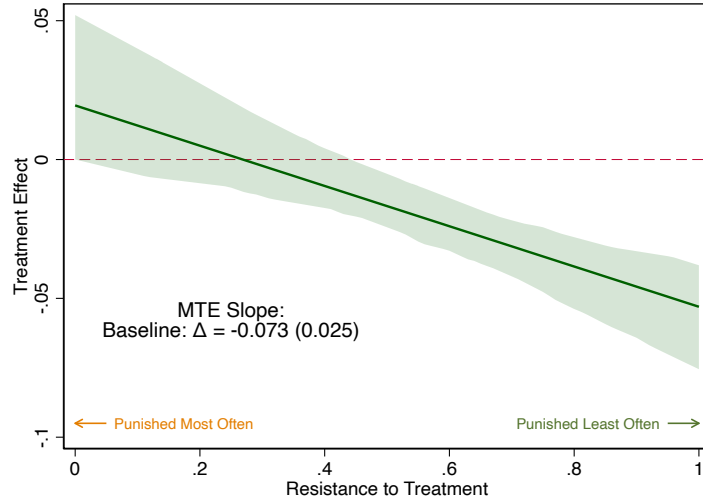
Table 3: Effect of Harsh Fines, IV Estimates

	(1) Lenient Mean	IV Estimates		
		(2) β_{IV}	(3) β_{IV}	(4) ϵ
Any Violation	0.347	-0.0177 (0.0017)	-0.0159 (0.0016)	-0.069 (0.007)
Speeding Violation	0.170	-0.0146 (0.0013)	-0.0144 (0.0013)	-0.127 (0.012)
Other Violation	0.256	-0.0119 (0.0016)	-0.0097 (0.0015)	-0.057 (0.009)
Crash Involvement	0.080	-0.0029 (0.0010)	-0.0021 (0.0010)	-0.040 (0.018)
Contest in Court	0.262	0.1125 (0.0014)	0.1093 (0.0014)	0.626 (0.008)
Controls		No	Yes	Yes
Beat-Shift FE		Yes	Yes	Yes
Observations		1693457	1693457	1693457

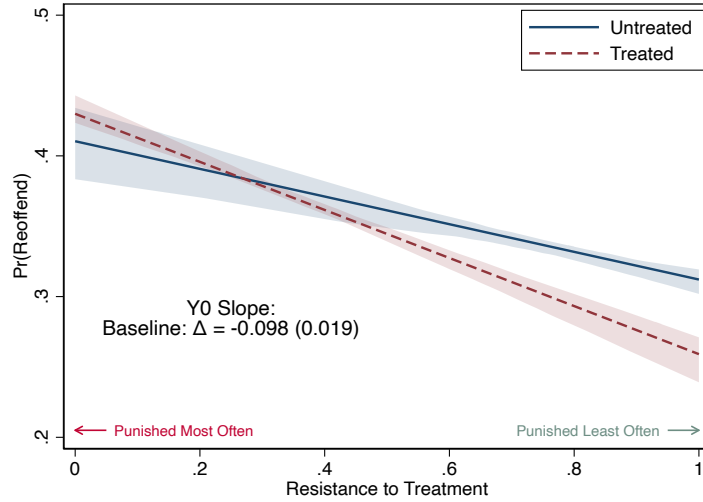
Notes: This table reports IV estimates of the impact of receiving a harsh fine on one-year reoffending. Standard errors clustered at the officer -level are in parentheses. First stage estimates are $\beta = 0.944$ (0.006) without controls and $\beta = 0.943$ (0.006) with controls. See table B-2 for the full set of first stage and reduced form estimates with and without controls. Implied elasticities are computed as $\hat{\beta}_{IV} \times \bar{fine}/\bar{y}$, where $\hat{\beta}_{IV}$ is estimated using the statutory fine as the treatment variable and the means are the lenient officer means.

Figure 5: Baseline MTR Estimates

(a) Marginal Treatment Effects

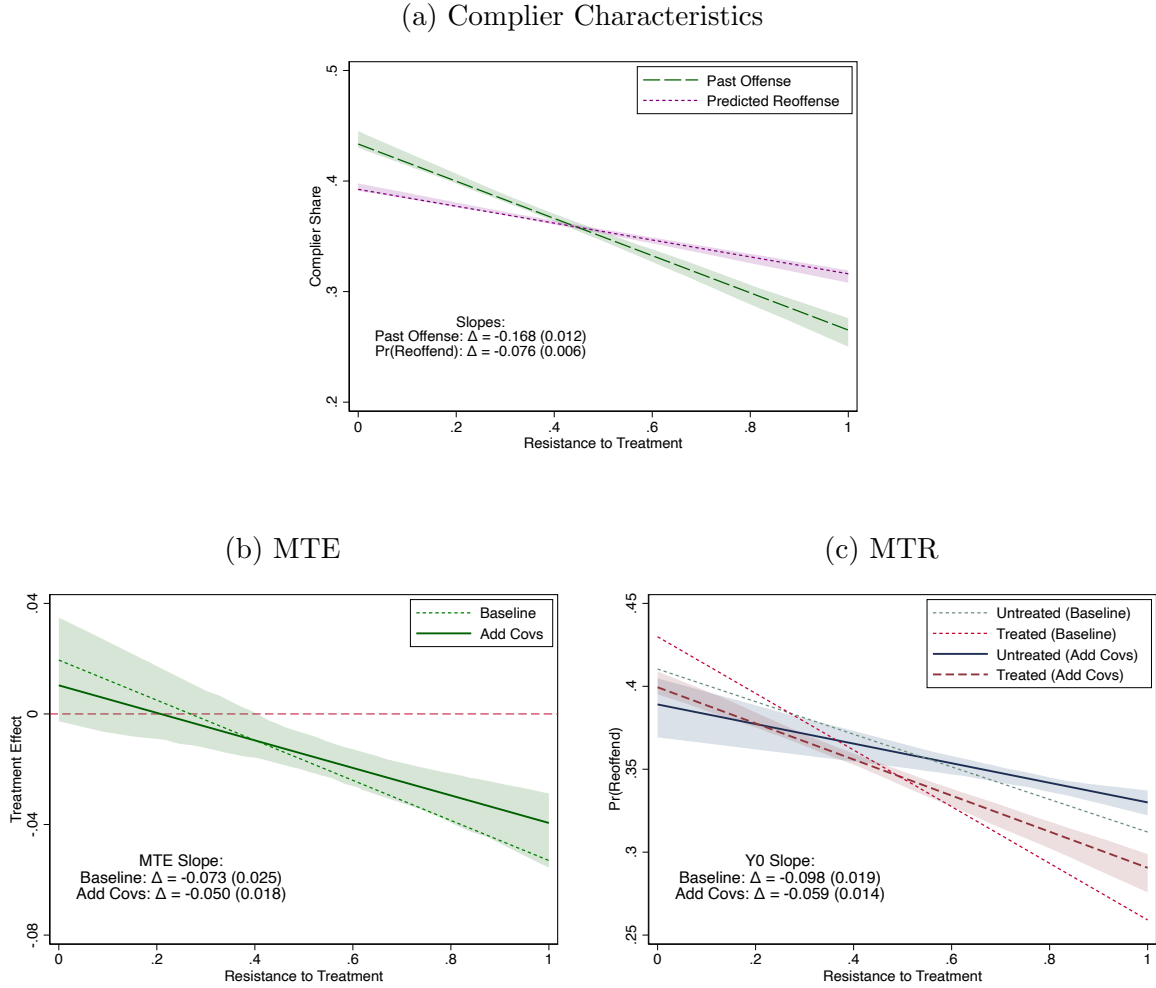


(b) Marginal Treatment Responses



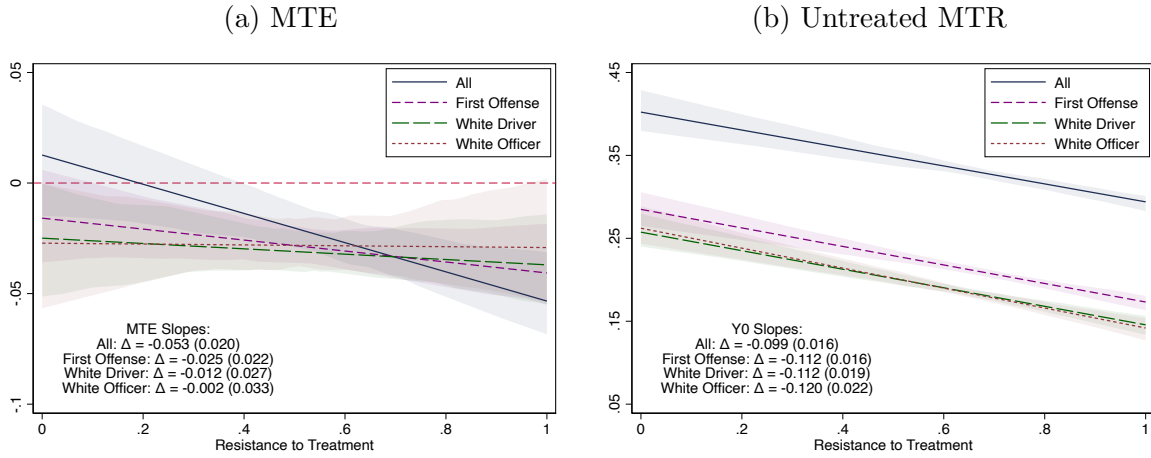
Notes: Outcome is any new traffic offense in the following year. Figures report estimated marginal treatment responses (panel a) and marginal treatment effects (panel b) obtained via the method described in section 5.1. Shaded regions denote 95% confidence intervals. Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level.

Figure 6: Sorting between and within Covariates



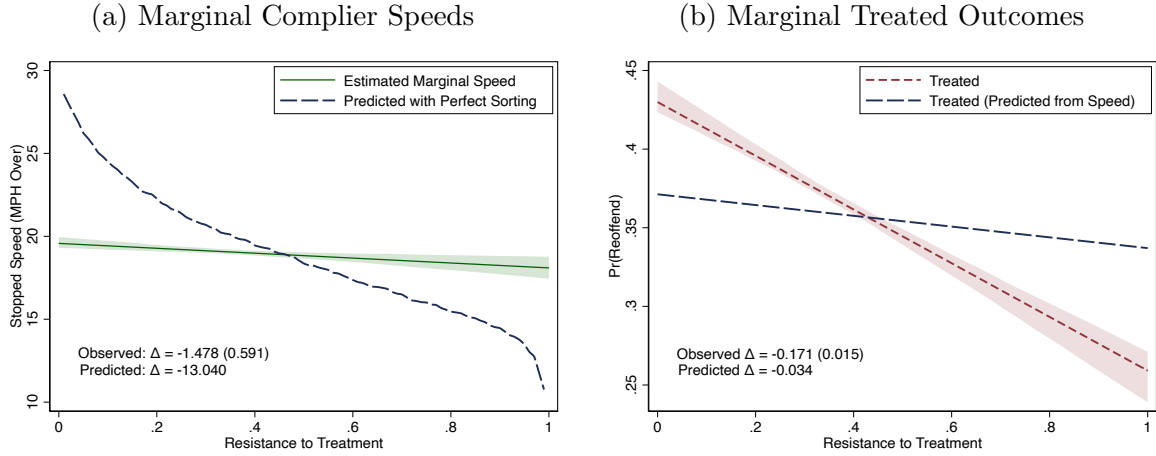
Notes: Panel (a) reports marginal complier characteristics estimated using the method detailed in section E-3. See section D-3 for details on the computation of our predicted reoffending measure (based on covariates). Panels (b) and (c) report the baseline MTE and MTR estimates (same as figure 5) as well as MTE and MTR estimates conditioning on driver covariates. Specifically, we estimate the MTE and MTR functions within covariate cells at the level of $\text{gender} \times \text{race} \times \mathbf{1}[\text{age} \geq 35] \times \mathbf{1}[\text{past offense}]$. Shaded regions denote 95% confidence intervals. Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level. See figure C-2 for MTR estimates with additional controls.

Figure 7: Sorting within Salient Subsamples



Notes: This figure reports estimated MTE (panel a) and untreated MTR (panel b) functions for salient subsamples. This figure relies on the sample of citations issued from 2010-2016 (the “late” sample). Blue solid lines are estimates for the full late sample. Purple dashed lines are estimates for drivers committing their first offense, which is defined as having no citations in the past five years. Green long-dashed lines report estimates for white, first-time offenders. Red short-dashed lines are estimates for white, first-time offenders cited by white officers. Shaded regions denote 95% confidence intervals. Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level.

Figure 8: Sorting on Stopped Speeds



Notes: Panel (a) explores the extent to which drivers are sorted based on stopped speed. The green solid line illustrates the stopped speed of marginal compliers estimated using the method explained in section E-3. The blue long-dashed line represents a simulated version of this curve under the assumption that drivers are sorted using only speed. Panel (b) explores the contribution of sorting based on speed to the slope of the treated MTR function. The maroon dashed line is our baseline estimated treated MTR function (same as panel b of figure 5). The blue long-dashed line a predicted version of the same line based on (i) the amount of stopped speed sorting based on the green line in panel (a) and (ii) an estimated relationship between stopped speed and the probability of reoffending. See E-4 for further details. Shaded regions denote 95% confidence intervals. Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level.

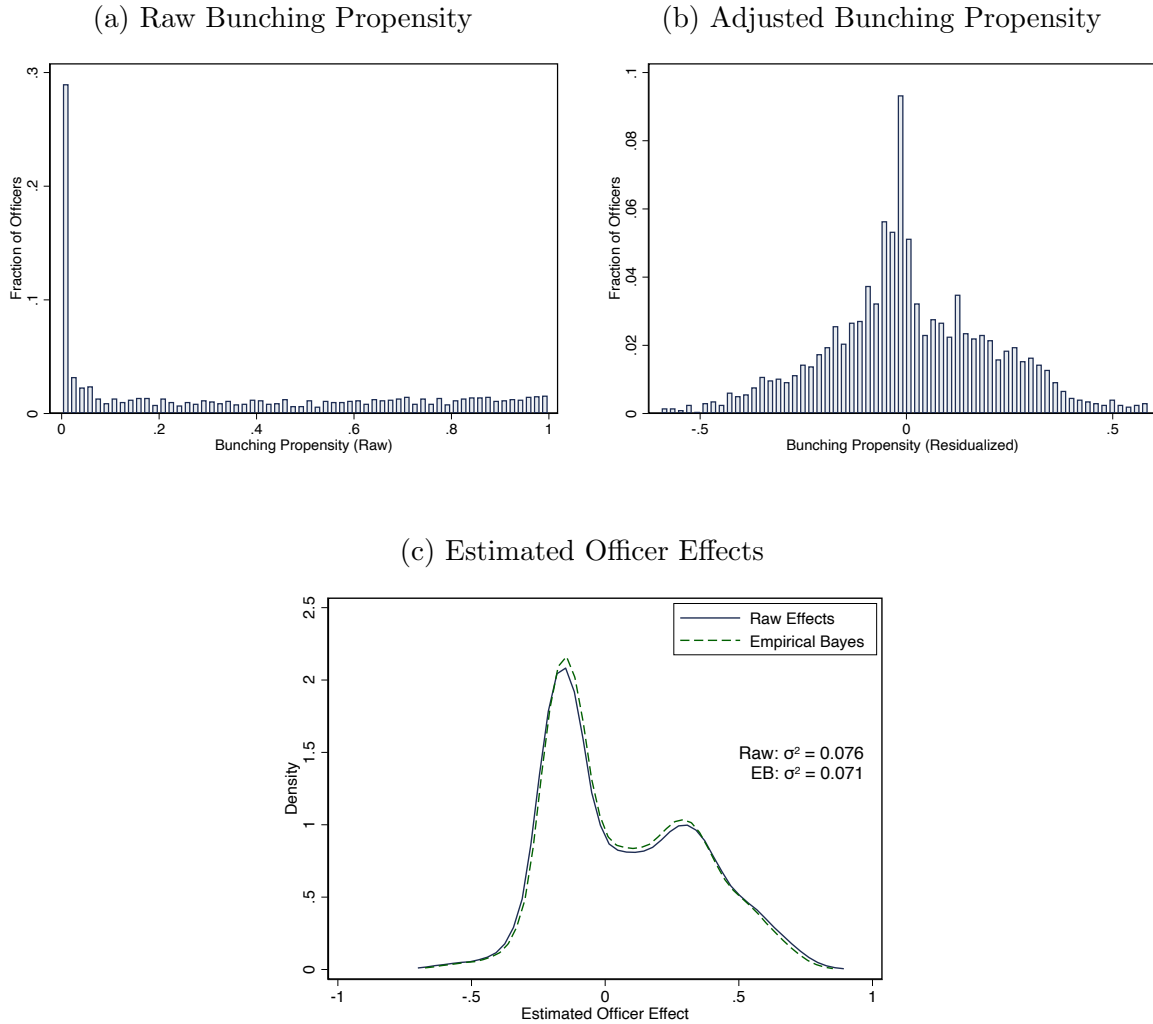
Table 4: Model Counterfactual

	Treated (Harsh Punishment)		Untreated (Lenient Punishment)		Overall
	Y_0	$Y_1 - Y_0$	Y_0	$Y_1 - Y_0$	Y
Baseline	0.376 [0.370, 0.381]	-0.001 [-0.006, 0.005]	0.327 [0.324, 0.329]	-0.038 [-0.043, -0.032]	0.358
Reverse Resistance To Treatment	0.342 [0.339, 0.343]	-0.027 [-0.030, -0.023]	0.392 [0.383, 0.398]	0.010 [0.004, 0.019]	0.341
$\lambda \rightarrow 0$ (pure deterrence objective)	[0.335, 0.338]	[-0.032, -0.026]	[0.362, 0.363]	[-0.010, -0.005]	[0.332, 0.337]

Notes: Table presents the expected level of offending and deterrability of drivers, separately by treated and untreated. The second row shows these numbers for the counterfactual where officers care solely about deterrability and not about level of offending.

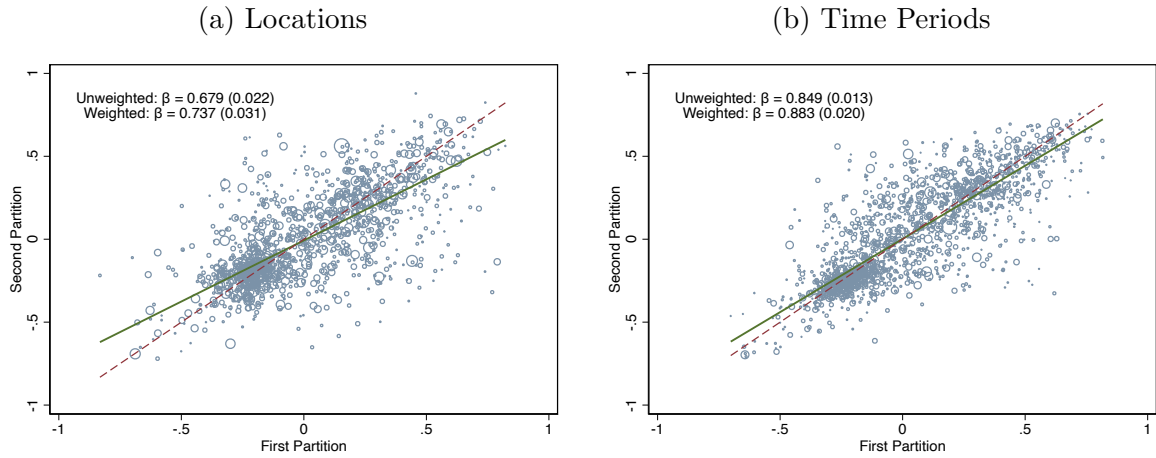
A Stringency Instrument

Figure A-1: Across-Officer Distribution of Bunching Propensity



Notes: Panel (a) plots the officer-level distribution of the share of tickets bunched. Panel (b) plots the officer-level distribution of bunching propensity, residualized of beat-shift fixed effects. Panel (c) reports estimated officer effects from a regression of $\mathbf{1}[bunch_{ijs}]$ on officer fixed effects, beat-shift fixed effects, and the full set of driver covariates, as described in section 2.4. The solid blue illustrates the distribution of raw officer effects and the dashed green line illustrates the distribution of effects after applying Empirical Bayes shrinkage (Morris, 1983).

Figure A-2: Within-Officer Correlation in Bunching Propensity



Notes: Each circle corresponds to an officer. Dashed red line is the 45-degree line. This figure splits each officer's sample of citations into two groups and illustrates the correlation in (residualized) bunching propensity across groups. In panel (a), the groups are constructed as location partitions, with each partition comprised of half of an officer's patrol locations. In panel (b), the groups are constructed as time partitions, with the x and y -axes corresponding to the officer's first and second half of tickets over time, respectively. Each figure reports the raw linear regression coefficient as well as the linear regression coefficient when weighting by the total number of citations. Another way to note the stability over time in an officer's bunching propensity is to regress $\mathbf{1}[bunch_{ijs}]$ on beat-shift fixed effects, officer fixed effects, and a quadratic in officer experience (in months). The p -value on each experience term is > 0.45 and the joint test p -value = 0.7855. In other words, after conditioning on officer identity, there is no experience profile in the likelihood of a bunched ticket.

Table A-1: Relationship between Lenience and Officer Characteristics

	Binary		Continuous		
	(1) Mean	(2) Lenient	(3) Raw	(4) Adjusted	(5) Weighted
Female	0.0893	-0.0704 (0.0366)	-0.0508 (0.0266)	-0.0356 (0.0161)	-0.00340 (0.00388)
Race = Black	0.143	-0.0916 (0.0297)	-0.0289 (0.0231)	-0.00799 (0.0142)	0.00550 (0.00291)
Race = Hispanic	0.169	-0.0933 (0.0693)	-0.0771 (0.0564)	0.0117 (0.0359)	0.00921 (0.00707)
Race = Other	0.191	-0.0120 (0.0655)	-0.0193 (0.0544)	-0.0186 (0.0350)	-0.00135 (0.00724)
Age	34.06	-0.0203 (0.0561)	-0.0236 (0.0478)	0.00982 (0.0288)	0.00741 (0.00631)
Experience	7.09	-0.117 (0.0388)	-0.0778 (0.0324)	-0.0345 (0.0214)	-0.000936 (0.00616)
Any College	0.319	-0.00798 (0.0213)	-0.0114 (0.0171)	-0.00229 (0.0108)	0.00490 (0.00279)
Mean Officers	— 1960	0.753 1960	0.353 1960	0.005 1960	0.006 1958

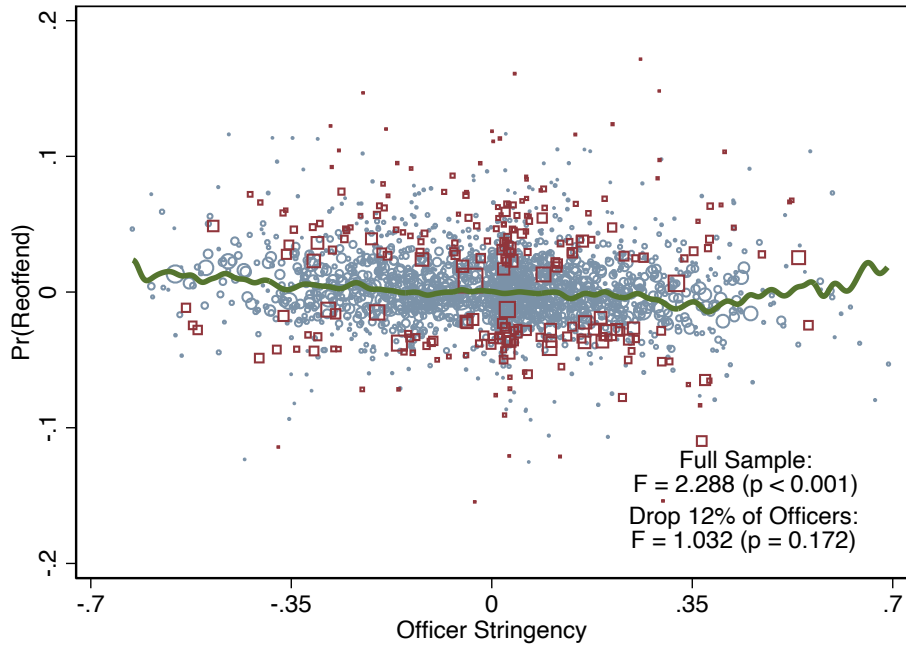
Notes: Robust standard errors in parentheses. Age and experience are in years/10 and are computed as of January 2007. Raw lenience is the fraction of an officer's tickets that are bunched and adjusted lenience is the fraction of an officer's tickets that are bunched, residualized of location-time fixed effects. In column 4, the regression is weighted by one over the variance of adjusted lenience. Regressions also included quadratic terms in age and experience, which are statistically insignificant in all cases.

Table A-2: First Stage Estimates Across Subsamples

	Subgroup	
	(1) = 1	(2) = 0
Female	0.970 (0.007)	0.928 (0.007)
Age > 30	0.957 (0.007)	0.927 (0.007)
Race = White	0.954 (0.008)	0.934 (0.007)
Race = Black	0.922 (0.010)	0.948 (0.006)
Race = Hispanic	0.923 (0.008)	0.948 (0.007)
Race = Other	0.916 (0.015)	0.945 (0.006)
Race = Unknown	0.964 (0.016)	0.941 (0.007)
County Resident	0.972 (0.007)	0.925 (0.007)
Zip Income > \$50,000	0.946 (0.007)	0.942 (0.007)
Vehicle > \$20,000	0.916 (0.009)	0.953 (0.006)
Citation Past Year	0.913 (0.007)	0.961 (0.007)

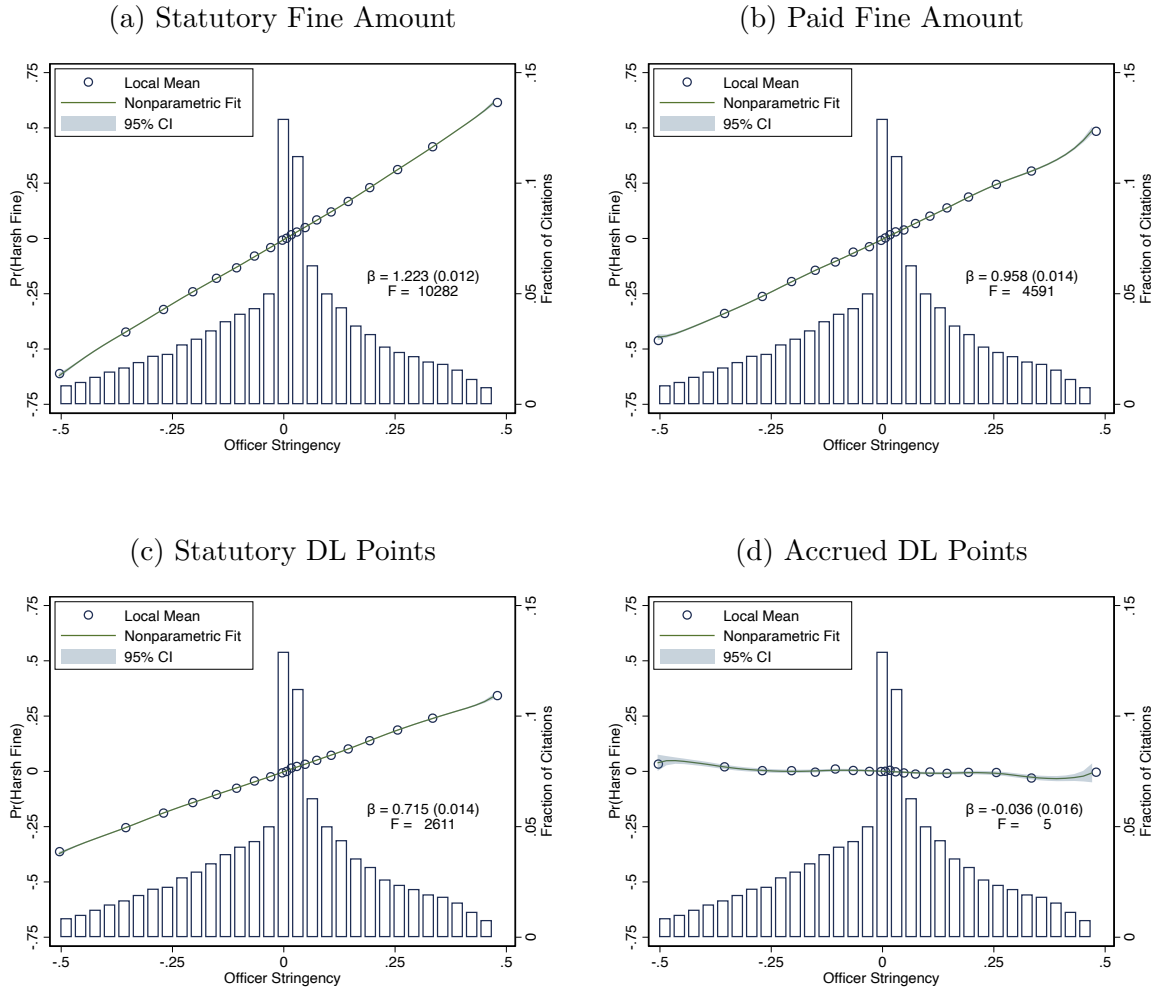
Notes: This table reports first stage estimates for subsamples. Each coefficient is from a separate regression of $\mathbf{1}[harsh]$ on the stringency instrument and beat-shift fixed effects using only the denoted subgroup of drivers, where the subgroups are the groups for which the denoted indicator variable = 1 (column 1) and = 0 (column 2). Standard errors clustered at the officer-level in parentheses. For reference, the first stage estimate in the full sample is $\beta = 0.944$ (0.006).

Figure A-3: Fit Test from Frandsen et al. (2019)



Notes: This figure illustrates results of the joint test of monotonicity and exclusion from Frandsen et al. (2019). Per the recommendation of the authors, we focus only on the fit component of the test because the slope component has minimal power in applications with many judges. Each circle (or square) represents an officer ($N = 1,960$) and plots the officer’s average stringency and reoffending rates, residualized of covariates and beat-shift fixed effects. Green line denotes a non-parametric fit and the figure reports the results of computing residuals from the non-parametric fit, regressing the residuals on officer dummies, and performing a joint significance test. Note that the resulting F -statistic from this “ad-hoc” test ($F = 2.3$) is biased upwards because we have not adjusted for estimation error in obtaining the residuals. The table also reports the results of a second iteration of the fit test after dropping the “worst” 12 percent ($N = 235$) of officers, as measured by the p -value associated with their regression coefficient in the first iteration. These dropped officers are denoted with red squares. Figure C-4 reports the deterrence IV estimate as well as the estimated MTR parameters when these officers are dropped from the analysis.

Figure A-4: First Stage Estimates, Sanction Measures



Notes: Each panel shows an identical to plot figure 3 but replaces the outcome variable with a different sanctions measure. In panel (a), the outcome is the statutory fine based on the charged speed. In panel (b), the outcome is the effective fine amount, taking into account the *ex-post* court outcomes of offenders. In panel (c), the outcome is statutory driver license points based on the points schedule. In panel (d), the outcome is accrued DL points, taking into account the *ex-post* court outcomes of offenders. See appendix section D-1 for details on the computation of effective sanction measures (paid fines and accrued points).

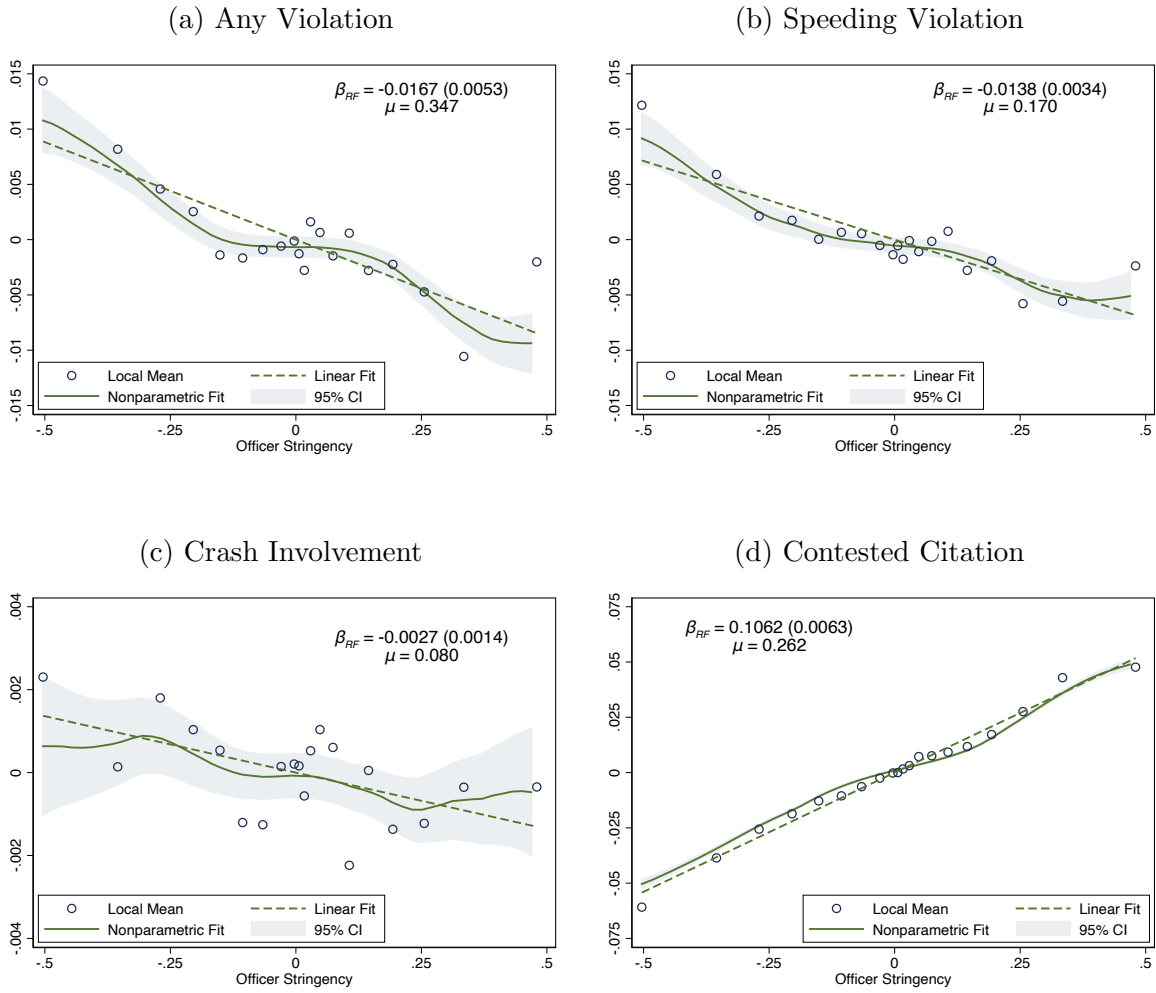
B Deterrence Effects

Table B-1: Naive OLS Estimates

	(1)	(2)	(3)	(4)
	Reoffend	Reoffend	Reoffend	Reoffend
Fine (\$100s)	0.0426 (0.00241)	0.0548 (0.00178)	0.0228 (0.00160)	0.0286 (0.00127)
Mean	0.325	0.325	0.325	0.325
Controls	No	No	Yes	Yes
Officer FE	No	Yes	No	Yes
Beat-Shift FE	Yes	Yes	Yes	Yes
Observations	1693457	1693457	1693457	1693457

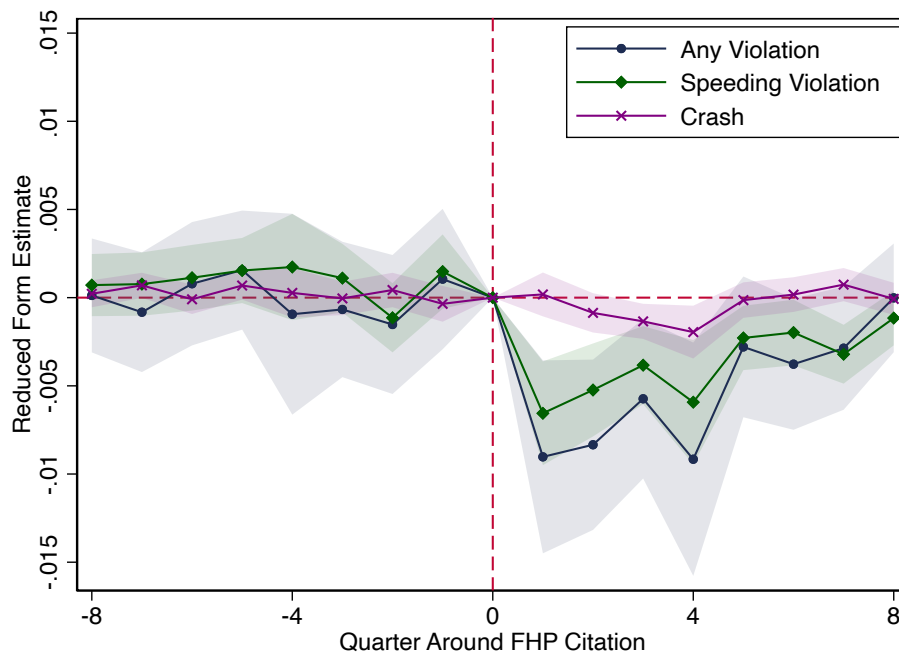
Notes: Standard errors clustered at the officer-level in parentheses. Dependent variable is an indicator for a new traffic offense in the next year. The reported mean is the mean for drivers cited at 9 MPH over the limit.

Figure B-1: Reduced Form Estimates



Notes: Same as figure A-4 except for reduced form outcomes.

Figure B-2: Dynamic Reduced Form Estimates



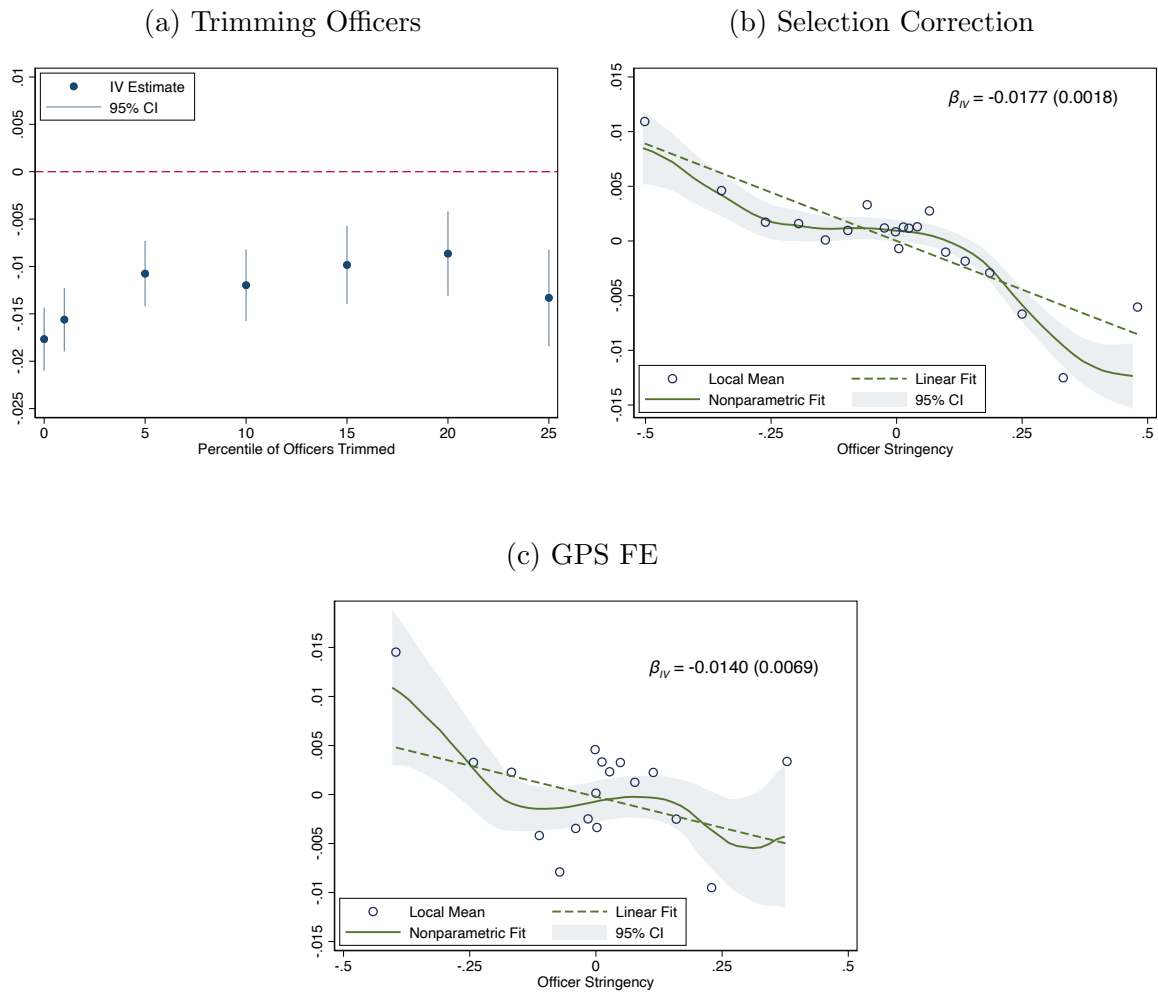
Notes: Same as figure 4 using any moving violation in a given quarter (blue circles), any speeding violation in a given quarter (green diamonds), and any crash involvement in a given quarter (purple x 's) as the outcome variable. Shaded regions denote 95 percent confidence intervals obtained from standard errors clustered at the officer-level.

Table B-2: First Stage and Reduced Form Estimates

	(1) Lenient Mean	(2) β	(3) β
<i>Panel A: First Stage</i>			
Harsh Fine	0.5573	0.9441 (0.0064)	0.9432 (0.0064)
Fine Amount	194.308	122.340 (1.206)	122.210 (1.190)
Fine Amount (Paid)	167.187	95.819 (1.414)	96.066 (1.406)
DL Points	3.416	0.7152 (0.0140)	0.7142 (0.0138)
DL Points (Accrued)	1.684	-0.0362 (0.0164)	-0.0273 (0.0154)
<i>Panel B: Reduced Form</i>			
Any Violation	0.3471	-0.0167 (0.0053)	-0.0150 (0.0034)
Speeding Violation	0.1702	-0.0138 (0.0034)	-0.0136 (0.0025)
Other Violation	0.2563	-0.0112 (0.0045)	-0.0092 (0.0028)
Moving Violation	0.2801	-0.0135 (0.0047)	-0.0128 (0.0031)
Non-Moving Violation	0.1602	-0.0117 (0.0035)	-0.0097 (0.0022)
Crash Involvement	0.0799	-0.0027 (0.0014)	-0.0020 (0.0011)
Contest in Court	0.2620	0.1062 (0.0063)	0.1031 (0.0060)
Controls		No	Yes
Beat-Shift FE		Yes	Yes
Officers		1960	1960
Observations		1693457	1693457

Notes: Standard errors clustered at the officer-level in parentheses. This table reports first stage and reduced form regression estimates with and without covariates. Each coefficient is from a separate regression of the denoted outcome on the stringency instrument and beat-shift effects, with (column 2) and without (column 3) controls

Figure B-3: Robustness, Sample Selection



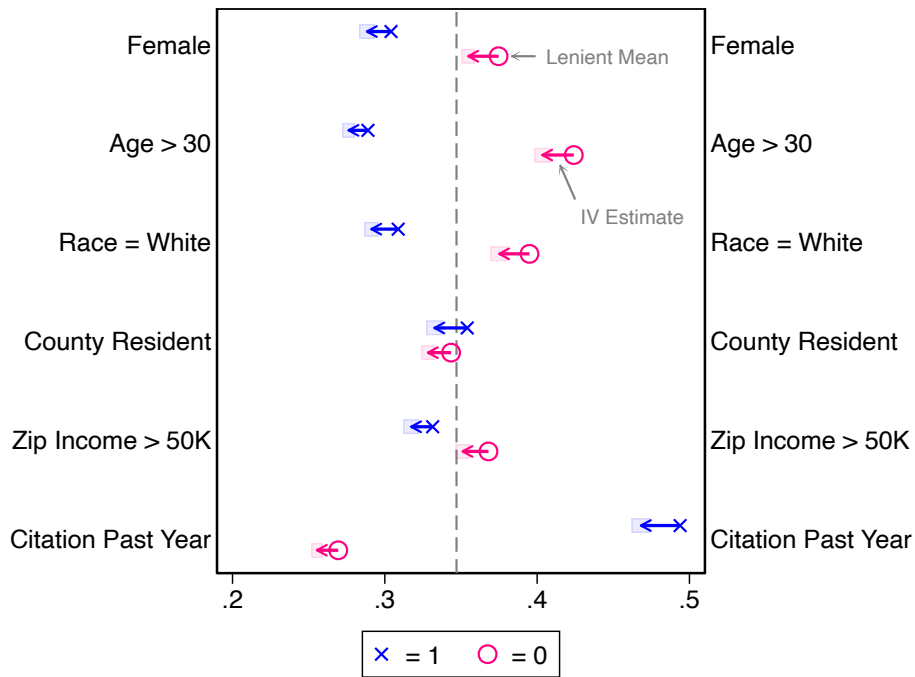
Notes: For comparison, our main IV estimate is $\beta_{IV} = -0.0177$ (0.0017). Panel (a) shows the sensitivity of our IV estimate to trimming officers with the most selected samples. Panel (b) plots the reduced form and reports the IV estimate using a Heckman (1979) selection correction based on officer ticketing frequency. Panel (c) plots the reduced form and reports the IV estimate using GPS road segment fixed effects for the subset of citations including GPS coordinates ($N = 244, 858$).

Table B-3: Robustness, Alternative Instruments

Instrument	(1) N	<i>F</i> -Stat		(4) β_{IV}
		(2) Balance	(3) FS	
Leave-out	1693457	2.702	21966	-0.0159 0.0016
Leave-Out (Residualized)	1693457	2.691	25278	-0.0191 0.0016
Leave-County-Out	1500479	1.475	626	-0.0176 0.0024
Binary	1693457	0.921	289	-0.0268 0.0039
Officer Dummies	1693404	2.778	323	-0.0183 0.0016
<i>Within Demographics</i>				
Race	1689414	5.111	24453	-0.0168 0.0016
Race \times Gender \times Age	1651212	4.579	21505	-0.0176 0.0016
Race \times Gender \times Age \times History	1587971	5.175	18469	-0.0182 0.0017
Race \times Gender \times Age \times History \times Income	1471826	3.884	14912	-0.0189 0.0017

Notes: This table shows how results vary under different computations of the stringency instrument. Columns 2 and 3 report F-statistics associated with a joint balance test and the first stage; column 4 reports the IV estimate for one-year speeding recidivism. Row 1 reports results corresponding to the main instrument. In row 2, the instrument is the leave-out-mean after residualizing out beat-shift effects (e.g., [Dobbie et al. 2018](#)). Row 3 computes the instrument as the leave-county-out mean. Row 4 uses a binary instrument and row 5 uses the full set of officer dummies as instruments. Rows 6-9 show results when the instrument is computed as the leave-out mean within demographic cells, defined according to four race groups (white, Black, Hispanic, other/unknown), gender, $\mathbf{1}[\text{age} \geq 35]$, $\mathbf{1}[\text{any citation in past year}]$ and $\mathbf{1}[\text{zip code income} \geq \$50,000]$. Regressions using the by-group instrument also include fixed effects at the relevant demographic cell-level.

Figure B-4: IV Estimate Heterogeneity by Driver Characteristics



Notes: This figure shows heterogeneity in IV estimates for one-year recidivism by driver characteristics. Each characteristic is denoted as a binary category; the x 's plot lenient means for the category = 1 subgroup and the o 's plot lenient means for the category = 0 subgroup. Arrows pointing away from the means indicate the IV estimate, and shaded region around the arrow denotes the 95 percent confidence interval. Vertical dashed line denotes the lenient officer mean recidivism rate for the full sample.

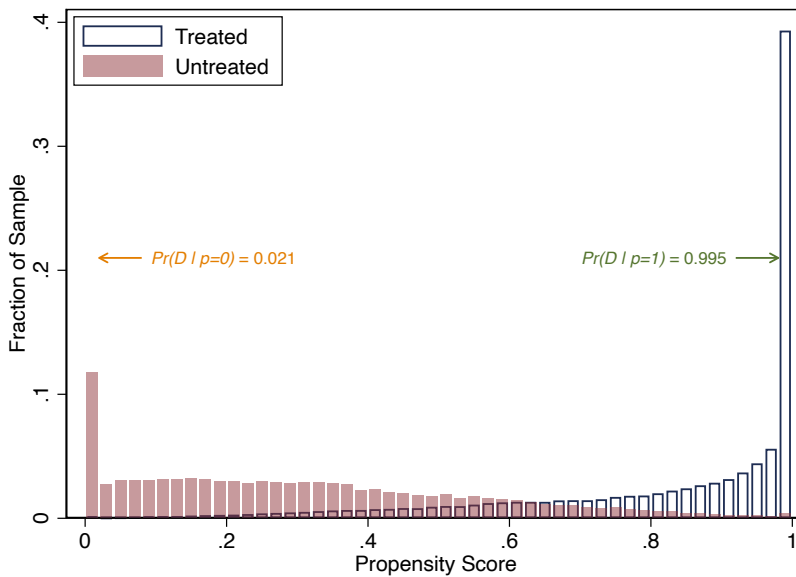
Figure B-5: Evidence of Driver Learning



Notes: Panel (a) illustrates reduced form relationships and reports IV estimates for motorists with and without past exposure to stringent (non-bunching) officers. To mitigate selection issues resulting from the fact that past exposure to stringency reduces the likelihood a driver reappears in the data, we focus on exposure at least one year in the past because treatment effects fade out after one year (see figure 4). Specifically, we take the subset of drivers with an FHP-issued citation at least one year prior and compare treatment effects for those with and without past tickets issued by stringent officers. In panel (b), we report treatment effects of harsh fines on the likelihood that drivers reoffend in the same county they were ticketed and in different counties from the one they were ticketed in. Both panels report the estimated difference (and associated standard error) in treatment effects.

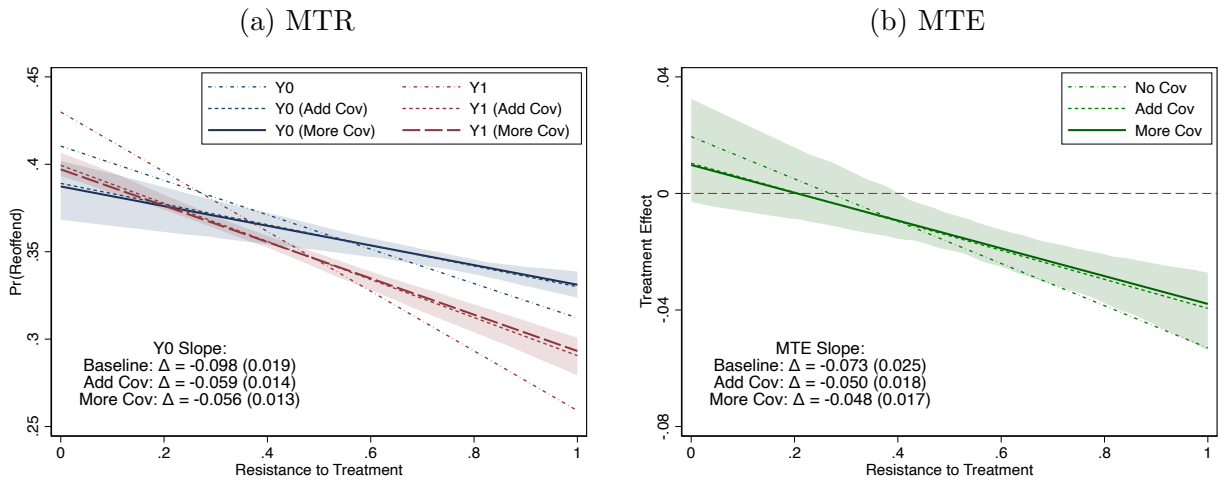
C MTR Estimates

Figure C-1: Common Support for MTR Estimates



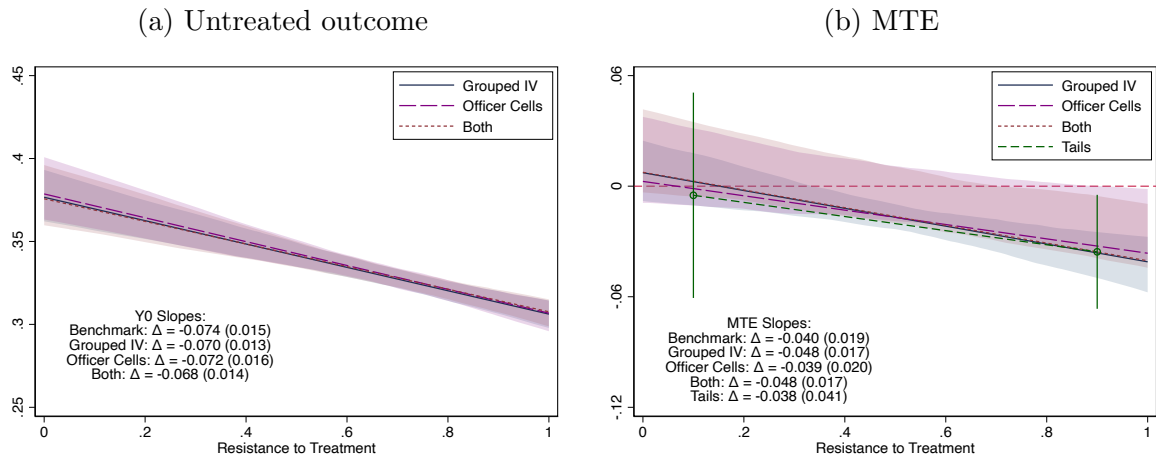
Notes: Figure plots the distribution of propensity scores for the treated (66%) and untreated (34%) subsets of sample, where treatment is defined as $\mathbf{1}[harsh]$. Following the text, the propensity score is estimated from a linear regression of $\mathbf{1}[harsh]$ on officer stringency and beat-shift fixed effects.

Figure C-2: Sensitivity of MTR Estimates to Covariates



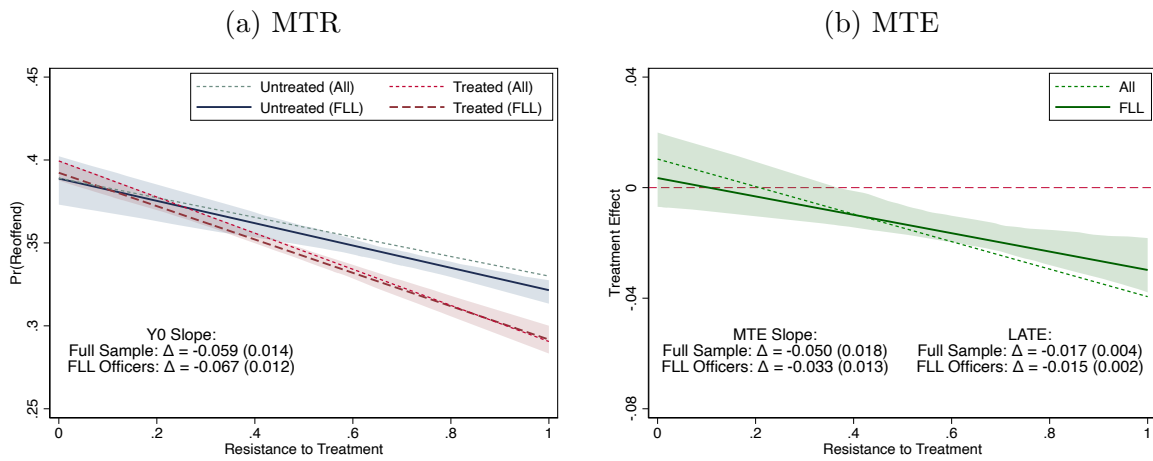
Notes: This figure reports MTR and MTE estimates using the full sample with and without covariates. In this figure, the baseline specification is no controls. Adding covariates refers to including our standard covariate cell controls at the level of race \times gender \times $\mathbf{1}[\text{age} \geq 35] \times \mathbf{1}[\text{past offense}]$ (32 cells). More covariates refers to including more detailed covariate cells at the level of race \times gender \times age quartile \times zip code income quartile $\times \mathbf{1}[\text{past offense}]$ (256 cells). Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level.

Figure C-3: MTR Robustness, Monotonicity



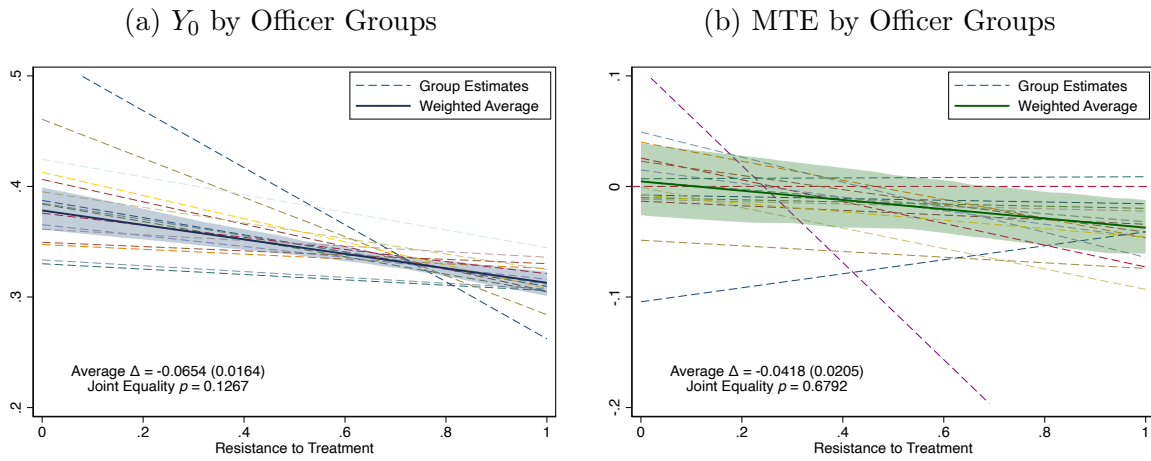
Notes: This figure reports the estimated untreated MTR and MTE using alternative specifications. The underlying sample is 2010-2015 citations (the “late” sample). The *grouped IV* specification recomputes the stringency instrument within our baseline covariate cells at the level of race \times gender \times $\mathbf{1}[\text{age} \geq 35] \times \mathbf{1}[\text{past offense}]$. The *officer cells* specification estimates separate MTR and MTE curves for each officer cell, defined by gender \times $\mathbf{1}[\text{nonwhite}] \times \mathbf{1}[\text{any college}] \times \mathbf{1}[\text{high experience}]$, where high experience is defined as an above median number of speeding citations issued over the period 2005–2009. *Both* refers to separate estimation by officer cells using the grouped instrument. See the main text and appendix E-2 for additional details on the *tails* approach to estimating the MTE. Reported *benchmark* estimates are those obtained applying our baseline MTR estimation for the same underlying sample (the “late” sample). Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level.

Figure C-4: MTR Robustness, FLL Sample



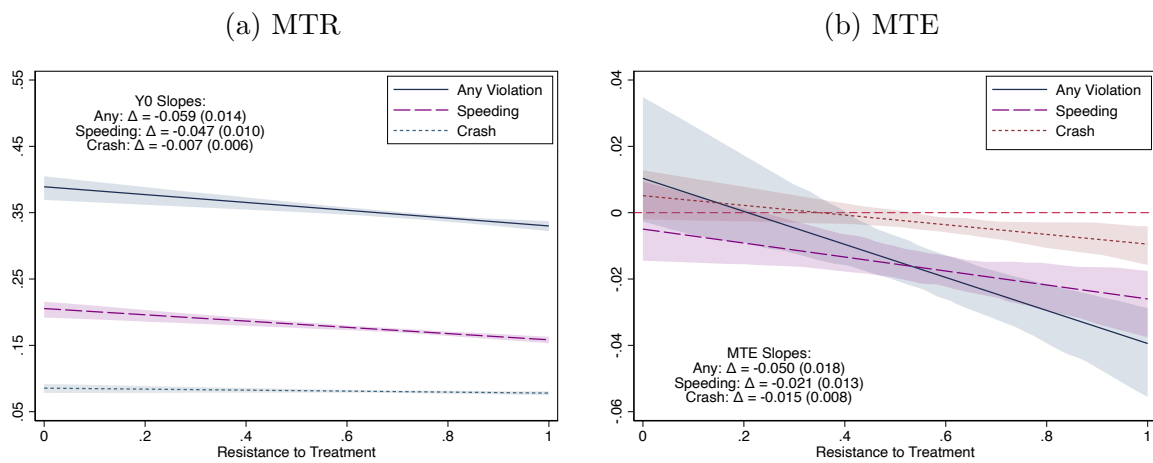
Notes: This figure reports our baseline MTR and MTE estimates with baseline covariate cells at the level of $\text{race} \times \text{gender} \times \mathbf{1}[\text{age} \geq 35] \times \mathbf{1}[\text{past offense}]$ (32 cells) for the full sample and for the subsample of officers that pass the Frandsen et al. (2019) fit test (see figure A-3). Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level.

Figure C-5: MTR Heterogeneity Across Officers



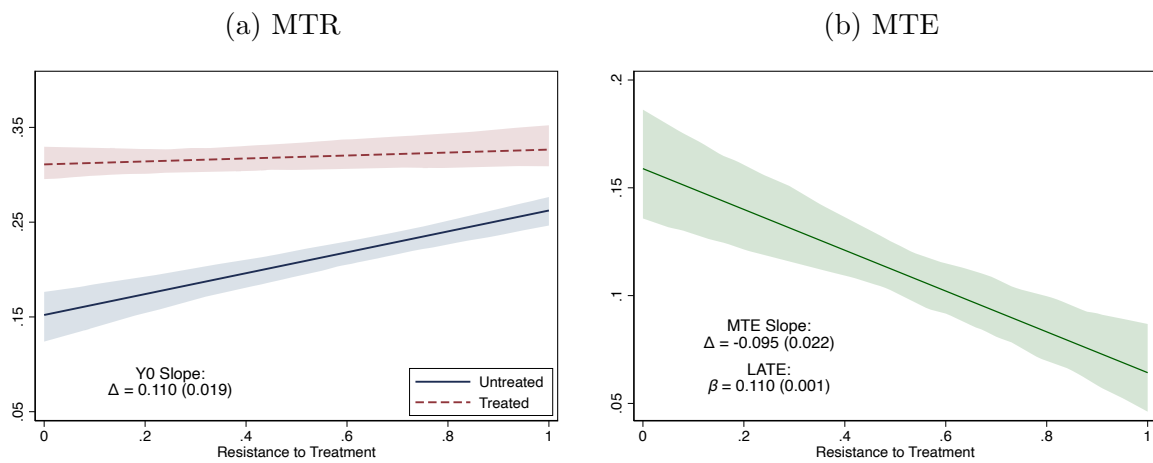
Notes: This figure reports heterogeneity in the estimated untreated MTR and MTE functions across officer groups using the sample of 2010-2015 citations (the “late” sample). The officer groups are cells at the level of gender \times $\mathbf{1}[\text{nonwhite}] \times \mathbf{1}[\text{any college}] \times \mathbf{1}[\text{high experience}]$, where high experience is defined as an above median number of speeding citations issued over the period 2005–2009 (same as in figure C-3). Figure reports the average slope weighting by each cell’s sample share as well as the p -value from a test of the null hypothesis that all slopes are equal. Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level.

Figure C-6: MTR Estimates, All Offending Outcomes



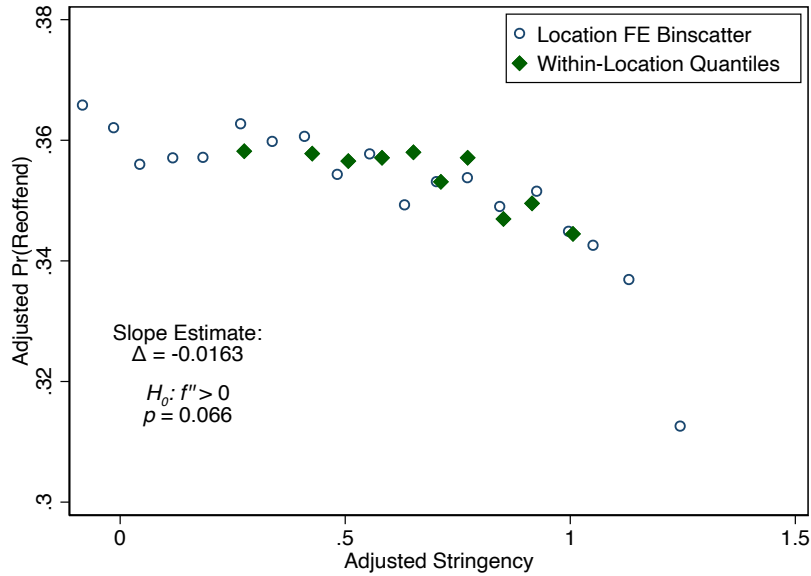
Notes: This figure reports our baseline MTR and MTE estimates with baseline covariate cells at the level of race \times gender \times $\mathbf{1}[\text{age} \geq 35] \times \mathbf{1}[\text{past offense}]$ (32 cells) varying the outcome of interest (any violation in the following year; any speeding offense in the following year; any crash involvement in the following year). Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level.

Figure C-7: MTR Estimates, Court Contesting



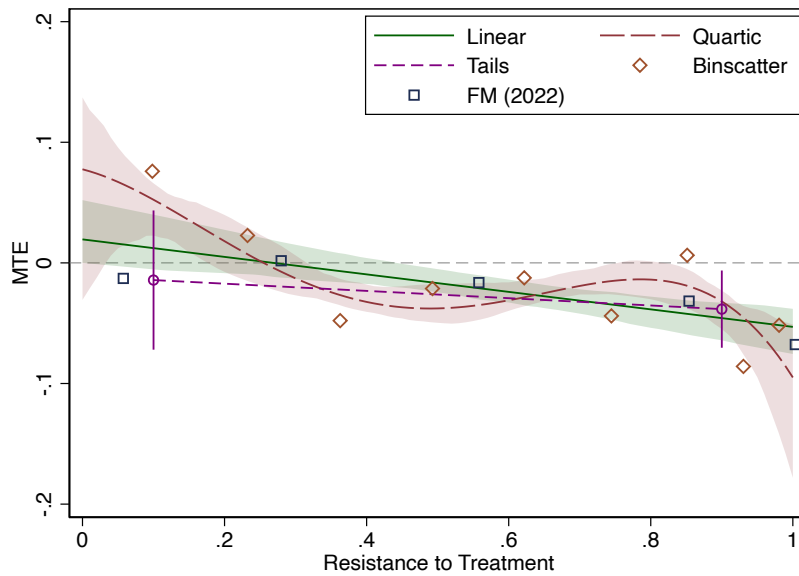
Notes: This figure reports our baseline MTR and MTE estimates with baseline covariate cells at the level of $\text{race} \times \text{gender} \times \mathbf{1}[\text{age} \geq 35] \times \mathbf{1}[\text{past offense}]$ (32 cells) where the outcome of interest is whether a citation is contested in court. See section D-1 for additional details on the computation of our contested ticket measure. Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level.

Figure C-8: Within-Locations Approach from Feigenberg & Miller (2022)



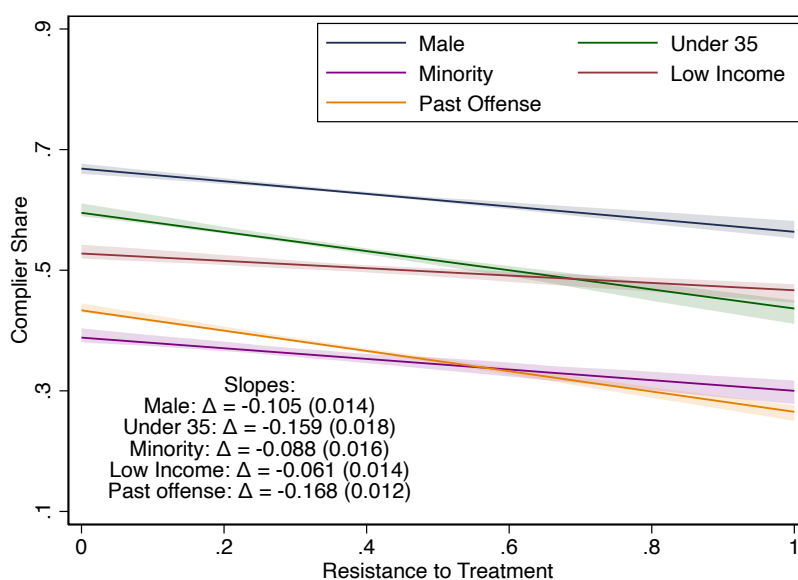
Notes: This figure illustrates the relationship between stringency and reoffending using the within-locations approach from Feigenberg & Miller (2022). See section E-6 for complete details on estimation. Figure reports the slope estimate, which is directly comparable to our baseline instrumental variables estimate ($\hat{\beta}_{IV} = -0.016$). The first derivative of the plotted relationship is directly comparable to our marginal treatment effect estimates. Figure also reports the p -value from a one-sided test of the null hypothesis that the underlying regression function has a positive second derivative (i.e., a test of the null that the MTE function is upward sloping) conducted via the `binstest` command from Cattaneo et al. (2021).

Figure C-9: Alternative MTE Estimates



Notes: Solid green line shows our baseline linear MTE (with associated confidence bands). Dashed red line shows a fourth-order polynomial estimate of the MTE (with associated confidence bands). Purple dots and connected dash line illustrate the tails approach. Orange diamonds plot an estimated nonparametric MTE obtained via a binscatter approach. Blue squares plot the estimated nonparametric MTE corresponding to the within-locations approach from Feigenberg & Miller (2022). All approaches condition on beat-shift fixed effects but not covariates except the Feigenberg & Miller (2022) where we include covariates per the approach outline in section E-6.

Figure C-10: Characteristics of Marginal Compliers



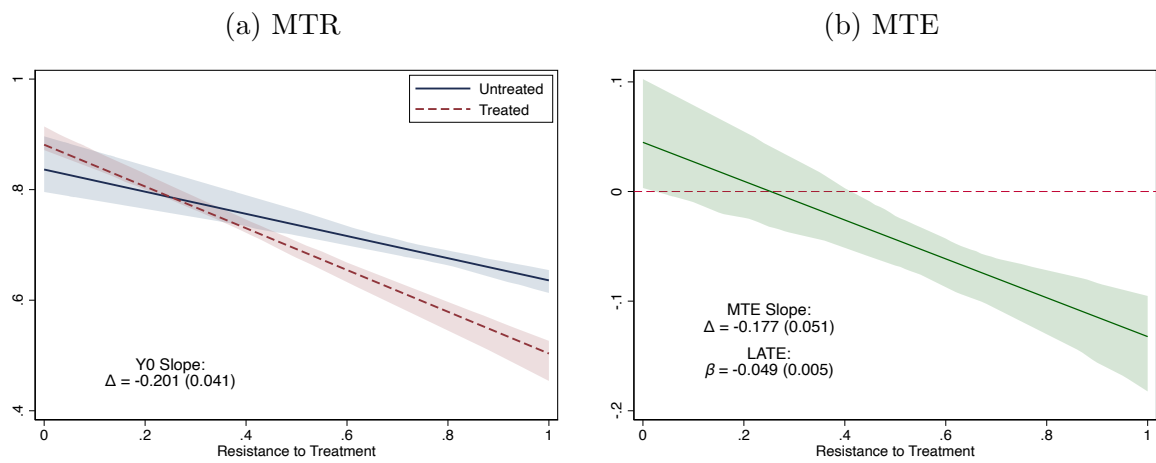
Notes: Figure reports estimated marginal complier shares using the method described in E-3. *Minority* is defined as race = Black or Hispanic and *Low Income* is defined as residing in a below-median income zip code. Past offense is defined as having received a traffic citation in the previous year. Plot reports estimated linear slopes with standard errors obtained via a bootstrap clustered by officer. Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level.

Table C-1: Robustness of Tails MTR Estimates

Tails Cutoff	N	Y_0	Implied Linear MTR Slope	Δ	Implied Linear MTE Slope
<i>A. Bottom: 0-10% v. 10-20%</i>	221736	0.411 (0.013)	-0.101 (0.019)	-0.005 (0.028)	-0.038 (0.041)
<i>Top: 80-90% v. 90-100%</i>	761987	0.330 (0.007)		-0.036 (0.016)	
<i>B. Bottom: 0-10% v. 10-35%</i>	392421	0.377 (0.007)	-0.059 (0.013)	0.019 (0.014)	-0.068 (0.025)
<i>Top: 75-90% v. 90-100%</i>	819636	0.334 (0.006)		-0.030 (0.011)	
<i>C. Bottom: 0-10% v. 10-50%</i>	570621	0.378 (0.005)	-0.066 (0.009)	0.008 (0.009)	-0.050 (0.016)
<i>Top: 50-90% v. 90-100%</i>	1113980	0.335 (0.003)		-0.024 (0.004)	
<i>Baseline Linear Estimation</i>			-0.059 (0.014)		-0.050 (0.018)

Notes: This table explores sensitivity of MTR estimates obtained via the tails approach (described in detail in section E-2) to the choice of comparison thresholds.

Figure C-11: MTR Estimates, Number of Offenses



Notes: This figure reports our baseline MTR and MTE estimates with baseline covariate cells at the level of $\text{race} \times \text{gender} \times \mathbf{1}[\text{age} \geq 35] \times \mathbf{1}[\text{past offense}]$ (32 cells) where the outcome of interest is whether a citation is the number of traffic offenses in the following year. Confidence bands and slope standard errors are computed via a bootstrap clustered at the officer-level.

Table C-2: Other Treatment Effect Parameters

	Treated	Untreated	Policy-Relevant Compliers
Y_0	0.379 (0.007)	0.330 (0.002)	0.355 (0.002)
$Y_1 - Y_0$	-0.004 (0.008)	-0.040 (0.006)	-0.009 (0.005)

Notes: This table presents the level of offending (Y_0) and treatment effect ($Y_1 - Y_0$) from a harsh fine for various groups. The values for treated individuals report estimates of $E(Y_0|X_i = \bar{X}, D = 1)$ and $E(Y_1 - Y_0|X_i = \bar{X}, D = 1)$, respectively, and the analogous values with $D = 0$ are reported for un-treated individuals. The policy-relevant compliers are individuals induced into treatment in the policy simulation considered in Section XX. Propensity scores are shifted from the estimated \hat{p}_i to $\tilde{p}_i = \min\{1, \hat{p}_i \times 1.1\}$. The level of offending for compliers is estimated as $\frac{E[-(1-D)Y_i|\tilde{p}] - E[-(1-D)Y_i|\hat{p}]}{E[\tilde{p}] - E[\hat{p}]}$, and the treatment effect for compliers (PRTE) is estimated as $\frac{E[Y_i|\tilde{p}] - E[Y_i|\hat{p}]}{E[\tilde{p}] - E[\hat{p}]}$.

C-1 Equity implications of officer behavior

Given sizable differences in criminal justice outcomes across racial groups more broadly, as well as in the context of speeding enforcement specifically (e.g., [Goncalves & Mello 2021](#)), another interesting question is the role that current officer sorting practices play in explaining racial disparities. In our sample, Black and Hispanic drivers are about 7.5 percentage points (12 percent) more likely than white motorists to be issued a harsh fine.

Figure [C-10](#), which presents demographic characteristics of marginal compliers, illustrates that younger, male, Black, and Hispanic drivers are prioritized for harsh sanctions. These gradients are significant; the least resistant motorists are about nine percent more likely to be Black or Hispanic than the most resistant offenders. Further, as shown in figure [B-4](#), Minority drivers reoffend at higher rates but exhibit similar responsiveness to harsh fines, suggesting that some of the racial gap in the likelihood of receiving harsh fines may be explained by officer sorting based on recidivism risk rather than deterrability.

A full examination of how officer objectives shape racial disparities in policing is beyond the scope of our paper. As a first step, however, we consider how the racial disparities in sanctions are affected in the simple counterfactual where officers sort drivers in reverse order. Results are presented in table [C-3](#). In the data, Black drivers are more likely to receive a harsh ticket than non-Black drivers in the same beat-shift (0.70 v. 0.65).¹⁸ This gap flips when drivers are sorted in reverse (0.617 v. 0.665). Similarly, Hispanic drivers are more harshly punished in our data than non-Hispanics (0.676 v. 0.653), which also reverses in our counterfactual (0.636 v. 0.663). Notably, these racial gaps are reversed while overall offending *declines*, as shown in table [4](#).

This analysis contributes to a recent literature considering equity and efficiency implications of criminal justice policy. [Rose \(2021\)](#) documents the racially disparate impact of probation rules that trigger prison sentences and shows that a narrowing of the set of offenses that trigger a sentence reduces the racial gap in incarceration at the cost of a small increase in overall offending. [Feigenberg & Miller \(2022\)](#), in the context of police vehicle searches for contraband, show that racial gaps in search rates can be narrowed while actually increasing the efficacy of searches, since the marginal search of minority drivers is less productive than the marginal search of white drivers. Similarly, our simple calculation illustrates the feasibility of changes in police practices that reduce racial gaps while simultaneously improving efficiency.

¹⁸These numbers are calculated by taking the values for $E(\text{Black}|X, U_D)$, where U_D is resistance to treatment, described in section [E-3](#). We then calculate $E(\text{Black}|X, \text{Treat})$ and $E(\text{Treat}|X, \text{Black}) = E(\text{Black}|X, \text{Treat}) \times \frac{Pr(\text{Treat})}{Pr(\text{Black})}$. We include in X the beat-shift fixed effects and report estimates for the average value of the fixed effects.

Table C-3: MTE Counterfactual and Driver Race

	Share Harsh Punishment	
	Baseline	Reverse Resistance to Treatment
Black	0.700	0.617
Non-Black	0.650	0.665
Hispanic	0.676	0.636
Non-Hispanic	0.653	0.663

Notes: Table presents the rates of harsh punishment for drivers, separately by driver race. The first column presents the rates of punishment in the data, and the second column presents the rates using the counterfactual where resistance to treatment is reversed. Further details are provided in Section C-1.

D Data Appendix

D-1 Traffic courts data

Traffic court dispositions associated with the citations from the *TCATS* database were also shared by the Florida Clerk of Courts. Citations were matched to disposition information using county codes and alphanumeric citation identifiers (which are unique within counties). Some citations have no associated disposition in the *TCATS* database, while others have multiple associated entries. Disposition verdicts can take on the following values:

1 = *guilty*; 2 = *not guilty*; 3 = *dismissed*; 4 = *paid fine or civil penalty*; 6 = *estreated or forfeited bond*; 7 = *adjudication withheld (criminal)*; 8 = *nolle prosequi*; 9 = *adjudged delinquent (juvenile)*; A = *adjudication withheld by judge*; B = *other*; C = *adjudication withheld by clerk (school election)*; D = *adjudication withheld by clerk (plea nolo and proof of compliance)*; E = *set aside or vacated by court*.

In practice, the verdicts 1, 3, 4, A, and C account for the vast majority of citations. Moreover, as confirmed in a phone conversation with Beth Allman at the Florida Clerk of Courts on July 24, 2018, several of the violation codes are difficult to interpret. In particular, it is very difficult in practice to infer the precise outcome of tickets with disposition codes 1, 3, A, or those with multiple dispositions in the TCATS database.

To construct an approximate measure of court contesting, we use any disposition not equal to 4 or C, which both imply that the individual paid their fine without contest, as an indicator that the driver contested a citation. To construct measures of *effective* sanctions, termed *paid* fines and *accrued* points in figure A-4, we adjust the statutory sanctions as follows:

- Replace fine = fine/2 if verdict = A
- Replace fine = 0 if verdict = 3
- Replace points = 0 if verdict $\in \{3, A, C\}$

Note that our measure of *paid* fines is likely conservative as it ignores court fees. Drivers contesting their tickets in court face a \$75 court fee in addition to their fine (the court fee can also be waived during the court process). See [Goncalves & Mello \(2021\)](#) and [Mello \(2021\)](#) for further discussion of the issues associated with working with the TCATS data.

D-2 Binary stringency measure

To identify officers who do not bunch, we use the [Frandsen \(2017\)](#) test for manipulation. In our setting, this test implies that, under the null hypothesis of no manipulation, the conditional probability of being found at the bunching speed is in a range around one third, $Pr(X = 9|x \in [8, 10]) \in [(1 - k)/(3 - k), (1 + k)/(3 + k)]$ where k is a restriction on the second finite difference, $\Delta^{(2)}Pr(S = 9) \equiv Pr(S = 8) - 2Pr(S = 9) + Pr(S = 10)$, such that $|\Delta^{(2)}Pr(S = 9)| \leq k(Pr(S = 9) - Pr(S = 10))$. Intuitively, if the distribution of

ticketed speeds is unmanipulated, the share of tickets at 9 MPH among those between 8 and 10 MPH should be approximately one-third, where the deviation k is due to curvature in the distribution of speeds. We calculate k by assuming the distribution $Pr(S)$ is Poisson and estimating the mean parameter λ using the empirical mean of ticketed speeds. We say that an officer is stringent (non-bunching) if we fail to reject that $Pr(S = 9|S \in [8, 10]) \leq (1 + k)/(3 + k)$ at the 99 percent confidence level.

To avoid the reflection problem, we randomly partition an officer’s stops into two halves and compute the binary measure separately for each half of the sample. We then use the officer’s binary measure in the *other half* as our binary stringency measure.

D-3 Predicted recidivism

At a few points in our analysis, we rely on a predicted reoffending measure based on covariates (e.g., figure 2). An important concern in constructing this measure is the possibility of contamination from treatment effects (e.g., if certain covariates are highly correlated with receiving harsh fines, the predicted recidivism for these covariate groups will be too low because of the treatment effect of harsh fines on future offending). To circumvent this concern, we construct our predicted recidivism index as follows. First, we randomly partition each officer’s stop into halves. Each officer \times partition is coded as lenient or nonlenient using the Frandsen (2017) described above. In each partition, we regress $\mathbf{1}$ [any new traffic offense] on driver covariates using only stops made by non-lenient (non-bunching) officers. We then use the coefficients from this regression to construct our \hat{Y} measure in the other partition.

E Technical Appendix

E-1 MTR and MTE Estimation Details

We are interested in identifying the distribution of counterfactual outcomes by resistance to treatment, $E(Y_j|U_D = u, X)$. We will follow the approach of Heckman & Vytlacil (2007), which considers the conditional expectation of Y separately by treatment status:

$$\begin{aligned}
 E[Y|P(Z) = p, X, D = 0] &= E[Y_0|U_D > p, X] \\
 &= \mu_0(X) + E[U_0|U_D > p, X] \\
 &= \mu_0(X) - \frac{p}{1-p} E[U_0|U_D < p, X] \\
 &= \mu_0(X) - \frac{1}{1-p} \int_0^p E[U_0|U_D = u, X] du
 \end{aligned}$$

Imposing the functional form assumption that the MTR is linear, we have that $E[U_0|U_D = u, X] = \alpha_0(p - 1/2)$, where $-1/2$ is needed so that $E(U_0|X) = 0$:

$$\begin{aligned}
 E[Y|P(Z) = p, X, D = 0] &= \mu_0(X) - \frac{1}{1-p} \int_0^p \alpha_0(p - 1/2) du \\
 &= \mu_0(X) - \frac{\alpha}{1-p} \left[\frac{p(p-1)}{2} \right] \\
 &= \mu_0(X) + \alpha_0 \frac{p}{2} \\
 &= X_i \beta + \alpha_0 \frac{p}{2}
 \end{aligned}$$

So we now have a functional form for what the conditional expectation of Y reflects when restricting attention to $D = 0$. With the assumption of a linear MTE, the regression is linear in $p/2$, and its coefficient reflects the shape of the potential outcome function for Y_0 . Note that, as shown by Brinch et al. (2017), the fact that the expectation is linear in p means that only a binary instrument is needed to identify the shape of the potential outcome function.

The steps we will take to estimate the potential outcome function are the following:

1. Estimate $P(Z)$ using the full sample, get \hat{p} for each observation.
2. Regress Y on X and $\frac{\hat{p}_i}{2}$, among those with $D_i = 0$.
3. Calculate $\hat{y}_i = X_i \hat{\beta} + \alpha_0 \frac{\hat{p}_i}{2}$ for all individuals (including with $D_i = 1$).
4. Calculate $\widehat{\mu_0(\bar{X})} = \frac{1}{N} \sum_i [\hat{y}_i - \hat{\alpha}_0 \frac{\hat{p}_i}{2}]$, then construct the potential outcome function for $Y = 0$:

$$E[Y_0|\widehat{U_D = u, \bar{X}}] = \widehat{\mu_0(\bar{X})} + \hat{\alpha}_0 \cdot (u - 1/2), \quad u \in [0, 1]$$

To identify the treated potential outcome function, we use a similar approach, and also

assume a linear MTR, $E[U_1|U_D = u, X] = \alpha_1(p - 1/2)$:

$$\begin{aligned}
E[Y|P(Z) = p, X, D = 1] &= E[Y_1|U_D \leq p, X] \\
&= \mu_1(X) + E[U_1|U_D \leq p, X] \\
&= \mu_1(X) + \frac{1}{p} \int_0^p E[U_1|U_D = u, X] du \\
&= \mu_1(X) + \frac{1}{p} \int_0^p \alpha_1(p - 1/2) du \\
&= \mu_1(X) + \frac{1}{p} [\alpha_1(p^2/2 - p/2)] \\
&= \mu_1(X) + \alpha_1(p/2 - 1/2)
\end{aligned}$$

E-2 Tails IV

The estimation of the marginal treatment response functions using the entire distribution of officer instrument values and the assumption of linear unobservable components allows for the precise estimation of counterfactual outcomes for individuals with both high and low resistances to treatment. However, the assumption of linearity may be violated, along with the strong assumption of monotonicity in the instrument's effect on treatment.

To avoid these concerns, we additionally estimate the potential outcomes for high and low resistance individuals using an alternative approach. For low resistance individuals, we will focus on officers with instrument value $Z \leq 0.2$. We will then construct a binary version of the instrument, $\tilde{Z} = Z > 0.1$. An IV regression of recidivism on receiving the full fine with the binarized instrument will reflect the treatment effect for complier individuals, who are those with $u_D \in [0.1, 0.2]$. In addition, [Abadie \(2002\)](#) shows how interacting the outcome with treatment status can identify the counterfactual outcome for the compliers:

$$\begin{aligned}
\frac{E[DY|X, \tilde{Z} = 1] - E[DY|X, \tilde{Z} = 0]}{E[D|X, \tilde{Z} = 1] - E[D|X, \tilde{Z} = 0]} &= E[Y_1|X, U_D \in [0.1, 0.2]] \\
\frac{E[(1 - D)Y|X, \tilde{Z} = 1] - E[(1 - D)Y|X, \tilde{Z} = 0]}{E[(1 - D)|X, \tilde{Z} = 1] - E[(1 - D)|X, \tilde{Z} = 0]} &= E[Y_0|X, U_D \in [0.1, 0.2]]
\end{aligned}$$

We will therefore run regressions of DY and $-(1 - D)Y$ on our treatment, where we restrict attention to officers with $Z < 0.2$ and instrument for treatment with \tilde{Z} . To identify the counterfactual outcomes of high resistance individuals, we will analogously restrict attention to individuals stopped by officers with $Z \geq 0.8$ and use the binarized instrument $\tilde{Z} = Z > 0.9$.

E-3 Characteristics of marginal compliers

We are interested in identifying the demographics of drivers who are at each level of resistance to treatment. We will denote by X_k some driver demographic variable k , and we denote by \tilde{X} the set of all location-time fixed effects. We are interested in identifying $E[X_k|\tilde{X}, U_D = u]$,

which we will impose to have an additively separable form with a linear term in u :

$$E[X_k | \tilde{X}, U_D = u] = \tilde{X}\alpha_k + \theta_k(u - 1/2)$$

We will identify α_k and θ_k using a procedure similar to the calculation of the MTR functions. We estimate the conditional expectation of X_k given the propensity score for the set of punished individuals:

$$\begin{aligned} E[X_k | \tilde{X}, P(Z) = p, D = 1] &= E[X_k | \tilde{X}, U_D < p] \\ &= \tilde{X}\alpha_k + \theta_k(E[U_D | U_D < p] - 1/2) \\ &= \tilde{X}\alpha_k + \theta_k(p/2 - 1/2) \end{aligned}$$

We will also use this approach to estimate the average stopped speed of drivers at each U_D . Denoting stopped speed by S_i^* and ticketed speed by S_i , we know by design that $S_i(D = 1) = S_i^*$ and $S_i(D = 0) = 9$. The procedure outlined above will identify marginal treatment response for $S_i^*(D = 1)$, which corresponds to stopped speed.

E-4 Predicted Re-offending From Stopped Speed

In section 5.3.2, we ask how much of the observed selection on levels can be explained by the stopped speed of drivers. To answer this question, we first calculate predicted reoffending based on stopped speed. For each county and stopped speed, and among officers who bunch no drivers, we calculate the share of drivers who reoffend. We assign this predicted offending value to all drivers who are given a harsh ticket and for whom we thus see stopped speed. We then conduct our complier characteristics analysis from section E-3 with predicted re-offending as the driver characteristic. This curve tells us the predicted re-offending of drivers at each level of resistance to treatment, based on their stopped speed. Note that the complier characteristics analysis of section E-3 intentionally restricts attention to drivers with a harsh ticket and does not suffer from any endogeneity concerns from this sample being selected.

E-5 Model Estimation Details

The model outputs that correspond to observed information in our data are the values of treated and untreated offending rates for individuals who are at the margin of punishment for officers at a given propensity to treat:

$$h_j(\theta) \equiv E[Y_j | \lambda\hat{Y}_0 - (1 - \lambda)(\hat{Y}_1 - \hat{Y}_0) = g^{-1}(\theta)], \quad j \in \{0, 1\}$$

where θ is a probability of punishment, and $g^{-1}(\theta)$ maps a probability of punishment to the cost of punishing that leads to that probability. In other words, this function identifies the average Y_j for drivers at the θ th percentile of the objective function.

These functions correspond to the marginal treatment responses we estimate in the data, $m_j(\theta) = E[Y_j | \bar{X}, U_d = \theta]$. We estimate these functions using linear specifications, $\hat{m}_j(\theta) = \hat{\alpha}_{0j} + \hat{\alpha}_{1j}(\theta - 1/2)$, and we aim to match the level and slope of these functions between the

model and data:

$$\hat{\alpha}_{0j}^{LB} \leq \int_0^1 h_j(u) du \leq \hat{\alpha}_{0j}^{UB} \quad (\text{E-1})$$

$$\hat{\alpha}_{1j}^{LB} \leq \int_0^1 \frac{\partial h_j(u)}{\partial u} du \leq \hat{\alpha}_{1j}^{UB}, \quad j \in \{0, 1\} \quad (\text{E-2})$$

To account for estimation error in our empirical estimates, we require only that our model moments fall within the 95% confidence intervals for our empirical moments $[\hat{\alpha}_{kj}^{LB}, \hat{\alpha}_{kj}^{UB}]$, $(k, j) \in \{0, 1\}^2$. We therefore have four moments to inform the model parameters. The model contains the weight parameter λ and the distribution of signals $F(\hat{Y}_{i0}, \hat{Y}_{i1})$. Unless we place substantial restrictions on the distribution of signals by parametrizing it with three or fewer parameters, the model parameters are not point identified from the marginal treatment responses. However, they may provide informative bounds on their true values. We will focus in particular on estimating the identified region for λ .

We will treat the estimation of λ as an optimization problem, where the above moment conditions must hold:

$$\lambda^u \equiv \max_{\lambda, F(\hat{Y}_{i0}, \hat{Y}_{i1})} \lambda \quad \text{s.t. (E-1) and (E-2) hold}$$

$$\lambda^l \equiv \min_{\lambda, F(\hat{Y}_{i0}, \hat{Y}_{i1})} \lambda \quad \text{s.t. (E-1) and (E-2) hold}$$

Our estimate of λ will be the region $[\lambda^l, \lambda^u]$.

We solve this pair of optimization problems using the genetic algorithm in matlab. The algorithm picks a set of starting points to evaluate the objective function. Depending on the value of the objective function at that point, the point has a probability of “survival.” If it survives, a new candidate point is generated nearby (the initial point’s “offspring”). In addition to these points, each generation has a random set of new guesses that do not originate with any points from the previous generation.

One of the key inputs for the problem is the choice of starting guesses for values of λ and $F(\hat{Y}_{i0}, \hat{Y}_{i1})$. In practice, we generate a grid of potential values $\{Y_0^k, Y_1^k\}$, and we specify a set of probabilities of each point on the grid.

The first guess we provide is a set of points that lie on the marginal treatment response functions, so that $Pr(Y_0^k, Y_1^k) \neq 0$ if $Y_0^k \in (Y_0|Y_0 = \hat{\alpha}_{00} + \hat{\alpha}_{10}u, \text{ for some } u \in [0, 1])$ and $Y_1^k \in (Y_1|Y_1 = \hat{\alpha}_{01} + \hat{\alpha}_{11}u, \text{ for some } u \in [0, 1])$, $Pr(Y_0^k, Y_1^k) = 0$ otherwise, and all non-zero probability points have the same likelihood. We provide 101 guesses with this grid of probabilities, with values for $\lambda = 0, 0.01, \dots, 1$.

The second set of guesses are off of the MTR functions. For values of $Y_{0\theta}$ and $Y_{1\theta}$ that lie on the MTR functions, we give non-zero probability to guesses $\hat{Y}_0, \hat{Y}_1, \hat{\hat{Y}}_0,$ and $\hat{\hat{Y}}_1$ that

satisfy the following, for a given set of ω and λ :

$$\begin{aligned}\omega\hat{Y}_0 + (1 - \omega)\hat{Y}_0 &= Y_{0\theta} \\ \omega(\hat{Y}_1 - \hat{Y}_0) + (1 - \omega)(\hat{Y}_1 - \hat{Y}_0) &= Y_{1\theta} - Y_{0\theta} \\ \lambda\hat{Y}_0 - (1 - \lambda)(\hat{Y}_1 - \hat{Y}_0) &= \lambda Y_{0\theta} - (1 - \lambda)(Y_{1\theta} - Y_{0\theta}) \\ \lambda\hat{Y}_0 - (1 - \lambda)(\hat{Y}_1 - \hat{Y}_0) &= \lambda Y_{0\theta} - (1 - \lambda)(Y_{1\theta} - Y_{0\theta})\end{aligned}$$

These guesses create a set of posteriors that average to a value on the marginal treatment response curves and that have the same objective function value as the guesses on the marginal treatment response curve.

Counterfactual Calculation – The counterfactual calculation requires identifying the set of parameter values $\lambda, F(\hat{Y}_{0k}, \hat{Y}_{1k})$ that satisfy the empirical moment inequalities and setting $\lambda = 0$ for each case.

We are interested in reporting the offending rate of treated individuals, $E[Y_0|D = 1]$. To calculate the range of possible values, we similarly perform a pair of optimization problems:

$$\begin{aligned}Y_{0,treated}^u &\equiv \max_{\lambda, F(\hat{Y}_{i0}, \hat{Y}_{i1})} E[Y_0|D = 1] \text{ when } \lambda \rightarrow 0 \\ &\text{s.t. (E-1) and (E-2) hold for } \lambda, F(\hat{Y}_{i0}, \hat{Y}_{i1}) \\ Y_{0,treated}^l &\equiv \min_{\lambda, F(\hat{Y}_{i0}, \hat{Y}_{i1})} E[Y_0|D = 1] \text{ when } \lambda \rightarrow 0 \\ &\text{s.t. (E-1) and (E-2) hold for } \lambda, F(\hat{Y}_{i0}, \hat{Y}_{i1})\end{aligned}$$

To calculate Y_0 for the untreated individuals in this counterfactual, we use the fact that $E[Y_0] = Pr(D = 0)E[Y_0|D = 0] + Pr(D = 1)E[Y_0|D = 1]$, where we observe $Pr(D = 0)$ empirically and we calculate $E[Y_0]$ from the value of $F(\hat{Y}_{i0}, \hat{Y}_{i1})$ that solves the min/max optimizations above. We take a similar set of steps to solve for range of values for $E[Y_1 - Y_0|D = 0]$ and $E[Y_1 - Y_0|D = 1]$.

E-6 Within-Location Approach from Feigenberg & Miller (2022)

Let j index officers, ℓ index counties, and τ index time categories, defined as shift \times 1[weekend]. Let t index time, defined as calendar year \times month. We construct an adjusted stringency measure for each officer \times location by estimating the regression:

$$D_{ij\ell\tau} = \phi_{j\ell\tau} + \gamma X_{ij\ell\tau} + \delta_\tau + u_{ij\ell\tau}$$

where the ϕ 's are fixed effects for each officer \times location \times shift category and $D = \mathbf{1}$ [harsh fine]. We estimate this regression separately for each location. We then aggregate to the officer \times location level as follows:

$$\tilde{D}_{j\ell} = \sum_{\tau} \left(\hat{\phi}_{j\ell\tau} + E[\hat{\gamma} X_{ij\ell\tau} + \hat{\delta}|\ell, \tau] \right) P(\tau|\ell)$$

We then repeat the exact same procedure, replacing D with $Y = \mathbf{1}[\text{reoffend}]$ to obtain an adjusted probability of reoffending for each officer \times location, $\tilde{Y}_{j\ell}$.

Following Feigenberg & Miller (2022), we document the relationship between $\tilde{Y}_{j\ell}$ and $\tilde{D}_{j\ell}$ in two ways. First, we show a simple binscatter, conditional on location fixed effects and weighting each officer \times location by number of stops via the `binsreg` command. We also estimate a linear slope corresponding to this approach by regressing $\tilde{Y}_{j\ell}$ on $\tilde{D}_{j\ell}$ and location fixed effects, weighting by the number of stops. Second, we construct location-specific quantiles of each and plot the relationship between quantiles.

In figure C-9, we show a version of the MTE based on the Feigenberg & Miller (2022) approach. This MTE estimate is obtained by estimating a binscatter of the derivative of the relationship between $\tilde{Y}_{j\ell}$ on $\tilde{D}_{j\ell}$, conditional on location fixed effects and weighting by the number of stops, where we estimate the derivative using `binsreg`.