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

Felipe Goncalves† Steven Mello‡

March 16, 2022

Abstract

We study police discretion over sanctions for speeding offenses. Relying on variation across officers in the propensity to issue harsh fines, we estimate the causal effect of sanctions and find that a 100 dollar increase in fines reduces the likelihood of a new speeding offense in the following year by about seven percent. We then use a marginal treatment effects approach to learn about officer objectives. The sorting of motorists into sanctions by officers exhibits selection on levels and inverse selection on gains, suggesting (i) that officers face a tradeoff between allocating sanctions to maximize deterrence and sanctioning the most-frequent offenders and (ii) that officers prioritize the latter, consistent with a retribution objective. We characterize how officers trade off deterrence and retribution goals by estimating a model of officer decisions, finding that officers weigh retribution as least as much as deterrence when allocating sanctions. We estimate that the reoffending rate is at least two percentage points, or six percent, lower in a counterfactual scenario where officers prioritize deterrence alone.

JEL Codes: K42

*We are grateful to Ilyana Kuziemko and Alex Mas for advice and encouragement on this project. We benefitted from helpful comments by Leila Agha, David Arnold, Eric Chyn, Will Dobbie, Ben Hansen, Peter Hull, Elisa Jacome, Emily Leslie, Erzo Luttmer, Emily Nix, Sam Norris, Rodrigo Pinto, Bruce Sacerdote, Yotam Shem-Tov, Doug Staiger, and Emily Weisburst; as well as seminar participants at Princeton, UCLA, Dartmouth, Case Western, UBC, Claremont McKenna, SDSU, and the SEA Annual Meeting. 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; fgoncalves@ucla.edu

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

1 Introduction

Two key principles, deterrence and retribution, underlie the design of criminal justice systems around the world. Under the first principle, deterrence, the central goal of a legal system is to maximize public safety. Accordingly, criminal sanctions should be set to equate the marginal social value of crime deterred and the marginal social costs of harsher punishments (Becker, 1968). To this end, a large empirical literature has evaluated the degree to which various criminal sanctions and law enforcement interventions reduce crime and whether deterrence benefits justify the fiscal and social costs of an expanded or harsher criminal justice system.[1]

The second principle, retribution, advocates that punishing offenders is socially desirable in its own right, independent of deterrence effects, for reasons such as fairness, state legitimacy, and morality (e.g., Kaplow & Shavell 2006; O'Flaherty & Sethi 2019; Moore 2019). Under retribution, optimal sanctions should "fit the crime" and are a theoretical question about society's preferences, rather than an empirical question about the effects of punishments. Still, an important question for scholars and policymakers is whether inherent tradeoffs exist between accomplishing deterrence and retribution goals.

In this paper, we study how these two objectives shape the behavior of street-level bureaucrats tasked with law enforcement. Given the reliance of criminal justice systems on agents wielding considerable discretion over punishments, how those agents use their discretion, and the public safety consequences of those discretionary choices, are questions of significant policy interest. Examining a setting where patrolling police officers exercise discretion over sanctions, speeding enforcement, our analysis proceeds in three parts. First, we document important deterrence effects of sanctions. Next, we show that officers face an explicit tradeoff between deterrence and retribution objectives in this setting. Finally, we characterize how officers weigh deterrence and retribution goals when allocating sanctions.

Speeding enforcement is a high stakes setting in terms of public safety. Traffic accidents are the second leading cause of death among individuals aged 15-34 in the United States. In 2020, there were nearly twice as many traffic fatalities ($\sim$ 39,000) as homicides ($\sim$ 22,000). 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). The National Highway Safety Administration estimates 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 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

[1]As examples, see Mueller-Smith (2015), Bhuller et al. (2020), and Rose & Shem-Tov (2021) on incarceration, Huttunen et al. (2020) and Finlay et al. (2021) on financial sanctions, and Chalfin & McCrary (2018) and Mello (2019) on police employment.

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 before 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).

Estimating causal effects of harsher speeding punishments on driver behavior is challenging due to the non-random assignment of sanctions. Speeding sanctions are tied to driving speeds relative to the posted speed limit, meaning that more severe transgressions are statutorily punished more harshly. The manipulation of sanctions by officers compounds endogeneity concerns. To circumvent these identification challenges, we rely on systematic variation across officers in bunching propensity. Specifically, we use the citing officer's propensity *not to bunch* other drivers, which we call officer stringency, as an instrument for the fine faced by a driver.

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. The key concern, then, for our identification strategy 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 measure of officer stringency is uncorrelated with driver characteristics that predict recidivism and with an officer's ticketing frequency. Moreover, we show that the citing officer's stringency cannot predict past offending. In contrast, officer stringency predicts a stark decline in traffic offending *just after* the citation.

Our main IV estimate suggests that a $100 increase in fines reduces the likelihood of a new speeding offense in the following year by 1.4 percentage points. Converting this point estimate into a fine elasticity gives $\epsilon = -0.13$ ($se = 0.01$); in other words, a doubling of the fine reduces the reoffending probability by 13 percent. We also find statistically significant, but smaller, effects of harsh fines on non-speeding traffic offenses ($\epsilon = -0.06$) and on crash involvement ($\epsilon = -0.04$) in the following year. Our estimated fine elasticities contribute to a large literature on criminal deterrence (Chalfin & McCrary 2017; Doleac 2021), the effects of sanctions on offender behavior (Mueller-Smith 2015; Bhuller et al. 2020; Huttunen et al. 2020; Rose & Shem-Tov 2021; Finlay et al. 2021), and the effects of traffic policing on road safety (Ashenfelter & Greenstone 2004; Makowsky & Stratmann 2011; DeAngelo & Hansen 2014; Luca 2014; Traxler et al. 2018).

In our setting, motorists ticketed harshly and leniently face the same sanctions for future offenses. Hence, the *ex post* responses we observe represent a specific deterrence effect, or the impact of the experience of punishment on offending (Nagin, 2013). Isolating specific deterrence effects is often empirically challenging due not only to identification issues but also the simultaneous presence of incapacitation or general deterrence effects (e.g., Hansen

2015). Our estimates contribute to a small literature isolating specific deterrence effects (e.g., Gehrsitz 2017; Libor & Traxler 2021). Exploring mechanisms further, we find stronger impacts of harsh fines for the subset of motorists encountering a stringent officer for the first time, consistent with drivers learning about punishments as an important mechanism.

Having shown that sanction decisions have important deterrence effects, we turn to the question of how deterrence and retribution goals shape officer decisions. Specifically, we consider how officers sort a set of offenders into harsh versus lenient sanctions. Deterrence goals dictate that officers should prioritize punishing motorists most responsive to harsh sanctions. We take the view that, to achieve retribution goals, officers prioritize allocating harsh sanctions to motorists most likely to reoffend, independent of their treatment responses.[2] To explore these hypotheses empirically, we use a marginal treatment effects framework, which allows for identification of heterogeneity in potential outcomes for drivers with varying propensities to be treated. If officers prioritize punishing deterrable drivers, the motorists most likely to be punished harshly, or those who are induced into harsh sanctions at low values of our stringency instrument, should exhibit the largest treatment effects. If officers prioritize retribution, we expect to see the highest reoffending rates among the motorists most likely to be punished. As these patterns are not necessarily mutually exclusive, it is an empirical question whether officers must choose between prioritizing levels and gains when sorting drivers into sanctions.

Estimating marginal treatment responses (MTR's) using the approach from Heckman & Vytlacil (2007a), we find that both treatment-specific potential outcomes and marginal treatment effects are decreasing in stringency.[3] In other words, the MTR's exhibit selection on levels, consistent with a retribution objective, and *inverse* selection on gains, inconsistent with deterrence maximization. Further, these selection patterns imply an explicit tradeoff between deterrence and retribution goals; because the most deterrable offenders are those with the lowest reoffending rates and vice versa, harsh punishments allocated to achieve a retribution goal have efficiency costs in the form of foregone deterrence.

Given the presence of this tradeoff, we write down and estimate 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 reoffending rates

[2] A broader definition (e.g., Kaplow & Shavell 2006) would allow for retribution to capture any non-deterrence preferences. While our decision to focus on reoffending levels may seem restrictive, our finding that officer behavior cannot be explained by deterrence goals alone is evidence that officers incorporate some notion of retribution into their punishment decisions, using this broad definition from Kaplow & Shavell (2006).

[3] As described in more detail in section 5.1, standard methods for MTE estimation rely on a potentially problematic strict monotonicity assumption. We leverage a unique feature of our setting, the full support of our stringency instrument, to additionally estimate a version of our marginal treatment effects that rely on a weaker assumption by using only variation in propensity scores near zero and one.

(levels) and expected reoffending responses (gains). The weight that an officer places on reoffending levels, or the weight placed on the retribution objective, is set-identified based on the moments of the estimated marginal treatment response functions.

The model estimates indicate that officers place at least as much weight on expected reoffending rates as expected offending responses when allocating harsh fines. Moreover, we cannot rule out that officers consider only reoffending rates when allocating sanctions. Our characterization of officer objectives advances a growing literature interested in the causes and consequences of discretion in the criminal justice system (e.g., Knowles et al. 2001; Weisburst 2017; Chalfin & Goncalves 2021; Goncalves & Mello 2021; Abrams et al. 2021).

To quantify the efficiency costs associated with retributive preferences, we estimate counterfactual reoffending rates under a different objective function. When officers are made to consider only deterrence and place no weight on accomplishing retribution goals, the overall reoffending rate declines by at least two percentage points (about six percent). In this counterfactual, the allocation of punishments becomes significantly less retributive: harshly sanctioned drivers have *lower* reoffending rates than those given lenient sanctions, further highlighting the inherent tradeoff between deterrence and retribution goals. We also find similar results with a simpler counterfactual that forces officers to sort drivers in reverse order of their current practice, leading to improved deterrence and reduced retribution.

Finally, we consider the equity implications of officers' retributive preferences. Given large racial disparities in a host of criminal justice outcomes, including speeding sanctions (Goncalves & Mello, 2021), we ask how racial differences in sanctions change with a reduction in retributive preferences. In a simple counterfactual with reversed driver sorting, the observed racial gaps in punishment are not only reduced but reversed, leading to lower punishment rates for Black and Hispanic drivers. This finding adds to the recent literature examining equity-efficiency tradeoffs in criminal justice policy (Feigenberg & Miller, 2021; Rose, 2021) and highlights the potential for changes to policing practices that reduce racial gaps in treatment while actually improving efficiency.

Our central contribution is to a literature considering the role of fairness concerns in the design of legal institutions (e.g., Kaplow & Shavell 2006; O'Flaherty & Sethi 2019; Moore 2019).[4] While this literature has theoretically examined the proper role of fairness considerations, we contribute novel empirical evidence to this debate. Our paper is the first to document the empirical relevance of the tradeoff between accomplishing fairness (retribution) and efficiency (deterrence) goals. Further, we empirically demonstrate a preference for fairness in the sanction decisions of criminal justice agents and quantify the efficiency costs associated with those retributive preferences.

Another contribution of our paper is to a broad literature on the allocation choices of economic agents. A common approach in this literature is to examine selection patterns

[4] See Kaplow & Shavell (2006) for specific applications to various legal areas, including contracts, torts, and criminal justice.

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 studies have nonetheless found decision-making 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 this type of decision-making in a new setting, criminal justice, and estimate the foregone deterrence associated with targeting sanctions based on offending levels rather than gains.

The rest of our paper proceeds as follows. Section 2 describes our data and section 3 provides the relevant institutional background. We devote section 4 to the estimation of causal effects of sanctions. Section 5 studies officer objectives and section 6 concludes.

2 Data

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 record a motorist's race as Hispanic if, based on census records, their surname is associated with Hispanic status for more than 80% of individuals.[5] 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.

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 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

[5]As discussed in Goncalves & Mello (2021), there are clear inconsistencies in the recording of Hispanic status in the UCT. 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.

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.

Beginning in 2013, a subset of FHP-issued citations in the UTC database include GPS coordinates. We match these citations to road segments and, in the appendix, replicate our main analyses controlling for precise location fixed effects.

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.[6] 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 1,693,436 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.

[6]We 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.

In our analysis, we focus on whether a driver receives a new speeding citation in the year following their focal FHP citation as our primary outcome of interest, but also show results for other types of traffic offenses as well as for crash involvement using the FDOT crash records. Again, reoffending and past offending are measured using all citations issued in the state rather than just the set of citations that comprise our focal FHP sample.

Worth noting here is the fact that our main outcome measure will capture whether a motorist is caught and ticketed for a new speeding offense, which itself could be subject to officer discretion. If anything, we expect that officer discretion at the recidivism stage will bias our estimates towards zero. Our IV estimates will compare reoffending rates for individuals receiving harsh and lenient fines and we find that those receiving harsh fines differentially reduce their offending rates. If officers are more likely to cite drivers with more severe offending histories or 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 reoffending rates of those who are sanctioned harshly or deflating reoffending rates of those who are given lenient sanctions.

3 Institutional background

3.1 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.

3.2 Discretion over sanctions

Panel (a) of figure 1 shows the speeding fine schedule in Florida and the 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 of 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), panel (b) of figure 1 shows that all bunching is attributable to a subset of *lenient* officers.[7] 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 illustrate 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 location and time fixed effects. Such variation is inconsistent with bunching due to driver

[7] See appendix B-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.

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:

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

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.[8] This regression has an $R^2 = 0.55$, with 0.31 (56 percent) attributable to the officer effects, 0.22 (41 percent) attributable to the beat-shift effects, and 0.018 (3 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 a full set of driver characteristics. Moreover, there is significant variation in the estimated $\hat{\alpha}_j$'s ($Var \approx 0.07$). Applying Empirical Bayes shrinkage (Morris, 1983) to adjust for estimation error has minimal impact on the dispersion of the estimated officer effects ($Var \approx 0.06$). 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 $\hat{\beta} = 0.57$ ($se = 0.03$). Next, we split an officer's citations in half temporally and perform the same exercise, which gives $\hat{\beta} = 0.82$ ($se = 0.03$).

3.3 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 effect of sanctions on offending, we find that a 125 dollar increase in fine (causally) increases

[8] The ψ's are the same fixed effects we use in our main analysis, described in section 4.1. They are at the level of county × $\mathbf{1}$[highway] × year × month × $\mathbf{1}$[weekend] × shift.

the likelihood that a driver contests a ticket in court by about 11 percentage points, or 35 percent relative to the mean (see figure A-4). 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.

Another possibility is that officers perceive a psychic or emotional cost to writing harsh charges, or equivalently experience utility from giving motorists a break. Variation across officers in the emotional costs of harshness or psychic benefits of lenience could then generate heterogeneity in bunching propensity. Importantly, our empirical design does not require explicit knowledge of why officers bunch drivers or why they vary in bunching propensity. However, as we discuss below, our design will require a monotonicity assumption that restricts the patterns across officers in which drivers are bunched.

4 Estimating causal effects of sanctions

4.1 Empirical strategy

Our first goal is to estimate the causal effect of sanctions on the future behavior of speeding offenders. The central identification challenge is that punishments are not randomly assigned. Statutory sanctions increase with the severity of the transgression, as shown in figure 1. Moreover, as discussed in section 3.2, officers frequently manipulate sanctions. Naive OLS estimates, shown in table A-2, 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 $\hat{\beta} = 0.019$ (se = 0.001), suggesting that harsher fines increase reoffending. Adding officer fixed effects *increases* the estimate to $\hat{\beta} = 0.031$ (se = 0.001), highlighting the nonrandom sorting of drivers into sanctions by officers.

Our empirical approach is to leverage systematic variation in lenience *across* officers for identification, mirroring a growing body of research using so-called examiner designs (e.g., Kling 2006, Dobbie & Song 2015, Maestas et al. 2013, Bhuller et al. 2020). Specifically, we compute the following stringency instrument:

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

where i indexes citations and j indexes officers. In words, Z_{ij} is the fraction of officer j's citations to all *other* drivers that are for speeds greater than 10 MPH over the posted limit; in other words, the fraction of unbunched citations. We then use Z_{ij} as an instrument to estimate regressions of the form:

$$Y_{ijs} = \beta harsh_{ijs} + \psi_s + \epsilon_{ijs}$$

where $harsh_{ijs} = \mathbf{1}[\text{speed}_{ijs} \geq 10]$. We also show results using a continuous sanctions measure as the explanatory variable of interest but focus on the binary specification because our modeling exercise will focus on the binary specification.

In the above regression, the ψ_s's are fixed effects at the level of county $\times$ $\mathbf{1}$[highway] $\times$ year $\times$ month $\times$ $\mathbf{1}$[weekend] $\times$ shift, which we term beat-shift fixed effects. 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. These fixed effects adjust for differential exposure of officers to pools of offenders across beat-shifts. In our regressions, we two-way cluster our standard errors at the officer and driver level.

4.2 Instrument validity

Our IV strategy will yield valid local average treatment effect (LATE) estimates under the following four assumptions:

1. *Relevance.* $H(Z)$ is a nontrivial function of Z.
2. *Exogeneity.* $\{Y_i(1), Y_i(0), H_i(Z)\} \perp Z \mid \psi$
3. *Exclusion.* $Y_i(H, Z) = Y_i(H)$
4. *Monotonicity.* $\forall w, j \in J$, either $H_i(w) \geq H_i(h)\ \forall i$ or $H_i(w) \leq H_i(h)\ \forall i$

where J indexes the set of officers, $H = \mathbf{1}[harsh]$, and $\{Y_i(1), Y_i(0)\}$ are the potential outcomes of driver i when sanctioned harshly ($H = 1$) and leniently ($H = 0$). The relevance assumption requires the existence of a first stage relationship between stringency and harsh fines, which is empirically testable. We discuss the exogeneity, exclusion, and monotonicity assumptions in turn below.

4.2.1 Exogeneity

Existing studies using similar empirical designs have appealed to the institutional quasi-random assignment of examiners (e.g., bail judges) to satisfy the exogeneity assumption. A central concern in our setting is that citing officers 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 of 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.

One potential threat to our empirical strategy 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 $Y_i(0)$, then $E[Y_i(0) \mid j = 1] > E[Y_i(0) \mid j = 2]$, 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. The first two panels plot the relationship between officer stringency, adjusted for beat-shift fixed effects, and ticketing frequency, using both a daily average number of tickets and a monthly average number of tickets, adjusted for beat-shift effects, as measures of ticketing activity. Panel (a) plots the relationship for all citations and panel (b) plots the relationship for only speeding citations. In all cases, regression coefficients are quantitatively small and statistically indistinguishable from zero. Panel (c) illustrates that there is no relationship between the stringency instrument and predicted recidivism 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 ($F = 29$). Motorist characteristics have considerably less ability to predict officer stringency ($F = 2.7$). While our test rejects the null of no relationship between observables and officer stringency, a joint significance test of $F = 2.7$ is quite small in a setting with no institutional random assignment and with $N \approx 1.7$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. However, we take sample selection concerns seriously and present an array of associated robustness checks. First, we show that estimates are nearly identical when using a binary instrument (stringent versus lenient) that easily passes a conventional randomization test, as shown in column 4 of table 2. Second, we show that results hold when dropping officers with relatively selected samples. Third, we implement a formal selection correction estimator *ala* Heckman (1979). Finally, we show similar results controlling for finer geographic detail and results relying only on within-driver variation, discussed in more detail in section 4.4.

4.2.2 Monotonicity and exclusion

In table A-3, we perform a standard check of the monotonicity assumption by estimating the first stage regression for subgroups of drivers. In the full sample, the first stage estimate is $\beta = 0.944$ ($se = 0.006$). Looking across subgroups, the first stage estimates range from $\beta = 0.913$ to $\beta = 0.972$, with standard errors ranging from 0.006 to 0.016.

Still, monotonicity violations are a natural concern in our setting given evidence of racial bias in officer leniency decisions (Anbarci & Lee 2014; Goncalves & Mello 2021). We have two approaches to address monotonicity concerns. First, we show results using a binary instrument which compares drivers ticketed by lenient (bunching) versus stringent (non-bunching) officers. This binary instrument satisfies monotonicity by construction because stringent officers never bunch drivers, which rules out the existence of defiers. Second, we recompute the continuous stringency instrument within demographic cells, which allows for monotonicity violations across, but not within, demographic groups. Our results, discussed further in section 4.4, are qualitatively similar using both approaches.

Importantly, Frandsen et al. (2019) show that examiner design IV estimates 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. However, our baseline estimates of marginal treatment effects will still rely on a strict monotonicity assumption. We defer a more thorough discussion of our approaches to estimating marginal treatment effects, and associated monotonicity issues, to section 5.1.

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. On the other hand, we feel that this 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.[9]

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 three DL points, while speeding 16-29 MPH over the limit carries four DL points. Points can increase car insurance premiums and drivers that accrue sufficient points can face DL suspensions. As shown in figure A-3, 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

[9]Moreover, the evidence is largely inconsistent with the court system playing an important role in generating the treatment effects. As shown in figure A-7, 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.

between stringency and points, again shown in figure A-3. Hence, the burden of accrued license points cannot explain the effects we observe.

To the extent that the monotonicity and exclusion assumptions are violated, we can still interpret our reduced form estimates as causal effects of officer stringency under the exogeneity assumption. Given that our estimated first stage is very close to one ($\beta_{FS} = 0.944$), our reduced form and IV estimates are quantitatively similar.

4.2.3 Understanding the LATE

Given the first stage, exogeneity, exclusion, and monotonicity assumptions, our IV estimates will recover a local average treatment effect (LATE) for the subgroup of marginal drivers. Intuitively, compliers are drivers about whom officers disagree, or drivers who would be punished harshly by some officers but given a break by others. Alternatively, in the continuous instrument setting, it is useful to think of compliers as those motorists who are neither always-takers, or those who would be fined harshly by any officer, nor never-takers, or those who would be bunched by any officer.

An interesting feature of our setting is the full support of our stringency instrument. In other words, our sample includes a subset of officers that *always* bunch drivers and a subset of officers that *never* bunch drivers. Given the LATE assumptions, the presence of officers that always issue harsh tickets implies that no driver is a never-taker. Similarly, the presence of officers that never issue harsh fines implies the nonexistence of always-takers.

Because all drivers are compliers for some value of the instrument, differences between naive OLS estimates and IV estimates should be interpreted as largely attributable to selection bias, rather than as driven by characteristic differences in the complier population. Moreover, our estimated LATE will be quite close to the average treatment effect.

Worth 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.

4.3 Results

Figure 3 plots the first stage relationship between officer stringency and $\mathbf{1}[harsh]$, laid over a histogram of the stringency instrument. The figure documents a linear and statistically precise relationship with an estimated first stage coefficient $\beta = 0.944$ ($se = 0.006$) and associated $F \approx 22,000$. Figure A-3 illustrates the estimated first stage for direct sanctions measures. In terms of fine amounts, shown in panel (a), the estimated first stage is $\beta = \$122$.

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

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

where $Y_{ijs\tau}$ is an indicator for whether driver i is cited for speeding in quarter τ, which are quarters *relative* 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 to predict offending over the previous eight quarters but predicts a stark decline offending immediately after the focal citation. Stringency is associated with a 0.7 percentage point decline in the likelihood a new speeding offense in the next quarter, which represents a 12 percent decline relative to the lenient officer mean. Effects persist over the first four quarters and fade out considerably thereafter.

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 to their offending rates.

Table 3 presents IV estimates for the full set of one-year offending outcomes.[10] 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 and 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 ($\sim$5 percent, $\epsilon = -0.07$). 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 almost 1.5 percentage points. The IV estimate is precisely estimated, with a 95 confidence interval of $(-0.017, -0.012)$. The point estimate 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 speeding recidivism by 13 percent.

Estimated impacts of harsh fines on non-speeding offenses are also statistically significant but less pronounced. We estimate fine elasticities of -0.06 for non-speeding offenses, -0.07 for moving violations, and -0.097 for non-moving violations. The finding that speeding sanctions reduce other traffic offenses is consistent with Gehrsitz (2017), who finds that short-term license suspensions imposed on speeders in Germany reduce all forms of traffic offending through a specific deterrence mechanism.

[10] In the appendix, we present graphical versions of the reduced form estimates (figure A-4), dynamic versions of the reduced form for other outcomes (figure A-5), and the full set of first stage and reduced form estimates with and without controls (table A-4).

Consistent with reductions in traffic offending implying a true behavioral response on the part of drivers, we also find that a harsh fine 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 involvment are statistically significant at the 10 percent level.

Finally, following our discussion in section 3.3, 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 hypothesis that court aversion motivates officer lenience.

4.4 Robustness

In the appendix, we present a battery of robustness checks. For simplicity, we focus our robustness checks on IV estimates of the impact of a harsh fine on the likelihood of a new speeding offense in the next year. Tables A-5 and A-6 show that our results are not sensitive to our choice of instrument or fixed effects. Estimated impacts on one-year speeding recidivism are, if anything, larger when using a residualized leave-out mean (e.g., Dobbie et al. 2018), the leave-county-out mean, the full set of officer dummies, and a binary measure as the stringency instrument. Table A-6 shows that estimates are comparable when using different specifications for the beat-shift fixed effects, including limiting comparisons to drivers stopped on the exact same day or exact same stretch of road.

As discussed in section 4.2, of particular interest are robustness checks for whether our estimates can be explained by selection of drivers that is correlated with our stringency measure. While the dynamic patterns in figure 4 assuage concerns in this front, we present four additional pieces of evidence in figure A-6.

Following Feigenberg & Miller (2021), we show that the estimated effect on one-year speeding recidivism is similar when dropping officers with selected samples based on driver observables. Specifically, we first regress a measure of predicted reoffending based on driver characteristics (same as plotted in figure 2) on beat-shift fixed effects and take the residuals. We then compute the average residual for each officer, which captures selection in an officer's sample relative to other officers patrolling the same beat-shifts. We then re-estimate our main IV estimate, dropping officers in the top and bottom q percent of the distribution of average residuals. As shown in panel (a), our estimate persists after dropping even the top and bottom 25 percent of "selected" officers.

In panel (b), we show selection-corrected results based on officer ticketing frequency. Suppose that all officers working in the same county-year face the same number of drivers on a given day, N. Officer j choose whether to ticket passing driver i, $S_{ij} \in \{0,1\}$ and officer j's daily rate of tickets in that county-year is $N_j = N \times Pr(S_{ij} = 1)$. We estimate

N by taking the 95th percentile of the N_j's and estimate $Pr(S_{ij} = 1)$ with N_j/N. Given an estimated $Pr(S_{ij} = 1)$, we can correct our estimates for selection by directly controlling for the inverse Mills ratio, $\frac{\phi(\Phi^{-1}(P(S_{ij}=1)))}{P(S_{ij}=1)}$, in our regressions (Heckman, 1979). This procedure yields a nearly identical IV estimate to our main specification.

Panel (c) of figure A-6 plots the reduced form and reports IV estimates for regressions that further interact our beat-shift fixed effects with stretch-of-road fixed effects, constructed by mapping the subset of geocoded tickets onto a map of all roads in Florida. Conditioning on a finer measure of geography allows us to more accurately adjust for the flow of offenders faced by each officer. This regression is based on a considerably smaller sample ($N = 219,470$) but yields a strikingly similar effect to our main IV estimate (-0.0140 versus -0.0146).

Finally, in panel (d), we show results using only within-driver variation. In principle, adding driver fixed effects to our main regression could be problematic given the causal effect of stringency on the likelihood that a driver reappears in the data. Drivers with multiple stops are either (i) disproportionately likely to have drawn a lenient officer during their first stop or (ii) highly selected as reoffenders despite receiving harsh sanctions. To circumvent these selection issues, we leverage the treatment effect fadeout observed in figure 4. First, we identify drivers with multiple citations in our main sample and then restrict to the within-driver pair of stops that are the farthest apart in time. We drop drivers whose remaining pair of stops are within one year of each other, leaving $142,910$ drivers (285,820 citations).[11] The idea of this approach is that, as we lengthen the time between stops, differential selection due to treatment effects tends to zero. We choose the one-year minimum as a compromise between preserving sample size and the treatment effect mostly, but not entirely, having faded out after one year. We then estimate our main IV regression using this sample of driver-stop pairs and including driver fixed effects. Panel (d) of figure A-6 illustrates the underlying reduced form and reports the IV estimate, $\beta_{IV} = -0.012$ (0.003), which is only slightly smaller than our baseline estimate (-0.0146).

To address concerns about monotonicity violations, table A-5 reports results using both a binary instrument and instruments recomputed within demographic cells. Recall that the binary instrument, which compares drivers stopped by bunching and non-bunching officers, satisfies monotonicity by construction because non-bunching officers never bunch drivers. Another advantage of the binary instrument is that it satisfies a conventional randomization test, as shown in table 2. Using the binary instrument gives $\beta_{IV} = -0.0213$ (0.003), which is about 50 percent larger than our baseline estimate. The bottom panel of table A-5 shows results when recomputing our instrument as the leave-out harsh fine rate within officer $\times$ demographic cells. IV estimates are comparable when computing the instrument within racial groups (-0.0124), and slightly smaller when computing by gender (-0.011),

[11] This restriction applies only to stops that enter our main sample of focal FHP citations; it does not restrict reoffending patterns generally. Drivers can still have reoffended after their first citation *outside of our main sample* at any time.

and smaller but still statistically significant when computing the instrument within race $\times$ gender interacted with age and income groups, (-0.008), with the caveat that our sample size shrinks significantly in this regression.

4.5 Mechanisms

One potential mechanism underlying our estimated effects of sanctions on future driving behavior is that harsher fines have deleterious effects on financial situations, causing changes in life circumstances or causing individuals to stop driving because they can no longer afford the associated costs (e.g., car payments, insurance, gas). However, as shown in figure A-7, offender responses to harsher fines are nearly identical for higher and lower income individuals, proxied with per capita income in the zip code of residence. Hence, a pure financial mechanism is inconsistent with 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 the dynamic patterns in figure 4 is driver learning or updating (Libor & Traxler, 2021). After facing a harsh fine, drivers update their beliefs about the expected costs of speeding and slow down accordingly, with the update fading out over time. Another prediction of such a model is that drivers with past exposure to harsh fines should respond less to subsequent harsh fines.

In practice, testing this prediction carries a few challenges. Differences between drivers with different citation or offending histories is unlikely to be informative as these drivers may differ on many dimensions. Instead, we want to compare drivers with varying (quasi-random) exposure to officer stringency in the past. However, the treatment effects we observe induce selection in these comparisons, because drivers cited by lenient officers are more likely to reappear in the data. To mitigate selection issues, we focus on exposure to stringency at least one year in the past, because treatment effects have largely faded as of one year following a harsh citation (see figure 4). Specifically, we take the subset of drivers with an FHP-issued citation at least one year prior ($N = 216,458$). We then compare treatment effects for drivers with and without past exposure to a stringent officer.

Results are shown in figure 5, which plots the reduced form and reports the IV estimate for each group. The estimated effect of a harsh fine on one-year speeding recidivism is about 50 percent larger for drivers without a prior FHP ticket but no past exposure to non-bunchers ($\beta_{IV} = -0.015$) than for those previously exposed to stringent officers ($\beta_{IV} = -0.01$). The effect for the unexposed group is more precisely estimated ($se = 0.004$ v. $se = 0.007$), and the effect for the exposed group is not statistically distinguishable from zero. While the difference in the IV estimates across groups is not statistically significant, we take this as evidence in favor of an updating hypothesis, similar to Libor & Traxler (2021).[12]

[12] Similarly, figure A-8 shows that harsh fines are associated with an 18 percent decline in recidivism in the same county and a 4 percent decline in recidivism in other counties, suggesting that

5 Deterrence and retribution

Having established that sanction decisions have important deterrence effects, we now turn to an analysis of how officers allocate sanctions. We study how officers sort a set of offenders into harsh versus lenient sanctions, with a particular interest in how the law enforcement principles of deterrence and retribution shape these decisions.

We begin by presenting and estimating a model of marginal treatment responses (MTR) and marginal treatment effects (MTE) to explore the potential reoffending outcomes of motorists with varying propensities to be fined harshly. Motorists with the highest propensities to be fined harshly, or those induced into treatment at the lowest values of our officer instrument, are those motorists most prioritized by officers for harsh sanctions. Hence, by examining marginal treatment responses, we can learn about which drivers are prioritized for harsh sanctions and, thereby, infer something about officer objectives. The logic of optimization tests based on marginal treatment responses originates with Roy (1951) and has been applied in a variety of settings to examine heterogeneous treatment effects (see Cornelissen et al. 2016 and Mogstad & Torgovitsky 2018 for a discussion of recent applications).

If officers sort drivers to maximize deterrence, they should prioritize issuing harsh sanctions to the most responsive offenders. This sorting behavior would generate selection on gains: we would expect to see the largest (most negative) treatment effects for the motorists most likely to be sanctioned harshly. In the same potential outcomes framework, we assert that officers maximizing retribution will prioritize punishing harshly the most frequent offenders, or those most likely to reoffend, regardless of their responsiveness to sanctions. Sorting to accomplish a retribution objective, then, generates selection on levels: motorists most likely to be sanctioned harshly will be those with the highest reoffending rates.[13]

5.1 Estimating marginal treatment responses

Our aim is to identify counterfactual reoffending outcomes for individuals at the margin of being treated (receiving a harsh fine) at each value of the stringency instrument. Each individual has a pair of potential reoffending outcomes Y_D that are a function of treatment status, $D \in \{0,1\}$, denoting whether they received the harsh punishment. The realized outcome can be written as 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

motorists update their area-specific beliefs about expected speeding costs.

[13]Note that the relevant law and economics literature does not map the notion of retribution to a potential outcomes framework. In some treatments (e.g., Kaplow & Shavell 2006), retribution captures any deviations from preferences for maximizing safety. We choose to model retributive goals as a preference for punishing frequent offenders, which captures the notion of culpability emphasized in the literature (e.g., Moore 2019). As note earlier, our finding that officer behavior cannot be explained by deterrence goals alone is evidence in itself that officers incorporate some notion of retribution into their punishment decisions.

is a counterfactual-specific vector of coefficients on characteristics X, and U_j is a random variable with $E(U_j|X) = 0$.

Treatment status follows a threshold crossing model, which is a function of characteristics Z (which include X and our excluded instrument) and unobservable U_D such that $D = \mathbb{I}(\mu_D(Z) > U_D)$. 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 by $P(Z)$. Each individual has a fixed value of U_D, which we call their "resistance to treatment." The higher this value, the greater the realized value $P(Z)$ must be for that individual to take up treatment.

Our goal is to identify the expected value of the counterfactual outcomes of 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 include in X our baseline set of beat-shift fixed effects.

We are interested in understanding how officers sort drivers into punishment, taking into account all information available at the time of the traffic stop. Hence, we also do not condition on driver covariates when estimating the marginal treatment effects. In figure A-12, we show that our findings are similar when including driver covariates in X, which results in MTRs and an MTE that reflect sorting within demographic groups rather than across all individuals.

We describe in detail how we estimate the MTR and MTE functions in appendix C-1, and we describe the approach here briefly. We estimate the MTRs function for Y_0 and Y_1 by taking the "separate" approach of Heckman & Vytlacil (2007b) and Brinch et al. (2017). We first estimate the propensity score p_i for each individual by regressing D_i on beat-shift fixed effects and the officer instrument in a linear probability model and then constructing a predicted treatment $\hat{D}_i \equiv p_i$. We then regress Y_i on beat-shift fixed effects and the estimated propensity score, and we *restrict the sample to either untreated or treated drivers,* $D_i = 0, 1$, corresponding to the Y_0 and Y_1 MTRs, respectively. The coefficients on the fixed effects give us the level of the MTR, and the coefficient on the propensity score gives us its linear slope term. We calculate the MTE as the difference between the two MTRs. We present the MTR and MTE functions for the average values of the fixed effects over all drivers in the sample.

5.1.1 MTR results

Panel (a) of figure 6 presents the estimated marginal treatment response functions, $E(Y|U_D)$. Motorists with lowest U_D, or those most likely to be fined harshly, have the highest reoffending rates in either treatment state. In other words, the sorting of drivers into sanctions by officers exhibits selection on levels, with officers most likely to issue harsh tickets to drivers with the highest overall offending rates. Panel (b) of figure 6 plots the marginal treatment effect, which is the difference between the untreated and treated counterfactual outcomes in panel (a). The MTE becomes more negative as resistance to treatment increases, indicating that motorists *least likely* to receive harsh fines actually are the *most deterred* by harsh fines. Deterrence impacts for the motorists most likely to be sanctioned harshly ($U_D = 0$) are marginally positive and statistically indistinguishable from zero.

The pattern of inverse selection on gains implied by the downward sloping marginal treatment effect curve rejects that officers sort drivers to maximize deterrence. The simultaneous presence of selection on levels, moreover, reveals that officers face an explicit tradeoff between maximizing deterrence and retribution objectives, because the most deterrable motorists are those with lower overall reoffending levels.[14]

5.1.2 MTR robustness

There are two important caveats associated with our marginal treatment response estimates. The first is that our baseline estimate imposes a possibly restrictive linear parametric structure on potential outcomes. In figure A-10, we present nonparametric estimates of marginal treatment effects, which exhibit a similar downward slope to our baseline parametric estimate in panel (b) of figure 6.

The second caveat 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). In our setting, this is a very strong assumption that warrants caution. To probe the sensitivity of our estimates to the monotonicty assumption, we exploit a unique feature of our setting, the full support of our stringency instrument, to estimate a version of the MTE that satisfies monotonicity by construction, which we describe in detail in appendix C-2.

The idea of our approach is to estimate the impact on offending of being counterfactually reassigned from an officer with $Z = 0$ to an officer with $Z = \epsilon$, where ϵ is small and positive. This yields an estimate of the marginal treatment effect at $Z \approx 0$ which satisfies monotonicity because $P(D|Z = 0) = 0$. We can 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. Figure

[14] Figure A-11 also rules out a competing hypothesis that officers sort motorists with goal of minimizing time in traffic court. Drivers with the lowest resistance to treatment have the largest (positive) treatment effects of harsh fines on court contesting.

A-10 plots the estimated MTE's from this "tails" approach. The slope of our baseline linear MTE is -0.066 and the implied MTE slope from the "tails" approach is -0.109. Hence, our conclusion that the MTE is downward-sloping, inconsistent with deterrence maximization by officers, is unaltered by this monotonicity correction.

5.1.3 Characteristics of marginal compliers

To further explore how officers sort drivers into sanctions, we estimate the characteristics of marginal compliers, or motorists on the margin of receiving a harsh fine 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 C-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 7 illustrates a stark, negative relationship between offending history and marginal compliance. Compliers at the lowest levels of resistance, or drivers most prioritized for harsh sanctions, are significantly more likely to have been cited in the past year than compliers at the highest levels of resistance of treatment. Panel (b) illustrates a similar relationship for offense severity, or the driving speed relatively to the posted limit. Recall that offense severity is unobserved for the subset of bunched drivers, but the average speed of marginal compliers can be estimated based on the speeds of the unbunched drivers. The priority that officers place on offending history and severity when issuing harsh fines further illustrates the notion of retributive preferences particularly well.[15]

Importantly, officer sorting of drivers based on offending history cannot solely explain the selection on levels that we observe in the estimated MTR's. In figure A-14, we show that the marginal treatment response functions exhibit strikingly similar patterns when examining only the subsample of drivers without a citation in the past year. In other words, even among motorists with no recent offending history, officers are able to prioritize likely recidivists for harsh punishments. Moreover, the similarity of the MTR's for this sample rules out a competing explanation for the shape of the MTR, which is the hypothesis that the most-punished individuals have smaller learning responses to harsh fines because of their prior experience with speeding sanctions.

5.2 Officer decision model

While the downward-sloping marginal treatment effect rules out that officers sort drivers on deterrability alone, officers may still consider a combination of deterrence and fairness benefits when sorting drivers into sanctions. In this section, we present a simple model of how officers decide whom to sanction harshly. We allow officers to value both deterring crime

[15]We also compute demographic characteristics of marginal compliers, presented in figure A-13, but defer a discussion of these results to section 5.4.

and punishing individuals with high offending rates, which they balance against a private cost of issuing harsh punishments. We then compare the implications of this model with our MTR's and ask which officer preferences are consistent with our empirical estimates.

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 $F(\hat{Y}_{i0}, \hat{Y}_{i1})$ and reflect the expected value of each outcome: $E(Y_{ik}|\hat{Y}_{ik}) = \hat{Y}_{ik}$, $k = 0, 1$.[16]

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

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

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 retribution, and $\lambda = 0$ indicates that they care only about deterrence. This weight is our main object of interest.

The utility of issuing the non-harsh punishment 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, $\theta = g(c)$, $c = g^{-1}(\theta)$.

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 that corresponds to these estimates are the values of treated and untreated offending rates for

[16]This 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 signals and officer inference.

individuals who are at the margin of punishment for officers at a given propensity to treat:

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

Because the MTR functions are estimated with linear specifications, 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 on the distribution of signals by parameterizing 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 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. The details of the model estimation are presented in appendix C-4.

We estimate that λ lies in the interval $[0.61, 1]$. For the lowest value, officers place a higher weight on retribution than deterrence 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.

5.3 Efficiency costs of retribution

We now consider the impact of changing how officers sort drivers into sanctions and ask how the reoffending levels and responses of treated and untreated motorists change. We conduct two related counterfactuals. First, we consider a case where officers sort drivers in reverse order, so that a driver's new (counterfactual) resistance to treatment is $\tilde{u} = 1 - u$. or 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. Second, we conduct a counterfactual using our estimated officer decision model, setting $\lambda = 0$. We do so for every set of parameter values in our partially identified set, meaning that our estimates will yield a range of counterfactual outcomes. In both calculations, we assume the overall probability of treatment (receiving a harsh fine) is the same as the observed probability.

These exercises speaks directly to the *efficiency costs* associated with the retributive preferences of officers by quantifying the additional number of traffic offenses that occur relative to counterfactual scenarios where drivers are sorted into sanctions in a way that generates selection on gains. Note that these counterfactuals differ from the typical calculations made with 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).

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 ($0 - 0.027$ 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. However, the treated individuals are now safer drivers than the untreated, with an offending rate of 0.342 versus 0.392.

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.

The improved targeting of harsh sanctions towards more deterrable drivers translates into fewer traffic offenses, with the overall offending rate declining from 0.358 to a range between 0.332 and 0.337. Hence, the weight that officers place on retribution goals 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 as a result of not prioritizing deterrence alone.

5.4 Racial implications of retributive preferences

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), an interesting question is the role that retribution versus deterrence goals 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 A-13, which presents the demographic characteristics of marginal compliers, illustrates that younger, male, Black, and Hispanic drivers are prioritized for harsh sanctions. These gradients are significant; at the lowest levels of resistance, Black (Hispanic) drivers make up 18 (21) percent of compliers, whereas at the highest levels of resistance, Black (Hispanic) drivers make up 13 (17.5) percent of compliers. Further, as shown in figure A-15, Black and Hispanic drivers exhibit higher rates of reoffending but nearly identical responsiveness to harsh fines, suggesting that some of the racial gap in treatment may be explained by officer preferences for retribution.

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 5. 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).[17] 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 relates to two recent papers that consider the equity and efficiency implications of criminal justice policy. Rose (2021) documents the racially disparate impact of probation rules that trigger a prison sentence. Further, he 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 (2021), 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 shows the possibility for changes in police practices that reduce racial gaps in treatment while simultaneously improving efficiency.

6 Conclusion

In this paper, we study how the law enforcement principles of deterrence and retribution shape the behavior of police officers wielding discretion over sanctions for speeding offenses. First, relying on variation across officers in the propensity to issue harsh fines, we show that sanction decisions have important deterrence effects. Comparing motorists cited at the same time in the same area 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.

We then use a marginal treatment effects approach to examine how reoffending levels and responses vary with the propensity to be sanctioned harshly. Based on an underlying Roy (1951) framework, this exercise yields insights about how officers sort offenders into harsh

[17] These numbers are calculated by taking the values for $E(\text{Black}|X, U_D)$, where U_D is resistance to treatment, described in section C-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.

versus lenient sanctions. Motorists most likely to be sanctioned harshly are those with the highest probability of reoffending, suggesting that officers prioritize frequent offenders when issuing harsh punishments. In other words, officers appear interested in targeting harsh fines to the "worst" offenders, which we interpret as consistent with retributive preferences.

On the other hand, we observe *inverse* selection on gains, with offenders least likely to be sanctioned harshly exhibiting the largest treatment effects. This pattern is inconsistent with deterrence maximization by officers; officers allocating sanctions to maximizing safety should prioritize punishing the most deterrable offenders. Moreover, the simultaneous presence of selection on levels and inverse selection on gains implies that officers face an explicit tradeoff between deterrence and retributive goals, because the subpopulations with the highest reoffending rates and the highest reoffending responses are different.

Given that officers face such a tradeoff, we write down and estimate a model of sanction choices. In the model, an officer observes noisy signals of a motorist's potential outcomes and chooses whom to sanction harshly based on a weighted average of expected reoffending rates and responses. We interpret the weight that officers place on reoffending rates as the weight they place on achieving retribution, as opposed to deterrence, goals. Matching the model parameters with the moments of our estimated marginal treatment response functions, we find that officers place at least as much weight on retribution as deterrence goals when issuing harsh sanctions. Further, we cannot rule out that officers care solely about retribution.

We use our model estimates to quantify the efficiency costs associated with officers' retributive preferences by asking how the reoffending rate changes in a counterfactual scenario where officers consider only deterrence goals when allocating harsh fines. We find that current officer practices forgo about a six percent decline in the reoffending rate that could be achieved by targeting harsh fines towards more deterrable drivers. Finally, we provide suggestive evidence that changes in officer preferences could increase efficiency while also reducing racial disparities in treatment.

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 wih a natural experiment. *Unpublished manuscript*.

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.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, *76*(2), 169–217.

Bhuller, M., Dahl, G., Loken, K., & Mogstad, M. (2020). Incarceration, recidivism, and employment. *Journal of Political Economy*, *128*(4), 1269–1324.

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.

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*.

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

Chalfin, A. & McCrary, J. (2018). Are us cities underpoliced? theory and evidence. *Review of Economics and Statistics*, *100*(1), 167–186.

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.

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.

Doleac, J. (2021). Encouraging desistance from crime. *Journal of Economic Literature, 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. (2021). Would eliminating racial disparities in motor vehicle searches have efficiency costs? *Quarterly Journal of Economics, forthcoming*.

Finlay, K., Gross, M., Luh, E., & Mueller-Smith, M. (2021). The impact of financial sanctions in the U.S. justice system: Regression discontinuity evidence from Michigan's driver responsibility program. *Unpublished manuscript*.

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. (2007a). 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.

Heckman, J. J. & Vytlacil, E. J. (2007b). 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. *Handbook of econometrics*, *6*, 4875–5143.

Huttunen, K., Kaila, M., & Nix, E. (2020). The punishment ladder: Estimating the impact of different punishments on defendant outcomes. *Unpublished manuscript*.

Imbens, G. W. & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.

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.

Knowles, J., Persico, N., & Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, *109*(1), 203–229.

Libor, D. & Traxler, C. (2021). Learning from Law Enforcement. *Journal of the European Economic Association, forthcoming*.

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 disourage 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. (2019). More cops, less crime. *Journal of Public Economics*, *172*, 174–200.

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.

Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. *Unpublished manuscript*.

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*.

O'Flaherty, B. & Sethi, R. (2019). *Shadows of doubt: Stereotypes, crime, and the pursuit of justice*. Harvard University Press.

Rose, E. (2021). Who gets a second chance? Effectiveness and equity in supervision of criminal offenders. *Quarterly Journal of Economics*, *136*(2), 1199–1253.

Rose, E. & Shem-Tov, Y. (2021). How does incarceration affect reoffending? Estimating the dose-response function. *Journal of Political Economy*, *129*(12), 3302–3356.

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*.

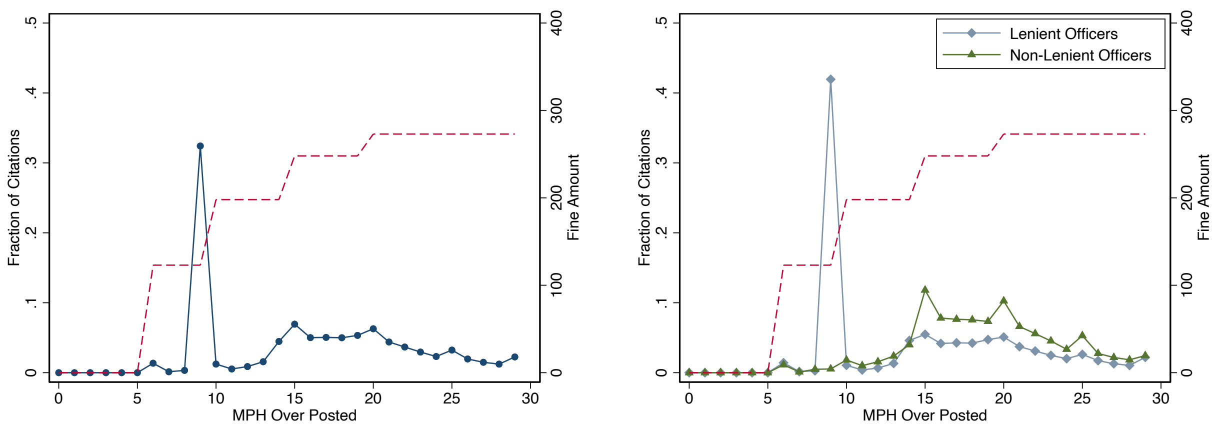
Table 1: Summary Statistics

	(1)	By Fines (2)	(3)
	All	Discounted	Harsh
Panel A: Demographics			
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
Panel B: Socioeconomic Status			
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
Panel C: Offending History			
Prior Prison Spell	0.0001	0.0001	0.0001
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
Panel D: Offense Characteristics			
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	1,693,436	579,755	1,113,681

*N*otes: This table reports sample means fo the main sample. See table A-1 for officer characteristics.

Figure 1: Distribution of Charged Speeds

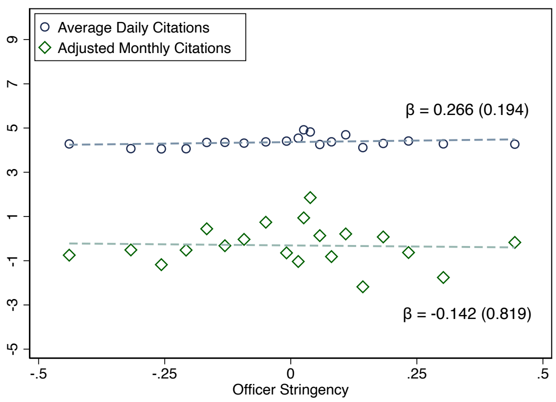
(a) Full Sample　　　　(b) By Officer Type



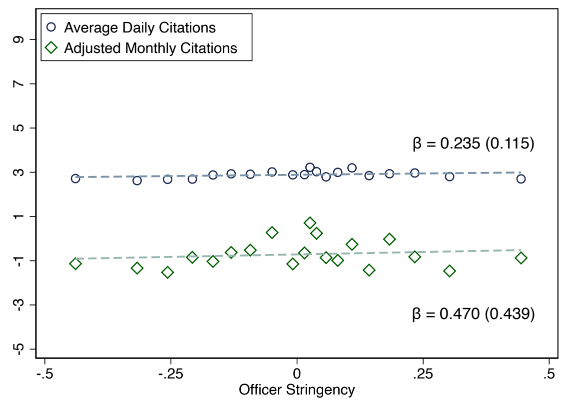
*N*otes: This figure plots the distribution of charged speeds on FHP-issued speeding citations in Florida. Dashed red line shows the fine schedule (right axis). Panel (a) shows the aggregate distribution, while panel (b) disaggregates the distribution by whether an officer is classified as lenient or non-lenient, using the method described in section 3.2.

Figure 2: Instrument Validity

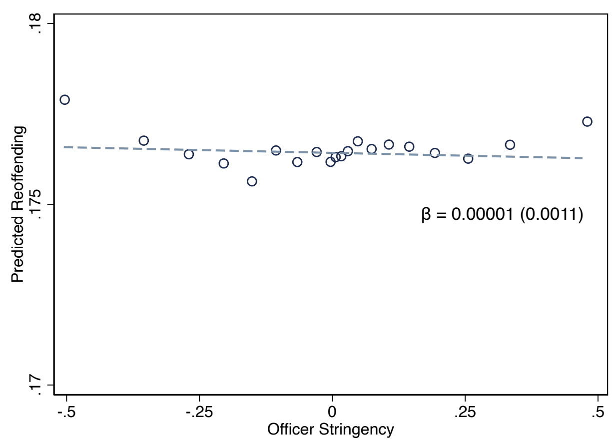
(a) Ticketing Frequency: All Offenses



(b) Ticketing Frequency: Speeding



(c) Predicted Reoffending



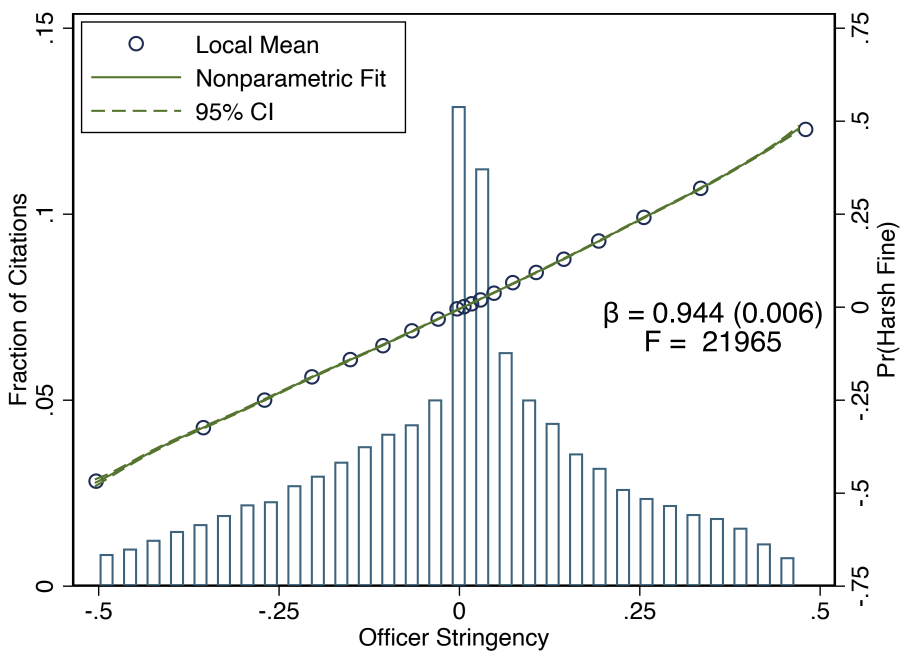
*N*otes: *Average daily citations* is an officer's mean number of citations issued per day. *Adjusted monthly citations* is an officer's mean number of citations per month, adjusted for beat-shift fixed effects. Predicted reoffending is the predicted probability of a new speeding offense in the next year obtained from a regression of reoffending on the full set of driver characteristics, using a holdout sample of only lenient (bunching) officers.

Table 2: Randomization Test

	(1) Reoffend	(2) Harsh Fine	(3) Stringency	(4) **1**[Stringent]
Female	-0.0328	-0.0235	-0.00115	0.00160
	(0.000692)	(0.00173)	(0.000930)	(0.00130)
Age	-0.00338	-0.00227	0.000273	-0.000539
	(0.000120)	(0.000272)	(0.000171)	(0.000256)
Age Squared	0.0000113	0.0000148	-0.00000296	0.00000586
	(0.00000132)	(0.00000267)	(0.00000165)	(0.00000258)
Race = Black	0.0244	0.0203	-0.000373	0.00408
	(0.00105)	(0.00270)	(0.00176)	(0.00242)
Race = Hispanic	0.0104	0.0344	0.00635	0.000649
	(0.00103)	(0.00290)	(0.00212)	(0.00295)
Race = Other	0.00696	0.0347	0.00583	0.00101
	(0.00156)	(0.00275)	(0.00176)	(0.00250)
Race = Unknown	0.00867	0.00506	0.00377	0.00130
	(0.00228)	(0.00550)	(0.00266)	(0.00374)
Prior Prison Spell	0.0700	-0.0135	0.0155	0.00844
	(0.0354)	(0.0267)	(0.0148)	(0.0230)
County Resident	0.0107	-0.0186	-0.00587	-0.00310
	(0.00107)	(0.00309)	(0.00265)	(0.00441)
Log Zip Income	0.00757	0.0102	0.00488	-0.00180
	(0.000782)	(0.00211)	(0.00164)	(0.00194)
Log Vehicle Price	0.0174	0.0198	0.00489	-0.00160
	(0.000716)	(0.00160)	(0.00123)	(0.00192)
Speeding Past Year	0.105	0.0247	0.000940	0.000603
	(0.00108)	(0.00158)	(0.000654)	(0.00110)
Other Past Year	0.0676	0.0150	-0.000780	0.00201
	(0.000857)	(0.00115)	(0.000728)	(0.00118)
Crash Past Year	0.0210	0.00737	0.000977	0.000723
	(0.00130)	(0.00132)	(0.000769)	(0.00113)
Mean	.172	.658	.658	.763
F-Stat	1736.41	29.27	2.67	.88
F-test	<.0001	<.0001	.0002	.5978
Beat-Shift FE	Yes	Yes	Yes	Yes
Officers	1960	1960	1960	1960
Observations	1693436	1693436	1693436	1693436

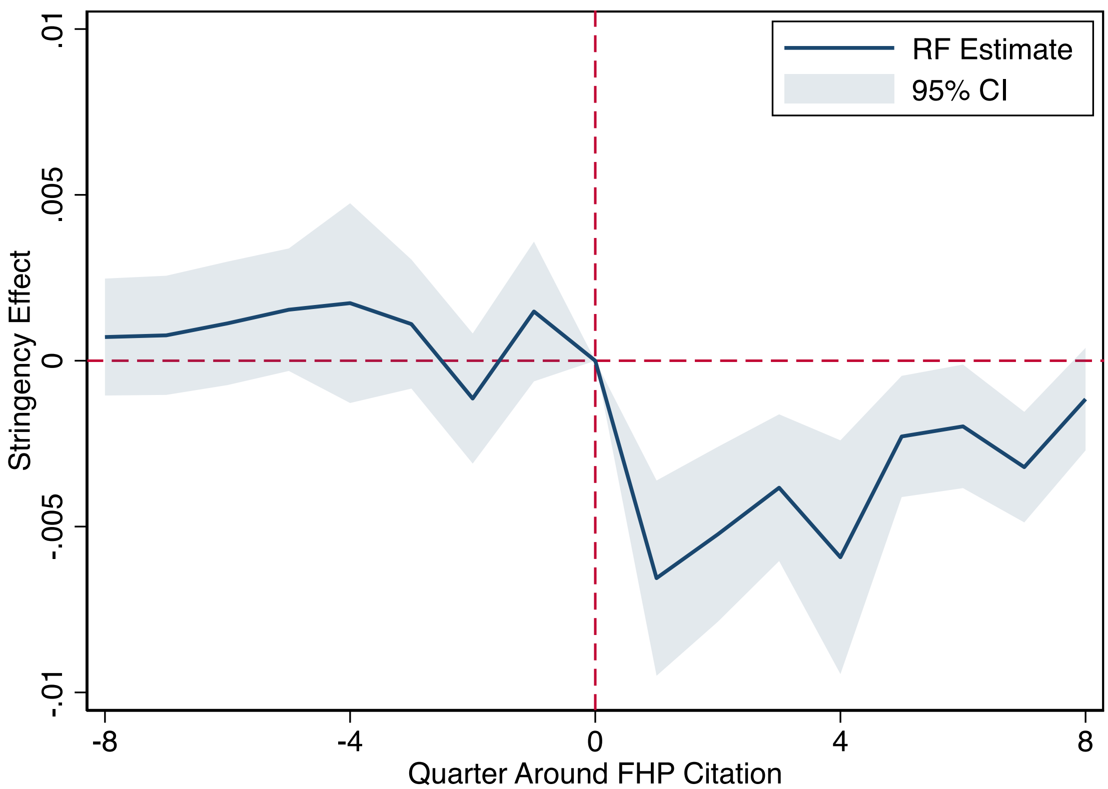
Notes: Standard errors two-way clustered at the officer and driver level in parentheses. Regressions also include indicators for missing age (<1%), missing zip code income (≈ 10%), and missing vehicle information (≈ 14%); joint significance tests include these variables.

Figure 3: Distribution of Instrument and First Stage



*N*otes: Figure illustrates a histogram of the officer stringency instrument, residualized of beat-shift fixed effects (left axis) and the first stage relationship between stringency and the probability of a harsh fine, both residualized of beat-shift fixed effects (right axis). Local means are denoted by blue circles and the green line shows a non-parametric fit, with 95 percent confidence intervals indicated by a dashed line. Graph reports the linear first stage estimate and associated F-statistic.

Figure 4: Reduced Form Over Time



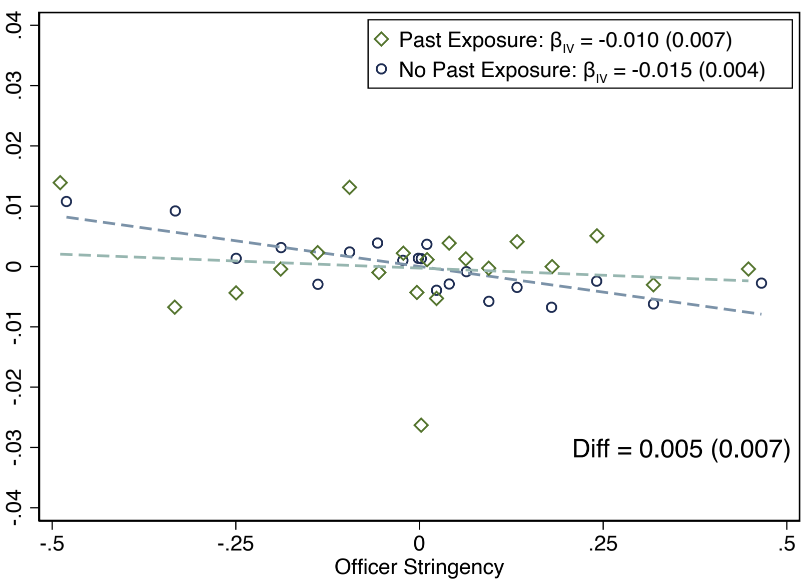
Notes: This figure reports coefficients on officer stringency from regressions of indicators for whether the driver received a speeding citation in each quarter relative to the date of their focal FHP citation. $\tau = 0$ denotes the exact date of the focal FHP citation (one day only, where all motorists 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 two-way clustered at the officer and driver level. First quarter effect is $\beta_{RF} = -0.0066$ (0.0015); lenient $\mu = .054$. Identical figures for other outcomes are shown in figure A-5.

Table 3: Effect of Harsh Fines, IV Estimates

	(1)	IV Estimates (2)	(3)	(4)
	Lenient Mean	β_{IV}	β_{IV}	ϵ
Any Violation	0.347	-0.0177	-0.0160	-0.069
		(0.0017)	(0.0016)	(0.007)
Speeding Violation	0.170	-0.0146	-0.0145	-0.128
		(0.0013)	(0.0013)	(0.012)
Other Violation	0.256	-0.0119	-0.0098	-0.057
		(0.0016)	(0.0015)	(0.009)
Moving Violation	0.280	-0.0143	-0.0136	-0.073
		(0.0016)	(0.0016)	(0.008)
Non-Moving Violation	0.160	-0.0124	-0.0104	-0.097
		(0.0013)	(0.0013)	(0.012)
Crash Involvement	0.080	-0.0029	-0.0022	-0.041
		(0.0010)	(0.0010)	(0.018)
Contest in Court	0.262	0.1125	0.1093	0.626
		(0.0014)	(0.0014)	(0.008)
Controls		No	Yes	Yes
Beat-Shift FE		Yes	Yes	Yes
Observations		1693436	1693436	1693436

Notes: This table reports IV estimates of the impact of receiving a harsh fine on one-year reoffending. Standard errors two-way clustered at the officer and driver level in parentheses. First stage estimates are $\beta = 0.944$ (0.006) without controls and $\beta = 0.943$ (0.006) with controls. See table A-4 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.

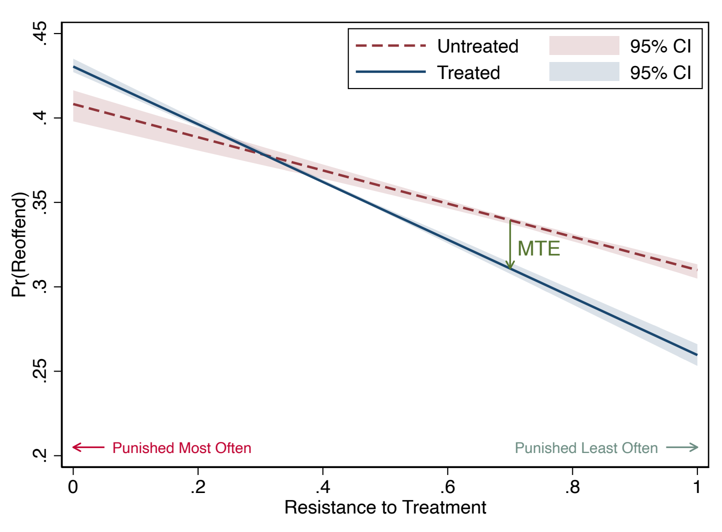
Figure 5: Impacts by Past Stringency Exposure



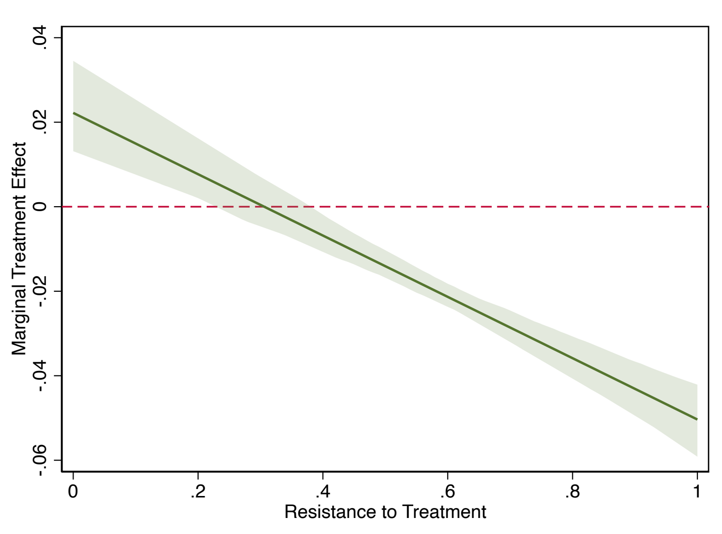
*N*otes: This figure is based on a subset of main sample citations where the driver has at least one FHP speeding citation with an identifiable officer more than one year ago ($N = 204,295$). Green diamonds and associated green line plot the relationship between stringency and one-year speeding recidivism (focal citation) for the subgroup with past exposure to a stringent (non-bunching) officer; blue circles and associated blue line plot the relationship between between stringency and one-year speeding recidivism (focal citation) for the subgroup with no past exposure to a stringent (non-bunching) officer.

Figure 6: Marginal Treatment Response Functions

(a) Marginal Treatment Responses



(b) Marginal Treatment Effects

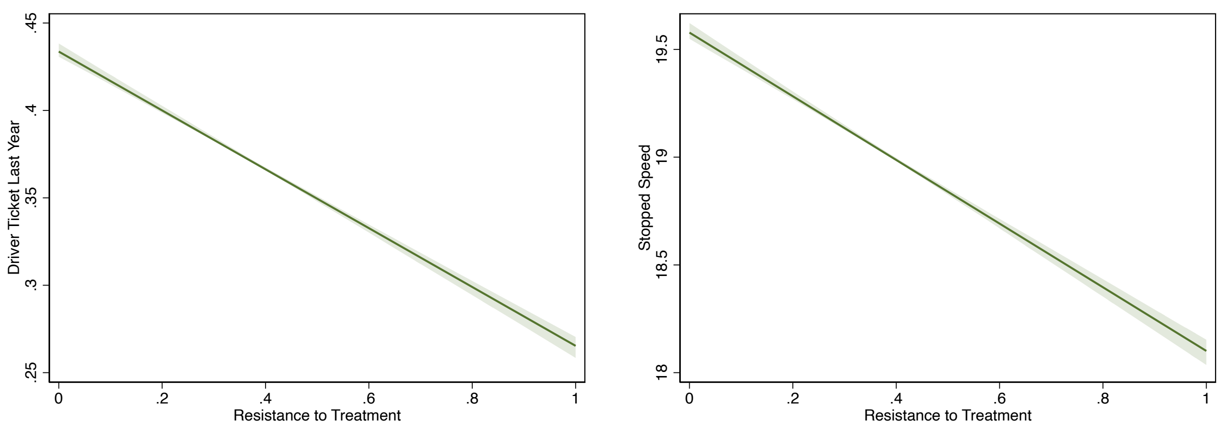


*N*otes: Outcome is any new traffic offense in the following year. Figures reports estimated marginal treatment responses (panel a) and marginal treatment effects (panel b) obtained via the method described in section 5.1. Shaded regions denote 95% confidence intervals computed via bootstrapping.

Figure 7: Marginal Complier Characteristics

(a) Citation in Past Year

(b) Stopped Speed (MPH)



*N*otes: Figure reports characteristics of marginal compliers; specifically the average characteristics of compliers at a given resistance to treatment, estimating using the method described in appendix C-3. Shaded region denotes 95% confidence intervals obtained via bootstrapping. Identical figures for motorist demographic characteristics are presented in figure A-13.

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	−0.001	0.327	−0.038	0.358
	[0.370, 0.381]	[−0.006, 0.005]	[0.324, 0.329]	[−0.043, −0.032]	
Reverse Resistance To Treatment	0.342	−0.027	0.392	0.010	0.341
	[0.339, 0.343]	[−0.030, −0.023]	[0.383, 0.398]	[0.004, 0.019]	
$\lambda \to 0$ (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.

Table 5: 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 5.4.

A Appendix Figures and Tables

Table A-1: Relationship between Lenience and Officer Characteristics

		Binary	Continuous		
	(1)	(2)	(3)	(4)	(5)
	Mean	Lenient	Raw	Adjusted	Weighted
Female	0.0893	-0.0704	-0.0508	-0.0356	-0.00342
		(0.0366)	(0.0266)	(0.0161)	(0.00388)
Race = Black	0.143	-0.0916	-0.0289	-0.00800	0.00550
		(0.0297)	(0.0231)	(0.0142)	(0.00291)
Race = Hispanic	0.169	-0.0933	-0.0771	0.0117	0.00921
		(0.0693)	(0.0564)	(0.0359)	(0.00707)
Race = Other	0.191	-0.0120	-0.0193	-0.0186	-0.00134
		(0.0655)	(0.0544)	(0.0350)	(0.00724)
Age	34.06	-0.0203	-0.0236	0.00982	0.00739
		(0.0561)	(0.0478)	(0.0288)	(0.00630)
Experience	7.09	-0.117	-0.0778	-0.0345	-0.000948
		(0.0388)	(0.0324)	(0.0214)	(0.00616)
Any College	0.319	-0.00798	-0.0114	-0.00228	0.00493
		(0.0213)	(0.0171)	(0.0108)	(0.00279)
Mean	—	0.753	0.353	0.005	0.006
Officers	1960	1960	1960	1960	1958

*N*otes: 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 are statistically insignificant in all cases.

Table A-2: Naive OLS Estimates

	(1) Reoffend	(2) Reoffend	(3) Reoffend	(4) Reoffend
Fine ($100s)	0.0201	0.0306	0.00905	0.0165
	(0.00153)	(0.00121)	(0.00108)	(0.000966)
Mean	0.161	0.161	0.161	0.161
Controls	No	No	Yes	Yes
Officer FE	No	Yes	No	Yes
Beat-Shift FE	Yes	Yes	Yes	Yes
Observations	1693435	1693435	1693435	1693435

*N*otes: Standard errors two-way clustered at the officer and driver level in parentheses. Dependent variable is an indicator for a new speeding offense in the next year. The reported mean is the mean for drivers cited at 9 MPH over the limit.

Table A-3: First Stage Estimates Across Subsamples

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

Notes: Standard errors two-way clustered at the officer and driver level in parentheses. First stage estimate for the full sample is $\beta_{FS} = 0.944$ (0.006). Each coefficient is from a separate regression of $\mathbf{1}[harsh]$ on the stringency instrument and beat-shift fixed effects using only the denoted subsample.

Table A-4: First Stage and Reduced Form Estimates

	(1) Lenient Mean	(2) β	(3) β
Panel A: First Stage			
Harsh Fine	0.5573	0.9441	0.9432
		(0.0064)	(0.0064)
Fine Amount	194.309	122.341	122.209
		(1.206)	(1.190)
Fine Amount (Paid)	167.187	95.819	96.062
		(1.414)	(1.407)
DL Points	3.416	0.7152	0.7142
		(0.0140)	(0.0138)
DL Points (Accrued)	1.683	-0.0362	-0.0275
		(0.0164)	(0.0154)
Panel B: Reduced Form			
Any Violation	0.3471	-0.0177	-0.0160
		(0.0017)	(0.0016)
Speeding Violation	0.1702	-0.0146	-0.0145
		(0.0013)	(0.0013)
Other Violation	0.2563	-0.0119	-0.0098
		(0.0016)	(0.0015)
Moving Violation	0.2801	-0.0143	-0.0136
		(0.0016)	(0.0016)
Non-Moving Violation	0.1602	-0.0124	-0.0104
		(0.0013)	(0.0013)
Crash Involvement	0.0799	-0.0029	-0.0022
		(0.0010)	(0.0010)
Contest in Court	0.2620	0.1125	0.1093
		(0.0014)	(0.0014)
Controls		No	Yes
Beat-Shift FE		Yes	Yes
Officers		1960	1960
Observations		1693436	1693436

*N*otes: Standard errors two-way clustered at the officer and driver level in parentheses. This table reports first stage and reduced form regression estimates. Each coefficient is from a separate regression of the stringency instrument on the denoted outcome.

Table A-5: Robustness, Alternative Instruments

Instrument	(1) N	F Stat (2) Balance	(3) FS	(4) β_{IV}
Leave-Out Mean	1693436	2.67	21965	-0.0146
				(0.00134)
Leave-Out Mean (Residualized)	1693436	2.51	25277	-0.0171
				(0.00134)
Binary	1693436	0.88	289	-0.0213
				(0.00321)
Officer Dummies	1693436	2.67	701289	-0.0168
				(0.00131)
Leave-County-Out Mean	1500469	1.54	626	-0.0164
				(0.00194)
Within Demographics				
Race	1364592	3.43	19162	-0.0124
				(0.00149)
Gender	1691486	3.76	22054	-0.0108
				(0.00134)
Race × Gender	1333296	3.44	18196	-0.00901
				(0.00151)
Race × Gender × Age × Income	320505	2.18	1670	-0.00774
				(0.00368)

*N*otes: 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 uses a binary instrument and row 4 uses the full set of officer dummies as instruments. Row 5 computes the instrument as the leave-county-out mean. Rows 6-9 show results when the instrument is computed as the leave-out mean within demographic cells. When computing the instrument within racial groups, we keep only white, Black, and Hispanic drivers. Age and income are binary categories split at the sample medians. Randomization statistics for within-covariate instruments exclude demographics.

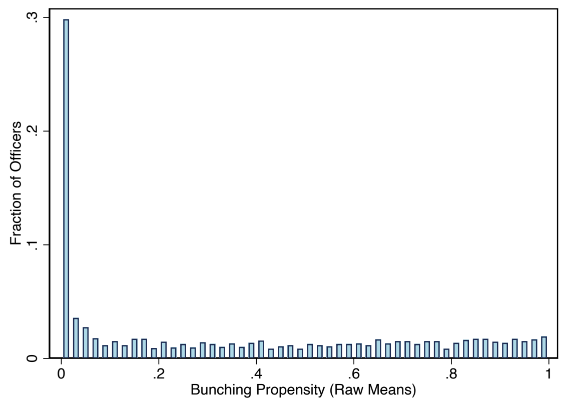
Table A-6: Robustness, Alternative Fixed Effects

Fixed Effects	(1) N	F Stat (2) Balance	F Stat (3) FS	(4) β_{IV}
Troop ×				
Highway × Year-Month × Weekend × Shift	1693436	5.893	41887	-0.0065
				(0.0011)
Posted × Year-Month × DOW × Shift	1668353	6.485	32464	-0.0113
				(0.0013)
County ×				
Year	1693436	2.869	22461	-0.0118
				(0.0012)
Highway × Year-Month × Weekend × Shift	1693436	2.666	21965	-0.0146
				(0.0013)
Highway × Year-Month × DOW × Shift	1661392	2.859	21532	-0.0152
				(0.0014)
Posted × Year-Month × DOW × Shift	1589646	3.164	17125	-0.0181
				(0.0016)
Posted × Exact Date × Shift	1429351	3.303	11702	-0.0193
				(0.0021)
Road Segment × County				
Highway × Year-Month × Weekend × Shift	219470	1.562	1217	-0.0140
				(0.0053)
Posted × Exact Date × Shift	164470	0.810	612	-0.0069
				(0.0098)

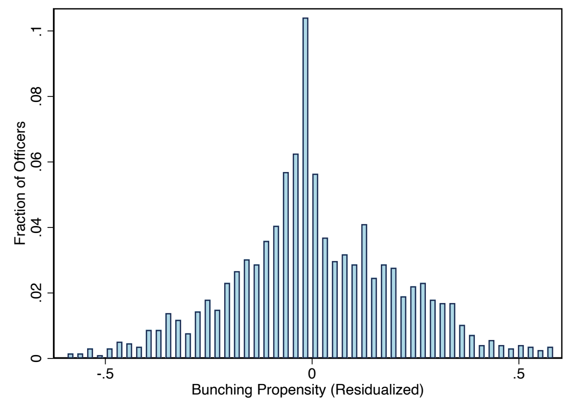
*N*otes: Same as table A-5 except that this table holds the stringency instrument constant and varies the level of the beat-shift fixed effects. *Posted* refers to the posted speed limit and $Highway = \mathbf{1}[posted \geq 55]$. Road segments coded only for the subset of the sample with GPS coordinates.

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

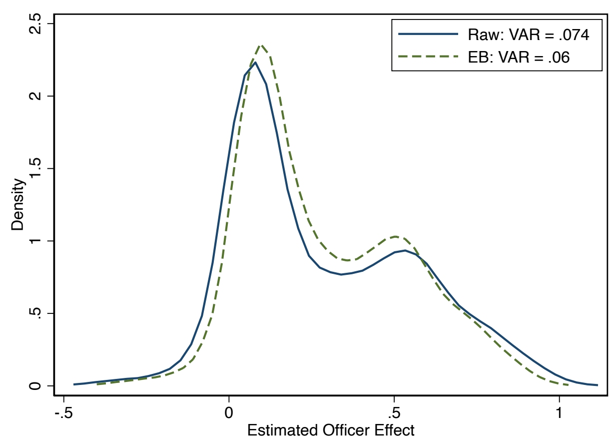
(a) Raw Bunching Propensity



(b) Adjusted Bunching Propensity



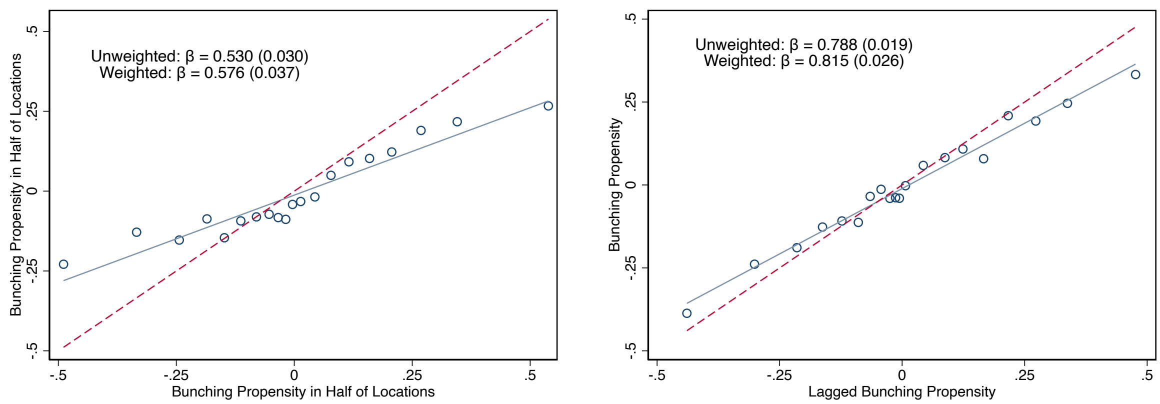
(c) Estimated Officer Effects



*N*otes: Panel (a) plots the officer-level distribution of the share of tickets bunched. Panel (b) plots the officer-level distribution of bunching propensity, adjusted for location-time effects. Panel (c) reports estimated officer effects from a regression of $\mathbf{1}[bunch_{ijs}]$ on officer fixed effects, location-time fixed effects, and the full set of driver covariates, as described in section 3.2. 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

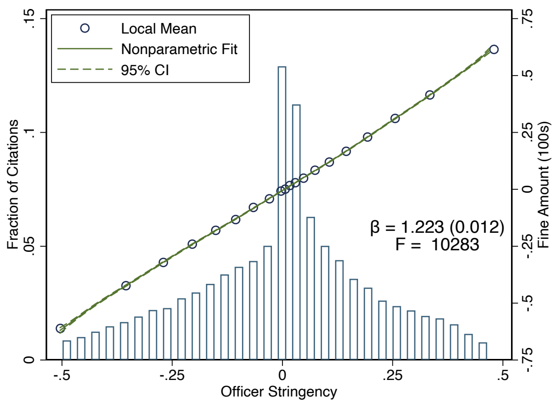
(a) Locations (b) Time Periods



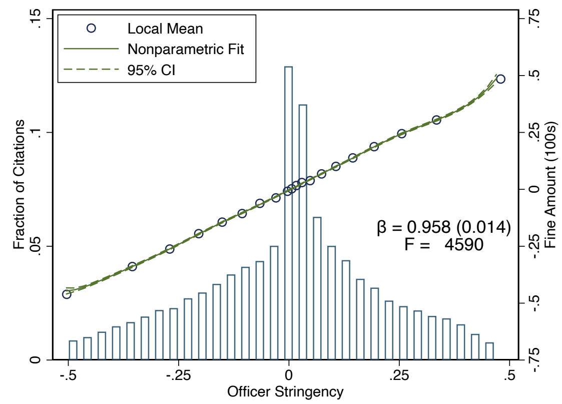
*N*otes: Dashed red line is the 45-degree line. This figure splits each officer's sample of citations into two groups and illustrates the correlation in (residualized) bunching propensity across groups. In panel (a), the groups are constructed as location partitions, with each partition comprised of 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.

Figure A-3: First Stage Estimates, Sanction Measures

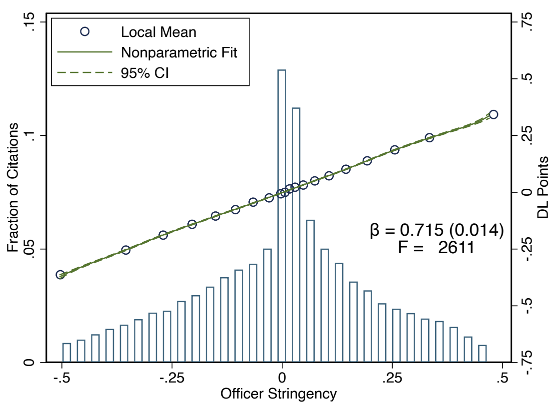
(a) Statutory Fine Amount



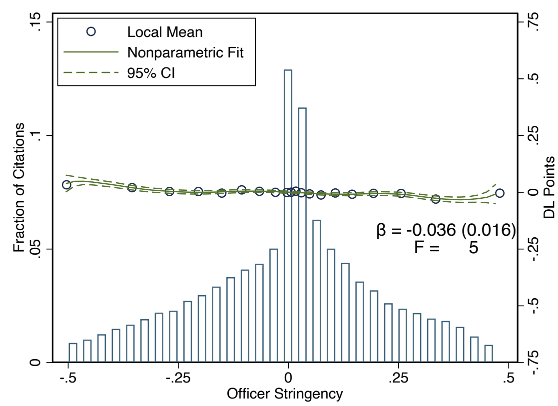
(b) Paid Fine Amount



(c) Statutory DL Points



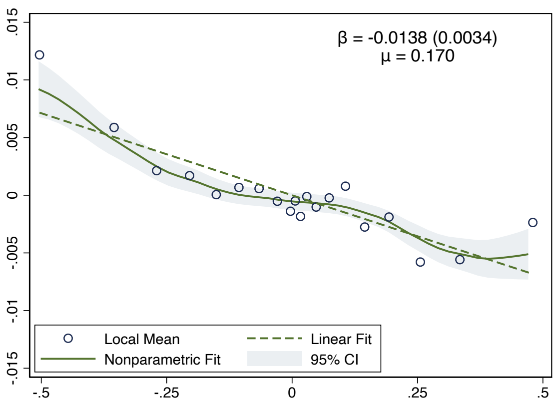
(d) Accrued DL Points



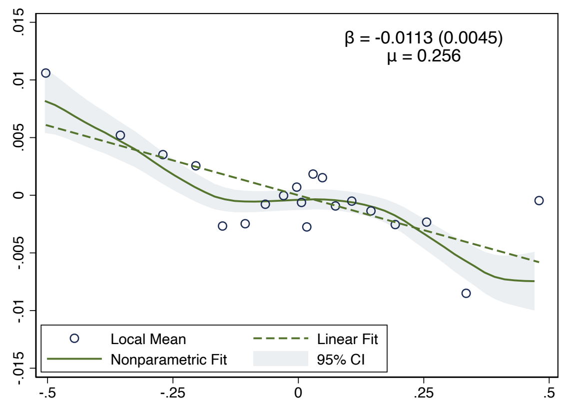
*N*otes: Each panel shows an identical to plot figure 3 but replaces the outcome variable with a different sanctions measure. In panel (a), the outcome is the statutory fine based on the charged speed. In panel (b), the outcome is the effective fine amount, taking into account the *ex-post* court outcomes 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 B-1 for details on the computation of effective sanction measures (paid fines and accrued points).

Figure A-4: Reduced Form Estimates

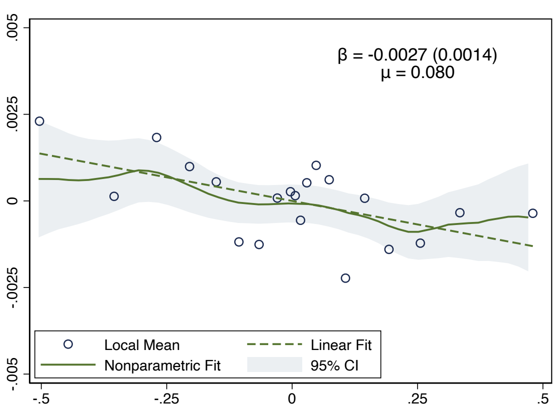
(a) Speeding Offense



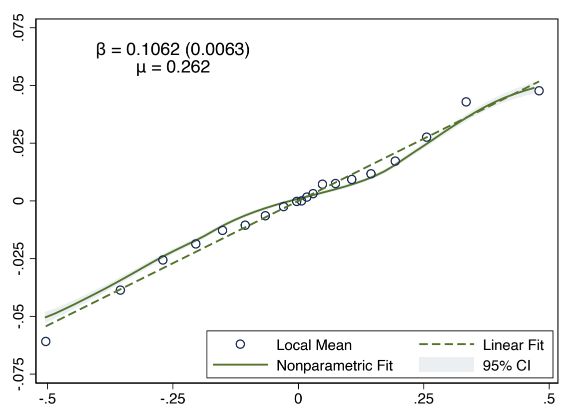
(b) Non-Speeding Offense



(c) Crash Involvement

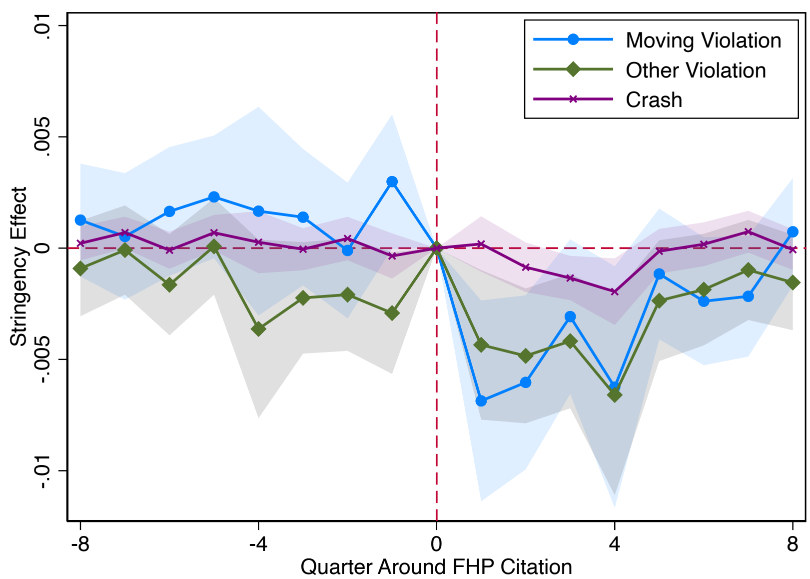


(d) Contested Citation



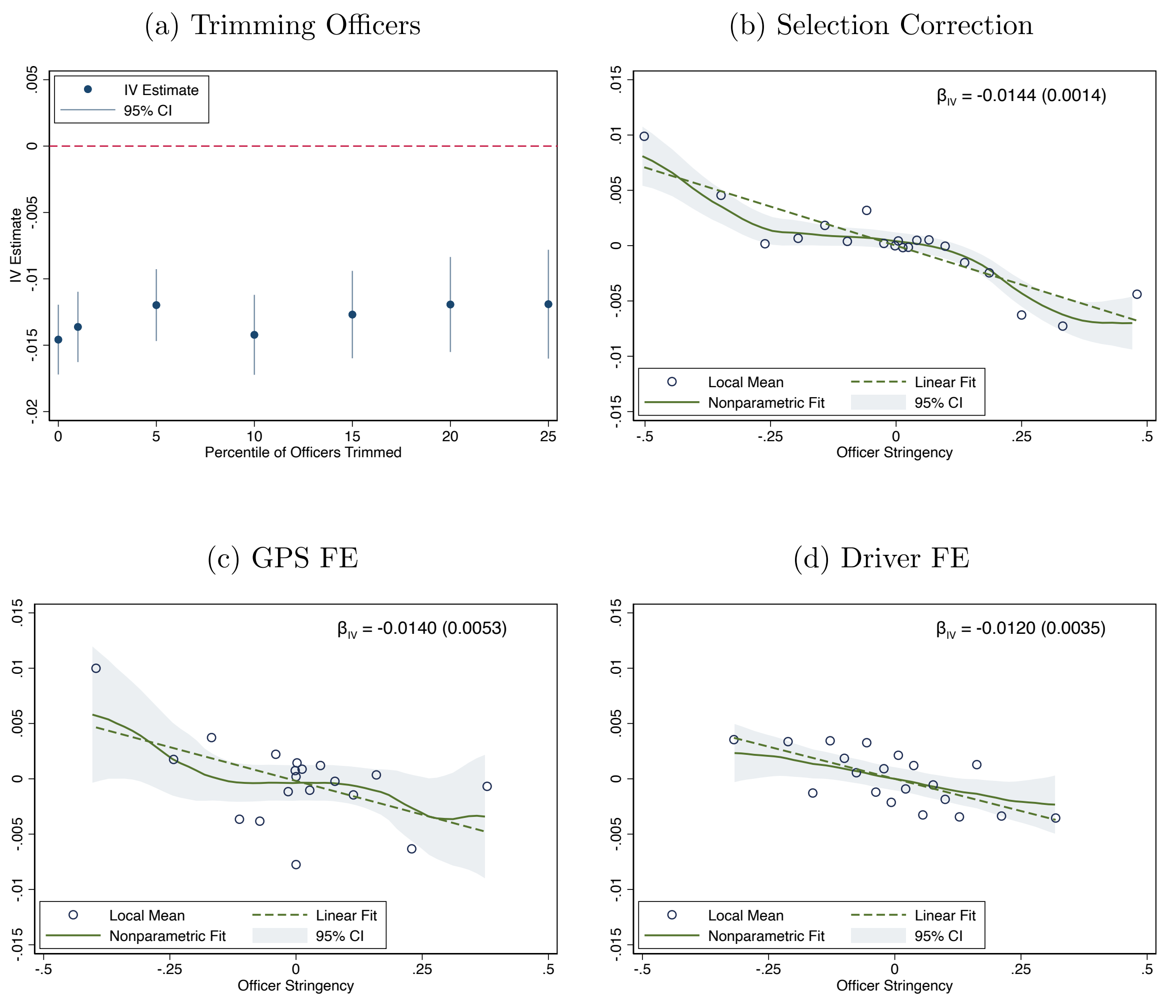
*N*otes: Same as table A-3 except for reduced form outcomes.

Figure A-5: Reduced Form Estimates Over Time



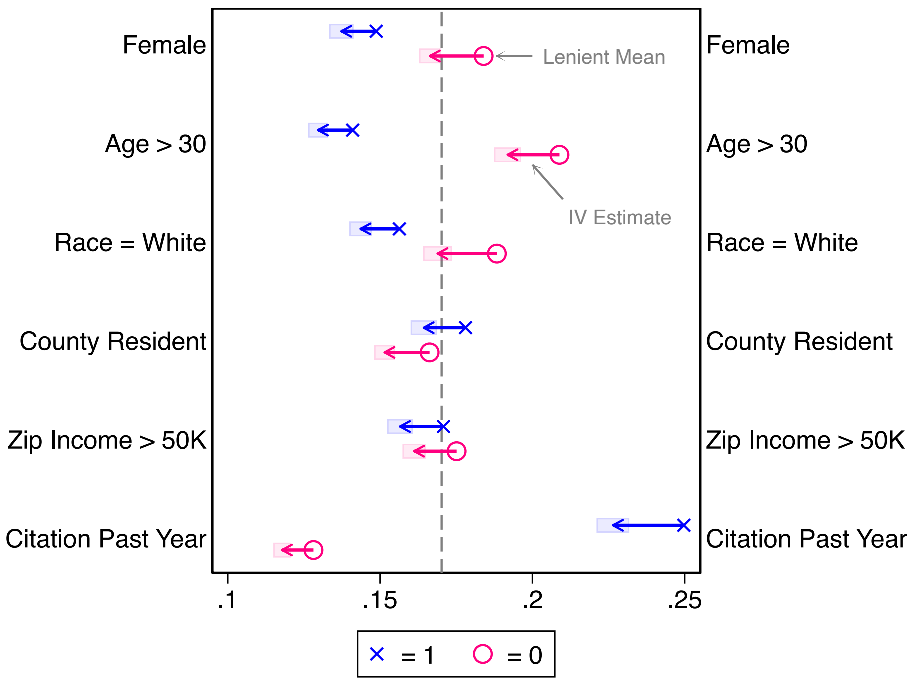
*N*otes: Same as figure 4 using any moving violation in a given quarter (blue circles), any non-moving violation in a given quarter (green diamonds), and any crash involvement in a given quarter (purple x's) as the outcome variable. Shaded regions denote 95 confidence intervals.

Figure A-6: Robustness, Sample Selection



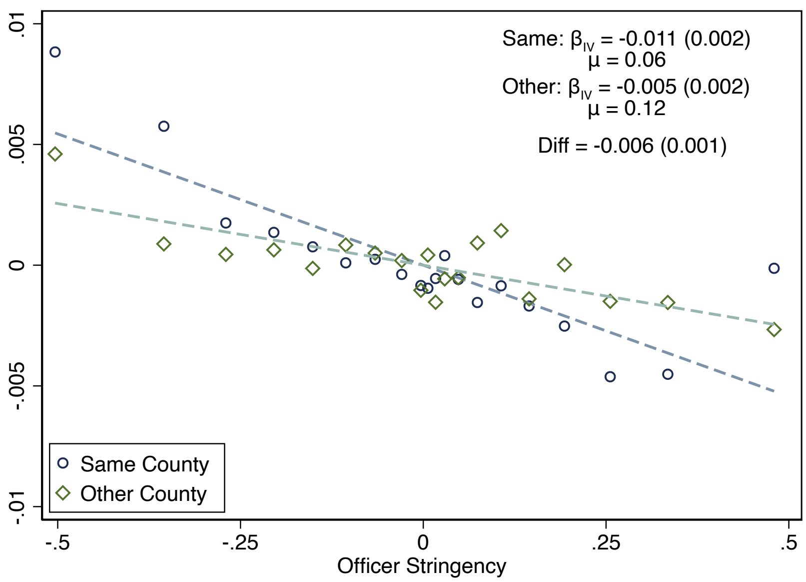
*N*otes: For comparison, our main IV estimate is $\beta_{IV} = -0.0146$ (0.0013). Panel (a) shows the sensitivity of our IV estimate to trimming officers with the most selected samples. Panel (b) plots the reduced form and reports the IV estimate using a Heckman (1979) selection correction based on officer ticketing frequency. Panel (c) plots the reduced form and reports the IV estimate using GPS road segment fixed effects. Panel (d) plots the reduced form and reports the IV estimate using a within-driver design.

Figure A-7: IV Estimate Heterogeneity by Driver Characteristics



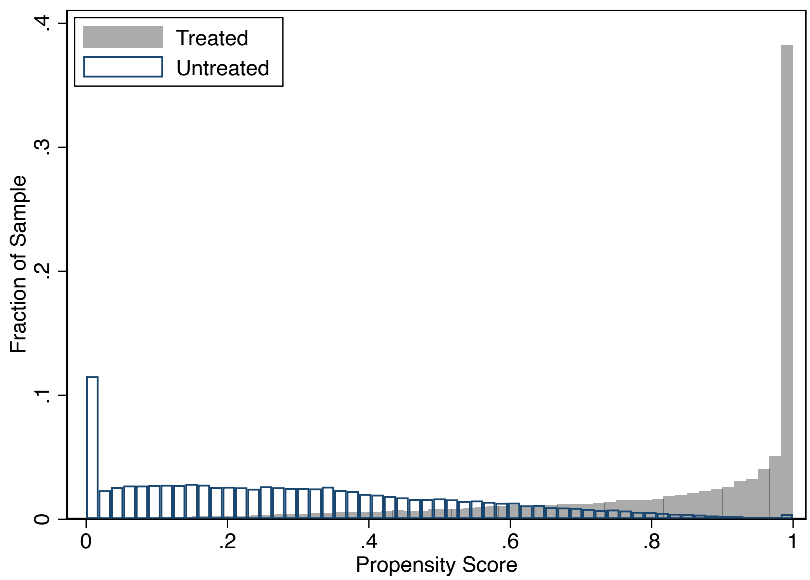
*N*otes: This figures shows heterogeneity in IV estimates for one-year speeding recidivism by driver characteristics. Each characteristic is denoted as a binary category; the x's plot lenient means for the category = 1 subgroup and the o's plot lenient means for the category = 0 subgroup. Arrows pointing 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 for the full sample.

Figure A-8: Localized Responses



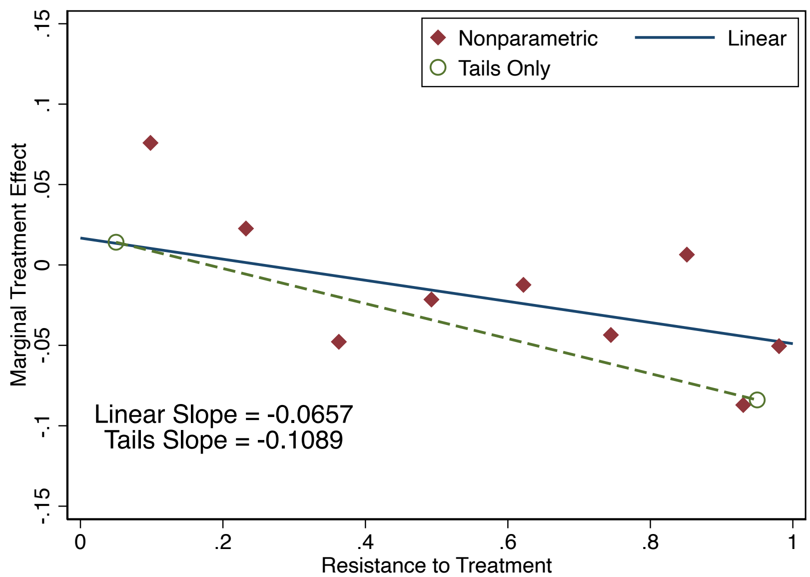
*N*otes: Figure plots the reduced form relationship between officer stringency and one-year speeding recidivism separately for whether a new speeding offense occurred in the same or different county relative to the focal FHP citation. Stringency and reoffending are adjusted for beat-shift fixed effects. IV estimates, control means, and the difference in the IV estimates are reported in the upper right corner.

Figure A-9: Common Support for MTE Estimates



Notes: Figure plots the distribution of propensity scores for the treated (66%) and untreated (35%) 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. For reference, $Pr(D = 1|p = 0) = 0.0209$ and $Pr(D = 1|p = 1) = 0.9947$.

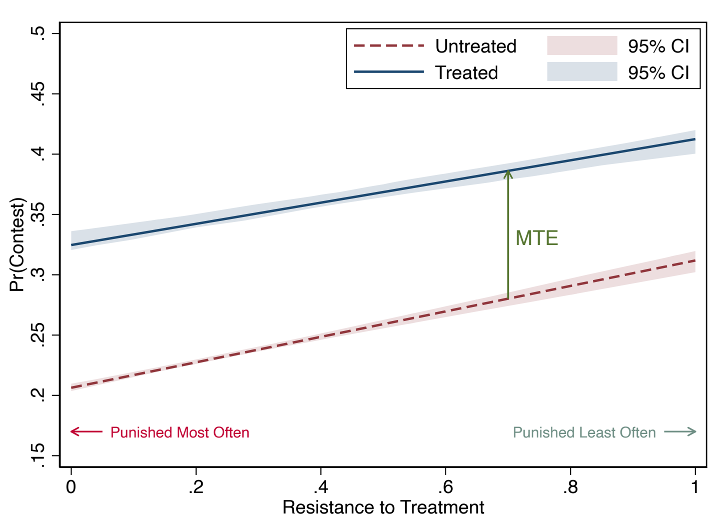
Figure A-10: Alternative MTE Estimates



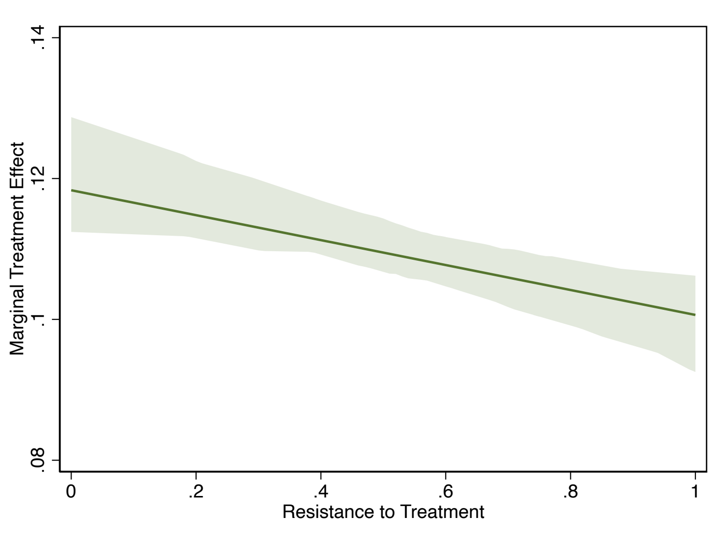
Notes: This figure shows marginal treatment effect estimates for any reoffending using three approaches. Blue line reports MTE estimates from the linear, separate approach (same as figure 6). Red diamonds report MTE estimates from a nonparametric approach. Green dots report MTE estimates using our monotonicity-robust approach relying only on extreme officers. See section 5.1 and appendix C for further details.

Figure A-11: Marginal Treatment Response Functions, Court Contesting

(a) Marginal Treatment Responses



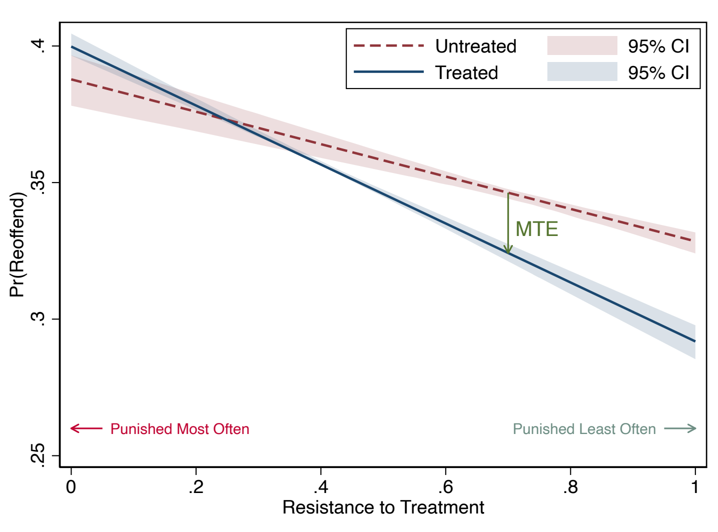
(b) Marginal Treatment Effects



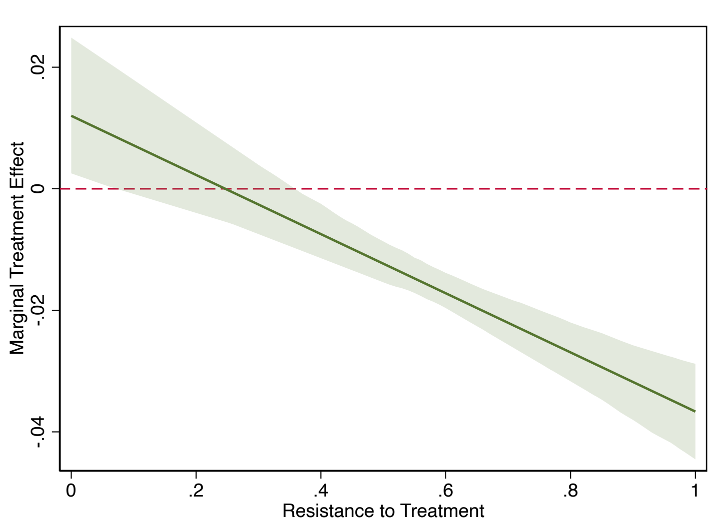
Notes: Outcome is whether citation is contested in traffic court, using the measure described in appendix B-1. Figures report estimated marginal treatment responses (panel a) and marginal treatment effects (panel b) obtained via the method described in section 5.1. Shaded regions denote 95% confidence intervals computed via bootstrapping.

Figure A-12: Marginal Treatment Response Functions with Driver Covariates

(a) Marginal Treatment Responses



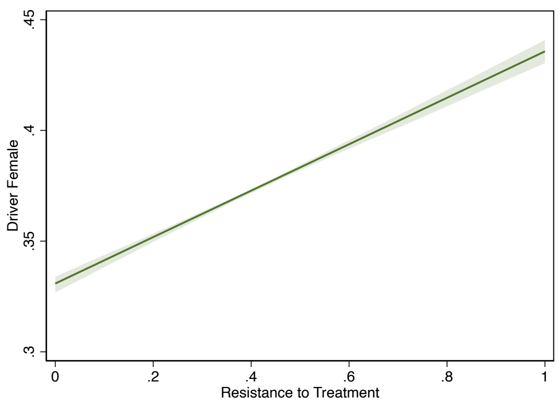
(b) Marginal Treatment Effects



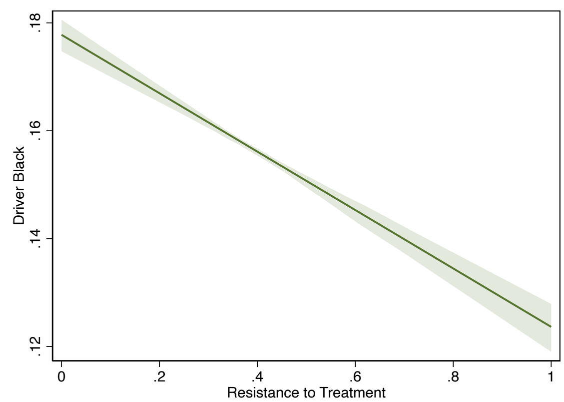
*N*otes: Outcome is any new traffic offense in the following year. Figures report estimated marginal treatment responses (panel a) and marginal treatment effects (panel b) obtained via the method described in section 5.1. Shaded regions denote 95% confidence intervals computed via bootstrapping. These specifications allow the level of Y_0 and Y_1 to vary with driver covariates.

Figure A-13: Demographic Characteristics of Marginal Compliers

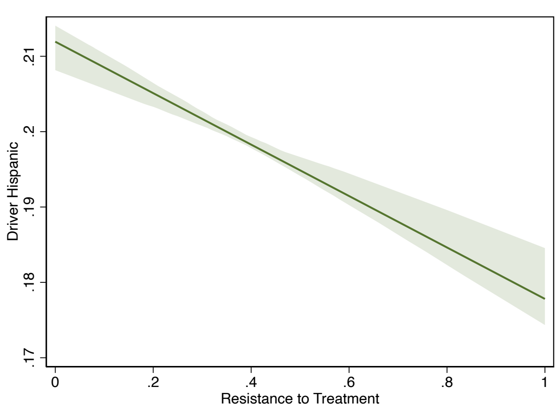
(a) Driver Female



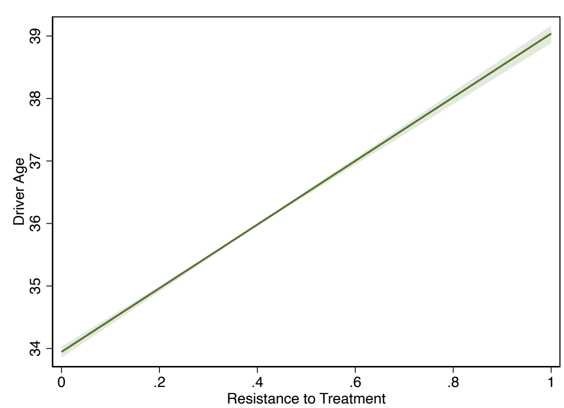
(b) Driver Race = Black



(c) Driver Race = Hispanic



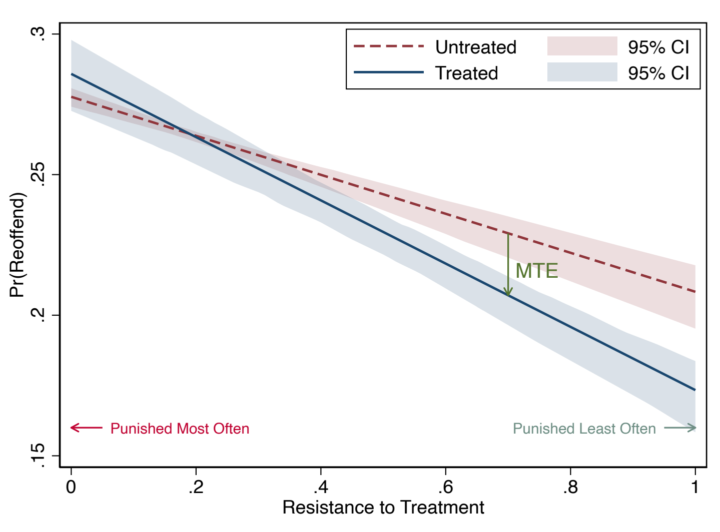
(d) Driver Age



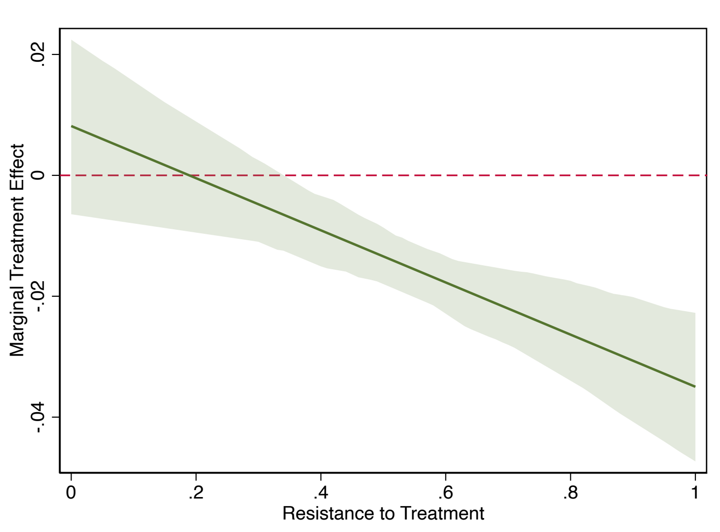
*N*otes: Figures report characteristics of marginal compliers; specifically the average characteristics of compliers at a given resistance to treatment, estimating using the method described in appendix C-3. Shaded region denotes 95% confidence intervals obtained via bootstrapping.

Figure A-14: Marginal Treatment Response Functions, Offenders with No Tickets in Previous Year

(a) Marginal Treatment Responses



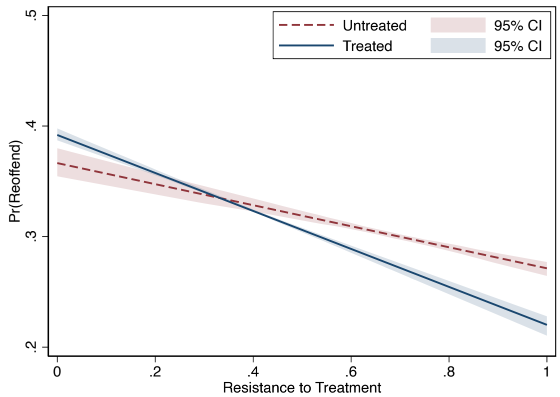
(b) Marginal Treatment Effects



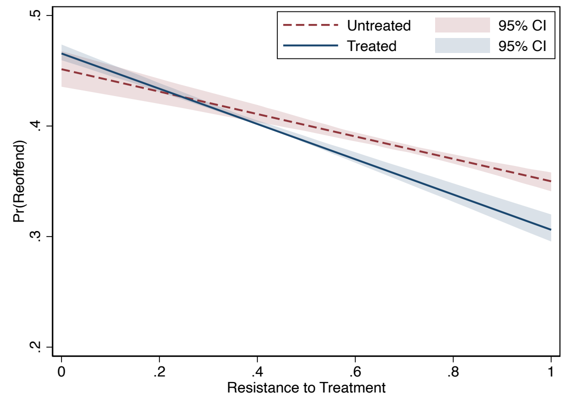
*N*otes: Same as figure 6 except using only the sample of offenders with no tickets in previous year.

Figure A-15: Marginal Treatment Response Functions by Driver Race

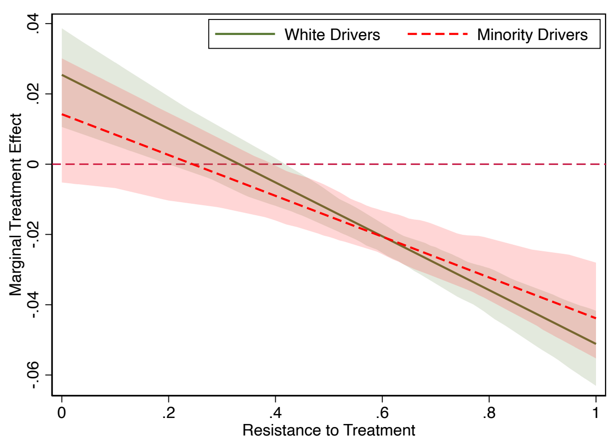
(a) MTR, White Drivers



(b) MTR, Minority Drivers



(c) Marginal Treatment Effects



*N*otes: Panels (a) and (b) plot marginal treatment response functions (same as figure 6) separately for white motorists and minority (Black and Hispanic) motorists. Panel (c) plots estimated marginal treatment effects separately for white and minority motorists.

B Data Appendix

B-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-3, 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.

B-2 Binary stringency measure

To identify officers who do not bunch, we use the Frandsen (2017) test for manipulation in bunching. 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.

C Technical Appendix

C-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 (2007a), 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 $E(U_0|X) = 0$:

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

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 $\widehat{\mu_0(\bar{X})} = \frac{1}{N} \sum_i \left[\hat{y}_i - \hat{\alpha}_0 \frac{\hat{p}_i}{2}\right]$, 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, $E[U_1|U_D = u, X] = \alpha_1(p - 1/2)$:

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

C-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 $Z \leq 0.1$. We will then construct a binary version of the instrument, $\tilde{Z} = Z > 0.05$. 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 $u_D \in [0.05, 0.1]$. In addition, Abadie (2002) shows how interacting the outcome with treatment status can identify the counterfactual outcome for the compliers:

$$\begin{aligned} \frac{E[DY|X, \tilde{Z} = 1] - E[DY|X, \tilde{Z} = 0]}{E[D|X, \tilde{Z} = 1] - E[D|X, \tilde{Z} = 0]} &= E[Y_1|X, U_D \in [0.05, 0.1]] \\ \frac{E[(1 - D)Y|X, \tilde{Z} = 1] - E[(1 - D)Y|X, \tilde{Z} = 0]}{E[(1 - D)|X, \tilde{Z} = 1] - E[(1 - D)|X, \tilde{Z} = 0]} &= E[Y_0|X, U_D \in [0.05, 0.1]] \end{aligned}$$

We will therefore run regressions of DY and $-(1 - D)Y$ on our treatment, where we restrict attention to officers with $Z < 0.1$ 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 $Z \geq 0.9$ and use the binarized instrument $\tilde{Z} = Z > 0.95$.

C-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.

C-4 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:

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

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 θth percentile of the objective function.

These functions correspond to the marginal treatment responses we estimate in the data, $m_j(\theta) = E[Y_j|\overline{X}, U_d = \theta]$. 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:

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

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

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:

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

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. For the sake of the following paragraph, label this guess $\widetilde{Pr}(\hat{Y}_{i0}, \hat{Y}_{i1})$. We provide 101 guesses with this grid of probabilities, with values for $\lambda = 0, 0.01, ..., 1$.

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:

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

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:

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

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 $E[Y_1 - Y_0|D=1]$.