How Do People React to Income-Based Fines? Evidence from Speeding Tickets Discontinuities*

Martti Kaila

University of Helsinki and *Helsinki GSE*

November 8, 2022

Job Market Paper
Latest Version

Abstract

This paper studies the impact of income-based criminal punishments on crime. In Finland, speeding tickets become income-dependent if the driver's speed exceeds the speeding limit by more than 20 km/h, leading to a substantial jump in the size of the speeding ticket. Contrary to predictions of a traditional Becker model, individuals do not bunch below the fine hike. Instead, the speeding distributions are smooth at the cutoff. However, I demonstrate that the size of the realized speeding ticket has sizable but short-lived impacts on reoffending ex-post. I use a regression discontinuity design to show that fines that are 200 euros larger decrease reoffending by 20 percent in the following six months. After 12 months, the effect disappears. My empirical results are consistent with an explanation that people operate under information frictions. To illustrate this, I construct a Becker model with misperception and learning that can explain all the empirical findings.

* *martti.kaila@helsinki.fi* *marttikaila.com* I am particularly thankful for Kristiina Huttunen, Emily Nix, Roope Uusitalo, and Markus Jäntti, who have provided guidance, help, and inspiration for this project. I thank Daniel Hauser, Ramin Izadi, Tuomas Kosonen, Dave Macdonald, Tuomas Matikka, Teemu Pekkarinen, Tuomas Pekkarinen, Mikko Silliman, and Tuuli Vanhapelto for helpful comments and discussions. In addition, I thank seminar participants at Aalto University, Helsinki GSE Labor and Public Economics Workshop, and Criminological Seminar at the University of Helsinki.

1 Introduction

One of the primary roles of government is to deter people from committing actions that cause harm to others. For example, since traffic accidents are the most common unnatural cause of death globally (WHO, 2017), governments worldwide use speeding tickets to discourage drivers from driving at speeds that pose a danger to others. Economic theory suggests that all else equal, increasing the severity of punishment, such as the size of speeding tickets, should decrease criminal activity (Becker, 1968). However, empirical tests yield mixed results on this key theoretical prediction (Chalfin and McCrary, 2017).

This paper studies the impact of punishment severity on crime using high-quality register data and a unique context where the magnitude of fines increases with a person's income. Specifically, in Finland, fines become income-dependent if the driver's speed exceeds the speeding limit by more than 20 km/h. Fines are allowed to increase without an upper limit at the cutoff. This policy means that high-income drivers have a substantial incentive to slow and bunch below the income-based threshold, since failing to avoid the income-based cutoff can be very expensive. For example, in 2019, the Police assigned NHL player Rasmus Ristolainen an income-based speeding ticket equal to around 120,000 euros.[1]

The first key result of the paper is that individuals ignore the discontinuous jumps in the severity of the punishment. This directly contradicts the seminal model of Becker (1968), which predicts that people should react to discontinues in the price of speeding at the 20 km/h cutoff by slowing down and bunching just below the income-based fine cutoff. Specifically, working with the Police in Finland, I obtained detailed police speeding ticket data which I link to administrative tax data. Using this data, I find zero excess mass below the income-based fine cutoff in the speeding distributions. Surprisingly, even high-income drivers fail to bunch below the cutoff, even though they face a considerable incentive to avoid the income-based fine.

The rest of the paper focuses on understanding why people ignore this stark discontinuity in punishments. I propose two potential explanations that could attenuate individuals' reactions to incentives, both of which have been previously applied in other non-crime contexts (Chetty, 2012; Kleven and Waseem, 2013). First, I examine the possibility of information frictions: drivers are unaware of the details of the system, leading to sub-optimal speed decisions around the cutoff. Second, I consider the possibility of adjustment costs that cause individuals not to react to the cutoff. The adjustment costs explanation posits that individuals operate under perfect information but neglect the cutoff since optimization

[1] Finnish and international media frequently cover the Finnish income-based fine system. See, for example, articles by The New York Times and The Atlantic

is mentally too costly relative to the expected punishment.

Both of these possible frictions can generate smooth speed distributions around the discontinuity *ex-ante*, but the two friction types imply very different predictions on how realized fines should affect recidivism *ex-post*. If individuals react to a realized fine, then this reaction suggests imperfect information, with individuals learning from the punishment. In contrast, under the adjustment cost story, the experienced fine should not change an individual's behavior since the ticket is just a realization of a rational bet.

The second main result of the paper is that people *do* react to the severity of realized punishments ex-post. Since the speeding distributions are continuous at the cutoff, I use a regression discontinuity design (RDD) to study how larger realized fines, due to the income-based system, affect reoffending. According to my RDD estimates, individuals assigned larger fines due to the income-based fine system are less likely to commit another traffic offense in the short term. The impact of larger fines on reoffending peaks around 4-8 months after the income-based fine. Those assigned, on average, a 200 euro larger fine are approximately 2-3 percentage points less likely to commit another traffic crime in the following six months. Compared to the average speeding behavior of the speeders who receive a smaller fixed fine, this estimate implies a 20 percent reduction in recidivism. In addition, reoffenders' speeding distributions evolve smoothly at the income-based fine cutoff. Hence, larger fines trigger an extensive margin response.

The RDD results lead to two conclusions. First, they suggest that individuals learn from punishments, implying that information frictions at least partly cause the smooth distribution around the discontinuity. Second, this learning response results in less speeding in the short term after fines are realized, but the fine hikes do not provide deterrence ex-ante.

Further, my results indicate that the drivers' learning experience increases with the size of the realized fine. I show that high-income individuals receiving larger fines at the income-based cutoff are less likely to reoffend in the short term than low-income individuals receiving smaller fines. To do this, I study how the RDD estimates differ by income quartiles. The fine jumps by around 50 (550) euros in the bottom (top) income category. For cumulative reoffending over six months, the point estimate is close to zero (0.1 pp.) for the bottom quartile, while in the top quartile the point estimate of -4.5 pp. is double the main estimate. In addition, the different income quartile groups respond similarly to small fixed-fine jumps. This suggests that the high-income individuals' stronger reaction to larger fines is mainly because of the larger fines and not because low- and high-income drivers differ on some other dimension.

Lastly, my RDD results show that the impact of the larger fine fades out over time. The effect of larger fines disappears 12 months after the initial speeding incident. This result suggests that despite the heftier fines leading to a substantial short-term learning reaction, drivers seem to forget in the long run.

What model could simultaneously explain smooth speeding distributions, reactions to realized fines, and long-run fade-out in the effect of fines? To rationalize these findings, in the last part of the paper, I construct a Becker model with misspecification and learning to capture drivers' behavior.[2] The model builds on the idea that if individuals find pricing schemes too complex to understand, they may use the so-called "ironing" heuristic, where people approximate marginal prices with average prices (Rees-Jones and Taubinsky, 2019).[3] Drawing from this insight, I assume that drivers find the actual penalty function too complex. Therefore, they make speeding decisions using a rule of thumb that the speeding ticket increases linearly with speed. Based on their speed, drivers may receive a fine. If the fine is larger than predicted, they conclude that the relationship between speed and fine is steeper than expected and update their beliefs accordingly.

I show that this Becker model with misperception and learning rationalizes all my empirical findings. First, the "ironing" heuristic causes drivers to ignore the fine discontinuity, leading to a smooth speeding distribution. Second, due to the jump in the true penalty function, drivers may receive larger fines than they expected. This steepens their linearized penalty function and reduces speeding in the next period. Furthermore, the larger the true discontinuity in the fine, the more significant the driver's reaction *ex-post*. Finally, I show that when individuals learn with the misspecified model, there are drivers whose beliefs and actions do not converge but follow cycles.[4] This is consistent with the fade-out over time of the impact of receiving a larger than expected fine at the discontinuity.

This paper contributes to several literatures. One of the key questions in the economics of crime is how punishments deter crime. However, the literature that studies the impact of punishment severity on crime is mixed (Chalfin and McCrary, 2017).[5] There is previous

[2]The theoretical framework builds extensively on the work of Traxler *et al.* (2018), and Dušek and Traxler (2021). In Section 5, I also present a Becker model with adjustment costs that can generate smooth speeding distributions but not reactions to realized fines.

[3]Bhargava *et al.* (2017) documents that people systemically select health plans that are financially dominated by other options. Ito (2014) provides evidence that individuals respond to average changes in electricity prices but not to marginal changes. Saez (2010) finds that wage earners do not respond to marginal changes in tax schedule induced by the US income tax system earned income tax credit. Bastani and Selin (2014) shows that wage earners do not react to a large marginal tax kink in Sweden.

[4]An extensive theoretical literature studies learning when an individual's model is misspecified (see, for example, Berk (1966), Bohren and Hauser (2021), Esponda and Pouzo (2016), Heidhues *et al.* (2018), or Nyarko (1991)). My result builds on Nyarko's 1991 idea that individuals' beliefs and actions may fail to converge if individuals' prior do not contain the true parameter value.

[5]Sentence enhancements for a specific crime or group of individuals seem to have a notable effect on crime(Bell *et al.*, 2014; Drago *et al.*, 2009; Helland and Tabarrok, 2007). The evidence on the impact of capital punishment on crime is inconclusive (Chalfin and McCrary, 2017). Papers that study the effects of harsher punishment arising from age discontinuities find, if anything, minimal effects (Lee and McCrary, 2017) However, Arora (2019) argues that these papers ignore dynamics effects that might be substantial

evidence showing that people may react to speeding ticket notches (Traxler *et al.*, 2018). I add to this literature by showing that individuals may completely ignore the discontinuous changes in the punishment severity due to information frictions.[6] This knowledge about the nature of frictions attenuating individuals' responses to punishments is crucial for designing efficient policies to prevent illegitimate actions, such as excessive speeding. The expected punishment may be too low if adjustment costs mute individuals' responses. If, on the other hand, people misperceive the punishment system, the penalty scheme should be more salient to be effective.

Second, I contribute to the literature that investigates how the experience of punishment deters recidivism ex-post. Criminologists characterize this reaction as a "specific deterrence" effect (Doleac, 2017). In contrast, economists often describe such behavior as information updating (Chalfin and McCrary, 2017). Identifying the specific deterrence effect is challenging since punishments may impact recidivism through other channels, such as peer effects.[7]

Of the papers that do isolate the specific deterrence effect, the most closely related to mine are by Dušek and Traxler (2021) and Goncalves and Mello (2017).[8] Dušek and Traxler (2021) finds that speeding tickets reduce reoffending at an extensive margin (no ticket vs. ticket) but not at an intensive margin (low vs. high ticket). In contrast, Goncalves and Mello (2017) observe that a speeding ticket hike decreases future speeding at the intensive margin. I add to these papers by showing that speeding ticket increases may reduce future speeding at the intensive margin, but the fine must be hefty enough. Furthermore, my results indicate that governments can use income-based penalties to close the disparity in speeding tickets between high- and low-income people.

Third, my paper relates to the literature studying individuals' reactions to discontinuities in the choice sets (for a review, see. Kleven (2016)) A common finding is that discontinuities,

[6]This result is in line with public finance and labor economics literature, which document that, because of the information frictions, individuals face considerable difficulty in reacting to discontinuous changes in the incentives (Saez, 2010; Kleven and Waseem, 2013; Kostøl and Myhre, 2021; Chetty *et al.*, 2013).

[7]Doleac (2017) reviews the literature studying how more severe punishments, such as prison sentences, impact reoffending. In these cases, the estimates may capture the impact of multiple mechanisms. A prison sentence may decrease recidivism through specific deterrence. On the other hand, a prison sentence could also have a criminogenic impact on an individual, for example through exposure and interaction with other criminal inmates. As a result of these various possible mechanisms, results on the impacts of prison sentences can be very context-specific. For instance, Aizer and Doyle (2015) and Mueller-Smith (2014) find that incarceration increases recidivism, whereas Bhuller *et al.* (2020) and Rose and Shem-Tov (2021) find that incarceration decreases reoffending. Good examples of papers that can identify specific deterrence effects are by Hansen (2015) and Gehrsitz (2017). Hansen (2015) studies the impact of larger fines or probation on the probability that an individual commits another drunk driving offense. Gehrsitz (2017) shows that a driver's license suspension reduces future traffic violations.

[8]Other notable papers that study learning in the context of crime are by Banerjee *et al.* (2019) and Philippe (2022). Banerjee *et al.* (2019) provides evidence that learning plays a key role in how individuals react to randomly assigned drunk-driving crackdowns. Philippe (2022) shows that individuals with first-hand experience with a reform that increases punishment severity learn from the realized punishment and adjust their behavior strategically.

for example, in tax rates, produce evident bunching responses. However, these bunching reactions often translate into small elasticities due to optimization frictions. I contribute to this literature by documenting that people may completely ignore substantial jumps in prices due to information friction but learn from price signals. This evidence is consistent with the findings in other contexts that people do not use marginal prices in decision-making (Rees-Jones and Taubinsky, 2019; Kostøl and Myhre, 2021; Ito, 2014).

My findings have implications for policies on criminal sanctions. My results suggest that speeding tickets curb excessive speeding. However, fines may work efficiently and have persistent effects only if the fine schedule is salient enough and individuals are reminded frequently. In particular, my results indicate that enforcement may need to be more aggressive and frequent to internalize the negative externalities of speeding than a simple Becker model would predict. This result could also hold more broadly for other types of criminal activity, and thus this paper provides a more complex model that I show can more accurately reflect choices in the real world.

The remainder of the paper is organized as follows. Section 2 provides institutional details and describes the data. Section 3 shows the speeding distribution. Section 4 investigates whether the size of the speeding ticket impacts recidivism. Section 5 introduces a framework that seeks to rationalize my findings, and Section 6 concludes.

2 Institutional Setting and Data

Speeding is a common crime. In Finland, 35 percent of crimes reported to the police are speeding incidents (See Appendix Figure A.1). These offenses are taken seriously across the world given the potentially large costs of driving recklessly. WHO (2017) estimates that traffic accidents are the most common cause of death globally among those aged 15-49 (see Appendix Figure A.2). In a large share of these lethal accidents, the primary cause is excess speed (WHO, 2018).

Finland offers an attractive context to study the impact of criminal sanctions on excess speeding for two reasons. First, the speeding punishment schedule and especially the income-based fine system in Finland, both of which are described in greater detail below, create substantial exogenous variation in the size of punishment across drivers and within a driver across very similar driving speeds. Second, the Finnish administrative data allows me to follow individuals over time and link them to other registers. For example, the linking to tax data to observe individual incomes is crucial since part of the variation in fines arises from an individual's income.

2.1 The Relationship Between Speed, Income, and Speeding Tickets

Figure 1 gives an illustrative example of the speeding punishment schedule when the legally mandated speed is 100 km/h. Other possible speeding limits are 20, 30, 40, 50, 60, 70, 80, and 120 km/h, and the implications are similar. Given the 100 km/h limit and a 6 km grace speed, so long as an individual drives at or below 106 km/h, there is no ticket given.

Two factors determine the size of the speeding ticket in Finland. First, if a speeding violation is minor, i.e. only 7-20 km/h above the posted limit, then the individual receives a fixed fine equal to 140-200 euros.

Second, if an individual's speeding violation is more than 20 km/h over the limit, the so-called day-fine system kicks in, and speeding tickets become income-dependent. Under the income-based fine system, the size of the assessed fine is determined by the product of the offense severity and an individual's income. Depending on the severity of the excess speeding, the individual receives 1-120 fine units called day-fines. In cases of speeding, the higher the observed speed, the more severe the offense, and hence the higher the number of day-fines the individual receives. The following rule governs the monetary value of a single day-fine

$$I = \frac{Y - 255}{60} - 3 \times D, \tag{1}$$

where I is the monetary value of a single fine, Y is the net monthly income, and D is the number of dependents.[9]

Finland introduced the income-based fine system in 1921 for two reasons. The first motivation was equity. The government wanted to design a fine system that treated everyone similarly regardless of income. Second, the government was concerned that the high inflation at the beginning of the 20th century would erode the deterrence effect of the fine system. One way to overcome this problem was to link the value of fines to income. The name "day-fine system" originates from the feature that the value of a single day-fine was initially set to be equal to the amount of salary a worker would lose if he had to spend one day in prison instead of working. (HE, 1920)[10]

Figure 1(a) provides a theoretical example of how the size of the fine changes at the income-based fine cutoff for different income decile groups.[11] The fixed fine amount is identical across all income groups and applies for 6-20 km/h above the speed limit. Once the

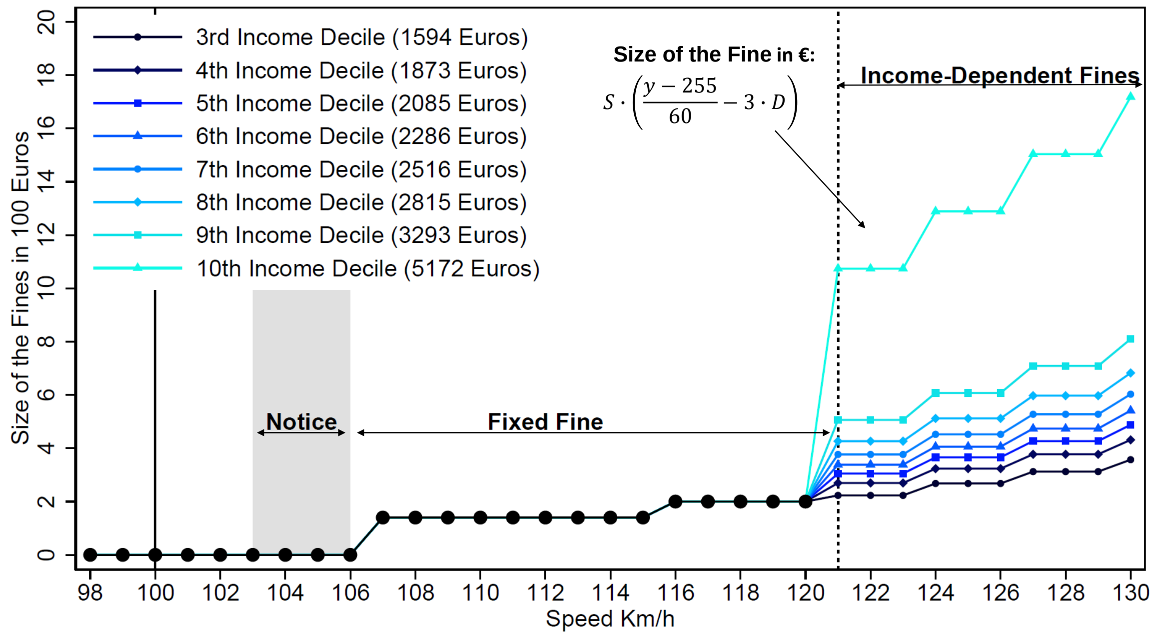
[9] Net income consists of the sum of all taxable earned income, capital income, employee benefits, and pensions, and most social benefits.

[10] The day-fine system is used in courts as well. In court cases, a judge decides the number of day-fines a defendant receives based on the severity of the crime. The equation 1 defines the size of each individual day-fine.

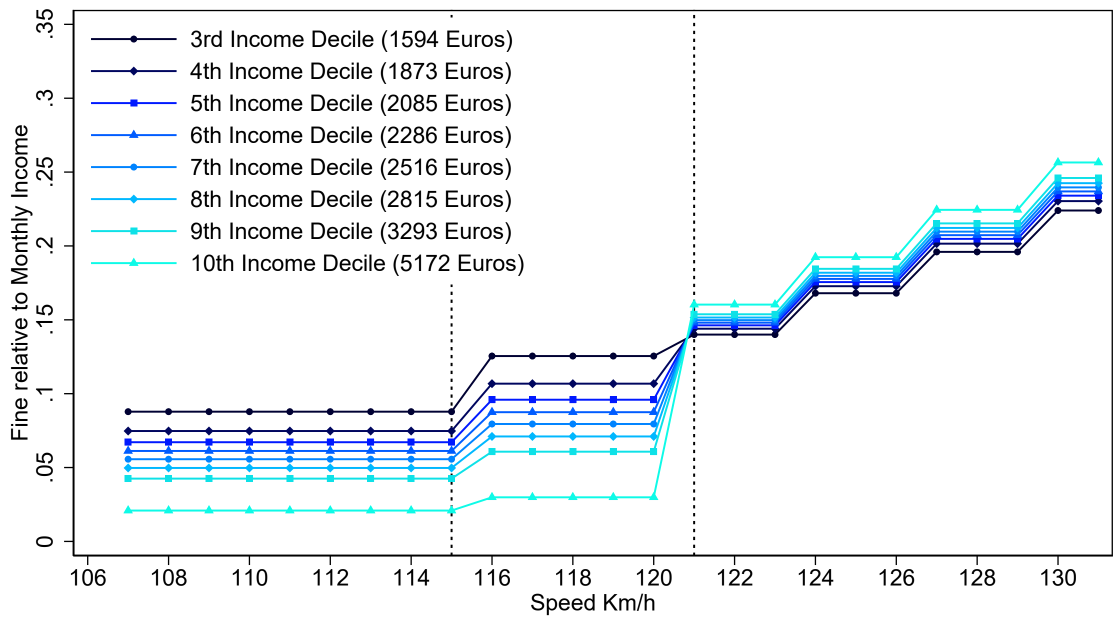
[11] The figure shows the fine schedule for deciles 3-10. The decile-specific income stated in the figure is the average net income over the years 2015-2017 calculated using microdata.

Figure 1: Relationship Between Income and the Size of Speeding Ticket

(a) Theoretical example when the speeding limit is 100 Km/H



(b) Size of the Fine wrt. Monthly Net Income



Notes: Figure Panel a shows the theoretical relationship between the size of the speeding ticket, speed, and net income for different income decile groups when the speeding limit is 100. Figure Panel b relates the speeding ticket's size to the decile's average net income. Average net incomes within deciles are calculated using microdata.

speed exceeds 20 km/h above the mandated limit, the income-based fine system begins and the ticket size can discontinuously increase enormously, depending on the driver's income. For an individual in the top income decile, the fine jumps by around 900 euros at the cutoff. In contrast, for an individual in the third income decile with a net monthly income of 1594 euros, the fine barely changes at the cutoff. Although Figure 1(a) provides an example using discrete income groups, in actuality the income-based fine system creates a continuum of discrete jumps in fines at the point when speed exceeds the limit by more than 20 km/h.

To illustrate the relative magnitude of fines across income groups, Figure 1(b) relates the size of the fine to an individual's net monthly income. Before the income-based fine cutoff, the relative magnitude of the fines differs a great deal between income groups. Specifically, with a flat fine imposing identical penalties for all individuals, those who are lower income pay the highest proportion of their incomes as the fine. However, after the income-based fine threshold, the fine equals around 15-20 percent of net monthly income for each income group. This is the sense in which an income-based fine system may be considered more equitable, as it imposes fines that constitute a similar proportion of incomes across different income groups.

Individuals just on the right-hand side of the income-based fine cutoff usually receive around 10-12 income-based fine units. After the cutoff, the number of income-based fine units increases in a stepwise manner. Appendix Table B.1 provides the exact punishment schedules that the police use in Finland to punish speeding.

Two additional aspects of the system are essential to this study. First, the income-based ticket cannot be lower than the fixed fine equal to 200 euros.[12] Second, the income-based speeding ticket schedule does not have an upper limit in euros. This has resulted in some extraordinarily large speeding tickets, with some of the largest tickets in the recent past equalling around 100,000-120,000 Euros. The current unofficial record holder is professional National Hockey League player Rasmus Ristolainen, who exceeded the 40 km/h speeding limit by 41 km/h and received a speeding ticket for 120,680 Euros in 2019.

2.2 Speeding surveillance

Speeding is monitored in two ways in Finland. First, the police use fixed surveillance stations or automatic traffic surveillance vehicles that measure the speeds of passing cars. If the surveillance station observes a speeding car, it takes a picture that the police use to assign a fine. Second, traditional police patrols monitor speeding with radar and laser speed guns. When a patrol observes a driver that exceeds the speeding limit, they may stop

[12]The fine stays constant at the cutoff for a person whose net income is around 1455 euros per month. If an individual earns less than this, the fine does not change at the cutoff.

the car and issue a notice or fine to the driver. I use only data created by the cameras, since their measurements are not affected by decisions such as whom they should stop.[13] Appendix Figure A.3 presents a picture of a typical police speeding surveillance camera.

After the surveillance machine has taken a picture, it is sent to the police Traffic Safety Centre, where an officer identifies the car owner based on the register number. Next, the police adjust the speed downwards due to possible measurement error and then decide, based on the adjusted speed, whether they assign a notice or fine to an individual.[14] All the numbers I report in this paper are adjusted speeds. Finally, if the police assign a fine or notice, the owner receives an electronic message and letter at home by post within 30 days from the date of the offense. The message notifies the individual of the fine and its size.[15]

The police calculate the amount of the income-based fines for each person who receives a ticket by using the individual's annual income from the most recent tax decision. If an individual's income has changed since the last tax decision, it may be corrected to reflect the current situation. I do not observe individuals' monthly earnings at the time of a speeding incident. Instead, I proxy an individual's monthly net earnings with yearly net earnings, which may create a modest measurement error.

Frequent speeding or extreme speeding can also lead to a temporary suspension of an individual's driver's license. Three speeding tickets within a year or at least four tickets within two years lead to a driving suspension lasting between 6-18 weeks. However, the suspension policy does not change at the income-based fine cutoff, which makes it easier to interpret the regression discontinuity design estimates I will estimate and describe.

2.3 Data

Data sources My primary data set comes from the National Police Board of Finland and includes all the speeding tickets that the police gave between April 2018 and May 2020.[16] The police data contains rich information about the speeding incident and the fine amount. I can observe the recorded speed, the speeding event's location, the prevailing speed limit, and whether an officer or an automatic camera system measured the speed. The data also includes the monetary value of a fixed fine, the monetary value of a single day-fine, and the number of day-fines an individual received for the speeding incident.

A key component of the police data is that each speeding ticket in the data includes a unique personal identifier, which I use to link the police speeding data to other admin-

[13] For example, Goncalves and Mello (2021) provides evidence that in Florida, U.S officers discriminate against minorities by not discounting their speed, which leads to larger fines for minority groups.

[14] The correction is 3 km/h if the speed is less than 100 km/h and 3% if the speed is larger than 100 km/h

[15] The Police always send the ticket to the owner of the car. The owner must request an administrative review if the owner was not the driver when the speeding incident occurred.

[16] Due to a reform, I do not observe fixed fines after May 2020.

istrative registers. First, I merge the speeding data into Statistics Finland's crime report register, which spans between 2006-2020. The crime report data contains the same speeding events as the police speeding data, but it does not include information about the exact speed. Thus I cannot use it to plot speeding distributions, but I can construct the recidivism outcomes using this data. This approach increases my sample size.[17]

Further, I merge the speeding data into Statistics Finland's FOLK data module, which contains a full population of Finnish residents between 1987-2019. From this data set, I observe individual's labor market outcomes such as labor earnings, income, and employment, and basic demographics like age, municipality of residence, marital status, and whether a household owns a car.

I also conduct analysis using traffic data collected by the Finnish Transport Agency using automatic traffic monitoring system (TMS) stations. There are around 500 TMS stations scattered around Finland. These TMS stations observe if a vehicle passes the station and records information such as time, speed of the vehicle, and vehicle class. Using the data collected by TMS, I can plot the speed distribution of all drivers, not just those caught speeding by cameras. Unfortunately, TMS data is anonymous. Hence, I cannot observe information on the identity of the driver.

Sample Restrictions The first restriction I impose is that I only focus on speeding incidents measured by automatic cameras. I ignore police patrol speeding data because prior evidence shows that police officers may manipulate the observed speed, complicating the interpretation of possible bunching (Goncalves and Mello, 2017). Second, I omit speeding tickets given in areas where the limit was 80 or 120 km/h. There are too few observations of speeding when the limit is 120. I observe some bunching when the limit is 80 km/h, but the bunching does not take place where it should. Finally, I only take individuals whose personal identifier numbers are available in the police data and who are at least 18 years old at the time of the speeding ticket.[18]

Outcome Variables In the second part of the paper, I study whether the size of the fine impacts recidivism. For this analysis, I use police crime report data to follow individuals for 12 months after the initial speeding ticket and calculate a cumulative reoffending probability for each month. I measure reoffending by constructing an indicator variable that takes the value one if an individual has committed a new traffic crime in the current month or any past months subsequent to the initial speeding ticket.[19] I am able to follow the same individuals for five months, but after 6 months my sample starts to decrease. However, in the robustness

[17] With this approach, I can use observations from the beginning of my sample period for which I would not observe any pre-outcomes without the crime report data

[18] Personal identifiers are missing for foreigners who do not live in Finland permanently.

[19] Appendix Figure A.17 shows that I obtain similar results if the outcome is just speeding.

section, I show that my results are insensitive to this limitation. I also use crime report data to build variables measuring non-traffic criminal activity. Like the primary outcome, the non-traffic crime indicator takes value one if an individual has committed a non-traffic crime 1-12 months after the initial speeding ticket.

Background Variables. Finnish register data let me observe a rich set of background variables such as an individual's gender, education level, age, and employment status. I use these background variables to check the validity of the setting, as control variables, and to perform heterogeneity analysis. The most important variable is the net yearly income, which I use to approximate an individual's net monthly income at the time of the speeding incident.[20]

Descriptive Statistics Table 1 provides descriptive statistics of the background variables. Column 1 reports the means of all background variables in the sample that I obtain after imposing the restrictions described above. Column 2 reports the same statistics but in the small window around the income-based fine cutoff.

To give an idea of how common speeding tickets are and who receives fines, Figure 2 plots a non-parametric relationship between crime and income among the Finnish adult population who live in a household with a car. The figure reveals two interesting patterns. First, speeding is a common crime. Around 15 percent of individuals in the median income percentile have received a speeding ticket over a two-and-a-half-year follow-up period. By contrast, when the suspected offense is not a traffic crime, only around 2 percent of the median-income individuals show up in the police crime report data. Second, the association between an individual's income and speeding tickets is very different compared to the relationship between income and non-traffic crimes. Higher-income individuals are much more likely to receive a speeding ticket than lower-income individuals. This fact suggests that speeding is likely a luxury crime. In contrast, there is a strong negative gradient between income and non-traffic crimes.[21]

In addition, Figure 2 decomposes the speeding tickets into fixed fines (small blue dots) and income-based fines (small grey dots). The decomposition shows that the income gradient is much steeper among fixed fines than income-based fines. The disparity in the speeding probability by income is consistent with the idea that if individuals are risk averse, fixed fines deter low-income people more than high-income people. Of course, other things may contribute to the gap. For example, wealthy individuals may drive more.

[20]The approximation, while imperfect, does a good job of recovering the true monthly income. For those who receive income-based fines, I can calculate their monthly net income from the income-based fine. Appendix Figure A.4 illustrates that the correlation between these two income measures is not perfect but still high (*corr=0.6)*

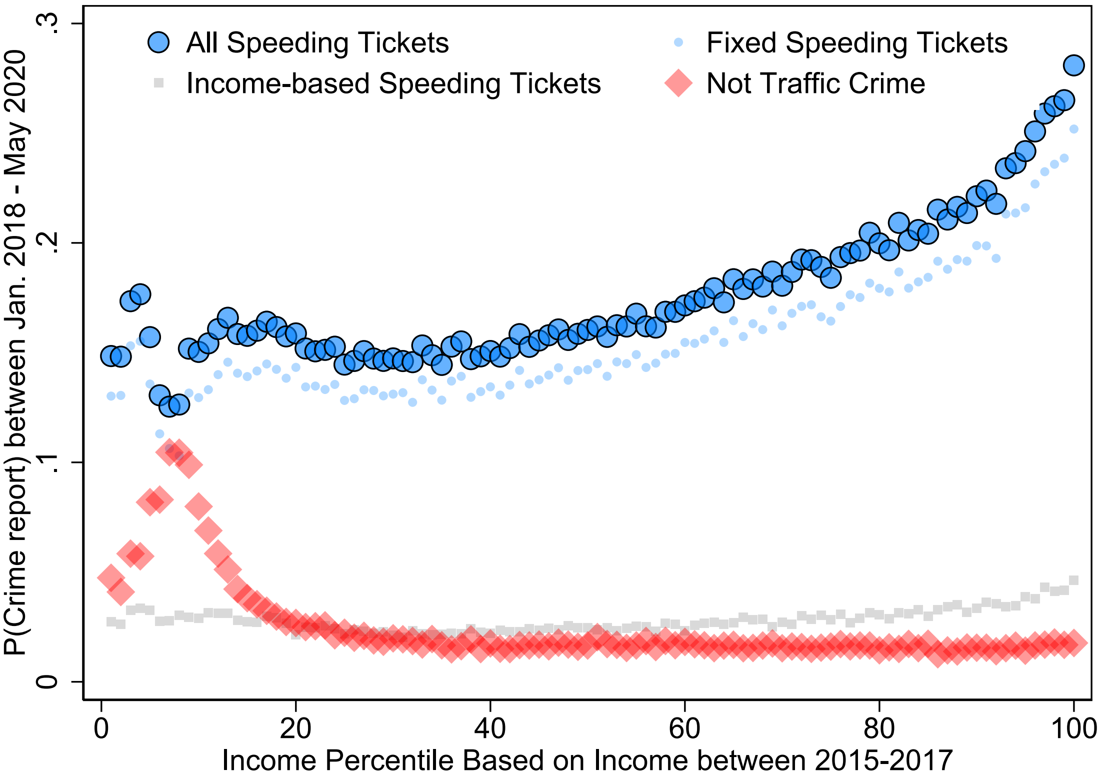
[21]Appendix Figures A.5 and A.7 show that the results are robust to multiple alternative ways to construct income percentiles.

Table 1: Descriptive Statistics and Balance Check

	All Mean (1)	**Window +-3** Mean (2)	**RDD** RDD (3)	**CCT RDD** RDD (4)
Traffic offence, months t-1 to t-6	0.122	0.134	0.014	0.018
			(0.011)	(0.013)
Non-Traffic offence, months t-1 to t-6	0.014	0.018	-0.002	-0.003
			(0.004)	(0.005)
Employed	0.706	0.708	0.002	0.004
			(0.013)	(0.016)
Unemployed	0.032	0.037	0.016	0.019
			(0.006)	(0.007)
Outside the Labor Force	0.262	0.255	-0.019	-0.024
			(0.012)	(0.015)
Monthly Net Income	2,589	2,572	-80.366	-79.738
			(41.980)	(52.735)
Primary Education Only	0.164	0.168	0.003	0.007
			(0.011)	(0.013)
Secondary Education	0.434	0.442	0.002	0.006
			(0.013)	(0.016)
Tertiary Education	0.402	0.391	-0.005	-0.008
			(0.014)	(0.018)
Female	0.331	0.298	-0.002	-0.001
			(0.011)	(0.014)
Age	47.296	45.926	-0.947	-0.895
			(0.432)	(0.536)
N of Children	0.834	0.847	0.062	0.081
			(0.037)	(0.046)
Finnish Speaking	0.886	0.884	0.020	0.023
			(0.010)	(0.012)
Married	0.501	0.481	0.000	0.006
			(0.014)	(0.017)
Urban Municipality	0.729	0.716	0.014	0.021
			(0.013)	(0.016)
Semi-urban Municipality	0.160	0.166	-0.015	-0.017
			(0.009)	(0.012)
Rural municipality	0.111	0.118	-0.002	-0.007
			(0.010)	(0.013)
Capital region	0.338	0.307	-0.003	0.002
			(0.012)	(0.016)
Observations	338,191	23,026		

Note: The table shows the means of predetermined characteristics among individuals in the speeding ticket sample and results from balance checks. Column (1) shows the means of background variables in the estimation sample. Column (2) presents the means of background variables in the estimation sample after restricting individuals within +-3 km/h from the income-based cutoff. Column (3) shows results from the balance check where the dependent variable is a predetermined characteristic. The column reports the estimates of β obtained using the equation 2. Column (4) reports results from the balance check conducted using the approach of Calonico *et al.* (2014a). I cluster standard errors at the individual level. See Section 2.3 for more details on sample construction.

Figure 2: The Relationship between Income, Probability of Speeding, and Crime Report



Notes: Figure presents the relationship between income percentile and the probability of observing the individual in the Police crime report statistics between January 2018 - May 2020. I construct the figure using every 18-65 year old in 2015 in Statistics Finland's total population labor market and demographics data. I calculate the income percentile by comparing individuals' mean income over 2015-2017 to those in the same birth cohort. Finally, I restrict to individuals who live in a household with a car. Blue dots with circles show the share of individuals receiving a speeding ticket in each percentile between January 2018 and May 2020. Red dots without the circle show the share of individuals in Police crime report data when the suspected crime is not a traffic offense between January 2018 and May 2020 for each percentile. Small dots decompose the speeding ticket averages into shares that arise from fixed fines (blue dots) and income-based fines (grey dots). Appendix Figure A.6 shows the average earnings within each percentile. Appendix Figures A.5 and A.7 show that the figure is robust in multiple ways.

3 Do People Bunch at the Income-based Fine Cutoffs?

This section investigates whether people react to income-based fine cutoffs by bunching. A simple Becker (1968) model of criminal behavior, in which individuals compare gains from illegal speeding to expected punishment, predicts that individuals should react to income-based fine cutoff by bunching. In the model, the cutoff makes speeds above it discontinuously more expensive, implying that some drivers should respond by slowing down and bunching below the threshold.[22] If this prediction holds, we should see excess mass below the income-based fine cutoff in the speeding distribution.

A common finding in other contexts is that people react to discontinuous changes in incentives by bunching. For example, Kleven and Waseem (2013) document sharp excess mass below tax notches in Pakistan, Einav *et al.* (2019) present a substantial bunching response in drug consumption to a discontinuous change in marginal price, and Best *et al.* (2019) shows large bunching below mortgage interest notches. Given this evidence, the prediction that individuals bunch below speeding ticket notches seems more than plausible.

However, this section documents that this intuitive bunching prediction does not hold up in the data in my setting. The speeding distributions in both the traffic and police camera data exhibit smooth distributions, with no excess mass below the thresholds where fines change discontinuously. This finding contradicts the existing theory of criminal behavior.

To demonstrate this counter-intuitive result, I first plot the speed distributions in the Finnish Transport Infrastructure Agency's (FTIA) traffic data. The automatic traffic monitoring stations scattered around Finland generate the data. These stations measure the speed of each passing car. The data allows me to see the total speed distribution, not only the speed distributions conditional on a driver going faster than the posted speed limit.

Speeding Distributions in Traffic Data Figure 3 shows the first key result of the paper. The figure plots total speed distributions in the FTIA data. Figure Panel 3(a) plots the raw distributions in different speed limit zones. Figure Panel 3(b) pools together all speed limit zones and normalizes the x-axis variable as a distance from the income-based fine cutoff. The figures reveal two interesting patterns. First, each speed distribution peaks around the speeding limit, suggesting that, on average, speed limits do strongly shape drivers' behavior.

Second, even though the speed distributions peak at the posted speed limits, Figure 3(b) shows that speed distributions exhibit no bunching around the points where the police's tolerance level (-14) ends, fixed fines increase (-5), or income-based fines turn on (0). The fact that we do not see any mass just below these fine thresholds suggests that drivers do

[22]Section 5 presents this result using a formal model.

not react to fine discontinuities as the simple Becker model would predict.

However, it is possible that high-income individuals react to fines, but the behavior of drivers whose incentives change very little at the cutoff masks the bunching by high-income individuals. Unfortunately, the FTIA data does not contain information on the driver. To overcome this shortcoming, next I move to use the police speeding data, which allows me to observe each driver's income.

Speeding Distributions in Police Data Figure 4 plots the speeding distributions using the police speeding camera data.[23] Figure 4(a) shows the distributions at different speed limit zones. The figure illustrates that speed distributions evolve strikingly smoothly. In other words, drivers do not bunch below the speeding ticket hikes highlighted by vertical lines, despite significant incentives to do. In Figure 4(b), I pool all the speeding distributions together and normalize the x-axis variable to measure the distance from the income-based fine cutoff. This approach makes the lack of excess mass below the cutoff even more evident.

Appendix Figure A.8 replicates Figure 4(b) but separately for the income quintile groups, which I define using predetermined income amounts. The figure shows a very similar pattern as in Figure 4(b). Even the top income quintile's speeding distribution evolves smoothly at the cutoff, although they have the largest incentives to bunch below the threshold.

Figures 4(a) and 4(c) show that the only place where we observe excess mass is at speed 97 when the speed limit is 80 km/h. However, this spike is unlikely to arise from the driver's behavioral response. The mass point is just after a fixed fine hike, implying that people do not bunch to avoid larger fines.

I cannot study whether individuals know the locations of police surveillance stations and slow down just before cameras to avoid speeding tickets. If this is the case, the speeding ticket sample contains individuals who either do not care about the cameras or do not know their location. However, the fact that the total speed distributions shown by Figure 3 behave continuously everywhere suggests that the smoothness of the speeding distributions is not just an artifact of the police camera data.

Starting from the next section, the rest of the paper tries to understand why people ignore the cutoff.

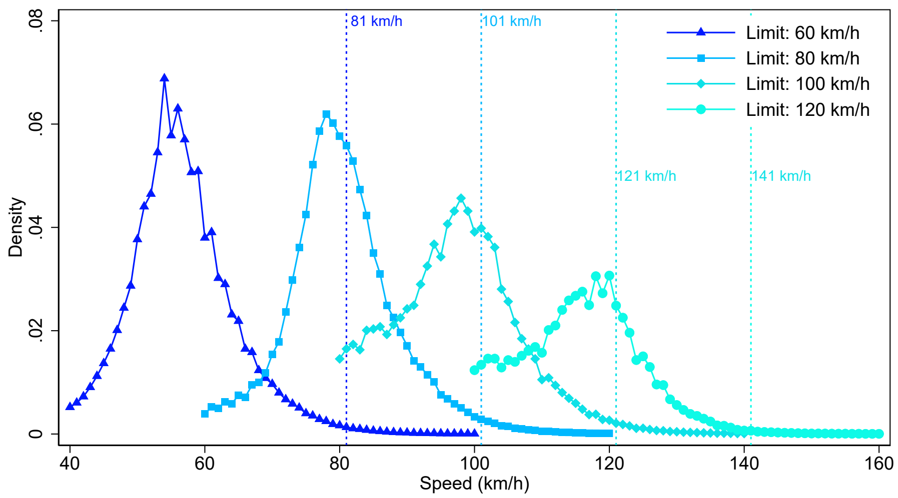
4 Does A Larger Speeding Ticket Influence Re-offending?

The previous section documents that drivers do not behave as a simple Becker model predicts. Drivers ignore the discontinuous changes in the price of speeding at the income-based

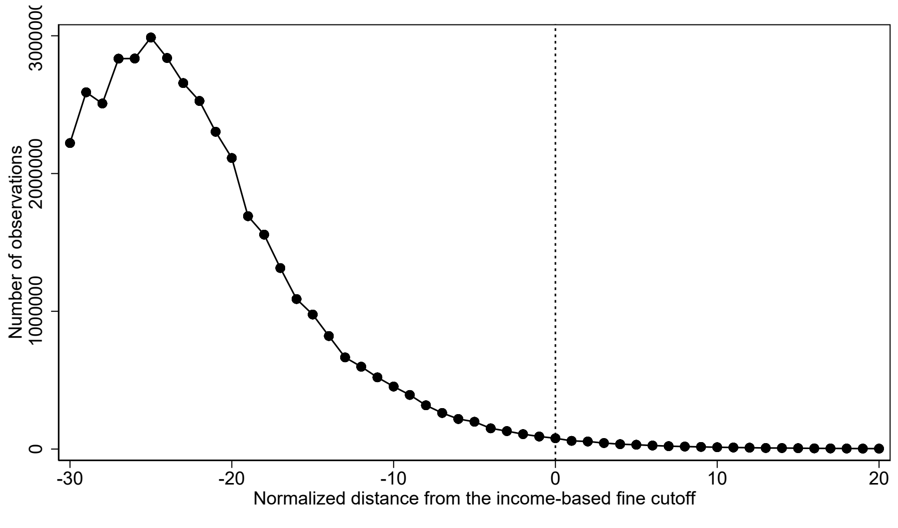
[23]The police assign a fine only if the driver's speed crosses the police's tolerance level. This means that the speed ticket data reveals the speed distribution's right tail.

Figure 3: The Speed Distributions in Traffic Data

(a) Raw Speed Distributions



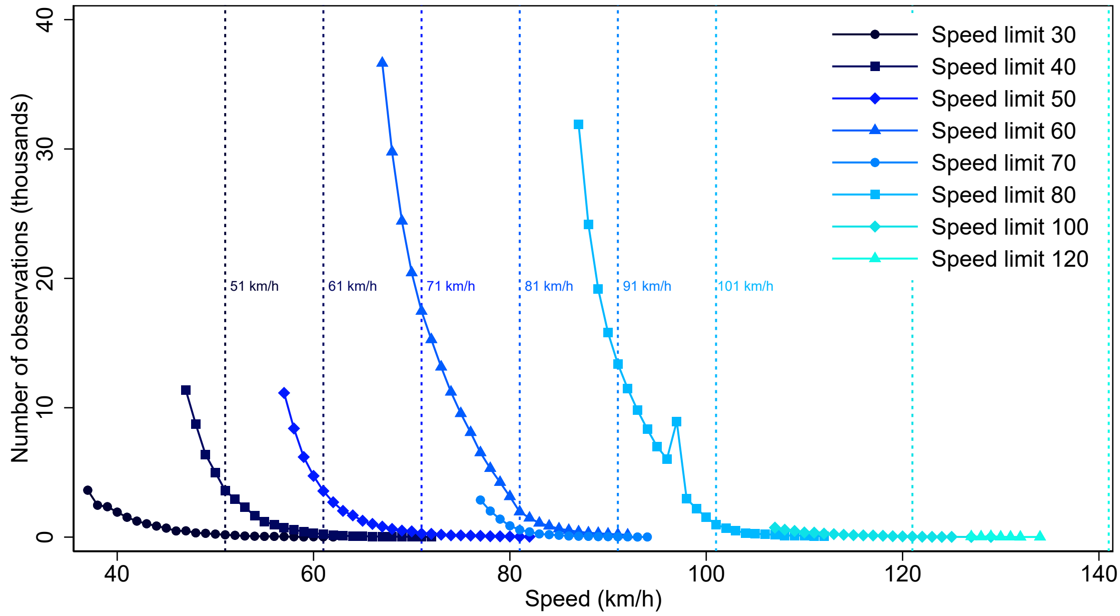
(b) Pooled Speed Distribution



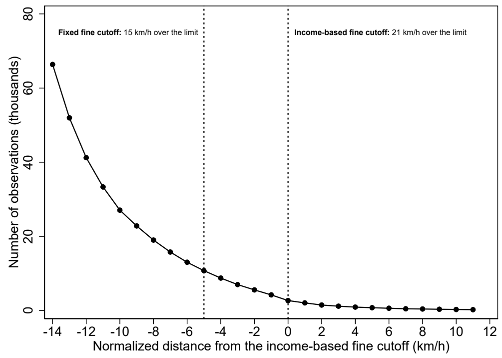
Notes: The figure presents the speed ticket distribution in the data created by the Finnish Transport Agency's automatic traffic monitoring stations. Figure panel (a) shows the speed distributions separately in different speed limit zones. Figure panel (b) pools all speed distributions together and normalizes the x-axis variable as a distance from the income-based fine cutoff.

Figure 4: The Speeding Distributions in Police Camera Data

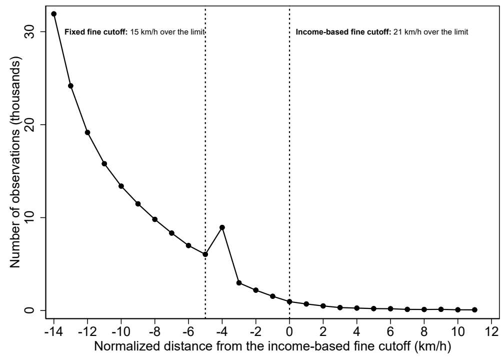
(a) Speeding Distributions



(b) Normalized Distribution (limit 80 km/h omitted)



(c) Normalized Distribution (limit 80 km/h included)



Notes: The figure presents the speeding ticket distributions in the sample containing speeding tickets assigned by cameras between Jan 2018 - May 2020. Figure Panel (a) shows the distributions in different speeding limit areas. The x-axis refers to the speed in km/h, and the y-axis to the number of speeding tickets per km/h. Vertical dashed lines indicate the points where the average speeding ticket jumps due to the income-based fine system. Figure Panel (b) pools all the speeding limit zones except 80 km/h together and normalizes the x-axis to measure the distance from the income-based fine cutoff. The first vertical line highlights the point at which the fine jumps a fixed amount. The second vertical line highlights the point at which the income-based fine system kicks in. Figure Panel (c) conducts the same exercise as panel (b) but uses only the limit zone of 80 km/h.

fine cutoff *ex-ante* and do not bunch as this theory would predict.

There are two possible explanations for the lack of bunching that I explore in this section. First, drivers may fail to bunch because of information frictions. In other words, people do not understand how the system works, leading to suboptimal decisions. Second, people may operate under perfect information but ignore the cutoff because of adjustment costs. An individual's optimization problem contains multiple parameters, such as the measurement error of a car's speedometer, the downward correction made by the police, and the exact location of the income-based cutoff, which varies with the speeding limit in absolute terms. Due to the complexity of the problem, individuals may feel that gains from bunching are smaller than the adjustment costs, implying that they neglect the cutoff.

Although both information frictions and adjustment costs can generate smooth speeding distributions, they make very different predictions on how realized speeding tickets should affect reoffending. If information frictions characterize an individual's behavior, then assigned fines should reveal information about the punishment system and affect recidivism. However, under adjustment costs, individuals should not react to fines since they already operate with perfect information.

This section asks whether being assigned a larger versus a smaller speeding ticket changes the driver's behavior *ex-post*. To answer this question, I use a regression discontinuity design to compare individuals just below and above the cutoff. The smoothness suggests that drivers do not manipulate whether or not they receive a large fine.

How individuals react to the size of the fine helps us understand why speeding distributions are smooth and how drivers generally make speeding decisions. Adjustment costs will likely explain the lack of bunching if the assigned speeding ticket size does not influence recidivism. However, if larger speeding tickets reduce reoffending, then information frictions and learning characterize the individual's behavior. In Section 5, I will present a more formal theoretical framework that rationalizes all of the empirical findings presented in this and the previous section.

4.1 Empirical Specification

I investigate the impact of larger realized fines on recidivism using a sharp regression discontinuity design (RDD). My RDD equation takes the form

$$Y_{il,t} = \beta_t Z_{il,0} + f(S_{il,0}) + f(S_{il,0}) \times Z_{il,0} + \alpha_{l,0} + \epsilon_{il,t}, \tag{2}$$

where the dependent variable $Y_{il,t}$ measures the cumulative recidivism t months after the initial speeding incident. The indicator variable $Z_{il,0}$ is equal to 1 if an individual i crossed the income-based fine cutoff in the speeding limit region l in period zero (normalized to be the time of the speeding incident). The running variable $S_{il,0}$ controls the distance from the fine cutoff. The interaction term $f(S_{il,0}) \times Z_{il,0}$ allows the relationship between the outcome and the running variable to change at the cutoff. Further, I control for speeding limit fixed effects $\alpha_{l,0}$. The error terms ($\epsilon_{il,0}$) are clustered at the individual level. I use triangular weights centered at income-based fine cutoffs, and select optimal bandwidth around the cutoff using methods of Calonico *et al.* (2014b).

The coefficient of interest, β_t, captures the effect of larger fines on recidivism. Interpreting this estimate as causal requires that individuals cannot manipulate the running variable precisely. If this identifying assumption holds, individuals at the cutoff are, on average, similar but receive speeding tickets of very different magnitudes.

Treatment Figure 5(a) clarifies the treatment for those who cross the income-based fine cutoff. Black connected dots in the figure reveal that those on the right-hand side of the cutoff (treatment group) receive a fine that is, on average, 200 euros larger than those on the left-hand side of the cutoff (control group). However, the size of the fine varies considerably with an individual's income (See, Appendix Figure A.20).

Validity of the Setting One of the greatest strengths of RDD is that its key assumption provides testable predictions that we can use to evaluate the validity of the setting. The first prediction is that if individuals do not have perfect control of the running variable, then the distribution of the running variable (in this case speed) should be continuous. Figure 5(a) provides evidence that this first prediction holds in my setting. Based on the figure, the number of observations decreases rapidly with speed, but there is no evidence of excess mass below the cutoff. In other words, we do not find evidence that individuals manipulate the speed perfectly.

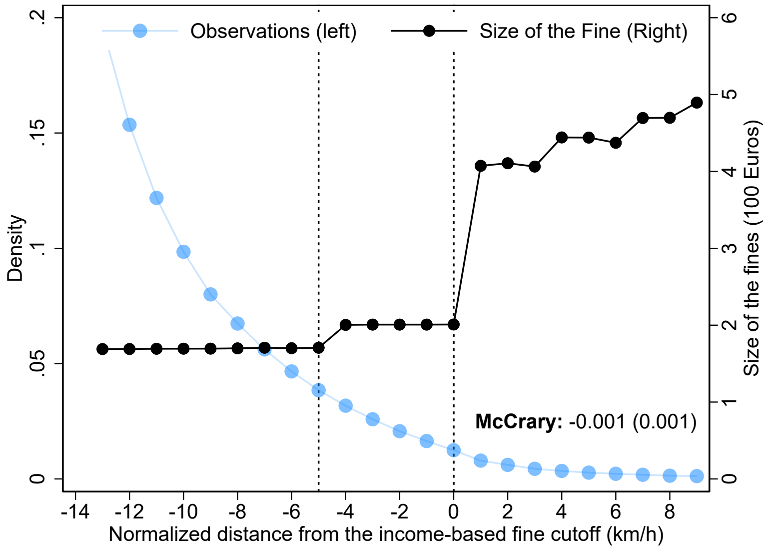
The second prediction is that if individuals do not manipulate the running variable perfectly, then drivers at the cutoff will be locally randomized to control and treatment groups, implying that people just below and above the cutoff should be similar on average. To determine whether pre-determined characteristics evolve smoothly at the cutoff, I investigate whether a propensity score measuring the probability of reoffending jumps at the cutoff.[24] The propensity score is an index that summarizes all background variables succinctly into one variable. I construct the propensity score in the following way. First, I regress a dummy variable for reoffending on a set of pre-determined variables, excluding speed. Then I use regression coefficients to calculate each individual's predicted value of reoffending.

Figure 5(b) illustrates that the propensity score behaves smoothly at the cutoff, provid-

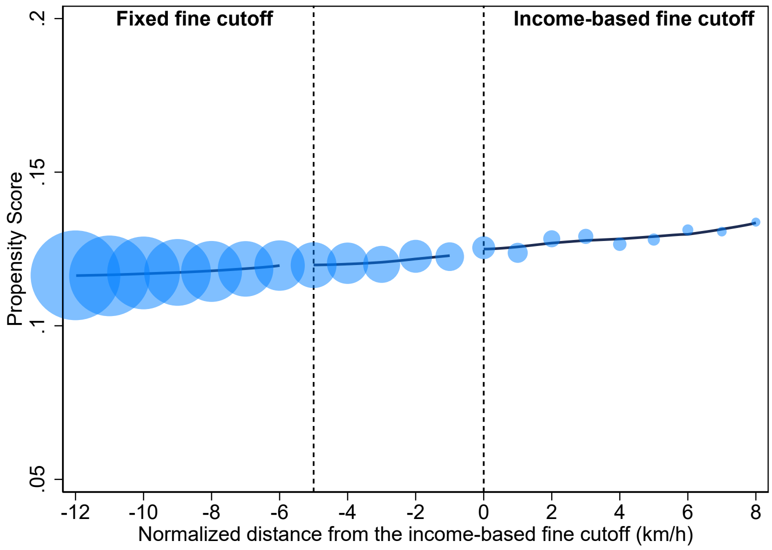
[24] This approach has been applied, for example, by Rose and Shem-Tov (2021)and Londoño-Vélez *et al.* (2020).

Figure 5: Graphical evidence

(a) Speeding Distribution and Size of the Fine

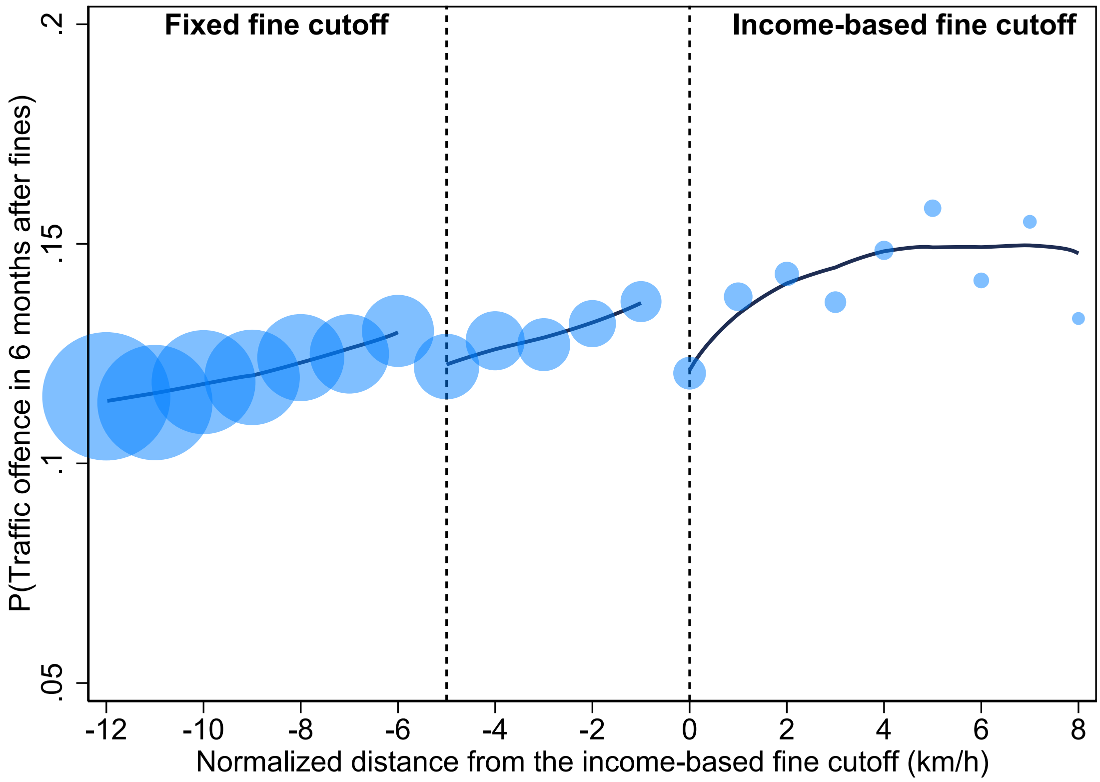


(b) Covariate Balance



Notes: The figure presents how speeding distribution (a), size of the fine (a), and predetermined background characteristics (b) evolve at the income-based fine cutoff. The x-axis measures the distance from the income-based fine cutoff in km/h. In Figure Panel (a), the left-hand side y-axis refers to the density per km/h, and the right-hand side y-axis measures the average size of the speeding ticket. Figure Panel (b) plots a summary index of predetermined covariates against normalized speed. The index is equal to the predicted values that I obtain by regressing the indicator variable, which takes value one if an individual commits another traffic offense within six months after the speeding tickets against covariates listed in table 1. The sample constructed as the section 2.3 describes.

Figure 6: Graphical Evidence: Recidivism Within 6 Months



Notes: The figure presents a recidivism outcome as a function of normalized speed. The x-axis measures the distance from the income-based fine cutoff in km/h. The y-axis refers to the probability of committing another traffic offense within six months after the initial speeding ticket. Vertical lines highlight the fine discontinuity points. The income-based (fixed) fine cutoff locates at 0 (-5). The sample is constructed as the section 2.3 describes.

ing another piece of evidence supporting the key assumption of RDD. In addition, I study how individual variables behave at the cutoff. Columns (3) and (4) of Table 1 show that most of the pre-determined background variables behave continuously at the income-based fine cutoff. Out of 18 variables tested, only two are statistically significant at a 5 percent level, which is slightly more than one would expect to show up due to randomness. However, these significant variables do not seem to be jointly related to the propensity to commit traffic crimes, since the propensity score for reoffending evolves so smoothly in Figure 5(b). Furthermore, in the robustness section 4.5, I show that the main estimates barely move when I add these background characteristics as controls in the estimation.

4.2 The Impact of Speeding Ticket Size on Recidivism

Descriptive Evidence Figure 6 presents the first evidence of how people respond to larger fines that arise from the income-based fine system. The x-axis of the figure measures the distance from the income-based cutoff in km/h. The dots plot the probability of committing another traffic offense within six months after a speeding ticket. We observe sharp drops in the probability of reoffending at the cutoffs where the fines jump discontinuously. The drop seems large at the income-based fine cutoff, but we also see some action at the fixed fine cutoff. Before and after the fine cutoffs, the probability of reoffending evolves smoothly. Under the key assumption I stated above, which seems plausible given the evidence, the discontinuous jump in the fines causes the observed drops in the probability of reoffending. Next, I use an RDD to quantify the size of the drop in reoffending at the income-based fine cutoff.

RDD Estimates Figure 7 presents the second key result of the paper. The figure shows how the larger speeding tickets impact the probability of committing another traffic offense in the future. In Figure 7(a), I graph the RDD estimates, i.e. the β_ts from equation 2. These estimates capture the impact of a larger speeding ticket on the cumulative reoffending probability. My main outcomes (red dots) track individuals' cumulative reoffending probabilities 1-12 months after a fine.[25]

Figure 7(a) shows that the short-term cumulative reoffending probability drops when a driver receives a larger fine due to the day fine (red dots). The impact peaks around 5-8 months after the speeding incident. At this time, drivers who received a larger fine are around 2.5 percentage points less likely to commit another traffic offense than individuals who received a fixed fine. For a table format of the results, see Appendix table B.5.

I also provide suggestive evidence that drivers do not react to larger realized fines by bunching below the income-based fine cutoff. Appendix figure A.9 plots the speeding distributions for individuals reoffending within a year. The figure shows that individuals who received an income-based fine and reoffended within a year do not bunch below the cutoff. These results suggest that a larger speeding ticket triggers only an extensive margin response.

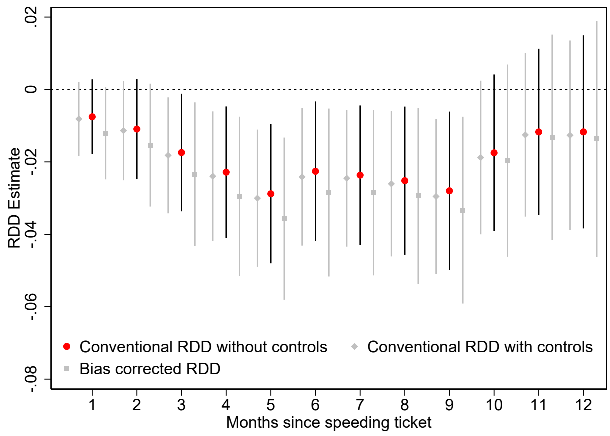
To understand the relative magnitudes of these estimates, Figure 7(b) plots cumulative reoffending probabilities in the control group, and Figure 7(c) relates the point estimates shown in Figure 7(a) to the reoffending probabilities in the control group. Reoffending is common: of those who are just below the income-based fine cutoff, more than 20 percent commit another traffic offense within a year. However, larger income-based fines decrease recidivism in the short term. Figure 7(c) shows that those who receive an income-based fine are around 20 percent less likely to commit another traffic offense in the following six months than individuals who receive a smaller fixed fine.

Lastly, Figure 7 illustrates that the decrease in speeding as a result of a larger speeding ticket starts to fade out around nine months after the initial ticket was assigned. Although

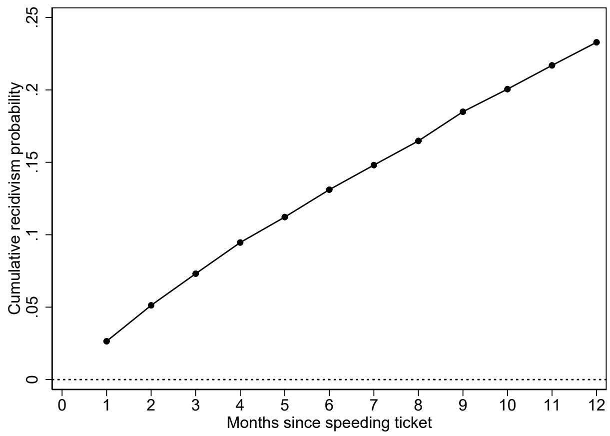
[25] I can follow the same sample 6 six months, after which the sample size starts to decrease. I show that my results are robust to this limitation in 4.5.

Figure 7: RDD Estimates at the Day Fine cutoff - 80 km/h limit omitted

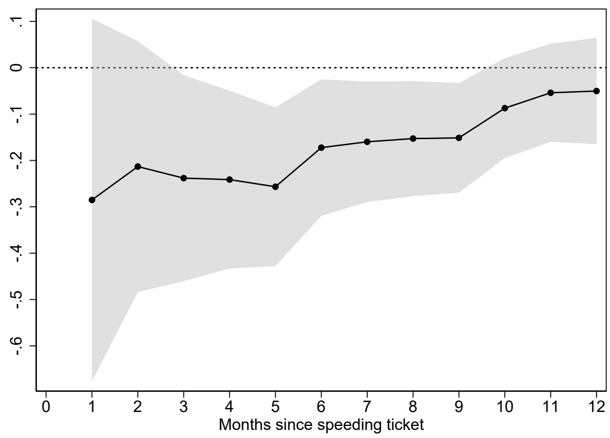
(a) RDD Point Estimates



(b) The Mean of Outcome in Control Group



(c) The Relative Magnitude of Point Estimates



Notes: Figure panel (a) plots the RDD estimates of $\beta_{il,t}$ obtained using Equation (2). Estimates plot the impact of the higher fine on the cumulative probability of committing a traffic offense in t months after the initial fine. Dots plot the estimates obtained using conventional local linear regression with a triangular kernel. Diamonds plot the estimates with controls. Squares report biased corrected RDD estimates obtained using the robust approach by Calonico *et al.* (2014b). Vertical solid lines are the 95 percent confidence intervals. Standard errors are clustered at the individual level. For bias-corrected estimates, the confidence intervals are given by the approach of Calonico *et al.* (2014b). Panel (b) shows the mean of the outcome variable in the control group, and panel (c) relates the point estimate shown by the diamonds in panel (a) to reoffending probabilities in the control group. Sample construction and data as defined in section 2.3.

the point estimate is still negative twelve months after the initial speeding ticket, it is not statistically significant from zero at a 95% confidence level small. This result suggests that if the target of larger fines is to achieve a permanent reduction in speeding, then fines must be assigned frequently to those who speed.

4.3 Does the Effect Vary by Income?

Next, I show that high-income individuals who cross the income-based fine cutoff and receive a larger fine react more strongly than low-income individuals who receive a smaller fine. Further, I show that these high- and low-income individuals react similarly to an equal-sized jump in the fine. These results suggest that the larger the jump in the fine, the more extensive the learning experience.

I start by splitting my sample into quartiles using individuals' annual net income one year before the ticketing year. Then I estimate the effect separately in each quartile group using equation 2. Figure 8(a) illustrates how the size of the fine develops as a function of speed for different income quartile groups. We see that different income quartiles experience very dissimilar speed ticket hikes at the income-based fine cutoff. The fine increases by around 500 euros for the top quartile, whereas the hike equals around 50 euros for the bottom quartile. For the bottom group (top group), the fixed fine equals around 20 (4) percent of net monthly income. In relative terms, the income-based fine is similar for each income group.

Figure 8(b) then plots RDD estimates for different income quartiles when the outcome is cumulative reoffending probability six months after the initial speeding ticket. Blue diamonds show the RDD estimates for different income quartiles at the income-based fine cutoff. Although the estimates are relatively imprecise, they suggest that high-income individuals react very strongly to their fines. The point estimates are close to zero for two bottom income quartiles, after which the estimates start to increase almost linearly. For the top income group, the RDD estimate (-0.045) is double the size of the main estimate that I presented in the previous section.

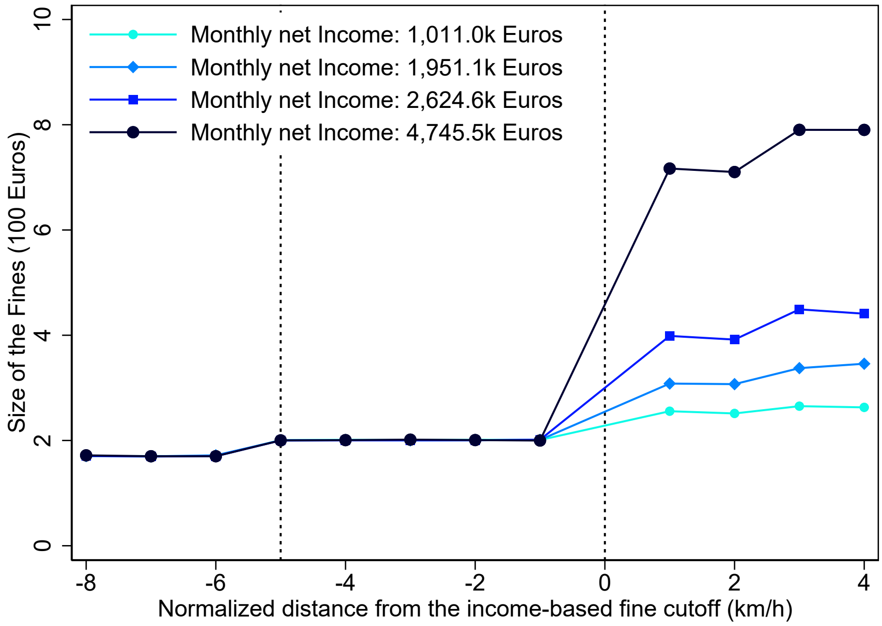
The black dots in the figure 8(b) present similar RDD estimates as above but use variation from the fixed fine cutoff where an individual's income does not affect the size of the fine.[26] Even though the estimates are imprecise, the magnitudes are smaller and more comparable between groups.

Together, these estimates suggest that individuals who receive larger fines are less likely to recidivate than individuals who receive small fines. Because all income groups react

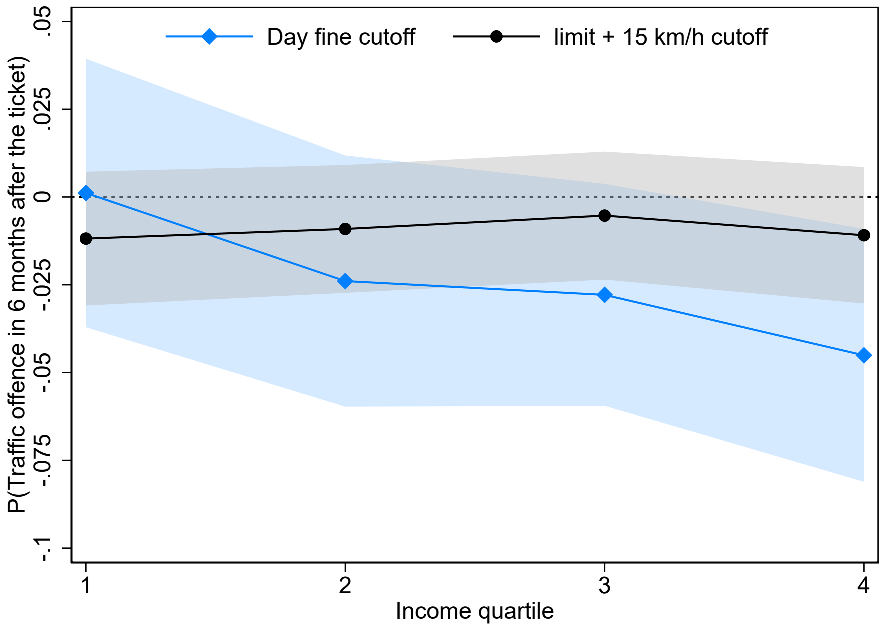
[26] Appendix Figure A.14 shows the impact of fixed fines on reoffending in the full sample. In the full sample, the fixed fine has a small and imprecise effect on reoffending.

Figure 8: Impact of Fines and Income Quintile

(a) Average Size of Fine Income Income Groups



(b) RDD Estimates by Income Groups



Notes: Figure Panel (a) shows a jump in the fine's size at the income-based fine cutoff for different income quartiles. The income quartiles are constructed using an individual's annual net income measured 1-2 years before the speeding incident. The Legend in the left upper corner shows the approximated monthly income obtained by dividing the individual's annual net income by twelve. The x-axis measures the distance from the income-based fine cutoff. Figure Panel (b) reports RDD estimates for different income quartiles when the outcome is the probability that an individual commits another traffic crime within six months after the original speeding incident. Blue diamonds (black dots) show the RDD results at the income-based (fixed fine) cutoff. The estimates are obtained using equation 2. The shaded bands show 95 percent confidence intervals. Standard errors are clustered at the individual level. Sample construction defined in 2.3.

similarly to fixed fine hikes, the sizable reaction to income-based fines by high-income people is due to larger fines and not because they are different on some unobserved dimension that impacts speeding behavior and is correlated with income.

4.4 A Dose-Response Parameter

Finally, I use the variation in the size of the fine stemming from income to estimate the "so-called" dose-response parameter. This parameter reveals the impact of a marginal change in the "dose" of fines on recidivism. In other words, the parameter answers the question of how recidivism changes when the realized fine increases by one euro.

The dose parameter is important because of two reasons. First, the dose parameter may be easier to apply for policy purposes. Continuous concepts, such as price and income elasticities, often define optimal policies like marginal tax rates in economic models (Hendren, 2016). Ideally, these continuous parameters should be estimated with a continuous variation (Callaway *et al.*, 2021). My main RDD estimates are identified from large fine jumps and may not be applicable in a situation where a policymaker wants to know how marginal change in realized fines affects recidivism. Second, the dose-response parameter result clarifies what we can learn from the RDD estimates presented in Figure 8(b).

I demonstrate in the Appendix Section D that a comparison of quartile-specific RDD estimates identifies the "so-called" dose-response parameter under the assumption that individuals with different incomes can respond differently to similar fines, but this difference in responses can not vary with the level of fine. If this assumption holds, then the difference in RDD estimates identifies the dose-response parameter that reveals how marginal changes in fines affect recidivism.

In the ideal case, I would split individuals into small groups and compare groups who receive marginally different fines due to differences in incomes. However, this is clearly not feasible given the imprecision of estimates I present in Figure 8(b).

Therefore, I estimate the dose-response parameter with the linear specification. The regression equation takes the form

$$Y_{il} = \alpha_l + \gamma S_{il} + \theta S_{il} \times Z_{il} + \beta Z_{il} + \delta_1 F_{il} + \delta_2 F_{il} \times Z_{ij} + \boldsymbol{X}'\psi + \epsilon_{il} \tag{3}$$

where Y_{il} is the outcome variable, which takes value one if an individual has reoffended within t months after the original speeding ticket, S_{il} is the running variable, and Z_{il} is the binary variable indicating whether the individual crossed the income-based fine cutoff. Variable F_{il} measures the predicted fines based on an individual's income, and $\boldsymbol{X}$ is a vector that may contain controls and other interactions. Under the assumption that individuals at the margin of receiving similar doses may react differently to similar fines, but this difference does not vary with the level of the fine, β captures how a marginal change in the expected fine affects recidivism. For more details, see Appendix Section D.

Table 2 panel B presents dose-response parameter estimates. To make the numbers more interpretable, I have multiplied all the estimates by 100. Hence, the dose-response parameter reveals how reoffending changes when fines increase by 100 euros. The second row of Table 2 reports the dose-response parameter with different specifications. I find that a 100 euro increase in fines decreases reoffending by 0.7 percentage points. The result is robust to adding controls and additional interaction terms.

Of course, it is a strong assumption that individuals with different incomes can react differently to similar fines but the difference does not change with the level of fine. However, Table panel 2 A adds confidence that this may hold in my setting. Table panel 2 A reports results from a similar analysis but uses variation from the fixed fine cutoff. Interestingly, when the change in the size of the fine is not linked to income, the relationship between recidivism and the predicted fine disappears. This finding suggests that the connection between predicted fines and recidivism, present in panel B, is mainly explained by larger fines and not by higher income.

4.5 Robustness and Additional Validity Tests

This section shows that my results do not arise from arbitrary RDD specification choices. First, one concern is that the selected specification approximates the unknown conditional expectation function poorly. To alleviate these concerns, Appendix Table B.9 shows that I obtain similar results when I use a specification with a second-order polynomial or conduct the analysis using the robust approach proposed by Calonico *et al.* (2014b). Further, Appendix Figure A.18 shows "honest" Armstrong and Kolesár (2020) confidence intervals that incorporate the potential bias in the estimates.

Second, Appendix Figure A.11 demonstrates that the point estimates are insensitive to the bandwidth choice. Although I tie my hands by selecting the optimal estimation window using the methods of Calonico *et al.* (2014b), the figure A.11 adds additional confidence that an arbitrary choice of window does not drive the results.

I also carry out a set of additional validity checks. First, if the size of the fine changes exogenously at the cutoff, RDD estimates should not change when we add pre-determined characteristics as control variables to the specification. Figure 7 and Appendix Table B.9 illustrate that estimates are insensitive to controls. Second, the RDD's key identification assumption implies that the conditional expectation function of potential outcomes should

Table 2: The Dose-response Parameter

Dependent variable: P(Another traffic offence in 6 months)	(1)	(2)	(3)	(4)
Panel A: Fixed Fine Cutoff				
Predicted Ticket	0.0026	0.0067	0.0021	0.0061
	(0.0026)	(0.0023)	(0.0013)	(0.0025)
Predicted Ticket × Z	-0.0001	0.0001	-0.0001	0.0002
	(0.0010)	(0.0010)	(0.0019)	(0.0019)
Controls		✓		✓
Running variable × Z			✓	✓
Bandwidth	4	4	4	4
Observations	99,195	99,195	99,195	99,195
Panel B: Income-based Cutoff				
Predicted Ticket	0.0007	0.0076	-0.0016	0.0055
	(0.0007)	(0.0041)	(0.0021)	(0.0045)
Predicted Ticket × Z	-0.0066	-0.0055	-0.0059	-0.0054
	(0.0017)	(0.0017)	(0.0033)	(0.0032)
Controls		✓		✓
Running variable × Z			✓	✓
Bandwidth	4	4	4	4
Observations	31,912	31,912	31,912	31,912

Note: The table report results dose-report parameters with different specification. The results are obtained using equation 3. Panel A (B) reports the fixed (income-based) fine cutoff results. The interaction term predicted income and Z identifies the average causal response parameter. In equation 3, this corresponds to δ_2. All the estimates and standard errors are multiplied by 100. Thus the dose-parameter reveals how reoffending changes when the fine increases by 100 euros. Standard errors are clustered at the individual level. Section 2.3 explains the sample construction.

evolve smoothly. If this is the case, we should not see discontinuities in the outcomes at the places where fines evolve smoothly. Appendix Figure A.12 shows results from a placebo test where I randomly reallocate the cutoff to places where fines do not jump and carry out RDD analysis.[27] The placebo estimates (black diamonds) are small and not statistically significant.

I also conduct an analysis where my outcome is the cumulative probability of committing a non-traffic crime. Since my main sample contains individuals who rarely commit non-traffic crimes, it is unlikely that the size of the speeding ticket would impact non-traffic crimes. Indeed, I find that the RDD point estimates are very small and not statistically significant when my outcome is the cumulative probability of non-traffic criminality. Appendix Figure A.15 reports these results.

5 What Explains the Smooth Distributions and Reactions to a Larger Fine?

In the previous sections, I presented four new empirical findings. First, I showed that there is no excess mass just below the fine discontinuity. This is surprising as it goes against a traditional Becker model of criminal sanctions, and most normal models of behavior. For example, in the tax literature, bunching is often observed just below the threshold where marginal taxes increase discontinuously (Kleven, 2016).

Second, I showed that when drivers receive a large speeding ticket after just crossing the threshold where speeding tickets can increase dramatically, the sudden increase in the assigned fine decreases recidivism. This reaction ex-post occurs despite the lack of bunching around the cutoff a priori. Third, the larger the realized fine, the larger the ex-post reaction by drivers. Specifically, higher-income drivers who experience the largest jumps in assessed fines experience the largest drop in recidivism. Fourth, the effect of reoffending fades out over time. By ten months after the higher fine amount, there is no longer a statistically significant reduction in reoffending.

In this section, I present a simple theoretical framework that seeks to rationalize these puzzling results, building on previous models in this and other related literature (Traxler *et al.*, 2018). My conceptual framework incorporates two types of optimization frictions that could reconcile the main results. These frictions have been explored in related literatures (Chetty *et al.*, 2011; Chetty, 2012; Kostøl and Myhre, 2021) to explain attenuated responses to incentive changes. First, I demonstrate that adjustment costs may attenuate drivers' response to fine hikes and result in little or no bunching below the cutoff, which indicates

[27] This is similar to the permutation test proposed by Ganong and Jäger (2018)

a discontinuous increase in the assessed speeding fine. If adjustment costs are high enough, the distribution may be completely smooth. However, under a model with adjustment costs, individuals should not react to realized fines. The intuition for this result is that drivers already know precisely how the system works, but because it is too costly to monitor speed perfectly, they choose not to do so. As a result, a realized fine does not bring any new information that could change his/her behavior next period.

Second, I present a model with information frictions that can generate the observed patterns in the data. The model builds on the idea that individuals use the so-called "ironing heuristic" and approximate marginal prices with average prices if the cognitive cost of perceiving complex pricing schemes is considerable (Rees-Jones and Taubinsky, 2019). In the model, drivers replace a complex true penalty scheme with a rule of thumb that the magnitude of the fine increases linearly with speed. Because of this heuristic, drivers ignore the jump in the marginal price of speeding at the cutoff, leading to smooth speed distribution, as I observe in the data. However, when drivers are hit with an assigned fine, then they may change their behavior in the future. If an individual receives a larger fine than expected based on the linear rule, she concludes that the relationship between the size of the ticket and speed must be steeper than she expected. She will use this steeper penalty function in the next period and drive more slowly. Finally, based on the idea of Nyarko (1991), I show that there are drivers whose beliefs about the slope of the penalty function will never converge. This may explain the fade-out in effect.

5.1 A Stylized Model of Speeding with Adjustment Costs

Individuals choose how fast over the speed limit to go $x \in [0, X]$. They value speed via function $u(x, \theta) = \theta x - x^2$ that is strictly concave in x and increasing in type θ. Individuals' taste for speed have continuous and smooth distribution $G(\theta)$ with density $g(\theta)$. With probability p, the driver is caught speeding and receives a fine, which is a step function where

$$f(x) = \begin{cases} 0 \text{ if } x = 0 \\ f^l \text{ if } 0 < x \leq x^h \\ f^h \text{ if } x > x^h. \end{cases}$$

Finally, optimization frictions ψ make adjusting speed costly.[28] Since my empirical part focuses on the intensive margin variation, I also ignore the extensive margin in the model.

[28]As demonstrated by Chetty (2012), Kleven and Waseem (2013) and Kostøl and Myhre (2021) optimization frictions may attenuate individuals' reactions to discontinues changes in marginal or average prices. In my case, the adjustment costs may arise, for example, from the fact that if individuals can not drive at the speed defined by their type, they have to monitor their speed constantly, which decreases utility.

Drivers' optimization problem takes the following form

$$\max_x EU = u(x,\theta) - p(f^l + f^h \cdot \mathbb{1}[x > x^h]) - \psi \mathbb{1}[x \neq x^*].$$

The last term of the problem implies that individuals pay an adjustment cost ψ if they do not locate in their interior optimum x^* given by the first-order condition.

Frictionless Model To illustrate the impact of fines and optimization frictions on optimal speed, let us first consider the problem without them. First, without fines, drivers locate at their interior optimum

$$x^* \quad \text{such that} \quad \partial u(x,\theta)/\partial x = 0.$$

Thus, speeding distribution $x^*(\theta)$ will be smooth due to smooth type distribution $G(\theta)$.

Next, let us introduce speeding tickets. When fines are imposed, a continuum of individuals who value speeding enough such that their interior optimum x^* is higher than the fine notch just above x^h, will slow down and bunch to the corner solution where speed is

$$x^c = x^h \quad \text{such that} \quad \partial u(x,\theta)/\partial x \neq 0$$

Because of bunching, we should see excess mass at x^h. However, some drivers do not react: individuals whose interior optimum speed is lower than the speed limit or whose taste for speed is high enough remain in their interior optimum.

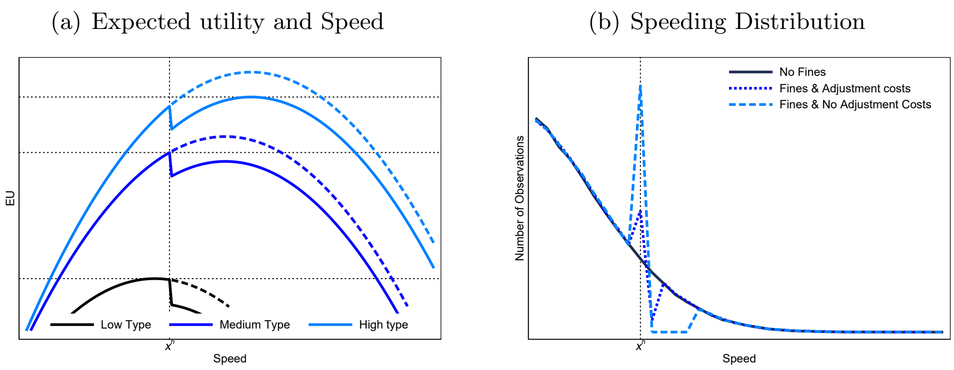
Adjustment Costs Finally, adjustment costs may attenuate the driver's response to fines and decrease the mass at the corner solution. Under adjustment costs, the driving speed is given by

$$x = \begin{cases} x^c, & \text{if } [u(x^c,\theta) - p \cdot f^l] - [u(x^*,\theta) - p \cdot (f^l + f^h)] \geq \psi \\ x^*, & \text{if } [u(x^c,\theta) - p \cdot f^l] - [u(x^*,\theta) - p \cdot (f^l + f^h)] < \psi. \end{cases} \tag{4}$$

The equation 4 demonstrates that only individuals whose gain from reallocating from the interior solution to the corner solution is larger than the adjustment cost change their behavior. In other words, fewer individuals react to fines when adjustment costs are present, implying that we should observe less mass at the fine cutoff compared to a frictionless world.

Figure 9 illustrates the results stated above. Figure 9(a) shows the expected utility function for three different types. The dashed (solid) line shows the utility from speed without (with) fines. The medium-type individual may increase utility by moving at corner solutions x^c to 80 km/h. However, low- and high-type individuals do not have an incentive to react to speeding tickets. Low-type's interior optimum is less than the speed limit,

Figure 9: Expected Utility, Speed, and Speeding Distribution



whereas the expected punishment does not deter the high-type enough.

Figure 9(b) demonstrates the impact of speeding tickets on speed distribution. Solid blue lines show the distribution without tickets. Black dots illustrates that the notch in expected punishment moves some people to the corner solution, creating excess mass just below the speeding limit and a hole above it. Finally, the dashed blue line plots the distribution when adjustment costs impact drivers' behavior. The mass concentration below the cutoff is smaller under adjustment cost.

The model predicts that we should see sharp bunching below when frictions do not exist. However, adjustment costs attenuate individuals' responses and bunching. In extreme cases, the speed distribution may be smooth if individuals consider the expected punishment small compared to adjustment costs.

5.2 A Model with Misspecification and Learning

Next, I present an alternative version of the model in which individuals make speed decisions using a misspecified model and learn from their past choices. Extensive empirical literature shows that individuals do not necessarily adjust their behavior to marginal price changes as a simple economic theory would suggest. One suggested explanation for the muted responses is that people use the "ironing" heuristic when facing complex non-linear price schedules. If people find the mental cost of understanding a non-linear price scheme to be high, they replace the complex version with a linear approximation. In the case of a speeding ticket, the "ironing" heuristic could arise from the fact that people do not remember the exact details of the system, or they are just concentrating on many things while driving. (Rees-Jones and Taubinsky, 2019)

Motivated by literature suggesting that people respond to average prices, I assume that people replace the complex non-linear penalty scheme with linear approximation. They start with a prior about the slope of the penalty function that they use to solve for optimal speed. Since the misspecified penalty function is linear, they cannot end up with a corner solution, ruling out bunching. The optimal speed determines the signal the individuals receive. After observing the signal, individuals construct a posterior, which works as the next period's prior. Individuals learn myopically, meaning that they do not experiment.

The optimization problem with misspecification and learning takes the form

$$\max_{x_t} EU_t = u(x_t, \theta) - p \cdot \hat{f}_t(x),$$

where $\hat{f}_t(x) = \beta x_t + f^l \mathbb{1}[x > 0]$ stands for individuals' perceived penalty function at time t. The slope β is an unknown random variable and according to individual's prior beliefs $\beta \sim \mathcal{N}[\mu_\beta, \tau_\beta^{-1}]$. The known constant equals to low fine f^l.

Each period, the individual chooses a speed such that $\partial u_t(x)/\partial x - p \cdot \partial \hat{f}_t(x)/\partial x = 0$. If they are caught speeding, they receive a fine that they interpret as a signal. The signal takes form $s_t = \beta_t^f + \eta_t, \eta \sim \mathcal{N}[\mu_s, \tau_s^{-1}]$. If speed is larger than x^h, individual receives a large fine (β_t^h). Otherwise, β_t^f is β_t^l. After seeing the signal, the individual forms a posterior for β using the Bayes rule. Since prior and signal are normal random variables, the posterior is a weighted average of prior and signal: $\hat{\beta} = \frac{\tau_\beta \mu_\beta + \tau_s s}{\tau_\beta + \tau_s}$ (Baley and Veldkamp, 2021). The relative precision of the prior and signal determine the weights.

The model with misspecification generates the following results related closely to my empirical exercise.

Proposition 1. *Drivers do not bunch below the fine cutoff at time $t = 1$, generating a smooth speeding distribution.*

For a proof, see Appendix Section C.

The proposition follows directly from the fact that drivers ignore the discontinuity because they use a misspecified model. In other words, all the drivers locate at their interior optimum where $\partial u(x, n)/\partial x = p \cdot \hat{\beta}$, leading to zero excess mass at the fine cutoff.

Proposition 2. *Under the assumptions of the model, a driver with type θ would drive slower at time time $= 2$ if she were located just above the cutoff x^h versus below it at time $t = 1$. Furthermore, the size of the reaction increases with the size of discontinuity.*

For a proof, see Appendix Section C.

This result means that if a person with type θ drives faster than the cutoff x^h, she will receive a higher signal compared to a situation in which her speed is lower than x^h. As a result, the perceived slope of the penalty function is steeper in the next period if the individual crosses the threshold. Thus, in the next period, she will slow down. The larger the discontinuity, the steeper the updated slope, and the larger the individual's reaction.

Proposition 3. *There exists a continuum of individuals* $[\theta^l, \theta^h]$ *whose beliefs and actions never converge but oscillate around the speeding ticket discontinuity.*

For a proof, see Appendix Section C.

Figure 10 illustrates some of the results stated above. Consider an individual at the time t who holds a prior represented by a dark blue line in Figure 10(a). Using this prior, the individual chooses a speed such that $\partial u(x,n)/\partial x = p \cdot \hat{\beta}$. Assume that that speed is x^*, and hence, she receives a low signal. Let us also suppose that the prior and signal are equally precise, implying that the posterior is just average between signal and prior. The dashed light blue line represents the posterior 10(a).

Next, consider another individual who happens to choose a speed equal to $80+\epsilon$ where ϵ is some tiny number. Due to this minimal difference, the second individual receives a larger signal, shown by a solid red line. The dashed red line plots the posterior for the second individual. Next, individuals choose new speeds using posteriors as their priors. The individual who received the higher signal due to discontinuity will drive at a slower speed than the individual who received the lower signal.

Interestingly, if individuals make decisions using a misspecified model and learn from their actions, there will be some individuals whose beliefs and actions do not converge but follow cycles (Nyarko, 1991). Figure 10(b) and 10(c) illustrate the result and show how the individual's beliefs and speed evolve over time. The individual starts with the belief that the slope of the penalty is low, chooses a speed over 80 km/h, receives a large fine, and forms a posterior higher than the prior. The individual chooses a lower speed and receives a lower signal in the next period. Over time, individual beliefs decrease back to a low starting level. However, when the individual again thinks the slope is low, she decides to speed again, leading to a new cycle.

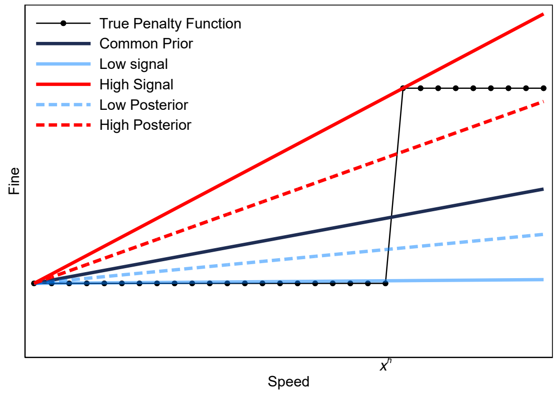
5.3 Predictions of the Models

Table 3 summarizes the predictions from both models. Both models can generate the zero-bunching result. However, only the model with misspecification and learning predicts that a larger speeding ticket (vs. smaller) reduces recidivism. Furthermore, under the misspecification and learning, we may see a fade-out in effect due to cycling beliefs.

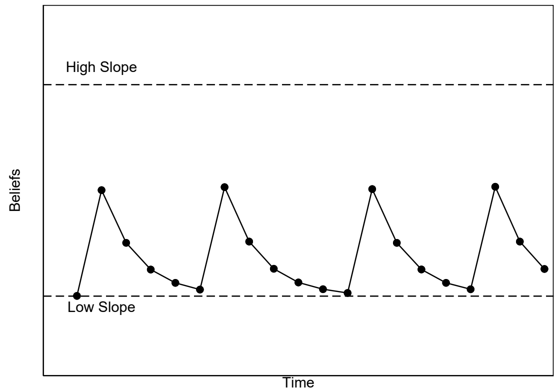
The evidence presented in the section 4.2 supports the story that drivers make decisions using a misspecified model, leading to a smooth speeding distribution. In addition to continuous speeding distributions, I find that size of the speeding ticket reduces the probability of reoffending, but the effect fades out over time. The model with misspecification and learning can explain both of these results.

Figure 10: The Evolution of Speed and Beliefs in the Model with Misspecification and Learning

(a) Prior, Signal, and Posterior



(b) The Evolution of Beliefs



(c) The Evolution of Speed

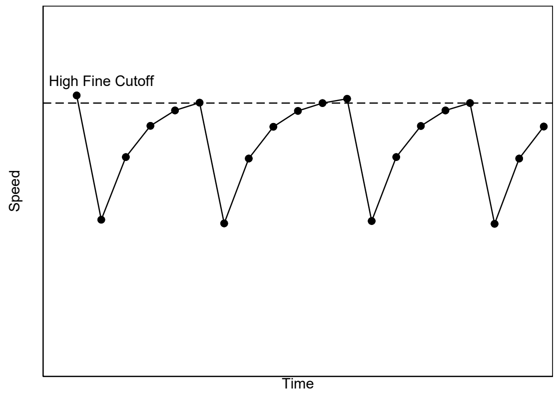


Table 3: Predictions from the Models

	Model with:	
Prediction	**Adjustment Costs**	**Misspecification**
Bunching	Possibly	No
Reactions to larger fine	No	Yes
Cycle	No	Possibly

The table lists the predictions from the models presented in the Section 5.

6 Conclusions

This paper studied drivers' reactions to discontinuous changes in the criminal sanctions created by the Finnish income-based fine system. First, I find that drivers do not slow and bunch below the speed at which the fine jumps because of the income-based fine system, leading to unexpectedly smooth speeding distributions. This result also holds for high-income drivers experiencing the largest jumps in the size of the fine. However, I show that drivers are less likely to recidivate in the short term when they are assigned a larger ticket vs. a smaller one. But the effect fades out after one year.

To reconcile these results, which are not consistent with traditional theory on the impacts of sanctions in the crime literature, I presented a new theoretical framework. I show that, theoretically, both adjustment costs and misspecification may explain the smooth speeding distributions. However, the reactions I document empirically to the size of the assigned fine and the fade out in effect are only consistent with the model that individuals make decisions and learn with a misspecified model.

My results have important implications for policies on criminal sanctions. First, the results suggest that punishments may only have intended consequences if the punishment schedule is salient enough. Second, individuals may have to be reminded frequently to have persistent effects.

References

Aizer, A. and Doyle, J., Joseph J. (2015). Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges *. *The Quarterly Journal of Economics*, **130** (2), 759–803.

Armstrong, T. B. and Kolesár, M. (2020). Simple and honest confidence intervals in nonparametric regression. *Quantitative Economics*, **11** (1), 1–39.

Arora, A. (2019). Juvenile Crime and Anticipated Punishment. *Working Paper*.

Baley, I. and Veldkamp, L. (2021). *Bayesian Learning*. Working Paper 29338, National Bureau of Economic Research.

Banerjee, A., Duflo, E., Keniston, D. and Singh, N. (2019). *The Efficient Deployment of Police Resources: Theory and New Evidence from a Randomized Drunk Driving Crackdown in India*. Working Paper 26224, National Bureau of Economic Research.

Bastani, S. and Selin, H. (2014). Bunching and non-bunching at kink points of the swedish tax schedule. *Journal of Public Economics*, **109**, 36–49.

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

Bell, B., Jaitman, L. and Machin, S. (2014). Crime deterrence: Evidence from the london 2011 riots. *The Economic Journal*, **124** (576), 480–506.

Berk, R. H. (1966). Limiting Behavior of Posterior Distributions when the Model is Incorrect. *The Annals of Mathematical Statistics*, **37** (1), 51 – 58.

Best, M. C., Cloyne, J. S., Ilzetzki, E. and Kleven, H. J. (2019). Estimating the Elasticity of Intertemporal Substitution Using Mortgage Notches. *The Review of Economic Studies*, **87** (2), 656–690.

Bhargava, S., Loewenstein, G. and Sydnor, J. (2017). Choose to Lose: Health Plan Choices from a Menu with Dominated Option*. *The Quarterly Journal of Economics*, **132** (3), 1319–1372.

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

Bohren, J. A. and Hauser, D. N. (2021). Learning with heterogeneous misspecified models: Characterization and robustness. *Econometrica*, **89** (6), 3025–3077.

Callaway, B., Goodman-Bacon, A. and Sant'Anna, P. H. C. (2021). Difference-in-differences with a continuous treatment.

Calonico, S., Cattaneo, M. D. and Titiunik, R. (2014a). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, **82** (6), 2295–2326.

—, — and — (2014b). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, **82** (6), 2295–2326.

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

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, **80** (3), 969–1018.

—, Friedman, J. N., Olsen, T. and Pistaferri, L. (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records *. *The Quarterly Journal of Economics*, **126** (2), 749–804.

—, — and Saez, E. (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the eitc on earnings. *American Economic Review*, **103** (7), 2683–2721.

Doleac, J. (2017). *Encouraging Desistance from Crime*. Tech. rep.

Drago, F., Galbiati, R. and Vertova, P. (2009). The deterrent effects of prison: Evidence from a natural experiment. *Journal of Political Economy*, **117** (2), 257–280.

Dušek, L. and Traxler, C. (2021). Learning from Law Enforcement. *Journal of the European Economic Association*, jvab037.

Einav, L., Finkelstein, A. and Schrimpf, P. (2019). Reprint of: Bunching at the kink: Implications for spending responses to health insurance contracts. *Journal of Public Economics*, **171**, 117–130, trans-Atlantic Public Economics Seminar 2016.

Esponda, I. and Pouzo, D. (2016). Berk–nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica*, **84** (3), 1093–1130.

Ganong, P. and Jäger, S. (2018). A permutation test for the regression kink design. *Journal of the American Statistical Association*, **113** (522), 494–504.

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. and Mello, S. (2017). *Does the Punishment Fit the Crime? Speeding Fines and Recidivism.* Tech. rep.

— and — (2021). A few bad apples? racial bias in policing. *American Economic Review*, **111** (5), 1406–41.

Hansen, B. (2015). Punishment and deterrence: Evidence from drunk driving. *American Economic Review*, **105** (4), 1581–1617.

HE (1920). Hallituksen esitys 36/1920.

Heidhues, P., Kőszegi, B. and Strack, P. (2018). Unrealistic expectations and misguided learning. *Econometrica*, **86** (4), 1159–1214.

Helland, E. and Tabarrok, A. (2007). Does three strikes deter? a nonparametric estimation. *The Journal of Human Resources*, **42** (2), 309–330.

Hendren, N. (2016). The policy elasticity. *Tax Policy and the Economy*, **30**.

Ito, K. (2014). Do consumers respond to marginal or average price? evidence from nonlinear electricity pricing. *American Economic Review*, **104** (2), 537–63.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, **8** (1), 435–464.

— and Waseem, M. (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan *. *The Quarterly Journal of Economics*, **128** (2), 669–723.

Kostøl, A. R. and Myhre, A. S. (2021). Labor supply responses to learning the tax and benefit schedule. *American Economic Review*, **111** (11), 3733–66.

Lee, D. and McCrary, J. (2017). The deterrence effect of prison: Dynamic theory and evidence. *Advances in Econometrics*, **38**, 73–146.

Londoño-Vélez, J., Rodríguez, C. and Sánchez, F. (2020). Upstream and downstream impacts of college merit-based financial aid for low-income students: Ser pilo paga in colombia. *American Economic Journal: Economic Policy*, **12** (2), 193–227.

Mueller-Smith, M. (2014). The Criminal and Labor Market Impacts of Incarceration. *Working Paper*.

Nyarko, Y. (1991). Learning in mis-specified models and the possibility of cycles. *Journal of Economic Theory*, **55** (2), 416–427.

Philippe, A. (2022). Learning by doing. How do criminals learn about criminal law? *Working Paper*.

Rees-Jones, A. and Taubinsky, D. (2019). Measuring "Schmeduling". *The Review of Economic Studies*, **87** (5), 2399–2438.

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

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, **2** (3), 180–212.

Traxler, C., Westermaier, F. G. and Wohlschlegel, A. (2018). Bunching on the autobahn? speeding responses to a 'notched' penalty scheme. *Journal of Public Economics*, **157**, 78–94.

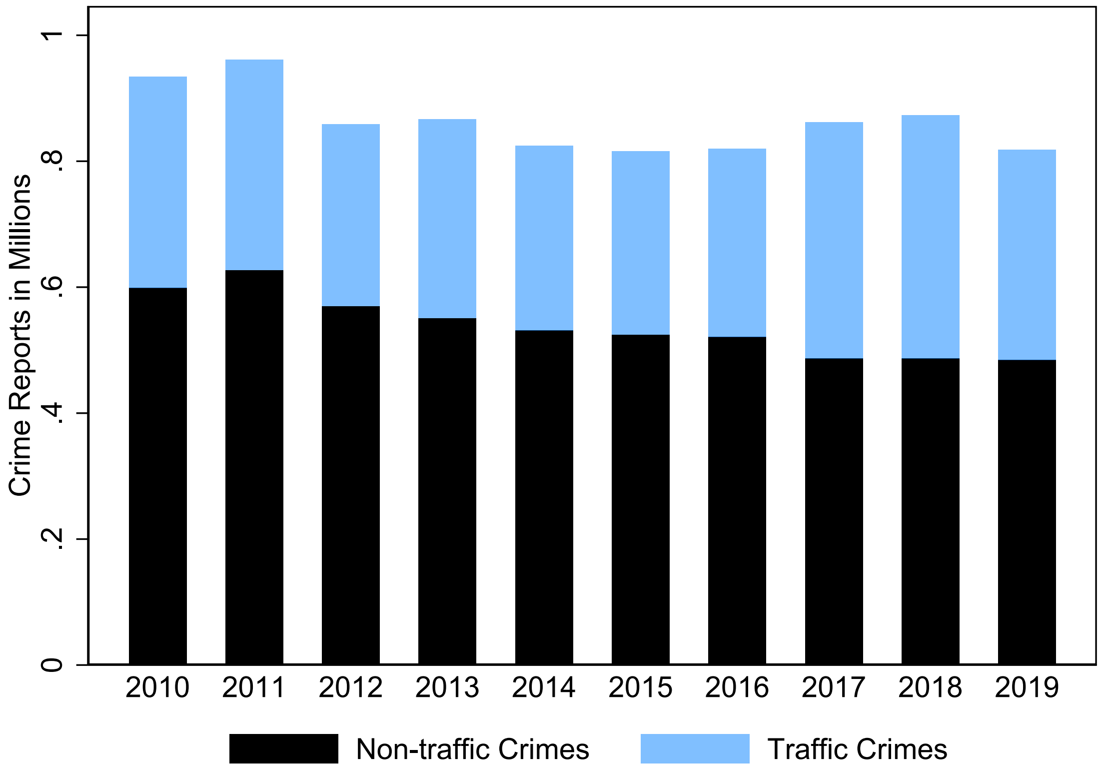
WHO (2017). *Managing Speed*. Report, World Health Organization.

— (2018). *Global status report on road safety 2018*. Report, World Health Organization.

Appendix

A Figures

Figure A.1: The Number of Crime Reports in Finland



Note: Figure presents the number of crime reports received by Finnish Police. Source: Statistics Finland

Figure A.2: The Most Common Causes of deaths for 15 to 49 year olds globally in 2019

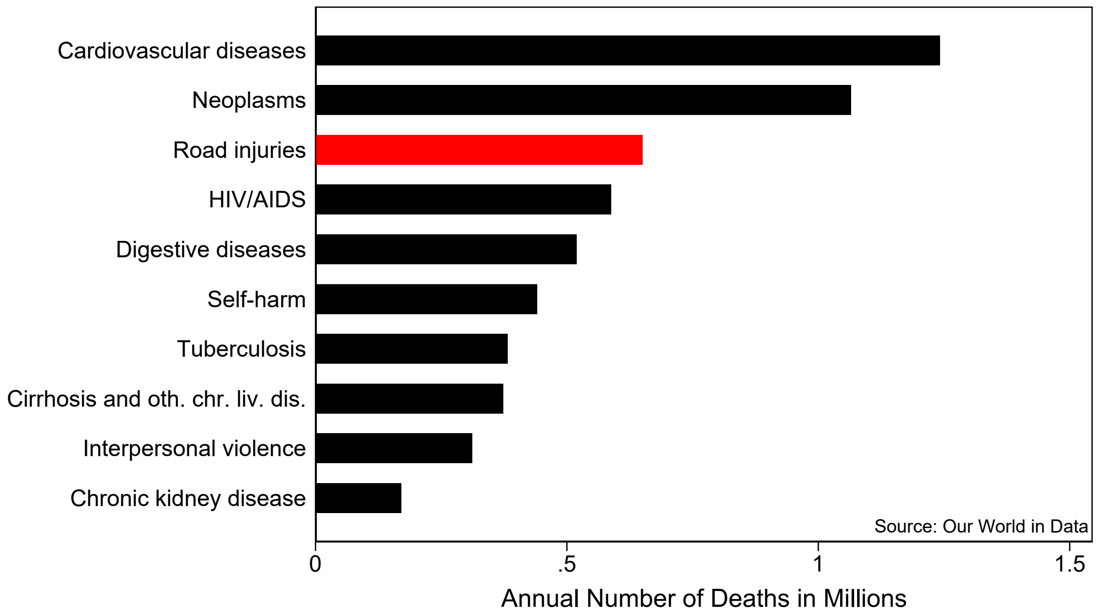
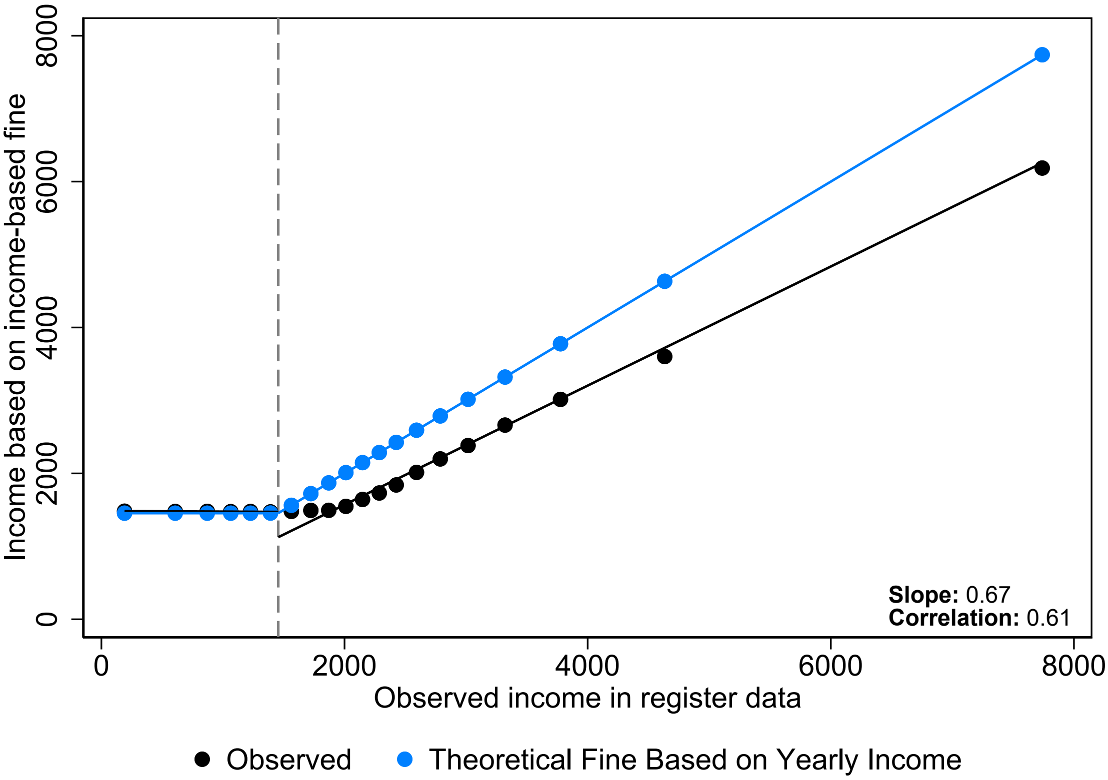


Figure A.3: Police Speeding Monitoring Camera

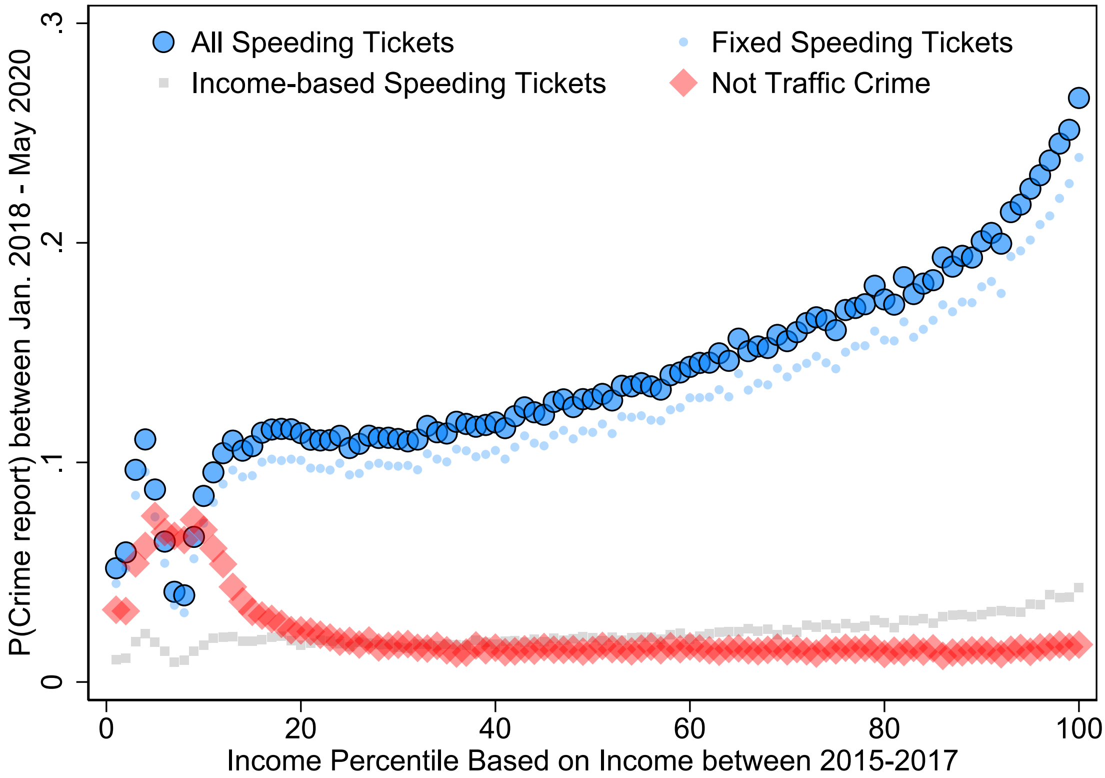
Note: Figure shows an automatic speeding camera used by the Finnish police to monitor speeding.

Figure A.4: Correlation Between Income Measures



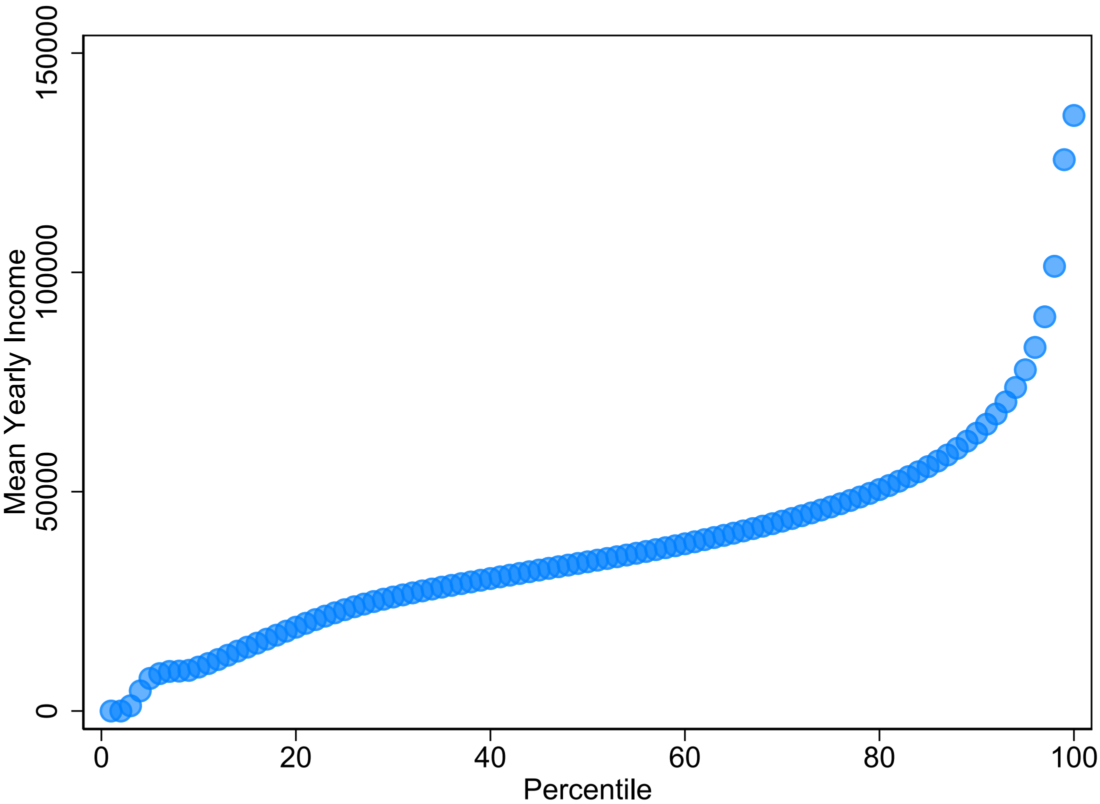
Note: The figure shows the correlation between income deduced from the size of the income-based fine and income obtained from the administrative data.

Figure A.5: The Relationship between Income, Probability of Speeding, and Crime Report



Note: Figure conducts the same exercise as Figure 2, but does not restrict to households with a car. The figure plots the relationship between income percentile and the probability of observing the individual in the Police crime report statistics between January 2018 - May 2020. To construct the figure, I take everyone 18-65 years old in 2015 in Statistics Finland's total population labor market and demographics data. I calculate the income percentile by comparing individuals' mean income over 2015-2017 to those in the same birth cohort. Blue dots with circles show the share of individuals receiving a speeding ticket in each percentile between January 2018 and May 2020. Red dots without the circle show the share of individuals in Police crime report data when the suspected crime is not a traffic offense between January 2018 and May 2020 for each percentile. Small dots decompose the speeding ticket averages into shares that arise from fixed fines (blue dots) and income-based fines (grey dots).

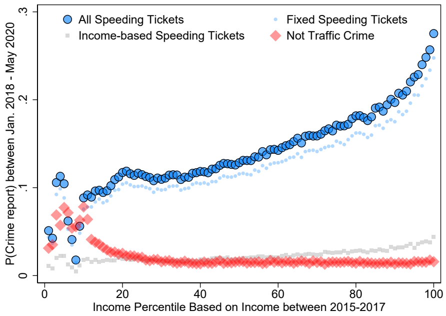
Figure A.6: The Relationship Between Mean Income and Income Percentile



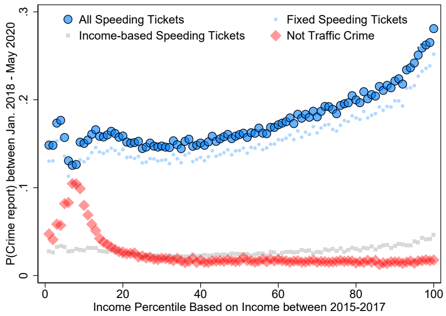
Note: The figure plots the relationship between mean income and income percentile. To construct the figure, I take everyone 18-65 years old in 2015 in Statistics Finland's total population labor market and demographics data. I calculate the income percentile by comparing individuals' mean income over 2015-2017 to those in the same birth cohort.

Figure A.7: The Relationship Between Income and Crime with Different Specifications

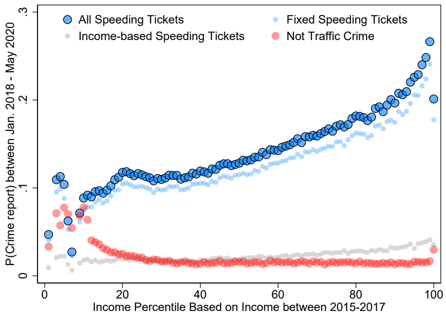
(a) Rank relative to all individuals



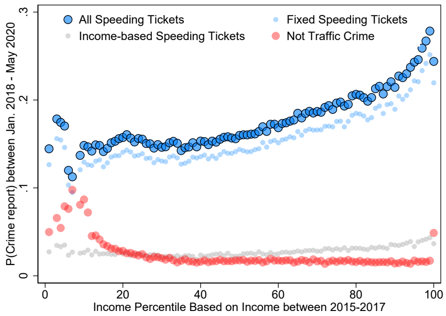
(b) Rank relative to all individuals and a car in the household



(c) Rank relative to everyone aged between 30-65 in 2015

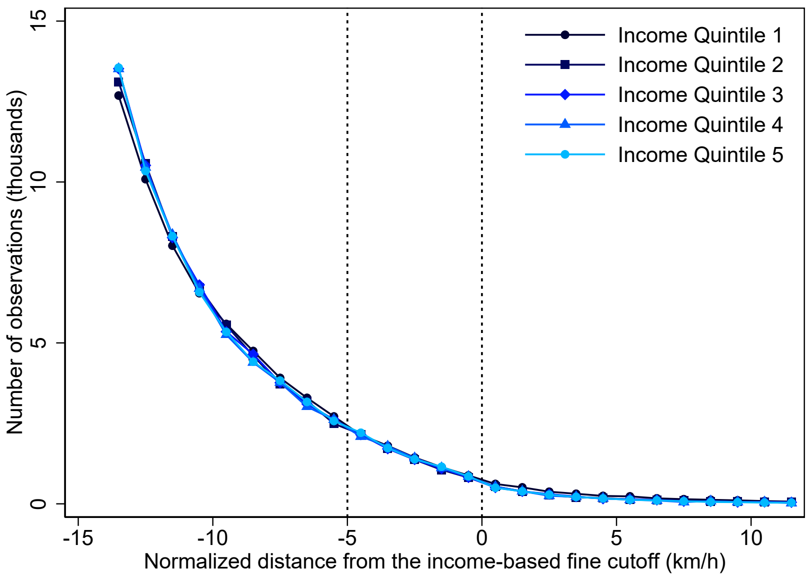


(d) Rank relative to everyone aged between 30-65 in 2015 with a car in the household



Note: Figure shows that the relationship between income and crime shown in the Figure 8 is robust to different specifications. Panel A measures the rank compared to everyone aged between 18-65 in 2015 but not within the age cohort, as is done in the main figure. Panel B constructs the percentiles as Panel A but restricts to households with a car. Panel C measures the rank compared to everyone aged between 30-65 in 2015 but not within the age cohort as is done in the main figure. Panel calculates the percentiles as Panel C but restricts to households with a car.

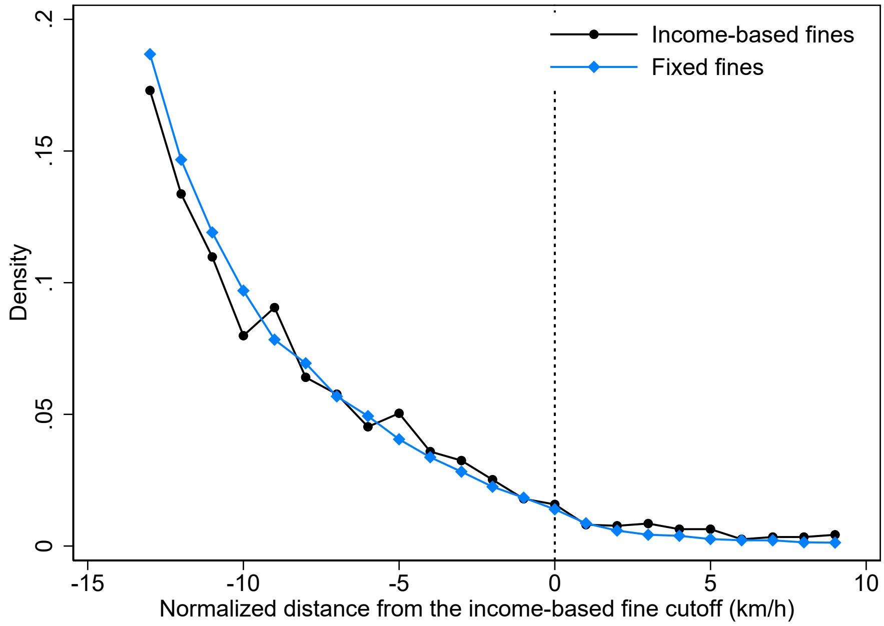
Figure A.8: Speeding Distributions for Different Income Quartiles



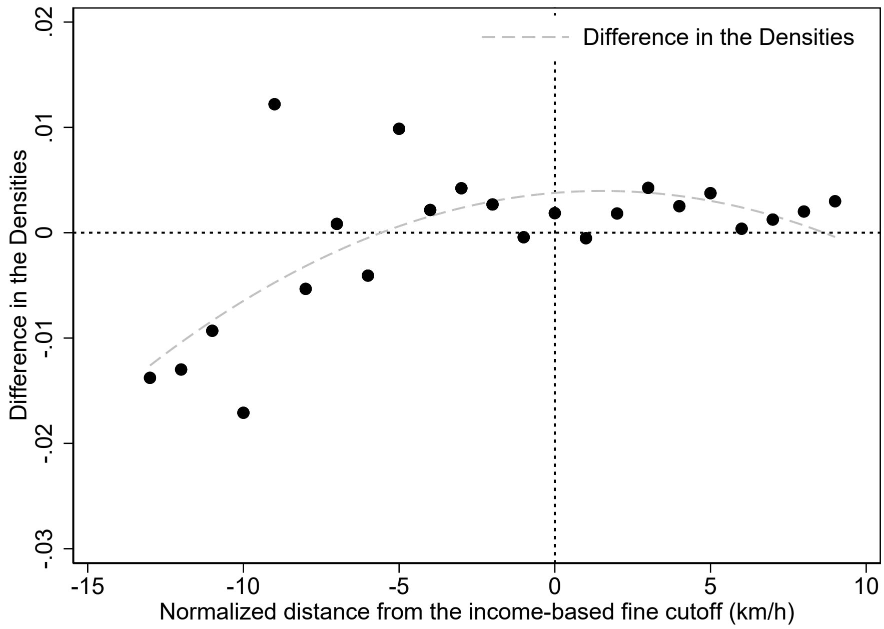
Note: Figure plots pooled speeding distribution separately for each income quartile. The sample contains speeding tickets assigned by cameras between Jan 2018 - May 2020, excluding the tickets assigned at the limit zone of 80 km /h. Income quartiles are defined using average net income measured over 1-2 before the speeding ticket. The x-axis measures the normalized distance from the income-based fine cutoff

Figure A.9: Speeding Distributions for Reoffenders

(a) Speeding Distributions for Reoffenders

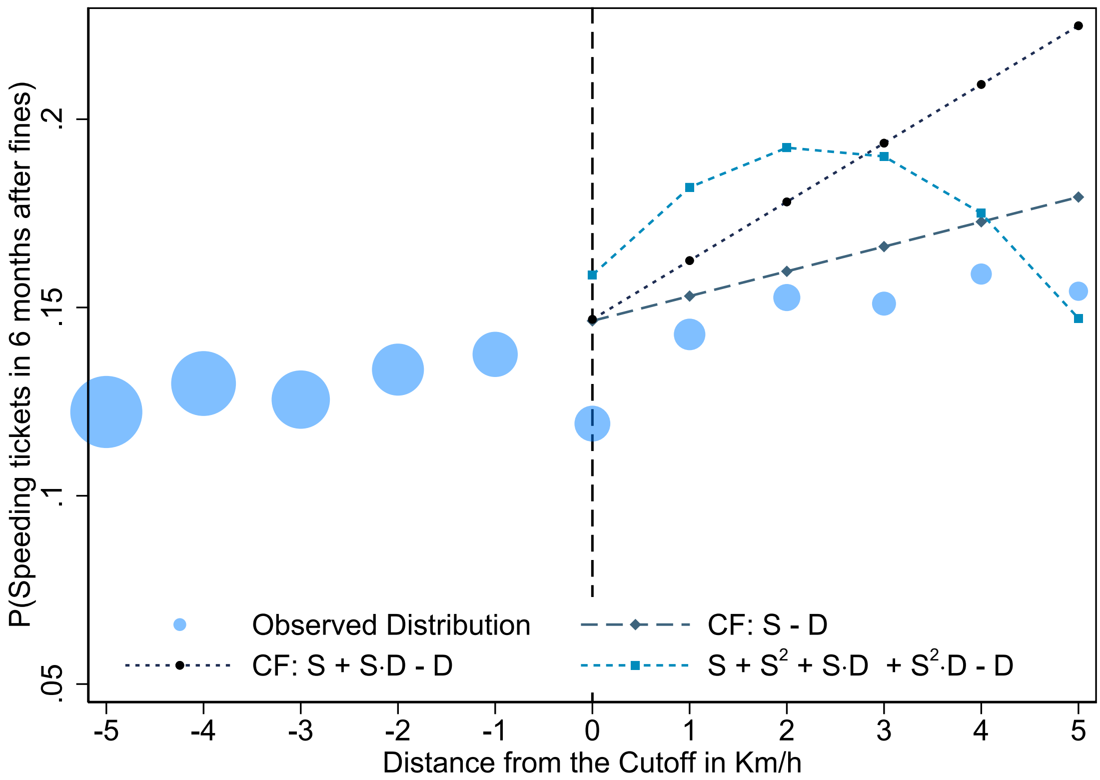


(b) Difference in Densities



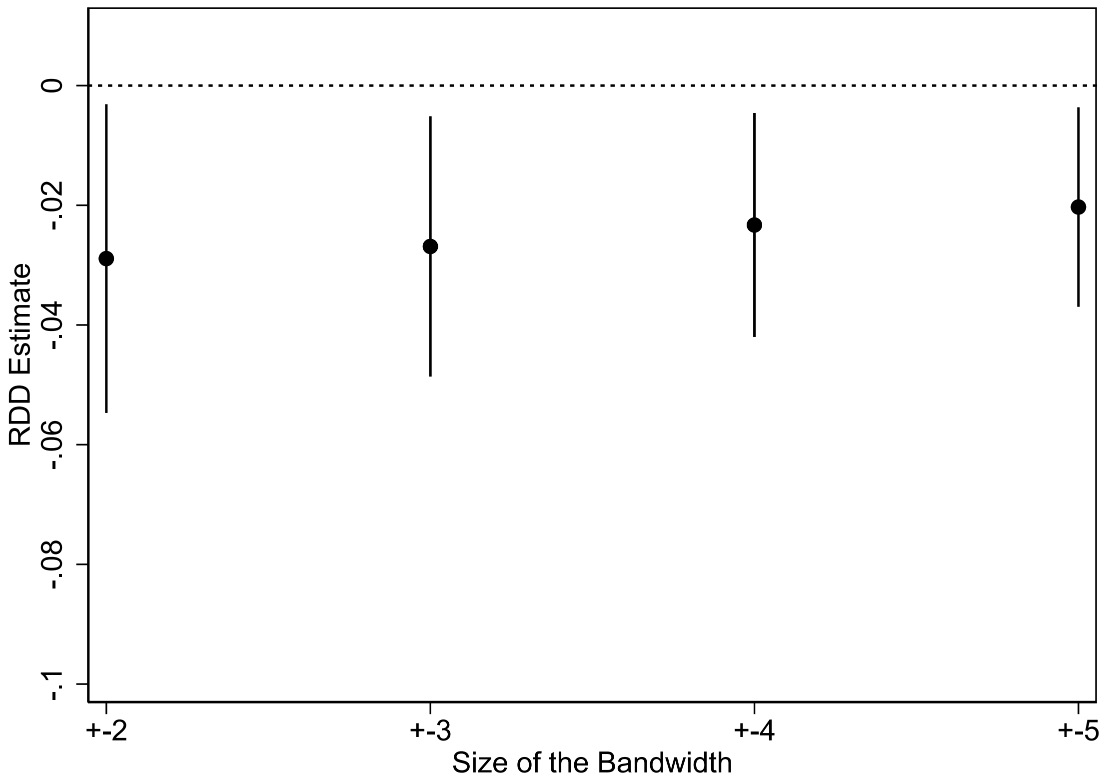
Note: Figure Panel a shows the speeding distributions for individuals with at least one speeding incident in the past 12 months. The distributions are plotted separately for those who received an income-based fine or fixed during the past 12 months. Figure panel b plots the difference between the distributions.

Figure A.10: Potential Counterfactual



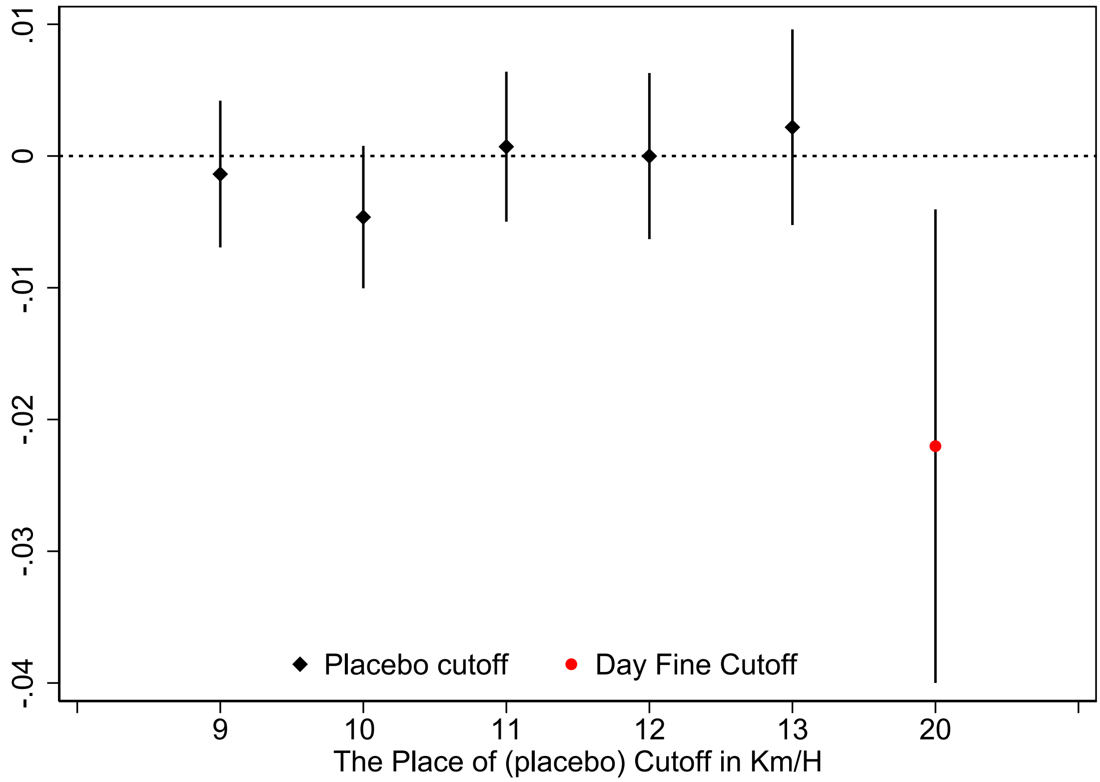
Note: This figure shows the relationship between observed speeding and future reoffending and potential counterfactuals. To construct the counterfactuals, I run different versions of equation 2, obtain predictions of the outcomes, and omit the contribution of binary variable Z on predicted values. The large blue bubbles show the actual data. Diamonds show the counterfactual when I control running variable linearly and similarly on both sides of the cutoff. Black dots plot the counterfactual when I let the linear relationship of the running variable vary on both sides of the threshold. Finally, blue squares demonstrate how the counterfactual looks when I add second-order polynomial.

Figure A.11: RDD Estimates with Different Bandwidths



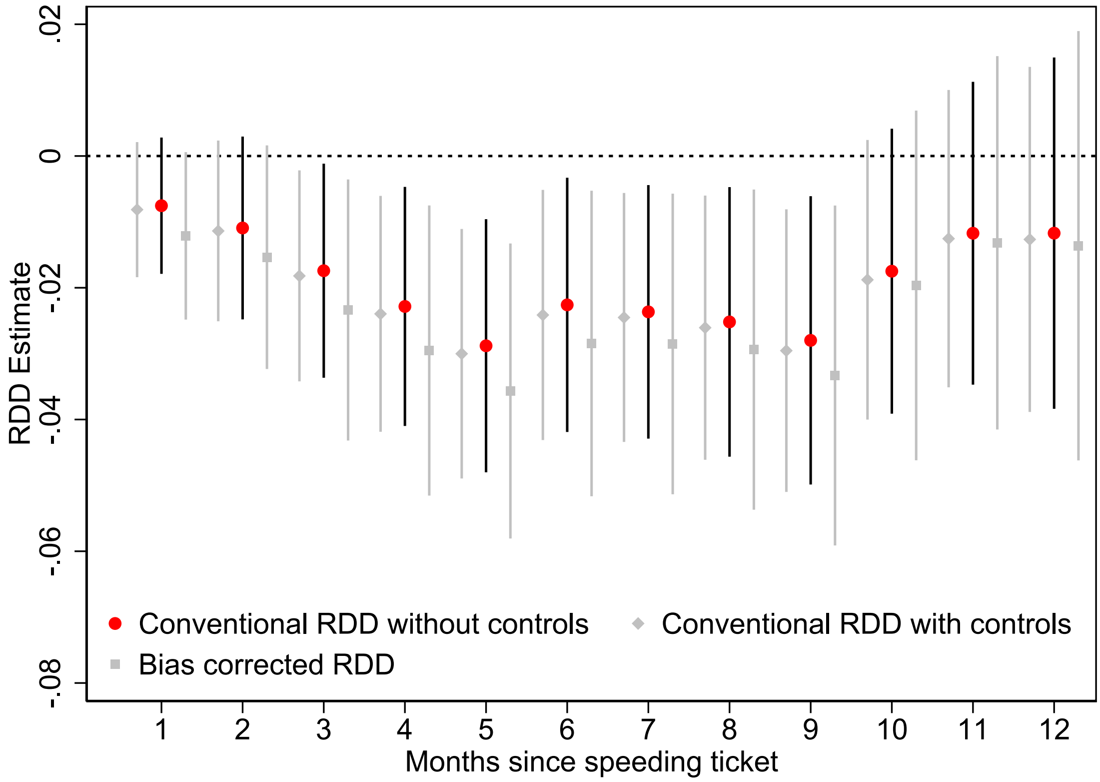
Note: The figure shows results from a robustness test in which I test whether the RDD estimates are sensitive to bandwidth choice. Black dots show the RDD estimates that are obtained using equation 2. Black spikes plot the 95 percent confidence intervals obtained using standard errors that are clustered at the individual level. The y-axis shows the size of the RDD estimate. The x-axis indicates the width of the estimation window. Section 2.3 explains the sample construction.

Figure A.12: RDD Estimates at the Placebo Cutoffs



Note: The figure plots RDD estimates I obtain from a permutation test similar to proposed by Ganong and Jäger (2018). In the test, I place placebo cutoffs at the non-discontinuity points and then run the analysis using the equation 2. If the setting is valid, we should not observe the impact on the outcome since the treatment is the same on both sides of the cutoff. The placebo cutoffs and bandwidths are always defined so that the individuals on the right-hand side of the income-based fine cutoff are not used in the placebo check. Black diamonds report the results from the placebo check. Red dot reports the main estimates. The spikes report 95 percent confidence intervals. Standard errors are clustered at the individual level. Section 2.3 defines the sample construction.

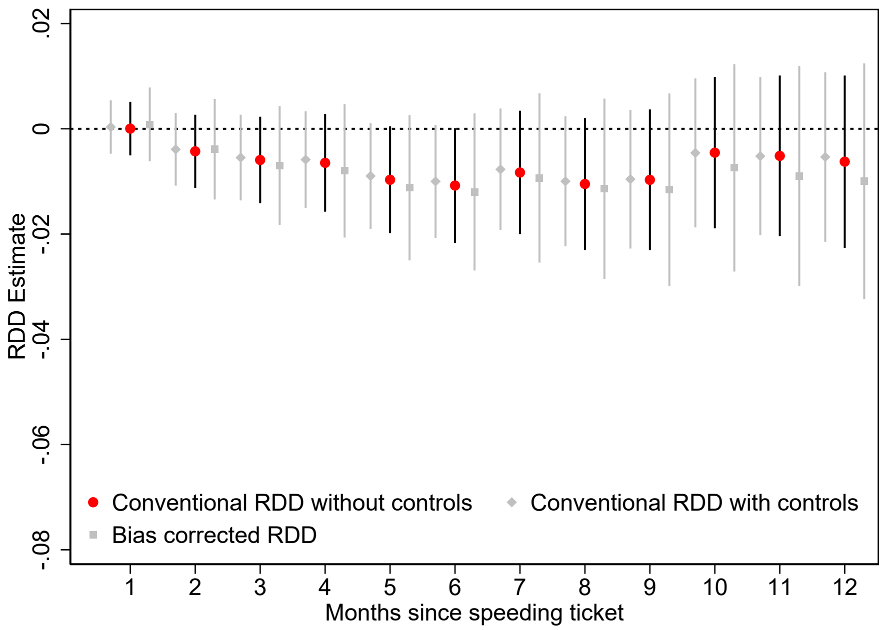
Figure A.13: RDD Estimates at the Day fine cutoff if all limits included



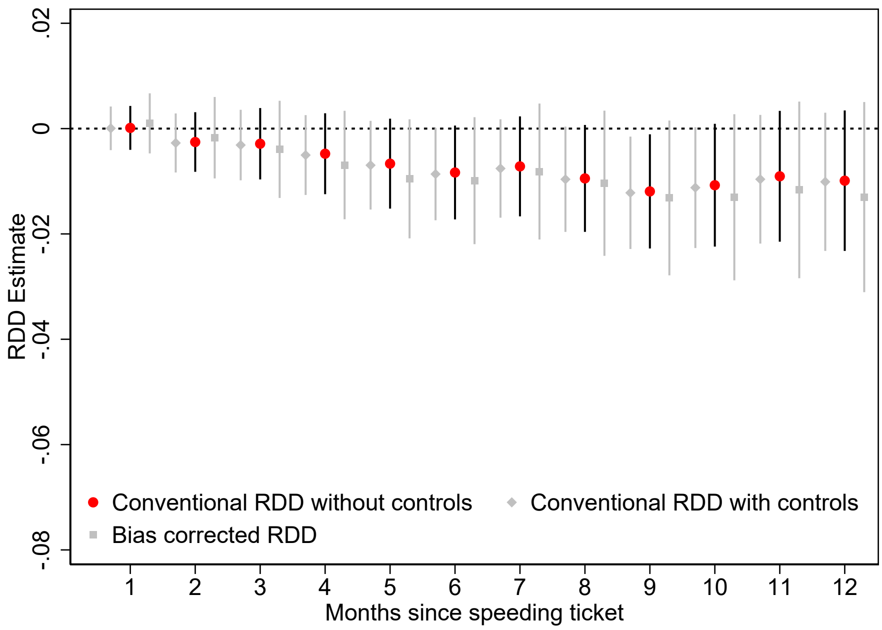
Note: Figure shows results from a similar analysis as Figure 7, but also includes the fines assigned when the speeding limit is 80 km/h. Figure plots the RDD estimates of β obtained using Equation (2). Estimates that are shown after the vertical dashed lines plot the cumulative impact of the higher fine on the probability of committing another traffic offense in t months after the initial speeding ticket. The estimates before the vertical dashed line show results from validity checks where the outcomes are pre-determined monthly probabilities of whether an individual has committed any traffic offense. Dots plot the estimates obtained using conventional local linear regression with a triangular kernel. Diamond plots the conventional estimates when controls are added. Squares report biased corrected RDD estimates obtained using the robust approach by Calonico *et al.* (2014b). Vertical lines behind the estimates show the 95 percent confidence intervals. Standard errors are clustered at the individual level in the conventional approach. For bias-corrected estimates, the confidence intervals are given by approach of Calonico *et al.* (2014b). Sample construction and data are defined in Section 2.3.

Figure A.14: RDD Estimates at the 15 Km/H cutoff

(a) limit 80 Km/H omitted

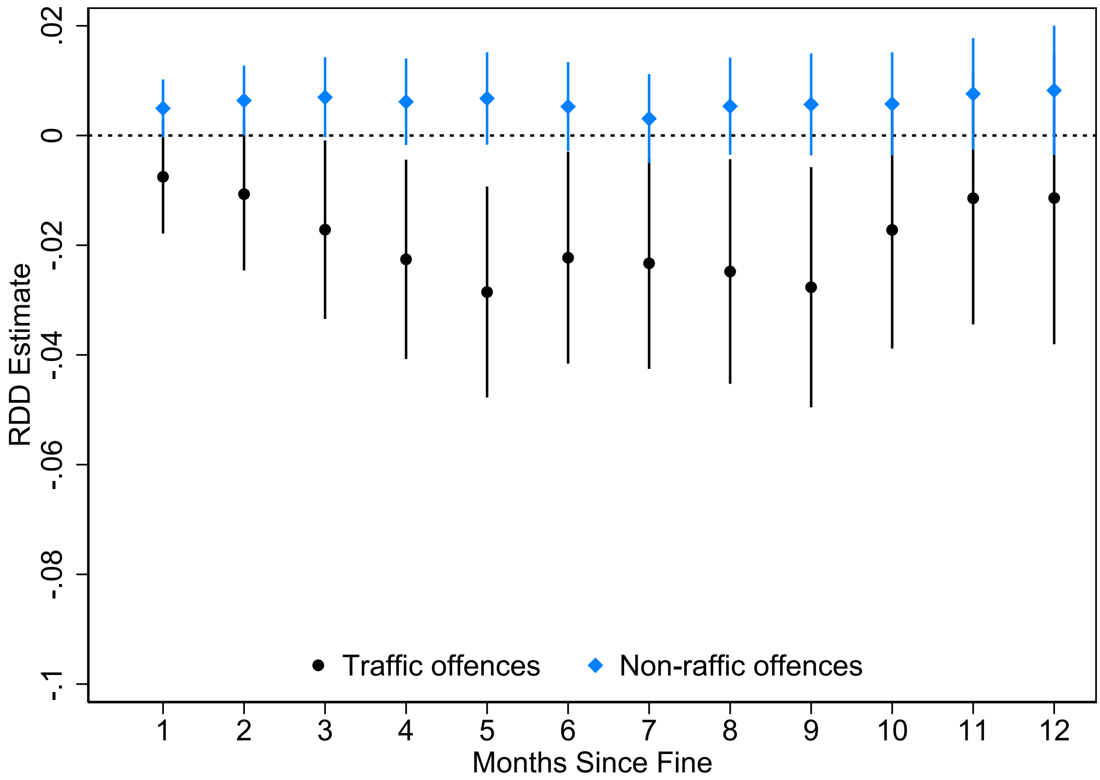


(b) limit 80 Km/H included



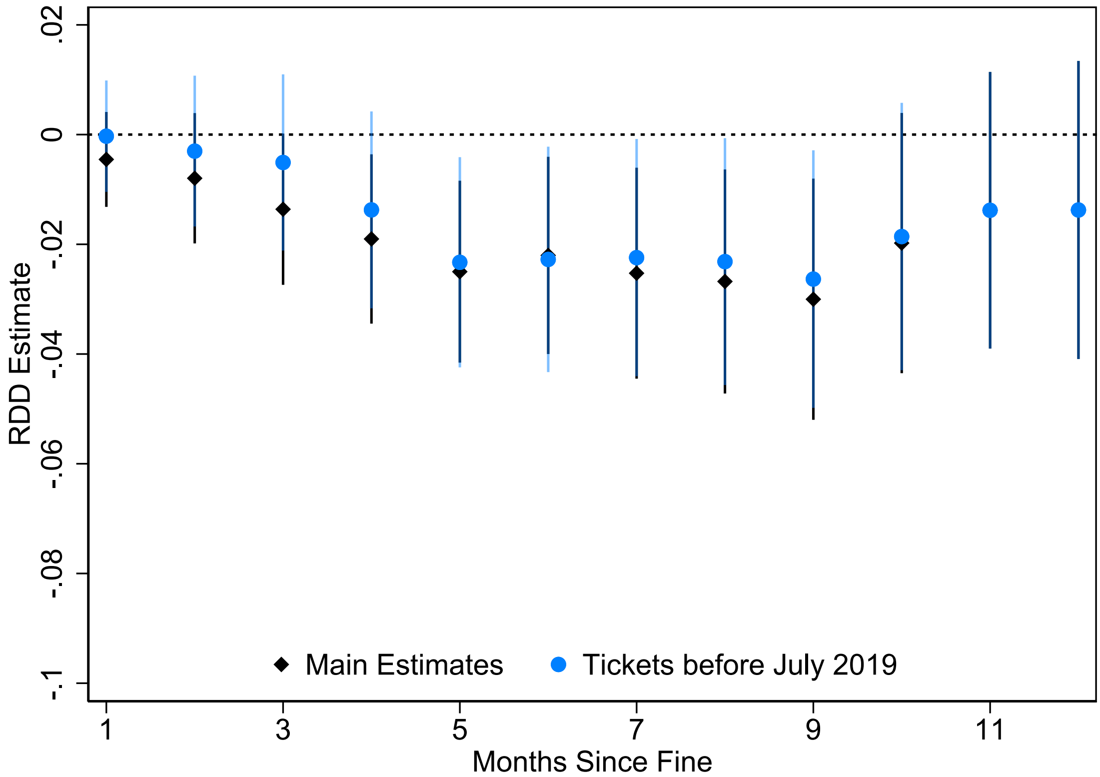
Note: The figure shows results from a similar analysis as Figure 7, but uses variation from the fixed fine cutoff where the hike does not relate to an individual's income. The figure panel (a) plots the RDD estimates of β obtained using Equation (2) when all limit zones except 120 km/h are included. Estimates that are shown after the vertical dashed lines plot the cumulative impact of the higher fine on the probability of committing another traffic offense within t months after the initial speeding ticket. The estimates before the vertical dashed line plot result from validity checks where the outcomes are pre-determined monthly probabilities of whether an individual has committed any traffic offense. Dots plot the estimates obtained using conventional local linear regression with a triangular kernel. Diamond plots the conventional estimates when controls are added. Vertical lines behind the estimates show the 95 percent confidence intervals. Standard errors are clustered at the individual level in the conventional approach. Sample construction and data as defined in Section 2.3. Figure panel (b) conducts the same analysis but drops the limit zone of 80 km/h, where I observe bunching.

Figure A.15: The Impact on Traffic and Non-traffic Crime



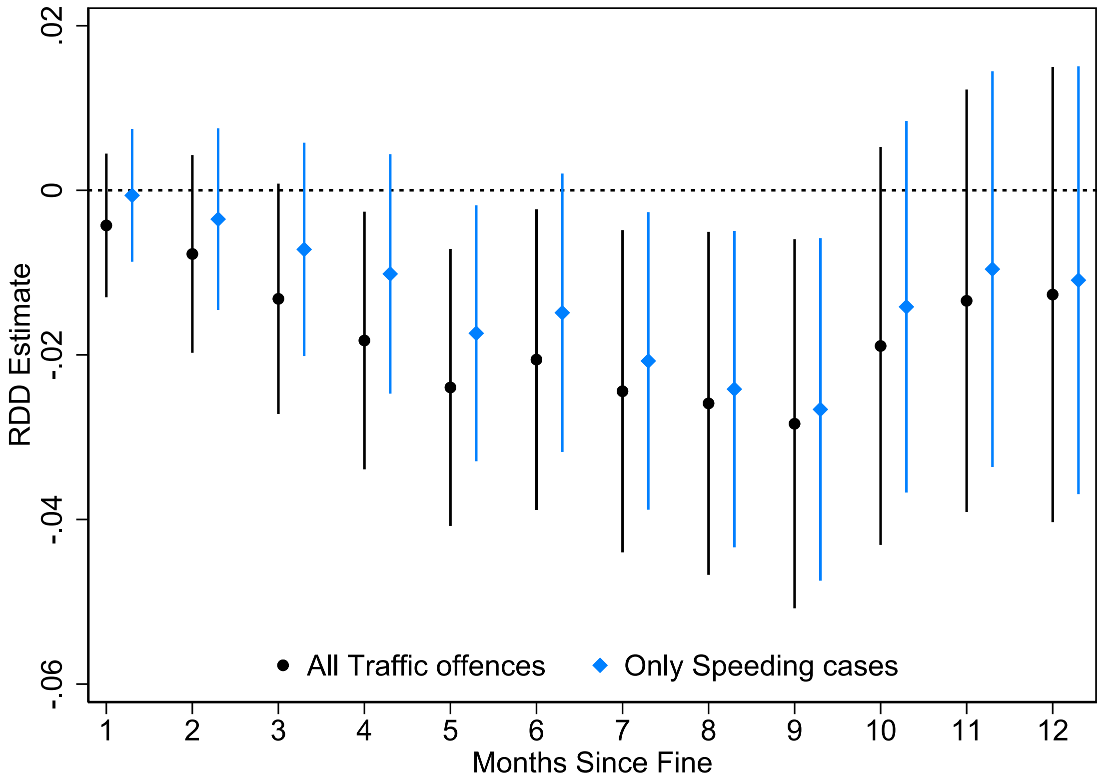
Note: The figure shows results from a similar analysis as Figure 7, but the primary outcome is cumulative non-traffic criminal activity. The figure plots the RDD estimates of β obtained using Equation (2) when all limit zones except 80 km/h and 120 km/h are included. I use local linear regression with a triangular kernel. Blue diamonds show the results when the outcome is a cumulative probability of committing a non-traffic crime. As a comparison, black dots show the results I present in the main figure Figure 7. Vertical lines behind the estimates show the 95 percent confidence intervals. Standard errors are clustered at the individual level. Sample construction and data as defined in Section 2.3.

Figure A.16: RDD estimates when restricting to individuals than can be followed 12 months



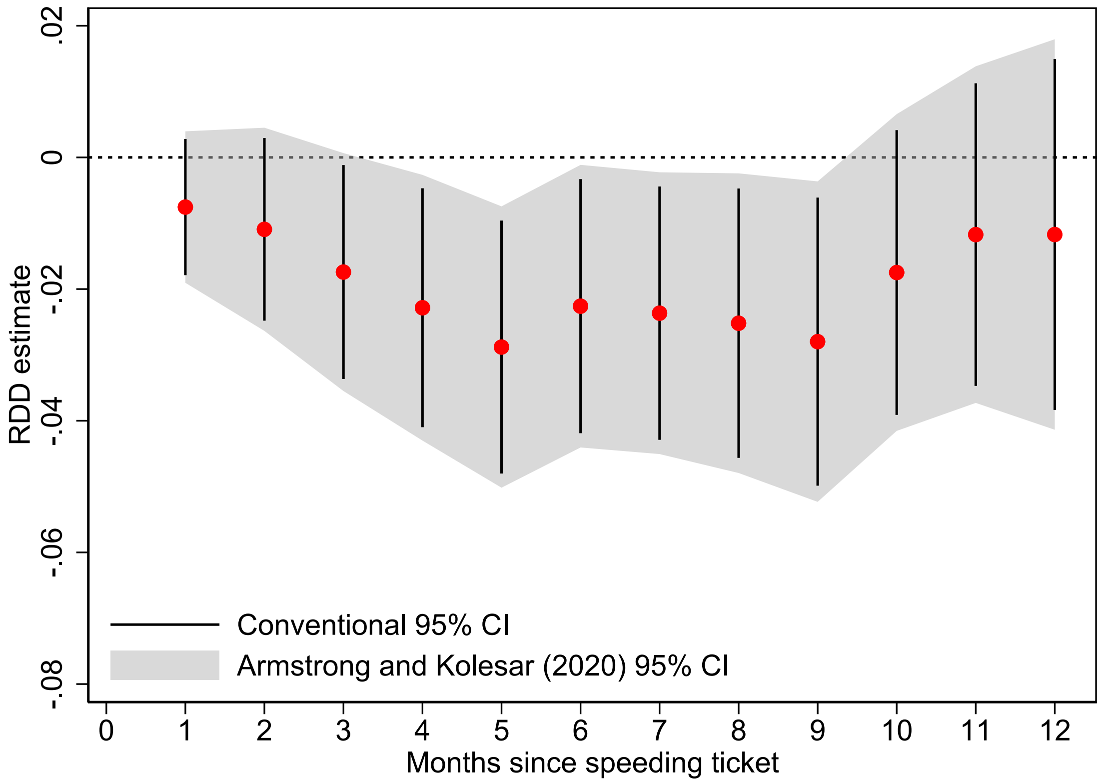
Note: The figure shows results from a robustness check that tests how sensitive my results are to the fact that I can not follow all drivers the entire 12 months. Black diamonds plot the main estimates shown in Figure 7. Blue dots restrict the sample to drivers I can follow for 12 months. All the estimates are obtained using equation 2. I use local linear regression with a triangular kernel. Blue diamonds show the results when the outcome is a cumulative probability of committing a non-traffic crime. As a comparison, black dots show the results I present in the main figure Figure 7. Vertical lines behind the estimates show the 95 percent confidence intervals. Standard errors are clustered at the individual level. Sample construction and data as defined in Section 2.3.

Figure A.17: RDD Estimates when Outcome Is Created Using Only Speeding Incidents



Note: Figure shows results from analysis where the outcome is constructed using only speeding incidents. Black dots plot the main estimates shown in Figure 7 when the outcome is constructed using all traffic offenses. Blue diamonds show the results when I restrict just to speeding violations. All the estimates are obtained using equation 2. I use local linear regression with a triangular kernel. Vertical lines behind the estimates show 95 percent confidence intervals. Standard errors are clustered at the individual level. Sample construction and data as defined in Section 2.3.

Figure A.18: Standard and Armstrong and Kolesár (2020) Confidence Intervals



Note: Figure plots the main RDD estimates with conventional and Armstrong and Kolesár (2020) 95 percent confidence intervals. Red dots plot the RDD estimates I obtain using equation 2. Vertical lines plot the conventional 95 percent confidence intervals. The shaded area shows the "honest confidence intervals" of Armstrong and Kolesár (2020). Standard errors are clustered at the individual level. Sample construction and data as defined in Section 2.3.

Figure A.19: Average Causal Response

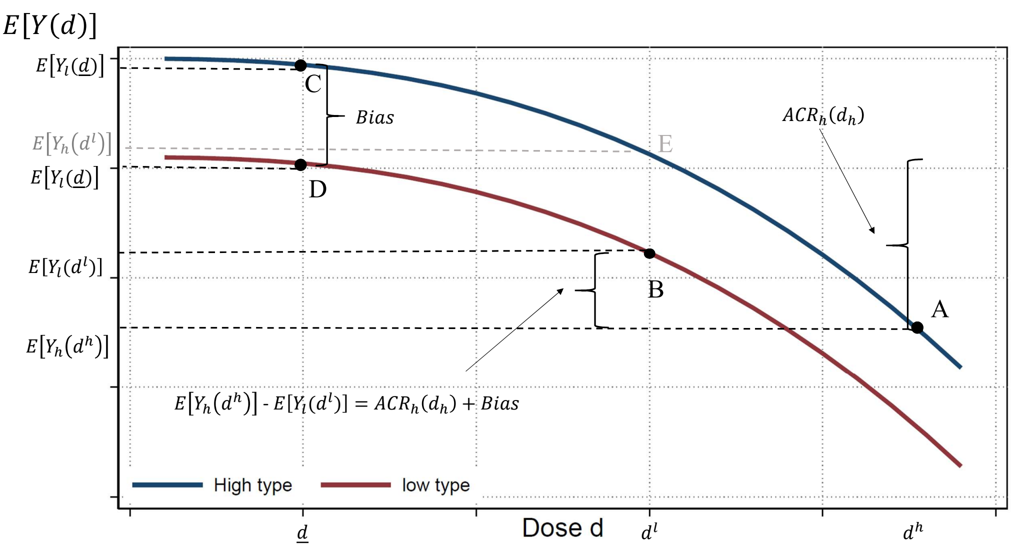
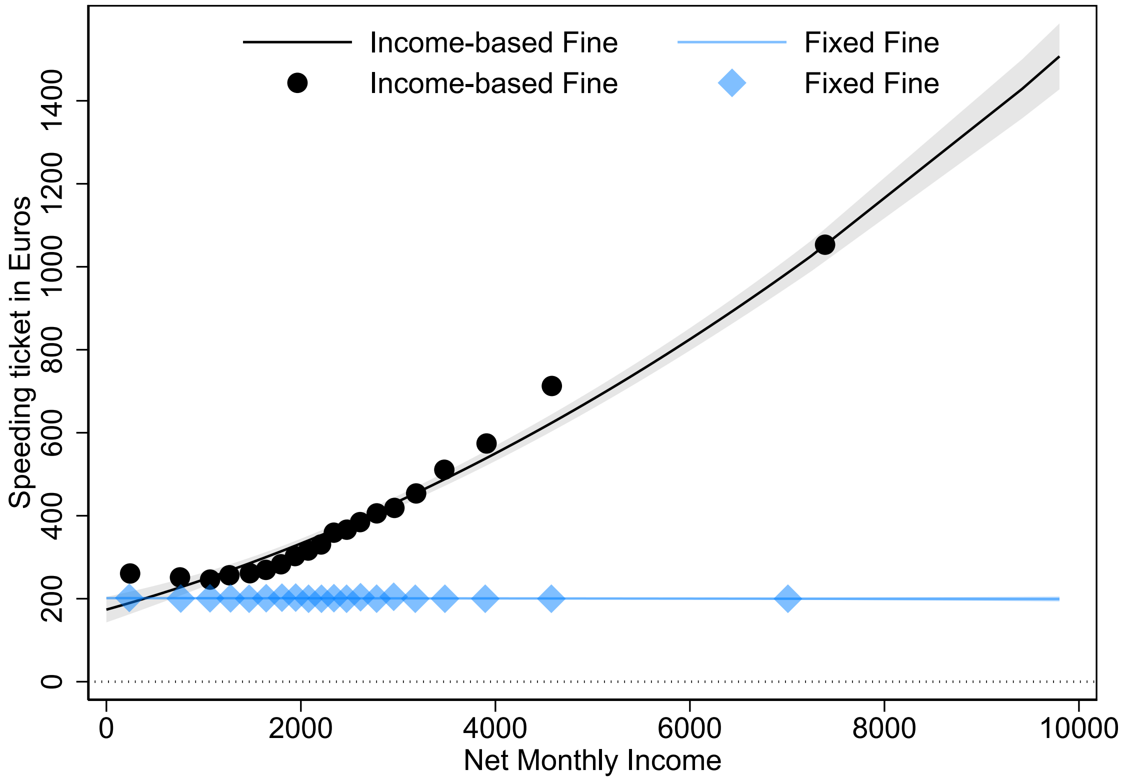
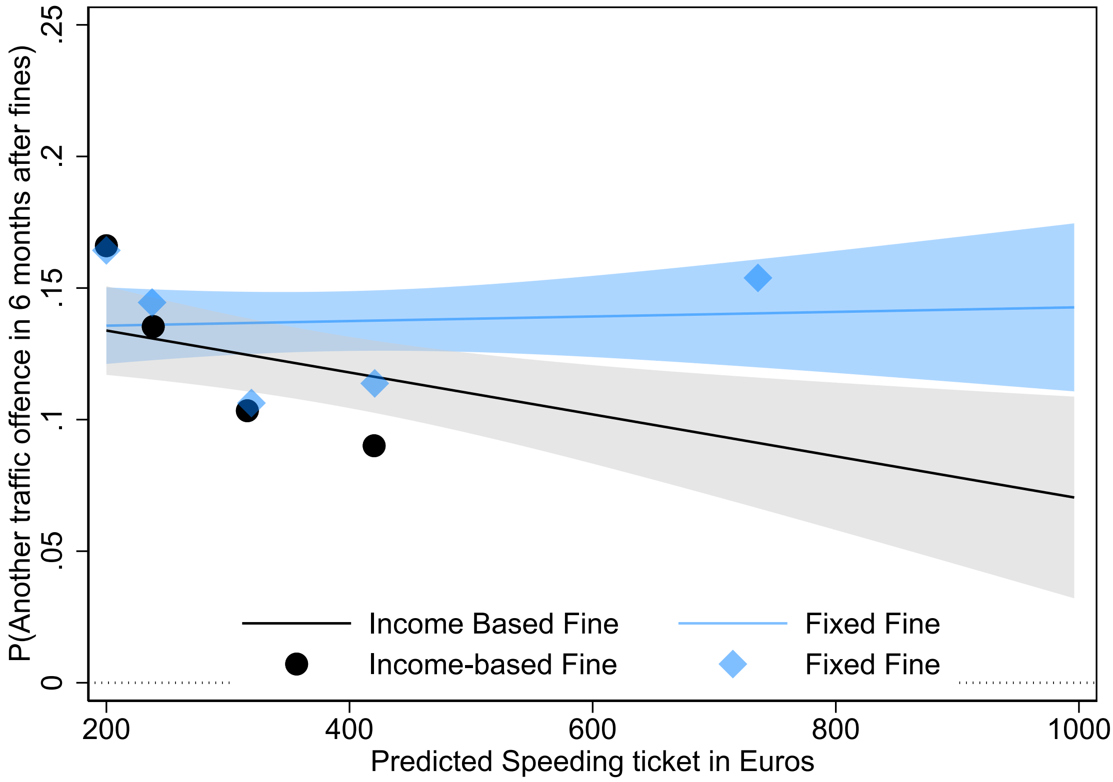


Figure A.20: The Relationship Between Income and Fine in Groups ± 1km/h from Day-Fine Cutoff



Note: The figure illustrates the link between income and speeding tickets within groups below and above the income-based fine cutoff. Black dots show the relationship in the treatment group whose speeding tickets vary with income. Blue circles show the same relationship in the treatment group, where everyone receives a similar fine. Section 2.3 describes the sample creation.

Figure A.21: The Relationship between Reoffending and Predicted Fines in Treatment and Control Group



Note: The figure illustrates the relationship between reoffending and the predicted income-based fines. The x-axis refers to a predicted income-based fine that an individual would receive based on her income if she crossed the income-based fine cutoff. The y-axis measures an individual's probability of reoffending within six months after the original speeding incident. Black dots show the non-parametric relationship between predicted income-based fines and reoffending for those who received an income-based fine. Blue diamonds show the same relationship for those receiving a fixed fine. The black (blue) line shows the predicted value from linear OLS where reoffending is regressed on predicted fines in a sample of individuals receiving an income-based fine (fixed fine). A change in slope between black and blue lines identifies the average causal response parameter under the assumption that income just shifts the level of offending.

B Tables

Table B.1: The Punishment Schedule

Speed Limit Violation (km/h)	**Number of Income Fines**	**Drift**	**Fixed Fine**
(1)	(2)	(3)	(4)
Panel A: Speeding limit $\leq$ 60 km/h			
-10	-	-	170
10-15	-	-	170
16-20	-	-	200
21-23	12	8 - 16	-
24-26	14	10 - 18	-
27-29	16	12 - 20	-
30-32	18	14 - 22	-
33-35	20	16 - 24	-
36-38	22	18 - 26	-
39-41	24	20 - 28	-
42-44	26	22 - 30	-
45-47	28	24 - 32	-
Panel B: Speeding limit $>$ 60 km/h			
-10	-	-	140
16-20	-	-	140
21-23	10	6-14	200
24-26	12	8-16	-
27-29	14	10-18	-
30-32	16	12-20	-
33-35	18	14-22	-
36-38	20	16-24	-
39-41	22	18-26	-
42-44	24	20-28	-
45-47	26	22-30	-
48-50	28	24-32	-

Note: The table shows the punishment schedule Finnish Police used during my study period. Column (1) shows the speeding limit violation in km/h. Column (2) presents the recommended number of income-based fines police assign for a given violation. Column(3) shows recommended range for the number of income-based. Column (4) indicates the size of the possible fixed fine in euros.

Table B.2: Descriptive Statistics and Balance check when limit zone 80 km/h is included

	All	**Window +-3**	**RDD**	**CCT RDD**
	Mean	Mean	RDD	RDD
	(1)	(2)	(3)	(4)
Traffic offence before ticket	0.121	0.135	0.007	0.010
			(0.009)	(0.011)
Non-traffic offence before ticket	0.014	0.020	-0.001	-0.001
			(0.004)	(0.005)
Employed	0.702	0.703	0.002	0.002
			(0.011)	(0.014)
Unemployed	0.032	0.037	0.011	0.014
			(0.005)	(0.006)
Outside the Labor Force	0.265	0.260	-0.016	-0.019
			(0.011)	(0.014)
Monthly Net Income	2,578	2,579	-25.217	-5.386
			(41.807)	(51.582)
Primary Education Only	0.164	0.168	0.009	0.011
			(0.009)	(0.011)
Secondary Education Only	0.439	0.444	-0.013	-0.017
			(0.012)	(0.015)
Tertiary Education Only	0.397	0.389	0.003	0.005
			(0.011)	(0.014)
Female	0.328	0.292	-0.011	-0.016
			(0.011)	(0.013)
Age	47.313	46.053	-0.758	-0.572
			(0.387)	(0.454)
N of Children	0.829	0.839	0.032	0.043
			(0.029)	(0.037)
Finnish Speaking	0.893	0.891	0.021	0.024
			(0.008)	(0.010)
Married	0.499	0.483	0.008	0.014
			(0.013)	(0.015)
Urban Municipality	0.719	0.716	0.017	0.023
			(0.012)	(0.014)
Semi-urban Municipality	0.161	0.163	-0.017	-0.021
			(0.008)	(0.010)
Rural municipality	0.120	0.121	-0.001	-0.005
			(0.009)	(0.011)
Capital region	0.324	0.295	0.008	0.009
			(0.011)	(0.014)
Observations	504,529	31,891		

Note: The table shows the predetermined characteristics of individuals in the speeding ticket sample and results from balance checks when a limit zone of 80 km/h is included. Column 1 shows the means of background variables in the estimation sample. Column 2 presents the means of background variables in the estimation sample after I restrict it to individuals within +-3 km/h from the income-based cutoff. Column 3 shows results from the balance check where the dependent variable is a predetermined characteristic. The column reports the estimates of β that I obtain using the equation 2. Column 4 reports results from the balance check I conduct using the approach of Calonico *et al.* (2014a). I cluster standard errors at the individual level.

Table B.3: Balance check at the 15 Km/H cutoff - all limits included

Dep. Var	**All** Mean (1)	**Window +-3** Mean (2)	**Conv. RD** RDD (3)	**Bias-corrected RD** RDD (4)
Employed	0.698	0.692	-0.005	-0.004
			(0.006)	(0.009)
Unemployed	0.036	0.040	-0.003	-0.003
			(0.003)	(0.004)
Outside the Labor Force	0.266	0.269	0.008	0.007
			(0.006)	(0.009)
Income in 1000s (€)	38.233	38.166	0.530	0.383
			(0.379)	(0.520)
Primary Education Only	0.166	0.166	-0.003	-0.004
			(0.005)	(0.007)
Secondary Education Only	0.439	0.436	0.001	0.002
			(0.007)	(0.010)
Tertiary Education Only	0.396	0.398	0.002	0.002
			(0.007)	(0.009)
Female	0.324	0.308	-0.004	0.002
			(0.007)	(0.009)
Age	47.285	47.008	0.090	-0.133
			(0.234)	(0.313)
N of Children	0.835	0.827	0.010	0.026
			(0.019)	(0.024)
Finnish Speaking	0.898	0.897	0.005	0.005
			(0.004)	(0.006)
Married	0.498	0.498	0.009	0.009
			(0.007)	(0.010)
Urban Municipality	0.707	0.708	-0.011	-0.017
			(0.007)	(0.009)
Semi-urban Municipality	0.166	0.166	0.007	0.013
			(0.005)	(0.007)
Rural municipality	0.126	0.126	0.003	0.003
			(0.005)	(0.006)
Capital region	0.310	0.296	0.002	0.006
			(0.006)	(0.009)
Observations	422,719	106,714		

Note: The table shows the predetermined characteristics of individuals in the speeding ticket sample and results from the balance checks from the fixed fine cutoff when all limit zones are included. Column 1 shows the means of background variables in the estimation sample. Column 2 presents the means of background variables in the estimation sample after I restrict it to individuals within +-3 km/h from the fixed fine cutoff. Column 3 shows results from the balance check where the dependent variable is a predetermined characteristic. The column reports the estimates of β that I obtain using the equation 2. Column 4 reports results from the balance check I conduct using the approach of Calonico *et al.* (2014a). I cluster standard errors at the individual level.

Table B.4: Balance check at the 15 Km/H cutoff - limit 80 Km/H omitted

Dep. Var	**All** Mean (1)	**Window +-3** Mean (2)	**Conv. RD** RDD (3)	**Bias-corrected RD** RDD (4)
Employed	0.702	0.698	-0.008	-0.012
			(0.008)	(0.011)
Unemployed	0.036	0.039	-0.004	-0.005
			(0.003)	(0.005)
Outside the Labor Force	0.263	0.263	0.012	0.017
			(0.008)	(0.011)
Income in 1000s (€)	38.461	38.088	0.689	0.677
			(0.477)	(0.663)
Primary Education Only	0.165	0.166	-0.001	-0.004
			(0.007)	(0.009)
Secondary Education Only	0.435	0.437	0.015	0.015
			(0.009)	(0.013)
Tertiary Education Only	0.400	0.397	-0.012	-0.010
			(0.009)	(0.012)
Female	0.325	0.310	-0.007	-0.006
			(0.008)	(0.011)
Age	47.276	46.872	0.330	0.266
			(0.282)	(0.392)
N of Children	0.840	0.839	-0.038	-0.030
			(0.021)	(0.029)
Finnish Speaking	0.892	0.890	0.010	0.013
			(0.006)	(0.008)
Married	0.501	0.496	0.007	0.008
			(0.009)	(0.012)
Urban Municipality	0.712	0.705	-0.013	-0.017
			(0.008)	(0.011)
Semi-urban Municipality	0.169	0.172	0.004	0.007
			(0.007)	(0.009)
Rural municipality	0.119	0.123	0.009	0.008
			(0.006)	(0.008)
Capital region	0.318	0.303	0.002	0.002
			(0.008)	(0.011)
Observations	260,762	63,367		

Note: The table shows the predetermined characteristics of individuals in the speeding ticket sample and results from the balance checks from the fixed fine cutoff when 80 km/h limit zone is omitted. Column 1 shows the means of background variables in the estimation sample. Column 2 presents the means of background variables in the estimation sample after I restrict it to individuals within +-3 km/h from the fixed fine cutoff. Column 3 shows results from the balance check where the dependent variable is a predetermined characteristic. The column reports the estimates of β that I obtain using the equation 2. Column 4 reports results from the balance check I conduct using the approach of Calonico *et al.* (2014a). I cluster standard errors at the individual level.

Table B.5: The RDD estimates from the Income-based Fine Cutoff (Limit 80 km/h omitted)

Dep. variable: P(Traffic offence between time 1-t)

Time (1)	**Mean** (2)	**RDD** (3)	**RDD** (4)	**RDD** (5)	**Obs.** (6)
1	0.028	-0.008	-0.008	-0.012	23,035
		(0.005)	(0.005)	(0.006)	
2	0.052	-0.011	-0.011	-0.015	23,035
		(0.007)	(0.007)	(0.009)	
3	0.074	-0.017	-0.018	-0.023	23,035
		(0.008)	(0.008)	(0.010)	
4	0.095	-0.023	-0.024	-0.030	23,035
		(0.009)	(0.009)	(0.011)	
5	0.113	-0.029	-0.030	-0.036	23,035
		(0.010)	(0.010)	(0.011)	
6	0.131	-0.022	-0.024	-0.028	31,924
		(0.010)	(0.010)	(0.012)	
7	0.148	-0.023	-0.025	-0.029	30,639
		(0.010)	(0.010)	(0.012)	
8	0.164	-0.025	-0.026	-0.029	29,335
		(0.010)	(0.010)	(0.012)	
9	0.183	-0.028	-0.030	-0.033	27,726
		(0.011)	(0.011)	(0.013)	
10	0.198	-0.017	-0.019	-0.020	34,922
		(0.011)	(0.011)	(0.014)	
11	0.215	-0.011	-0.013	-0.013	33,076
		(0.012)	(0.012)	(0.014)	
12	0.232	-0.011	-0.013	-0.014	30,604
		(0.014)	(0.013)	(0.017)	
Controls			✓		
CCT Estimator				✓	

Note: The table shows the main RDD estimates from the income-based fine cutoff plotted in figure 2 when the limit zone of 80 km/h is omitted. The outcome is the probability that an individual commits another traffic offense 1 to t months after the initial speeding ticket. Column(2) shows the mean of the outcome for those just on the left-hand side of the cutoff. Column (3) presents the RDD estimates obtained using equation 2. Column (4) shows results from a similar analysis but with controls. Column(5) reports biased corrected RDD estimates obtained using the robust approach by Calonico *et al.* (2014b). Column (6) shows the number of observations used in the analysis. In columns (3) and (4), standard errors in the parentheses are clustered at the individual level. For bias-corrected estimates, the confidence intervals are given by the approach of Calonico *et al.* (2014b). Sample construction and data as defined in Section 2.3.

Table B.6: The RDD estimates from the Income-based Fine Cutoff (Limit 80 km/h included)

Dep. variable: P(Traffic offence between time 1-t)

Time (1)	**Mean** (2)	**RDD** (3)	**RDD** (4)	**RDD** (5)	**Obs.** (6)
1	0.029	-0.005	-0.005	-0.008	31,904
		(0.005)	(0.005)	(0.006)	
2	0.055	-0.010	-0.011	-0.013	31,904
		(0.006)	(0.006)	(0.007)	
3	0.076	-0.017	-0.018	-0.023	31,904
		(0.007)	(0.007)	(0.009)	
4	0.097	-0.018	-0.019	-0.022	51,101
		(0.007)	(0.007)	(0.008)	
5	0.116	-0.025	-0.026	-0.032	31,904
		(0.009)	(0.008)	(0.010)	
6	0.133	-0.017	-0.018	-0.022	49,518
		(0.008)	(0.008)	(0.010)	
7	0.149	-0.019	-0.020	-0.024	47,691
		(0.009)	(0.009)	(0.010)	
8	0.166	-0.019	-0.020	-0.024	45,341
		(0.009)	(0.009)	(0.011)	
9	0.185	-0.025	-0.027	-0.032	42,971
		(0.011)	(0.011)	(0.013)	
10	0.204	-0.022	-0.023	-0.027	25,121
		(0.012)	(0.012)	(0.014)	
11	0.219	-0.015	-0.016	-0.020	38,010
		(0.011)	(0.011)	(0.014)	
12	0.237	-0.016	-0.018	-0.022	35,021
		(0.013)	(0.013)	(0.016)	
Controls			✓		
CCT Estimator				✓	

Note: The table shows the RDD estimates from the income-based fine cutoff when the limit zone of 80 km/h is not omitted. The outcome is the probability that an individual commits another traffic offense 1 to t months after the initial speeding ticket. Column(2) shows the mean of the outcome for those just on the left-hand side of the cutoff. Column (3) presents the RDD estimates obtained using equation 2. Column (4) shows results from a similar analysis but with controls. Column(5) reports biased corrected RDD estimates obtained using the robust approach by Calonico *et al.* (2014b). Column (6) shows the number of observations used in the analysis. In columns (3) and (4), standard errors in the parentheses are clustered at the individual level. For bias-corrected estimates, the confidence intervals are given by the approach of Calonico *et al.* (2014b). Sample construction and data as defined in Section 2.3.

Table B.7: The RDD estimates from the Fixed Fine Cutoff (Limit 80 km/h omitted)

Dep. variable: P(Traffic offence between time 1-t)

Time (1)	**Mean** (2)	**RDD** (3)	**RDD** (4)	**RDD** (5)	**Obs.** (6)
1	0.026	-0.000	0.000	0.001	74,347
		(0.003)	(0.003)	(0.004)	
2	0.049	-0.004	-0.004	-0.004	74,347
		(0.004)	(0.004)	(0.005)	
3	0.069	-0.006	-0.005	-0.007	74,347
		(0.004)	(0.004)	(0.006)	
4	0.090	-0.007	-0.006	-0.008	74,347
		(0.005)	(0.005)	(0.006)	
5	0.108	-0.010	-0.009	-0.011	74,347
		(0.005)	(0.005)	(0.007)	
6	0.126	-0.011	-0.010	-0.012	71,867
		(0.006)	(0.005)	(0.008)	
7	0.143	-0.008	-0.008	-0.009	68,858
		(0.006)	(0.006)	(0.008)	
8	0.159	-0.011	-0.010	-0.011	65,818
		(0.006)	(0.006)	(0.009)	
9	0.174	-0.010	-0.010	-0.012	62,164
		(0.007)	(0.007)	(0.009)	
10	0.191	-0.005	-0.005	-0.007	57,704
		(0.007)	(0.007)	(0.010)	
11	0.207	-0.005	-0.005	-0.009	54,656
		(0.008)	(0.008)	(0.011)	
12	0.225	-0.006	-0.005	-0.010	50,535
		(0.008)	(0.008)	(0.011)	
Controls			✓		
CCT Estimator				✓	

Note: The table shows the main RDD estimates from the fixed fine cutoff when the limit zone of 80 km/h is omitted. The outcome is the probability that an individual commits another traffic offense 1 to t months after the initial speeding ticket. Column(2) shows the mean of the outcome for those just on the left-hand side of the cutoff. Column (3) presents the RDD estimates obtained using equation 2. Column (4) shows results from a similar analysis but with controls. Column(5) reports biased corrected RDD estimates obtained using the robust approach by Calonico *et al.* (2014b). Column (6) shows the number of observations used in the analysis. In columns (3) and (4), standard errors in the parentheses are clustered at the individual level. For bias-corrected estimates, the confidence intervals are given by the approach of Calonico *et al.* (2014b). Sample construction and data as defined in Section 2.3.

Table B.8: The RDD estimates from the Fixed Fine Cutoff (Limit 80 km/h Included)

Dep. variable: P(Traffic offence between time 1-t)

Time (1)	**Mean** (2)	**RDD** (3)	**RDD** (4)	**RDD** (5)	**Obs.** (6)
1	0.026	0.000	0.000	0.001	117,456
		(0.002)	(0.002)	(0.003)	
2	0.049	-0.002	-0.003	-0.002	117,456
		(0.003)	(0.003)	(0.004)	
3	0.070	-0.003	-0.003	-0.004	117,456
		(0.003)	(0.003)	(0.005)	
4	0.090	-0.005	-0.005	-0.007	117,456
		(0.004)	(0.004)	(0.005)	
5	0.108	-0.007	-0.007	-0.010	117,456
		(0.004)	(0.004)	(0.006)	
6	0.126	-0.008	-0.009	-0.010	113,547
		(0.005)	(0.004)	(0.006)	
7	0.143	-0.007	-0.008	-0.008	109,111
		(0.005)	(0.005)	(0.007)	
8	0.159	-0.009	-0.010	-0.010	103,621
		(0.005)	(0.005)	(0.007)	
9	0.174	-0.012	-0.012	-0.013	98,219
		(0.006)	(0.005)	(0.007)	
10	0.191	-0.011	-0.011	-0.013	91,835
		(0.006)	(0.006)	(0.008)	
11	0.207	-0.009	-0.010	-0.012	86,674
		(0.006)	(0.006)	(0.009)	
12	0.224	-0.010	-0.010	-0.013	79,935
		(0.007)	(0.007)	(0.009)	
Controls			✓		
CCT Estimator				✓	

Note: The table shows the main RDD estimates from the fixed fine cutoff when the limit zone of 80 km/h is not omitted. The outcome is the probability that an individual commits another traffic offense 1 to t months after the initial speeding ticket. Column(2) shows the mean of the outcome for those just on the left-hand side of the cutoff. Column (3) presents the RDD estimates obtained using equation 2. Column (4) shows results from a similar analysis but with controls. Column(5) reports biased corrected RDD estimates obtained using the robust approach by Calonico *et al.* (2014b). Column (6) shows the number of observations used in the analysis. In columns (3) and (4), standard errors in the parentheses are clustered at the individual level. For bias-corrected estimates, the confidence intervals are given by the approach of Calonico *et al.* (2014b). Sample construction and data as defined in Section 2.3.

Table B.9: RDD Specification Checks

Dep. Variable	**RDD** (1)	**RDD** (2)	**CCT** (3)	**RDD** (4)
Jump in fines	203.156	205.605	213.546	213.274
	(8.076)	(7.574)	(10.310)	(12.045)
P(Reoffends)	-0.022	-0.024	-0.027	-0.034
	(0.010)	(0.010)	(0.014)	(0.015)
Controls		✓	✓	✓
Polynomial				✓
Bandwidth	4	4	4.020	4
Mean at the left side	0.131	0.131	0.131	0.131
Observations	31,920	31,899	31,899	31,899

Note: The table shows the RDD estimates with different specifications. In each regression, the outcome is the probability of reoffending within six months. Column (1) shows the main RDD estimates obtained using equation 2. Column (2) presents results when controls are added to the specification. Column (3) shows the results obtained using the approach of Calonico *et al.* (2014b). Column (4) shows results when I add a quadratic term to the RDD specification. Standard errors clustered at the individual level. Sample construction and data as defined in Section 2.3

C Proofs

Proof of Proposition 1

Proof. Individuals solve

$$\max_{x_t} EU_t = u(x_t, \theta) - p \cdot \hat{f}_t(x) \cdot x_x.$$

First order condition giving the optimal speed x^* is

$$\partial u(x,\theta)/\partial x - p \cdot \partial \hat{f}_t(x,\theta)/\partial x = 0 \tag{5}$$

since drivers ignore the discontinuity and use linear approximation of the penalty function. Thus, all individuals locate in their interior optimum where the equation 5 holds. Because θ has a continuous distribution, speeding distribution will be continuous, implying that there exists zero excess mass at the cutoff x^h

□

Proof of Proposition 2

Proof. Individuals solve

$$\max_{x_t} EU_t = u(x_t, \theta) - p \cdot \hat{f}_t(x) \cdot x_x.$$

□

where $\hat{f}_t(x,) = \beta x + f^l \mathbb{1}[x > 0]\}$ is individuals' perceived penalty function at time t. The slope β is an unknown random variable and according to individual's prior beliefs $\beta \sim \mathcal{N}[\mu_\beta, \tau_\beta^{-1}]$. The known constant equals to low fine f^l. When an individual is caught, she receives a fine that they interpret as a signal. The signal takes the form $s_t = \beta_t^f + \eta_t, \eta \sim \mathcal{N}[\mu_s, \tau_s^{-1}]$. If speed is higher than x^h, then the individual receives a large fine (β_t^h). Otherwise, β_t^f is β_t^l. The posterior is given by: $\hat{\beta} = \frac{\tau_\beta \mu_\beta + \tau_s s}{\tau_\beta + \tau_s}$

Proof of Proposition 3

Proof. The optimization problem may be stated as

$$\max_x EU = \theta x - \frac{x^2}{2} - \hat{\beta} \cdot x.$$

Note I have dropped the time subscript since I am considering the steady state. The true penalty function is

$$f(x) = \begin{cases} 0 \text{ if } x = 0 \\ f^l \text{ if } 0 < x \leq x^h \\ f^h \text{ if } x > x^h. \end{cases} \tag{6}$$

If actions converge, they converge to a steady state where two conditions have to hold.

$$\text{First order condition:} \quad \theta - x^* - \beta = 0 \tag{7}$$

$$\text{Surprise function:} \quad \Gamma(x) = f(x^*) - \hat{f}(x^*) = 0, \tag{8}$$

where $\hat{f}(x,) = \beta x + f^l \mathbb{1}[x > 0]$. Equation 7 is just a standard first order condition. Equation 8 implies that in the steady state the fine individual receives has to be equal to the predicted one. Otherwise, she will update the penalty function, and hence the selected speed can not be steady state. In other words, the steady state has to locate in a place where perceived penalty function and true penalty function cross.

We can combine equations 5 and 8 to obtain

$$\theta = x^* + \frac{f^h - f^l}{x^*} \mathbb{1}[x > x^h] \tag{9}$$

It follows directly from equation 9 and continuity of types that there is a continuum of types whose beliefs do not converge, and follow cycles. Note that θ^l is a type whose optimal speed is exactly at x^h

$\square$

D Dose-Response Parameter

My main RDD estimate identifies the average change in reoffending when the fine increases, on average, by 200 euros. The estimate is important and has causal interpretation under standard RD assumption. However, applying the estimate for policy purposes may be challenging. For example, if the question is how much does the reoffending change if the fine increases by one euro, the beta does not offer a direct answer since the variation comes from much larger hikes.

Figures 8(a) and 8(b) demonstrates that high-income individuals experience a larger jump in the size of the fine at the income-based fine cutoff than low-income individuals. Further, after receiving the larger fine, high-income individuals are less likely to recidivate than low-income individuals. The interesting question is, what can we learn from this heterogeneity?

Next, I provide a result clarifying which assumptions guarantee that the comparison of income group-specific RDD estimates identifies the dose-response parameter, which reveals how a marginal change in the size of the fine affects reoffending. My approach closely follows the paper by Callaway *et al.* (2021). More formally, I define that my dose-response parameter is equal to their average causal response on the treated parameter $ACRT(d, \underline{d})$

as

$$ACRT(d_j, \underline{d}) = \mathbb{E}[Y(d_j) - Y(\underline{d})|d = d_j],$$

where d refers to size of the does and $d_j > \underline{d}$. Now assume that we estimate RD parameters δ_h and δ_l for groups whose incomes are h and l such that $h > l$. The difference between δ_h and δ_l identifies the ACRT parameter under standard RD assumptions and one new assumption, which states that individuals who are at the margin of receiving similar doses may react differently to similar doses. However, this difference has to be constant between different levels of treatment. In other words, the new assumption says income is just a level shifter. More formally, the assumption says:

Assumption 1. *For marginal individuals $j, j-1$:*
$\mathbb{E}[Y_j(d^{j-1})] - \mathbb{E}[Y_{j-1}(d^{j-1}))] = \mathbb{E}[Y_j(d^i)] - \mathbb{E}[Y_{j-1}(d^i)] \quad \forall i, .$

The following proposal shows that we can decompose the difference between income group-specific RD estimates into an average causal response parameter and two other differences that capture how individuals with different incomes react to equal fines.

Proposition 4. *Under assumption1 and for doses d^h, d^l, $\underline{d}$, and RDD estimates δ^h, δ^l*

$$\delta^h - \delta^l = \underbrace{\mathbb{E}[Y_h(d^h)|S = s_0] - \mathbb{E}[Y_h(d^l)|S = s_0]}_{\textbf{\textit{Average Marginal Causal Response}}}$$
$$+(\mathbb{E}[Y_h(d^l)|S = s_0] - \mathbb{E}[Y_l(d^l)|S = s_0]) - (\mathbb{E}[Y_h(\underline{d})|S = s_0] - \mathbb{E}[Y_l(\underline{d})|S = s_0]))]$$

Proof. Under the standard sharp RDD assumptions, RDD estimate may be written as

$$\begin{aligned}
\delta &= \mathbb{E}[Y_i(d) - Y_i(\underline{d})|S_i = s_0] \\
&= \mathbb{E}[Y_i(d)|S = s_0] - \mathbb{E}[Y_i(\underline{d})|S = s_0] \\
&= \lim_{s \downarrow s_0} \mathbb{E}[Y_i|S_i = s] - \lim_{s \uparrow s_0} \mathbb{E}[Y_i|S_i = s]
\end{aligned}$$

where S is the running variable, s_0 is the income-based fine cutoff, and $d > \underline{d}$ are doses of fines. Assume we have two income groups, d and l such that d^l. Based on the above, sharp RD estimates for income groups may be written as

$$\begin{aligned}
\delta^h &= \mathbb{E}[Y_h(d^h)|S = s_0] - \mathbb{E}[Y_h(\underline{d})|S = s_0] \\
\delta^l &= \mathbb{E}[Y_l(d^l)|S = s_0] - \mathbb{E}[Y_l(\underline{d})|S = s_0].
\end{aligned}$$

Furthermore, the difference between sharp RDD estimates maybe be written as

$$\delta^h - \delta^l =$$

$$\mathbb{E}[Y_h(d^h)|S = s_0] - \mathbb{E}[Y_h(\underline{d})|S = s_0] - (\mathbb{E}[Y_l(d^l)|S = s_0] - \mathbb{E}[Y_l(\underline{d})|S = s_0])$$

Next, let us add and subtract $\mathbb{E}[Y_h(d^l)|S = s_0]$, and reorder the terms

$$\begin{aligned}
\delta^h - \delta^l &= \\
\mathbb{E}[Y_h(d^h)|S = s_0] &- \mathbb{E}[Y_h(d^l)|S = s_0] \\
+(\mathbb{E}[Y_h(d^l)|S = s_0] &- \mathbb{E}[Y_l(d^l)|S = s_0]) \\
-(\mathbb{E}[Y_h(\underline{d})|S = s_0] &- \mathbb{E}[Y_l(\underline{d})|S = s_0])]
\end{aligned}$$

□

The proposal implies that if the assumption 1 holds, then the comparison of income-specific RD estimates identifies the ACRT parameter. Figure A.19 illustrates the result of proposition 4. We want to compare points A and E, but we do not observe point E. We know that the comparison of points A and B is biased. The RD estimates identify the comparisons A-C and B-D. Suppose it holds that groups h and l react differently to similar doses. However, this difference does not vary with the dose level. In that case, comparisons between differences between A-C and B-D identify ACRT parameters.

There are a few reasons why comparing income-specific RD estimates may be problematic. First, the assumption 1 is strong. Nevertheless, Figure 8(b) shows individuals' reaction to a small fixed fine hike does not systematically vary with income, providing supportive but suggestive evidence for assumption 1. The second concern is that I would have to split the sample into tiny groups to get marginal changes in the size of the fines between groups. However, when I split the sample into quartiles based on income, there is still a large variation in the speeding ticket hikes between groups, but estimates are already relatively imprecise.

To get around the second problem. I utilize a different approach that takes advantage of the fact that an income-based fine system creates a continuum of fines on the right-hand side of the cutoff. I have individuals who are locally randomized to treatment and control groups at the income-based fine cutoff under the standard RD assumptions. In the treatment group, an individual's income sets the size of the fine dose. Hence, observed differences in reoffending between individuals who receive a small and large fine may reflect $ACRT_h(d^h)$. In the control group, everyone receives an equal dose. Thus, I can use the control group to eliminate bias under the assumption 1.

Figures A.20 and A.21 illustrates the idea. Figure A.20 plots the size of the speeding ticket in the control and treatment groups. Next, I use an individual's net annual income to calculate predicted income-based speeding tickets for each individual in the control and treatment groups. The predicted income-based speeding ticket tells what the size of the speeding ticket would be if an individual crossed the income-based fine cutoff. Then, I run OLS regression, where I explain reoffending with predicted speeding tickets separately in the treatment and control groups.

Figure A.21 plots the predicted values from the regression. The black line shows the estimates for individuals who receive an income-based fine. The line has a negative slope implying that those who receive larger speeding tickets are less likely to recidivate. However, it is unclear what part of the association is caused by dose-response and what is due to bias. The blue line plots the same relationship in the control group, where everyone, in reality, receives an equal dose. Since individuals are randomly assigned to treatment and control groups at the cutoff, the change in the slope between the control and treatment groups captures the $ACRT$ if assumption 1 holds.

I can identify the change in the slope at the cutoff using the following regression

$$Y_{il} = \alpha_l + \gamma S_{il} + \theta S_{il} \times Z_{il} + \beta Z_{il} + \delta_1 F_{il} + \delta_2 F_{il} \times Z_{ij} + \boldsymbol{X}'\psi + \epsilon_{il} \tag{10}$$

where Y_{il} is the outcome variable that takes value 1 if an individual has reoffended within t months after the original speeding ticket, S_{il} is the running variable, and Z_{il} is a binary variable indicating whether individual crossed the income-based fine cutoff. Variable F_{il} measures the predicted fines, and $\boldsymbol{X}$ is a vector that may contain controls and other interactions. Our main interest is on parameter β_i, which captures the change in the relationship between predicted fine and reoffending at the cutoff. In other words, it identifies $ACRT(d)$ that captures how the probability of reoffending changes when the fine increases marginally.