The Impact of Sentencing Ranges Design on Sentencing Decisions: An Empirical Analysis*

January 20, 2025

Tereza Burýšková, CERGE-EI†

Abstract

Sentencing ranges, which set the limits of imprisonment length for different crimes, are a common tool to improve fairness in the legal system and reduce unjustified disparities in sentencing. However, recent law and economics literature suggests that these ranges may unintentionally influence sentencing decisions and even introduce new forms of disparity. This paper examines the relationship between sentencing ranges design and sentencing decisions in the Czech legal context, leveraging a recent reform that adjusted sentencing ranges for theft and other property offenses. I use differences in differences and regression discontinuity design to identify the causal impacts of sentencing range design. Consistent with previous research, I find that sentencing ranges significantly influence sentencing outcomes. Specifically, I find evidence for a *severity effect* where cases receive harsher sentences simply due to being placed in a higher sentencing range as well as a *reference effect* where cases within the same range are used as a comparative baseline. I contribute to the current literature by providing evidence obtained with court data for phenomena that so far have been described only experimentally. Furthermore, the results could deepen the understanding of the motivations and mechanisms that drive sentencing decisions, representing an important step in the discussion of optimal sentencing range design.

JEL classification: K14, K42

Keywords: sentencing, sentencing ranges, empirical analysis of court data, sentencing disparities

*This paper is an extended and revised version of the author's master's thesis submitted to Charles University, Faculty of Law in 2024 under the same title.

†née Hofrichterová, Email: tereza.buryskova@cerge-ei.cz, CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, 11121 Prague, Czech Republic.

1 Introduction

Consistent and principled sentencing is an essential feature of the right to a fair trial. However, the sentencing process may be shaped by many biases leading to unjustified disparities in sentencing. These disparities, along with potential solutions to address them, have been a central focus for many legal scholars and economists (e.g., Frankel, 1972; Sporer & Goodman-Delahunty, 2009; Anderson, Kling, & Stith, 1999; Skugarevskiy, 2017; Cohen & Yang, 2019; Tuttle, 2019).

In the Civil law legal system, the policymakers try to overcome sentencing disparities by introducing a system of sentencing ranges. The sentencing range determines the lower and the upper limit for the years of imprisonment that the judge can sentence the offender to. Typically, the sentencing ranges are related to the severity of the particular case (e.g., the damage caused, some characteristics of the victim, or the amount of drug possessed). While these ranges are intended to promote fairness, their design might introduce new complexities into sentencing decisions.

The impact of sentencing ranges on sentences has been studied mostly in the US context, where minimum imprisonment lengths have been linked to both reducing sentencing disparities (Anderson et al., 1999) and increasing racial discrimination (Hofer, 2019; Tuttle, 2019). The US guidelines represent an important example of a legal norm initially introduced to overcome sentencing disparities, which, in the end, opened up another channel that causes them. Another approach to studying sentencing ranges involves comparing cases near a threshold that determines sentencing severity. Bjerk (2017) and Skugarevskiy (2017) examine the cases of drug possession and show notable differences in sentences for offenders with case characteristics just above and below such thresholds. However, the external validity of these findings is limited due to differences in legal systems. In Russia, for example, sentencing ranges are non-overlapping, which creates a sharp distinction between sentences for cases with different legal classifications. In contrast, Czech sentencing ranges typically overlap, allowing the judge to impose similar sentences even for cases around the thresholds.

My paper builds mainly on Drápal and Šoltés (2023), a pioneering analysis of the impact of sentencing ranges on sentences in the Czech environment. Similarly, as Skugarevskiy (2017) and Bjerk (2017), they focus on the around-threshold cases of drug possession and theft. They develop a simple behavioral model of the sentencing process, introducing two main effects driving the decision of the judge — reference and severity effect. They run an online experiment with 200 Czech prosecutors to test the model. In this experiment, they set up several scenarios describing theft or drug possession cases, differing only in the amount of damage caused or drug possessed. These values were conveniently set around a sentencing range threshold. Then, they asked the prosecutors to recommend a sentence for each scenario. They find that the sentences recommended for cases just above the threshold are 10 to 50 percent harsher than those recommended for cases just below the threshold. My contribution extends their work by using actual court data to confirm these effects and analyzing a recent reform of the sentencing range system.

In this paper, I examine the impact of sentencing ranges on sentences using a rich dataset of theft cases judged by Czech courts. First, I exploit a 2020 reform that reclassified certain types of theft, resulting in a quasi-exogenous change in sentencing ranges. Using difference in differences (DD), I find that judges respond to the reform of sentencing ranges by decreasing the sentences, even for cases that remained in the same sentencing range. This indicates that apart from the sentencing ranges serving as severity categories, they also play the role of a reference group the case are compared to.

Second, I run a regression discontinuity (RDD) analysis to estimate the discontinuity in sentencing upon crossing the sentencing range threshold. The results for the before-reform show a substantial upward jump in sentences, which is consistent with the findings of Drápal and Šoltés (2023). For the after- reform sample, the pattern remains unclear.

The key contribution of this paper is the empirical evidence from court data on the impact of sentencing ranges in the Czech Republic. My research not only contributes to the broader literature on criminology and economics of sentencing but also offers insights into judicial decision-making in the Czech context, potentially informing debates on the optimal design of sentencing ranges.

The rest of this paper is organized as follows: Section 2 briefly describes the recent legal context and the impact of the reform; Section 3 explains the intuition about the underlying mechanisms; Section 4 introduces the dataset used in the empirical analysis; Section 5 presents the results of the empirical analysis, and Section 6 discusses their implications. The main contribution is summarised in the Conclusion.

2 Legal context

2.1 Punishment of theft

Since the Czech Republic belongs to the Civil law legal system, the Czech Criminal Code, Act No. 40/2009 Coll. (referred to as the Criminal Code henceforth) determines the punishment for different crimes. Therefore, only limited discretion is left to the judge, who has to comply with a broad set of rules prescribed by the Criminal Code.

Theft is the most common crime in the Czech Republic, with 394,182 cases reported between 2006 and 2023. In general, theft is committed by misappropriating another person's property. In most theft cases, the sentencing range is determined solely by the damage caused; I refer to these as *ordinary cases*. For completeness, let me note that the Criminal Code also identifies several special qualification circumstances — such as pickpocketing, burglary, or recidivism —

that warrant harsher penalties, irrespective of the damage caused. However, for simplicity, my empirical analysis focuses exclusively on *ordinary cases* of theft.

For completeness, it should be noted that the Criminal Code allows the judge to choose an alternative punishment or to conditionally suspend the sentence to imprisonment. I describe how I deal with cases that were punished by alternative means in Section 5.1.1.

2.2 The 2020 reform

In October 2020, the provision of the Criminal Code was significantly modified. In particular, the definition of the terms determining the extent of damage shifted towards higher values of actual damage. The rationale behind this reform was to incorporate inflation (which has not been incorporated since the Criminal Code was passed in 2009) and adjust the thresholds to the current price level.

The change of term definitions implies different legal classifications for cases before and after the reform. For example, a case with damage of 75k CZK would be classified as a case with *larger damage* before the reform and would be punished by 1-5 years of imprisonment. However, after the reform, the damage would be classified only as *not small* and would be punished by 0-2 years of imprisonment only.

In this paper, I interpret this reform as a shift in sentencing ranges, abstracting from the fact that it, strictly speaking, changed the legal classification by redefining damage quantifiers. Table 1 summarises the punishment for the ordinary cases before and after the 2020 reform. The sentencing ranges are always designed to overlap at their thresholds. This, in principle, allows the judge to impose the same sentence for cases just above and just below the threshold.

Table 1: Sentencing ranges for ordinary theft cases in the Criminal Code (own summary based on the Criminal Code).

	sentencing range	
damage (CZK)	till September 2020	after October 2020
less than 5k	not a criminal offense	not a criminal offense
5k-10k	0-2 years	
10k-50k		0-2 years
50k-100k	1-5 years	
100k-500k		1-5 years
500k-1m	2-8 years	
1m-5m		2-8 years
5m-10m	5-10 years	
10m		5-10 years

Table 1 shows that the effects of the 2020 reform were twofold. For some cases, the sentencing range shifted (e.g., cases with damage 500k-1m CZK face a sentencing range of 1-5 years instead of 2-8 years). For other values of damage, the sentencing range itself did not change; however, more severe cases were added to that sentencing range (e.g., cases with damage 100k-500k CZK face the same sentencing range of 1-5; however, the same sentencing range now relates also to cases with damage 500k-1m CZK, which are relatively more severe). I denote these two groups as Treatment A cases (sentencing range shift) and Treatment B cases (addition of more severe cases), respectively.

This distinction aligns with the concepts of the severity and reference effects that influence the sentencing process (see Section 3 for an explanation). Analyzing changes in sentencing patterns for these two groups separately will help identify these effects and provide insight into the broader decision-making process.

Table 2: Two different types of ordinary theft cases in terms of reform effects

	sentencing range	
damage (CZK)	before reform	after reform
Treatment type A		
5k-10k	0-2 years	not a criminal offense
50k-100k	1-5 years	0-2 years
500k-1m	2-8 years	1-5 years
5m-10m	5-10 years	2-8 years
Treatment type B		
less than 5k	not a criminal offense	not a criminal offense
10k-50k	0-2 years	0-2 years
100k-500k	1-5 years	1-5 years
1m-5m	2-8 years	2-8 years

3 Mechanisms in sentencing: Severity and reference effect

In this section, I describe the mechanisms behind sentencing within a system of sentencing ranges. Specifically, I build on Drápal and Šoltés (2023), who introduce a theoretical model of sentencing that introduces severity and reference effect and demonstrates how these effects shape the sentences for cases close to a sentencing range threshold. Nevertheless, since the main contribution of my paper is the court data analysis, I do not expand their theoretical model further. Instead, I explain their intuition and apply it to formulate predictions for my empirical setting.

Drápal and Šoltés (2023) build mostly on Leibovitch (2017), extending her notion of statistical curving. They interpret the sentencing ranges as both — categorical indicators of the approximate severity of the crime and, at the same time, reference groups within which the cases are compared to each other. They denote these different roles of sentencing ranges as severity effect and reference effect, respectively.

The authors illustrate the severity and the reference effect using the around-threshold cases as an example. For instance, consider two theft cases. Case C with damage 99k CZK and Case D with damage 101 CZK. Assume that apart from the damage, these cases are identical in all other characteristics. Then, the current provision of the criminal code assigns C to the sentencing range of 0-2 years, and D to the sentencing range of 1-5 years. When determining the punishment for these two cases, the higher sentencing range for case D signals its increased severity which increases the sentence imposed (severity effect). On the other hand, case D is compared to more severe cases falling into the same sentencing range, which decreases the sentence (reference effect). Clearly, in this example, these two effects work against each other. Potentially, one can determine which effect prevails by comparing the sentences for cases C and D. If the sentence for case C is higher than the sentence for case D, we could conclude that the reference effect dominates. Conversely, if the sentence for case D exceeds the one for case C, it demonstrates the dominance of the severity effect. In this paper, I apply this intuition to study discontinuities at sentencing range thresholds using the RDD approach.

Furthermore, I extend this intuition to a novel empirical context of the reform and its dual effects on theft cases.

First, consider a group of cases whose legal classification (and the corresponding sentencing range) remains unchanged; however, more severe cases are added to that range. Then, since the legal classification is the same, the impact of the severity effect remains constant; however, the original cases now seem to be less severe compared to the cases that were added. Thus, the reference effect kicks in and should result in a decrease in sentences for the original group. This is exactly the case for the Treatment B cases defined in Section 2 and used in the empirical analysis. In light of the intuition presented, we should observe a decrease in sentences for this group.

Second, consider a group of cases where the sentencing range decreased, and at the same time, less severe cases were added to the sentencing range. This situation aligns with the Treatment A cases discussed in Section 2. Since these cases are now classified as less severe, we would expect to see a decrease in their sentences. However, because these cases now represent the most severe within their new range, the reference effect could drive sentences upward. Thus, in this case, the severity and reference effects are working in opposite directions and the change in sentences depends on their relative importance.

Table 3 summarizes the intuitive predictions for different exercises that I conduct in my empirical analysis.

Table 3: A summary of how severity and reference effect shape the sentences.

	Severity effect	Reference effect
	Impact on Mean Sentence	
Treatment A *(Sentencing Range Downward Shift)*	↓	↑
Treatment B *(Addition of More Severe Cases)*	0	↓
	Discontinuity in Mean Sentence	
Around threshold cases	+	−

Nevertheless, when applying the intuition developed for experimental settings to observational data analysis, one needs to be careful about the underlying assumptions. Here, the main concerns are related to other circumstances of the case. In the experimental setting leveraged by Drápal and Šoltés (2023), it was possible to set these circumstances to be absolutely identical for the cases considered. Here, I am dealing with cases differing in multiple dimensions. In this paper, I overcome this issue by introducing controls (by which I deal with the observable variables) and employing DD and RDD identification strategies.

4 Court data

In the Czech Republic, all criminal cases are well-documented, and the case-level data is available for research purposes. This dataset contains information on many aspects of the case that are potentially useful for my research. In particular, three main sets of variables are available for each case. First, there is data about the criminal procedure itself, including the court and the senate that passed the sentence and all important procedural steps; second, the data about the offense — mainly its legal classification and corresponding section and paragraph in the Criminal Code and the damage caused where relevant; third, data about the defendant (ethnicity, gender, etc.). Since this data is directly reported by the court officers and captures the evaluation of all evidence presented, it should be of sufficient quality without much systematic bias.

Technically, the dataset captures the period 2006-2023. However, the damage caused, which is central to my analysis, has been reported only since 2019. Appendix Figure A.1 shows that a stable report rate of around 40 % emerged by the beginning of 2020.

Table 4 presents the descriptive statistics of the dataset. By recidivist, I denote the offenders where the court counted the offender's previous convictions as an aggravating circumstance.[1] The year range is already limited to 2019-2023.

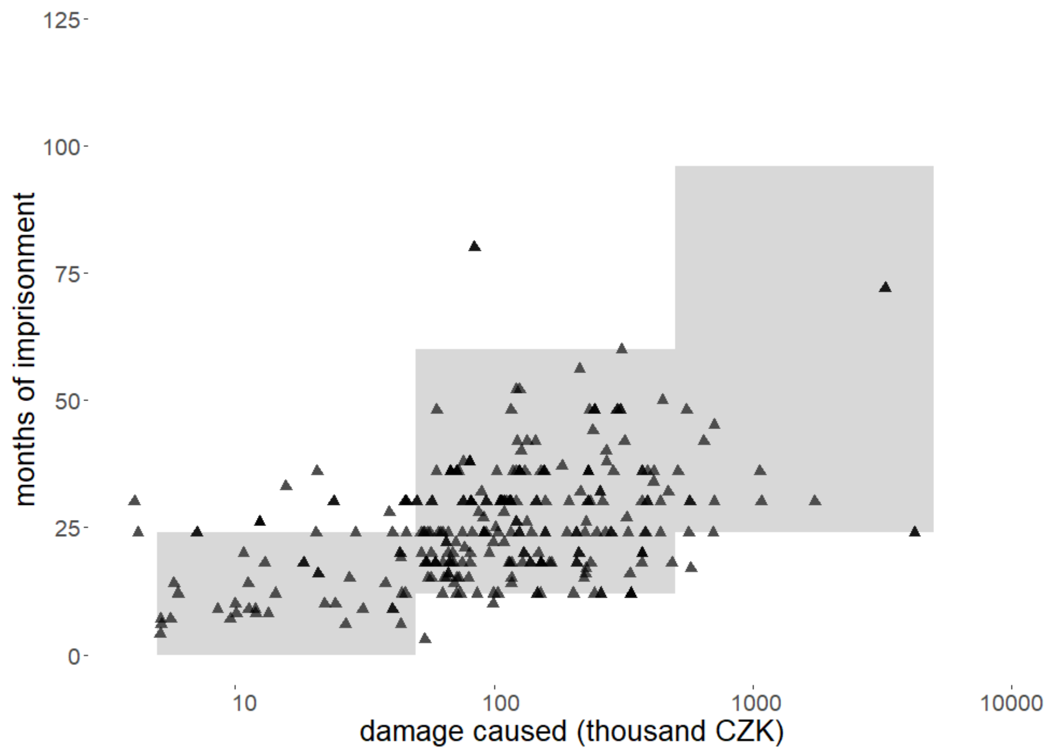
Table 4: Descriptive statistics of 2019-2023 theft cases used in the main empirical analysis. Ordinary cases are defined as cases where the criterion determining the sentencing range was the damage caused.

	all		ordinary cases	
	before	after	before	after
n	22,371	35,672	7,047	8,611
n damage filled	2,921	16,252	936	4,095
n unconditional imprisonment	9,080	16,027	1,951	2,669
n conditional suspension of imprisonment	6,783	8,992	3,477	3,838
n other punishment	6,508	10,653	1,619	2,104
damage (thousand CZK)	68.1	70.4	122.4	159.5
unconditional imprisonment (m)	17.16	15.97	25.60	24.95
conditionally suspended imprisonment (m)	10.16	10.51	9.89	10.41
offender age	32.4	33.4	33.0	33.8
recidivist (%)	11.2	12.2	6.7	7.2
offender male (%)	83.0	84.7	79.7	82.5

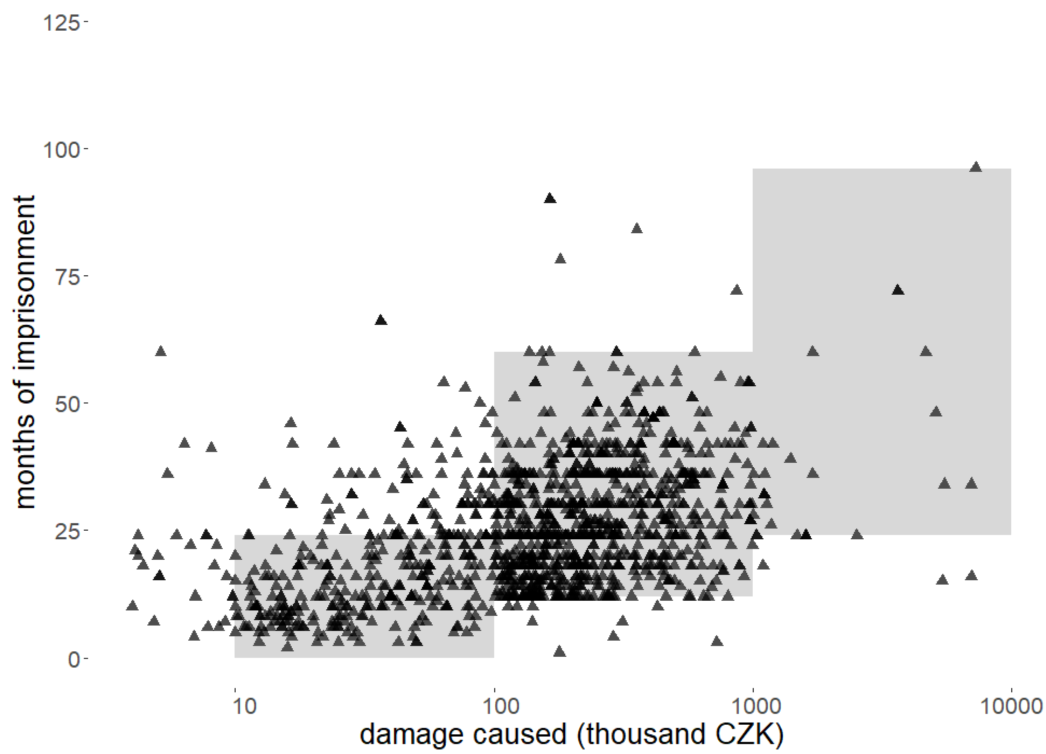
This table shows that theft is often punished by alternative means of punishment, including a conditionally suspended sentence. Moreover, in line with the previous literature (Drápal, 2023), I find that the conditionally suspended sentences are systematically lower than the standard conditional sentences suggesting that judges may percieve these as distinct punishment types. For a detailed discussion of how I incorporate conditional and unconditional imprisonment, see Section 5.1.1.

I conclude the description of the data with Figure 1, which provides the general relationship between damage caused and the length of unconditional imprisonment.

[1] Under the provision of the Criminal Code, it is the discretion of the court whether to consider the previous conviction as an aggravating circumstance.



(a) Before reform (n=494)



(b) After reform (n=2215)

Figure 1: Imprisonment length for ordinary cases as a function of damage caused. Only cases punished by unconditional imprisonment are shown. Rectangles represent the statutory sentencing ranges.

5 Results

5.1 DD

In the main part of my empirical analysis, I rely on the standard econometric method of difference in differences (DD). This method requires the introduction of a control group that is as similar as possible to the treatment group and has not been affected by the reform. Unfortunately, the 2020 reform was quite massive and influenced the sentencing ranges for all cases of theft and all remaining crimes against property. Therefore, I use the obstruction of justice and obstruction of a sentence of banishment (§ 337 of the Criminal Code). This crime represents the second most frequently committed crime overall. Typically, it is committed when the offender acts contrary to some decision of the court or some other authority (for instance, driving after receiving a driving ban, etc.). The legal definition of this offense is absolutely independent of the damage caused; thus, arguably, criminal cases should not be influenced by the 2020 reform. The main advantage of considering this control group is that (similar to theft), these cases are quite frequent, and judging them is part of the judges' routine.

Given that most cases punished by unconditional imprisonment lie between 100k and 1 m, I focus on this subsample of cases. I denote cases with damage between 500k and 1m as the Treatment A sample (sentencing range shift) and cases with damage between 100k and 500k as the Treatment B sample (addition of more severe cases into the given sentencing range). However, I confirm the same patterns also with different sample choices (see Appendix Section A.3).

The DD method relies on the validity of the parallel trends assumption (that the trends in treatment and control groups would be parallel). I address this assumption by carefully examining the sentence-related variables before and after the treatment. If the parallel trends assumption holds, we should not observe any apparent differences in the evolution of sentences between the treatment and control groups in the before-reform period.

5.1.1 Parallel trends on the intensive and extensive margin

The DD method relies on the validity of the parallel trends assumption. I address this assumption by examining the sentence-related variables before and after the treatment. If the parallel trends assumption holds, we should not observe any apparent differences in the evolution of sentences between the treatment and control groups in the before-reform period. I plot the evolution of unconditional imprisonment sentences for each group to check for the general trends in sentencing in both treatment groups and the control group in the before-treatment period (Figure 2). This simple visual comparison shows that the monthly averages of sentences for both control groups seem to be quite noisy; nevertheless, there are no apparent trends in the before-treatment period.

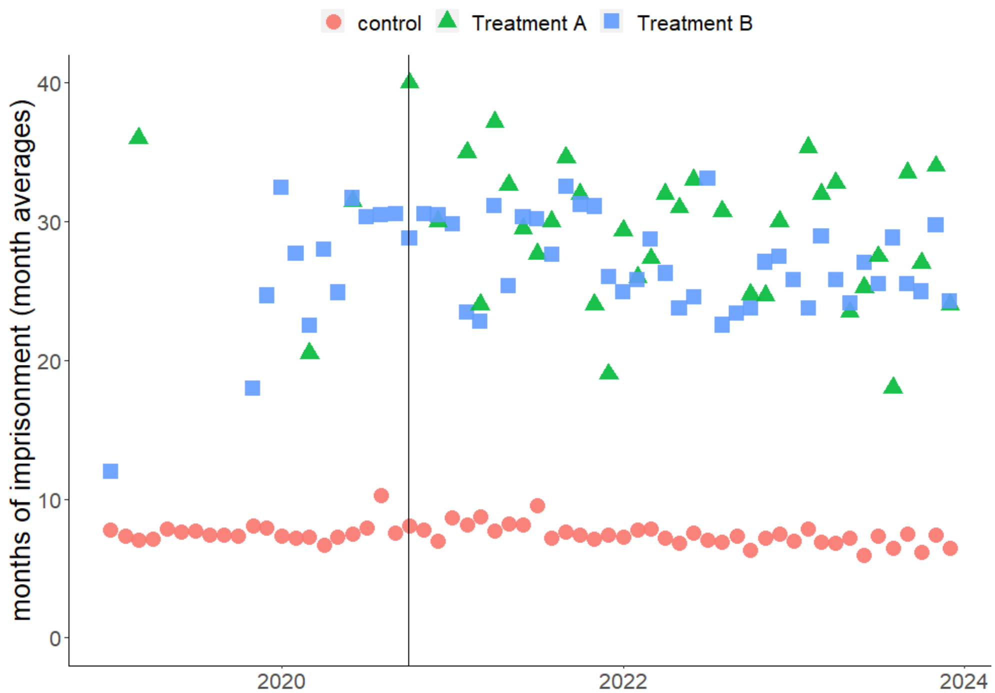


Figure 2: Monthly average unconditional sentence evolution for control and both treatment groups. The black vertical line represents the 2020 reform.

Since a considerable number of cases are punished by other types of imprisonment or the imprisonment is conditionally suspended, I need to decide how to deal with such cases in my empirical analysis. Therefore, I examine the extensive margin of sentencing in more detail.

First, I plot the evolution of the rates for cases punished by unconditional imprisonment sentences and conditionally suspended sentences in Figures 3 and 4.

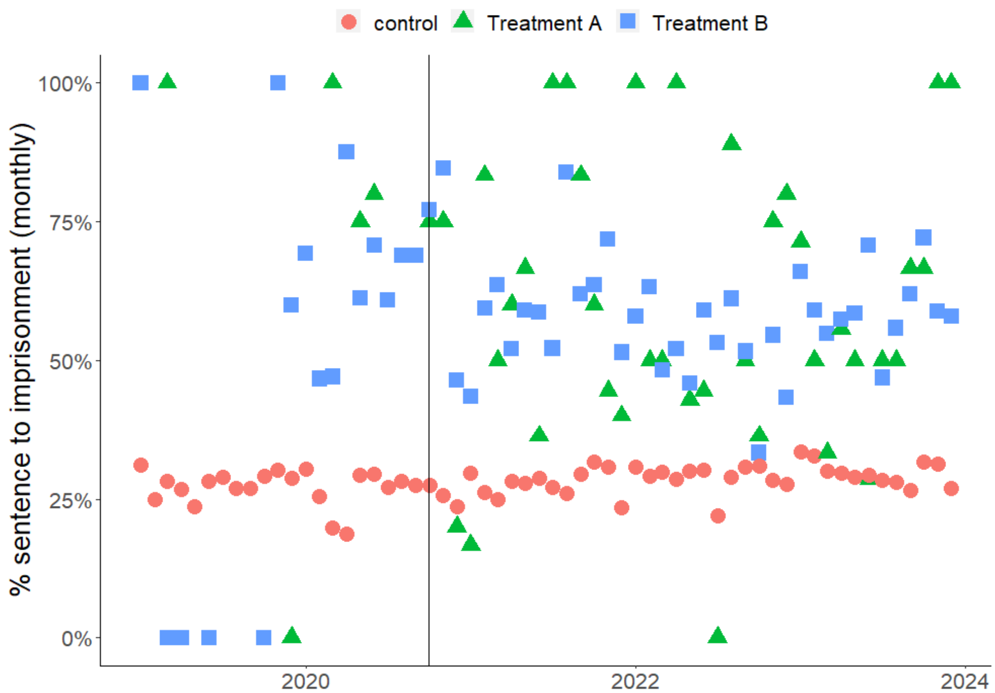


Figure 3: Monthly unconditional sentences to imprisonment rate for control and both treatment groups. The black vertical line represents the 2020 reform.

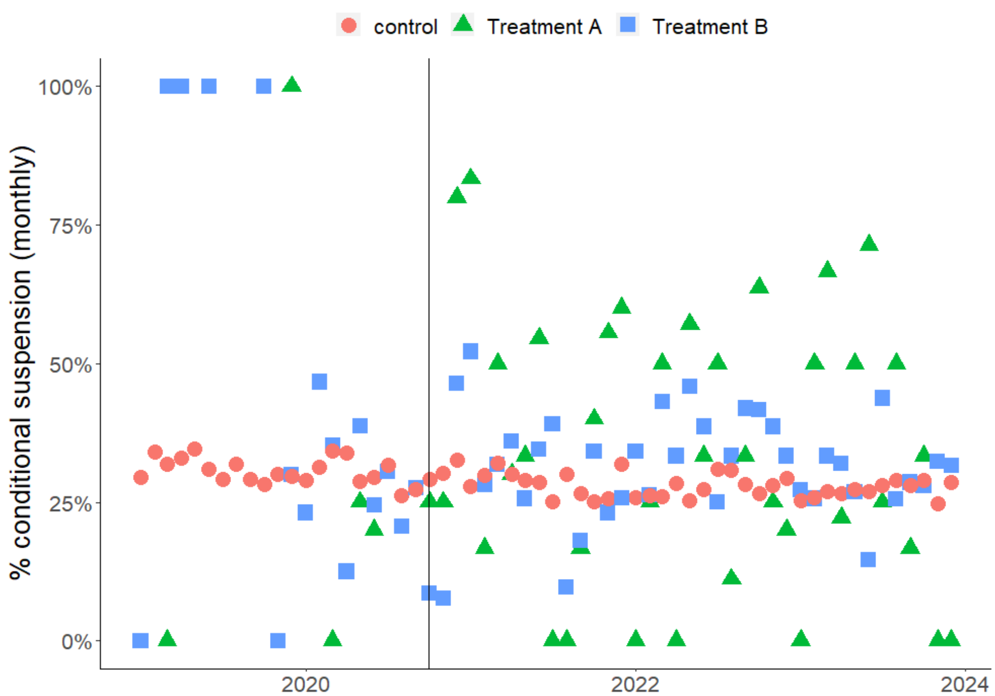


Figure 4: Monthly conditional suspension of sentence rates for control and both treatment groups. The black vertical line represents the 2020 reform.

The imprisonment rate in the control group appears to be stable throughout the time period. However, the imprisonment rates in the treatment groups are quite noisy, and it is not very clear how and whether the reform affected them. There seems to be a slight pattern of an unconditional imprisonment rate decrease after the reform.

Similarly, when plotting the conditional suspended imprisonment rates, the estimates remain noisy; however, the conditional suspension rate seems to grow slightly in the after-treatment period.

To explore the sentencing patterns further, I plot the evolution of the conditionally suspended sentence lengths (Figure 5). Overall, the conditionally suspended sentences seem to be substantially less severe than the unconditional sentences.

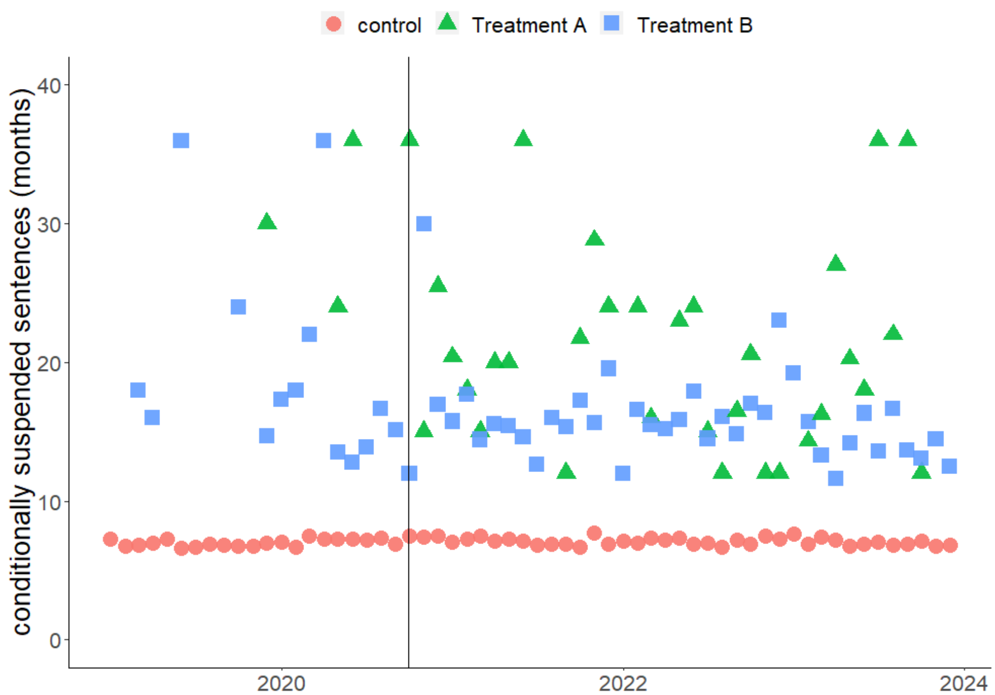


Figure 5: Monthly average unconditional sentence evolution for control and both treatment groups. The black vertical line represents the 2020 reform.

Given these descriptive analyses and previous research performed by legal professionals (Drápal, 2023), I conclude that the judges may not apply the same measures when sentencing to conditional or unconditional imprisonment. Instead, it seems that the judges perceive these as two separate types of punishment, with the first one being less severe.

Therefore, to ensure credibility and a clear interpretation of my results, I deal with the alternative means of punishment in the following manner. For cases that received alternative punishments or conditionally suspended sentences, I assign a sentence value of 0. For cases punished with unconditional sentences, I use the non-zero sentences recorded in the data. This approach accounts for situations where judges may reduce punishments to the extent that incarceration is no longer necessary. Additionally, I focus on a higher damage range to ensure that a significant proportion of cases receive unconditional imprisonment sentences.

5.1.2 DD regression

First, I estimate the treatment effects using the standard DD regression with a dummy after-treatment period indicator (Table 5).

$$S_i = \beta_0 + \beta_1 P_i \cdot T_i + \beta_2 T_i + \beta_3 P_i + \sum_{j=4}^{k} \beta_j X_{ij} + e_i, \tag{1}$$

where S_i is the unconditional sentence (defined as described in Section 5.2.1), P_i is the dummy indicating the after-treatment period and T_i is the treatment, X_i represents a set of covariates.

The DD estimates show a decrease in sentences in both treatment groups after the reform. The average sentence for Treatment A cases (sentencing range decrease) dropped by 5 months, whereas the drop for Treatment B cases (addition of more severe cases) is around 1 month. These results are statistically significant with and without controls.

Table 5: Estimates of the treatment effect obtained using DD approach. Controls include judge fixed effects, age, number of previous convictions, number of different punishments for the given crime, and concurrence, recidivism, juvenile and gender dummies of the offender.

	Dependent variable:	
	sentence	
Panel A: Treatment A		
	(1)	(2)
After:Treatment	-5.758^{***}	-4.506^{***}
	(0.492)	(0.463)
Intercept	Yes	Yes
Controls	No	Yes
Observations	45,161	45,160
R^2	0.023	0.163
Adjusted R^2	0.023	0.154
Residual Std. Error	5.373 (df = 45157)	5.001 (df = 44634)
Panel B: Treatment B		
	(1)	(2)
After:Treatment	-0.995^{***}	-0.912^{***}
	(0.356)	(0.335)
Intercept	Yes	Yes
Controls	No	Yes
Observations	46,544	46,543
R^2	0.0002	0.142
Adjusted R^2	0.0001	0.132
Residual Std. Error	5.274 (df = 46540)	4.914 (df = 46015)
Note:		*p<0.1; **p<0.05; ***p<0.01

5.1.3 Event study

Additionally, I estimate the DD model using an event-study approach. In particular, I divide the cases according to the quarter when the sentence was passed.[2]. Then, for each quarter q, I run this regression

$$S_i = \alpha_q + \beta_q T_i + \sum_{j=1}^{k} \gamma_{jq} X_{ij} + e_i, \tag{2}$$

where T_i is the dummy indicating treatment group, X_i represents a set of covariates. The covariates include judge fixed effects, number of previous convictions, age of the offender, concurrence dummy, and the number of different punishments for the given crime. My interest falls on the coefficients β_q. I normalize these coefficients by taking the coefficient one period before the reform β_{-1} as the baseline level.

Figure 6 shows the main event study plot for both treatment groups. The coefficients are summarised in Table 6. In the before-treatment period, the coefficients for Treatment B do not differ significantly from zero which speaks towards the validity of the parallel trends assumption. Conversely, for Treatment A, the before-treatment coefficients are quite noisy and often significantly lower than zero. Nevertheless, this pattern may be driven by a lack of data for the before-treatment period.

In the after-reform period, both treatment groups showed a clear decreasing trend. For Treatment A, the sentences dropped instantly with the introduction of the reform, while for Treatment B, the drop was more gradual.

Overall, the results of the DD analysis confirm a significant drop in sentences in both treatment groups. This finding can be interpreted through the lens of the mechanisms described in Section 3. The decrease associated with sentencing range downward shift provides evidence for the severity effect of sentencing ranges, whereas the decrease associated with the addition of more severe represents a sign of a reference effect.

In the Appendix, I present several robustness checks of my DD results. First, in Section A.2, I replicate the DD analysis using two different control groups. Even though the event study plots become quite noisy, I still replicate the main pattern using the binary treatment indicator. Additionally, in Section A.3 I pool different samples of treated cases, which seems to confirm the results obtained with the original sample.

[2]Unfortunately, a finer partition was not possible due to the amount of data before reform.

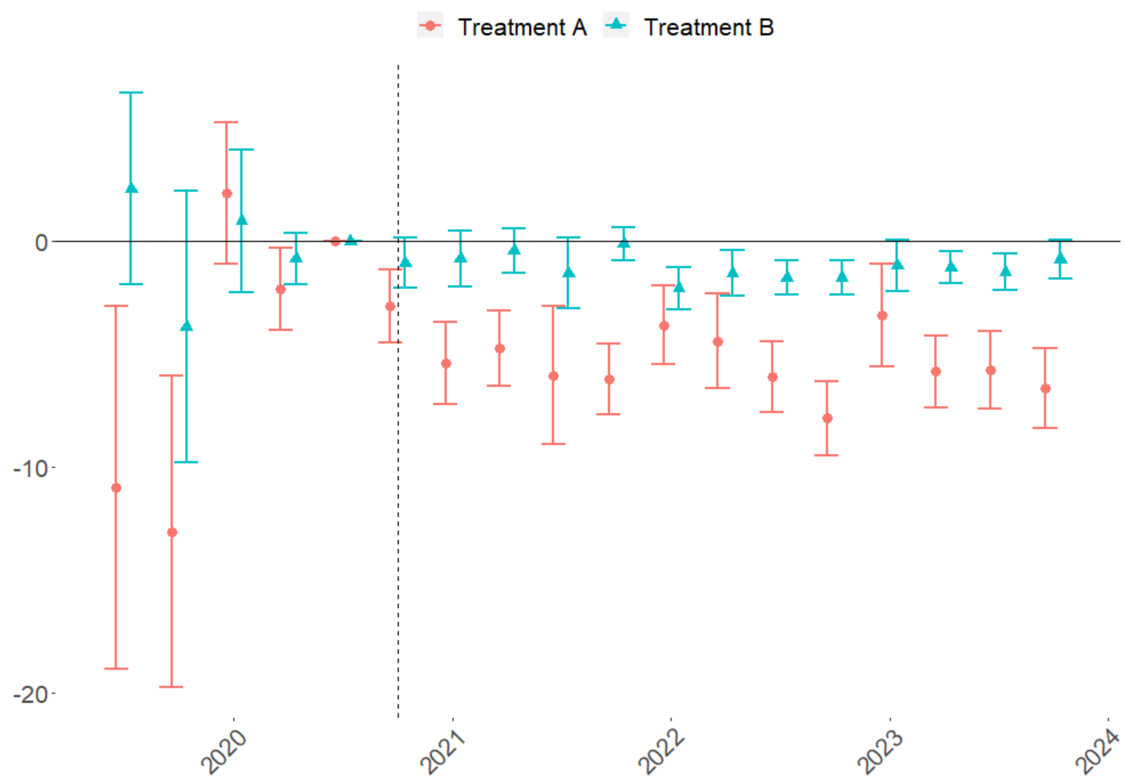


Figure 6: The quarterly effects on sentence. The baseline rate corresponds to Q3 2020. The dashed vertical line represents the 2020 reform. The regressions control for judge fixed effects, the number of previous convictions, the age of the offender, and the number of different punishments for the given crime. 95 percent confidence intervals are plotted.

Table 6: The quarterly effects on sentence. The baseline rate corresponds to Q3 2020. The regressions control for judge fixed effects, number of previous convictions, age of the offender, concurrence dummy, and the number of different punishments for the given crime.

	β_q	
	Treatment A	Treatment B
Q2 2019	-14.079***	-1.011
	(4.458)	(4.170)
Q3 2019	-10.887***	2.354
	(4.105)	(2.161)
Q4 2019	-12.862***	-3.773
	(3.522)	(3.061)
Q1 2020	2.132	0.921
	(1.603)	(1.617)
Q2 2020	-2.093**	-0.759
	(0.919)	(0.568)
Q3 2020	baseline	
Q4 2020	-2.863***	-0.949*
	(0.832)	(0.562)
Q1 2021	-5.376***	-0.746
	(0.928)	(0.634)
Q2 2021	-4.709***	-0.409
	(0.855)	(0.506)
Q3 2021	-5.916***	-1.390*
	(1.548)	(0.800)
Q4 2021	-6.089***	-0.095
	(0.794)	(0.371)
Q1 2022	-3.707***	-2.065***
	(0.894)	(0.473)
Q2 2022	-4.412***	-1.388***
	(1.075)	(0.525)
Q3 2022	-5.981***	-1.590***
	(0.797)	(0.398)
Q4 2022	-7.831***	-1.584***
	(0.840)	(0.381)
Q1 2023	-3.273***	-1.051*
	(1.157)	(0.578)
Q2 2023	-5.754***	-1.152***
	(0.818)	(0.356)
Q3 2023	-5.676***	-1.344***
	(0.874)	(0.420)
Q4 2023	-6.485***	-0.775*
	(0.900)	(0.444)
Note:	*p<0.1; **p<0.05; ***p<0.01	

5.2 RDD

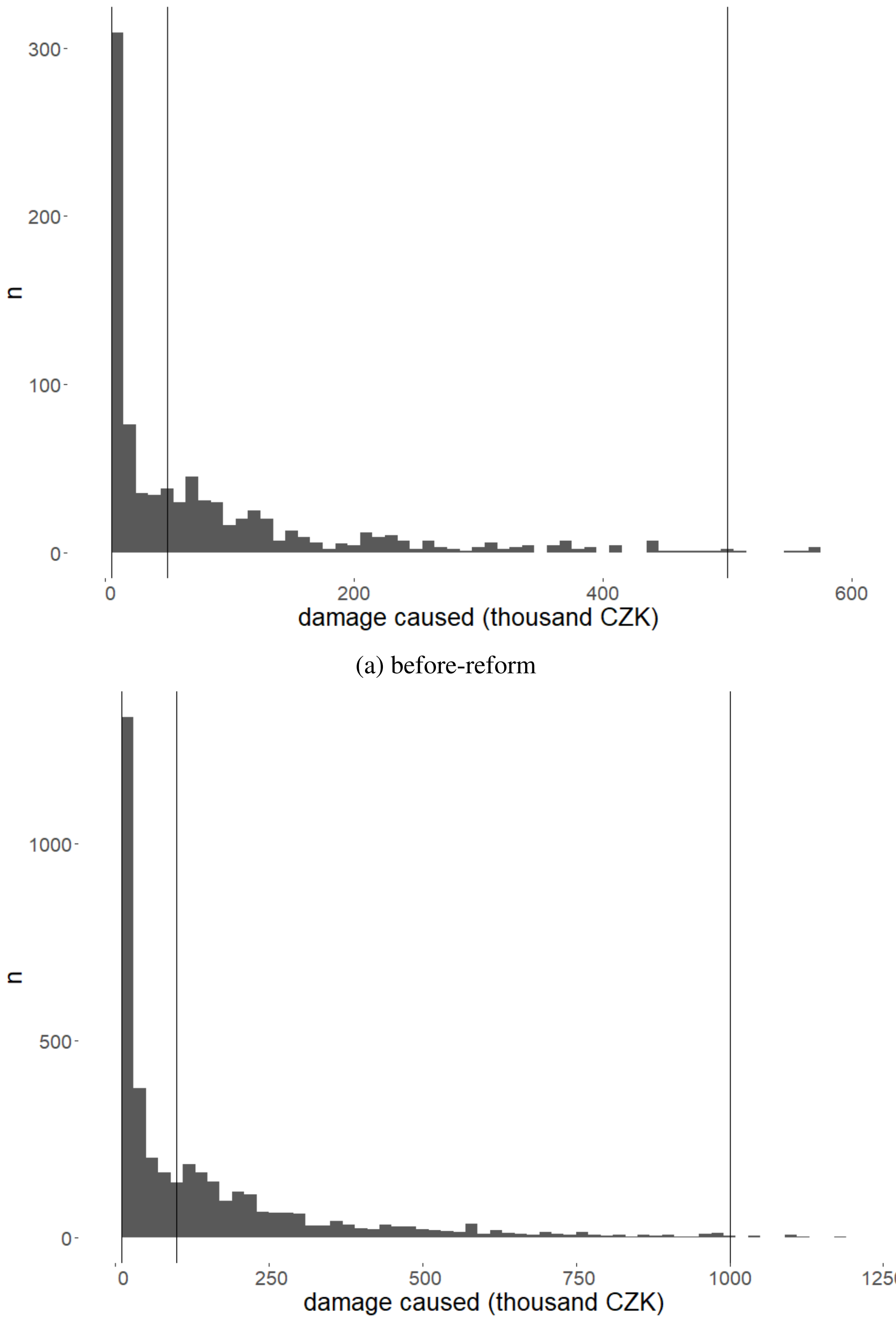
I use the regression discontinuity design (RDD) as complementary evidence to the reform analysis. I focus on the before- and after- reform periods separately. Following the standard RDD setup, I use the damage caused as a running variable, which has thresholds that split the sample into a different type of treatment (in this case, a different sentencing range). The RDD approach estimates the discontinuity in the mean sentence at the sentencing range threshold.

5.2.1 Underlying damage distribution

This approach assumes that the reported damage in cases is not systematically manipulated around the relevant thresholds. This concern is particularly significant given that variables in crime reports have been documented to suffer from manipulation in some contexts (Travova, 2023). Figure 7 shows the histogram of damage in my dataset. The distribution of cases implies that there are only two reasonable thresholds with a sufficient amount of cases around them for each group of cases. The density peaks around 10k (and 5k), where the theft becomes a criminal offense. This peak may be caused by the officials manipulating the value of damage so that it becomes a criminal offense or a survivorship bias — many cases with damage below 10k are not even reported to the police. Nevertheless, in my analysis, I do not focus on the 10k threshold, so the bias there does not affect the validity of my results.

The assumption of no manipulation at the threshold can be tested using McCrary (2008) sorting test. The rationale is to test the hypothesis that the probability density function of damage has a discontinuity at the threshold. Figures 8, 9 present the results.[3] Importantly, I do not find any significant discontinuity in damage distribution at either threshold.

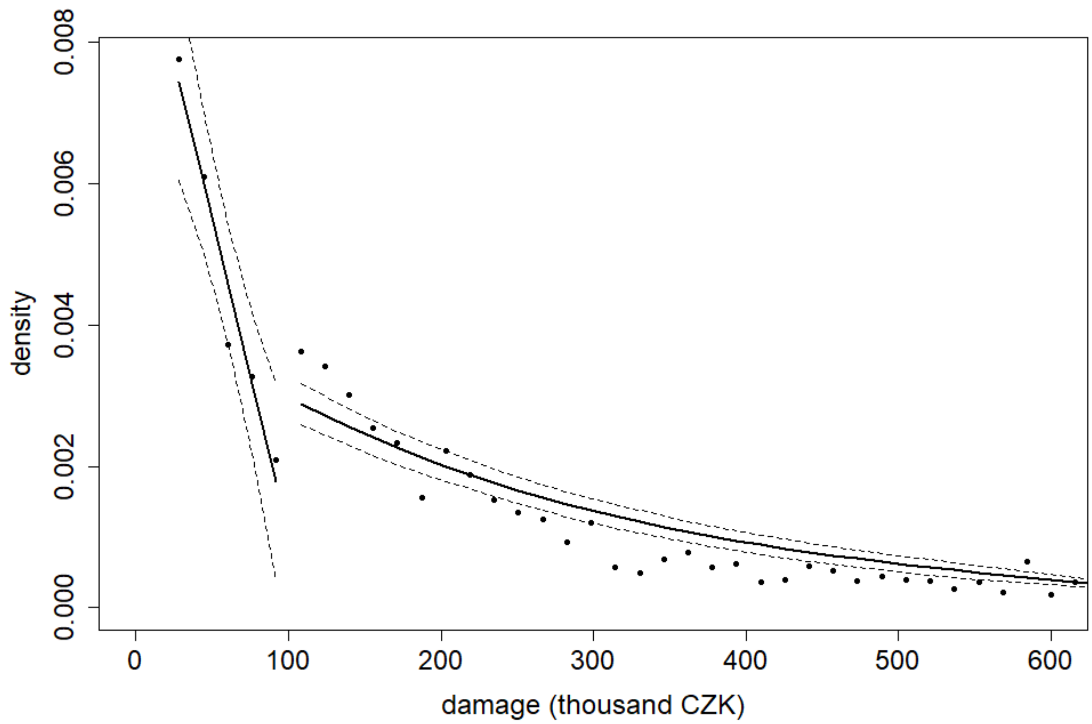
[3]To avoid any bias caused by cases with very low damage, I dropped cases with damage $<$25k CZK for the 100k threshold; for the threshold 1m, I dropped cases with damage $<$ 300k CZK.



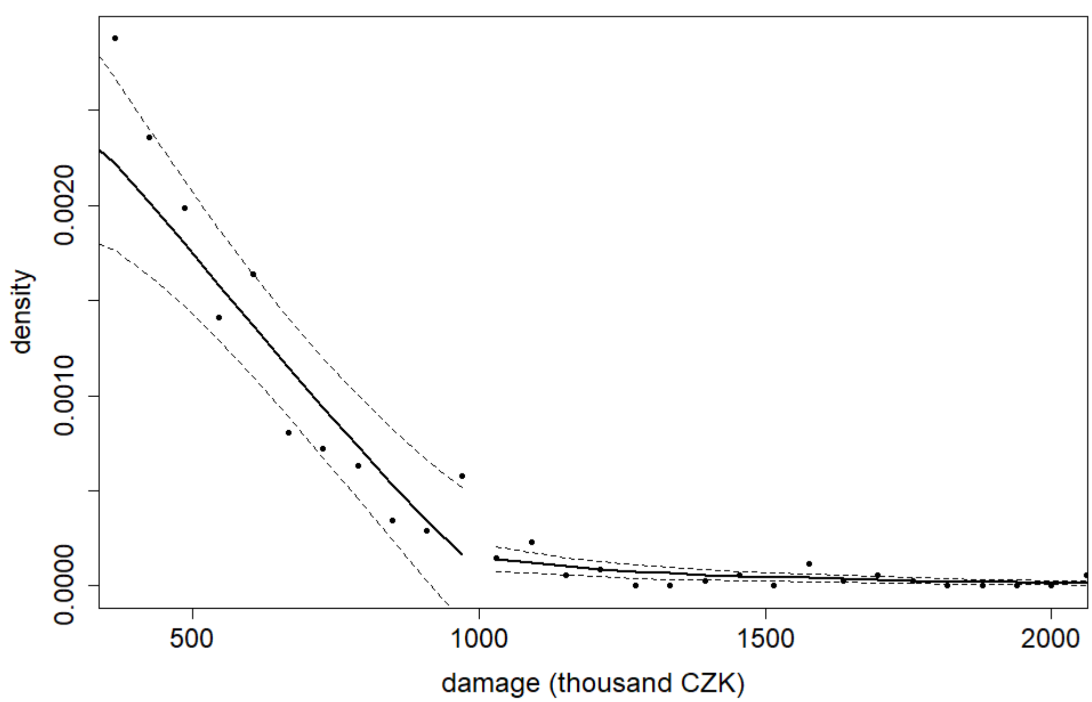
(a) before-reform

(b) after-reform

Figure 7: The histogram of damage for the before- and after- reform period

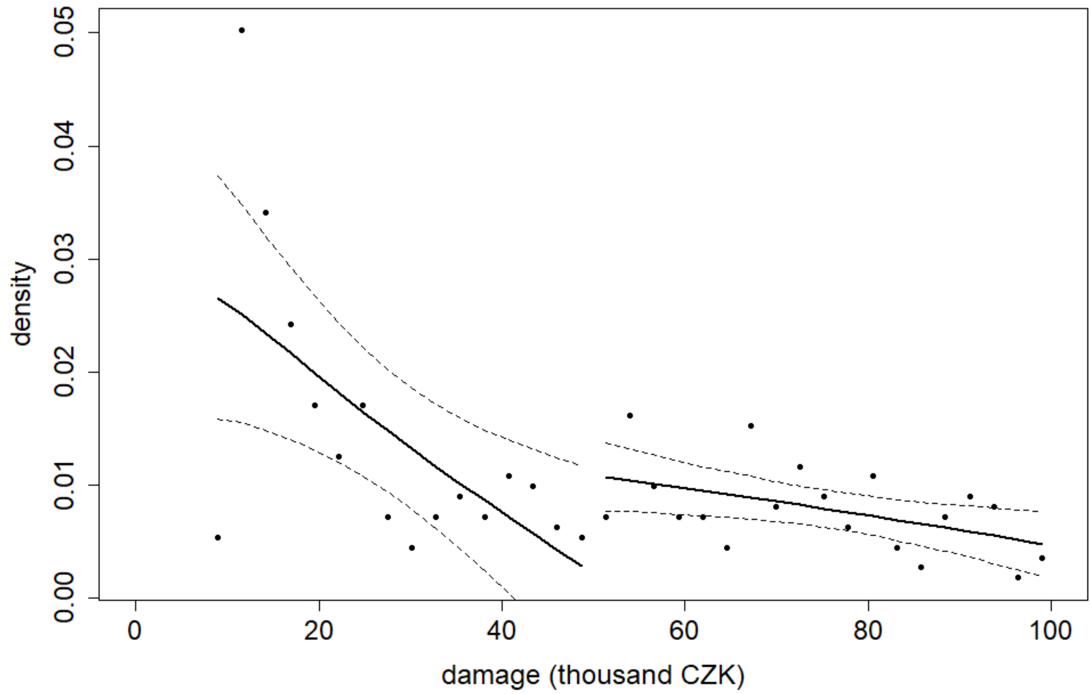


(a) 100k CZK threshold, binsize=16, bandwidth=500, McCrary's test p-value 0.14

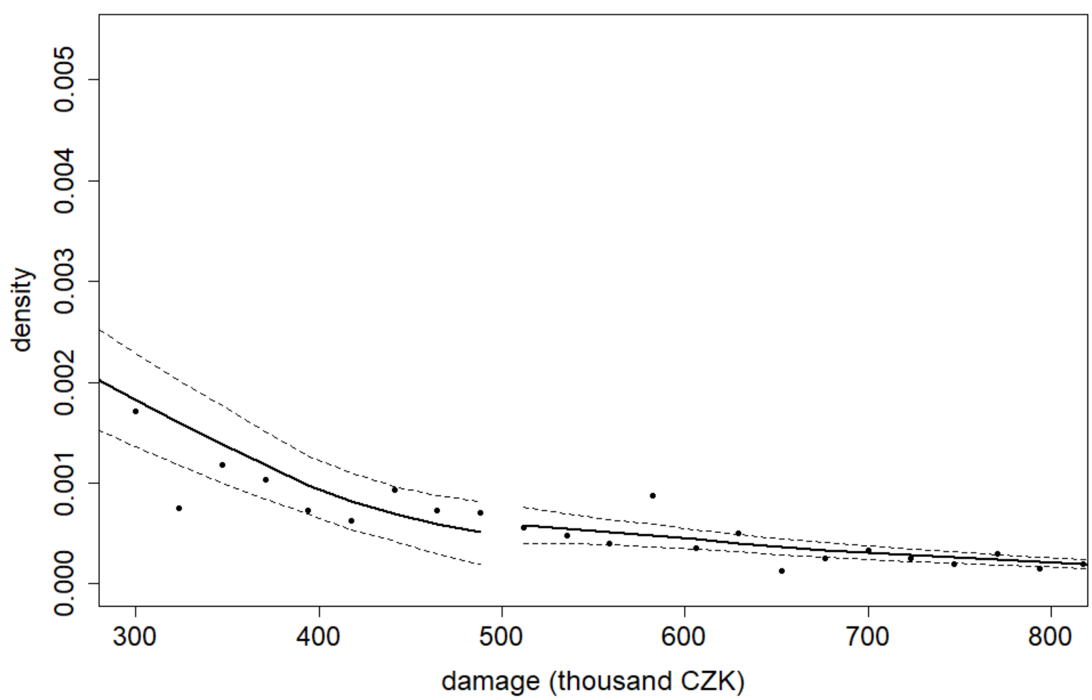


(b) 1m CZK threshold, binsize=61, bandwidth=750, McCrary's test p-value 0.11

Figure 8: Sample probability density of damage caused for after-reform cases around two sentencing ranges thresholds.



(a) 50k CZK threshold, binsize=2.6, bandwidth=100, McCrary's test p-value 0.40



(b) 500k CZK threshold, binsize=24, bandwidth=300, McCrary's test p-value 0.25

Figure 9: Sample probability density of damage caused for before-reform cases around two sentencing ranges thresholds.

5.2.2 Discontinuity estimation

After examining the underlying damage distribution, I run the standard RDD regression. In particular, I nonparametrically estimate the sentence as a function of damage and other covariates X_j on both sides of the sentencing range threshold and compute the difference of left and right limit at the threshold

$$\gamma_c = \lim_{d \longrightarrow c^+} \mathbf{E}\left[S|D = c, X\right] - \lim_{d \longrightarrow c^-} \mathbf{E}\left[S|D = c, X\right], \tag{3}$$

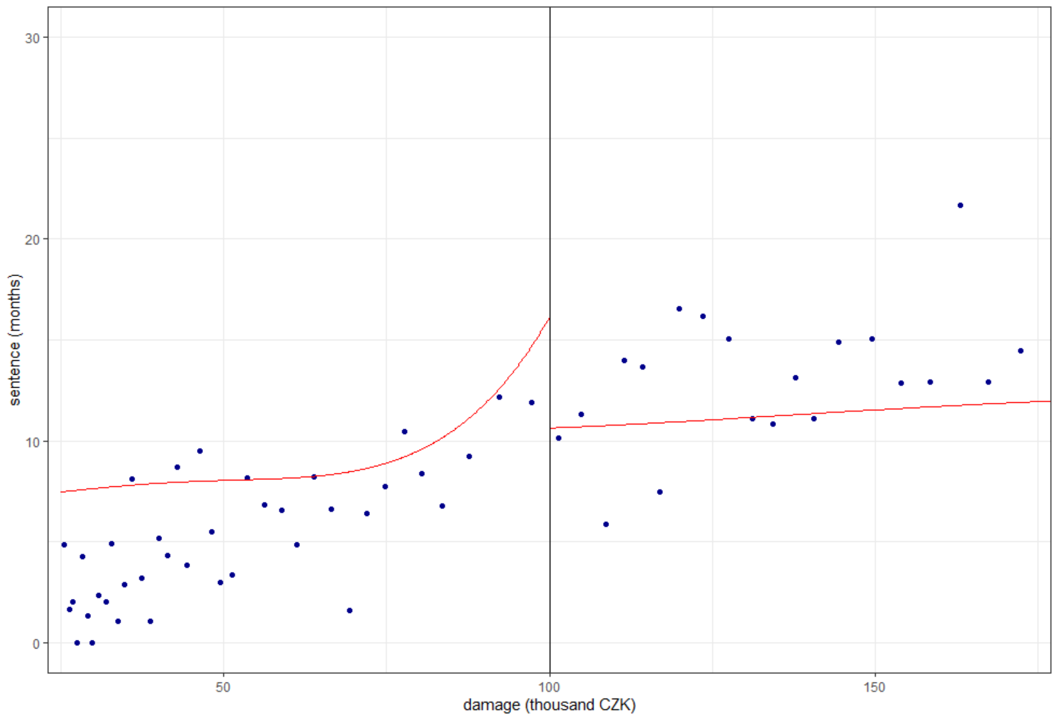
where D is the damage, c is the threshold of damage where the sentencing range switches, and γ_c represents the causal effect of the sentence threshold c.

Figures 10 and 11 plot the sentence around the given thresholds for after- and before-reform cases. On each side of the threshold, the plot is fitted using non-linear regression with controls. Similarly to the DD estimation, I impose a zero sentence for cases that were not punished by unconditional imprisonment. Tables 7 and 8 show the estimates of γ_c.

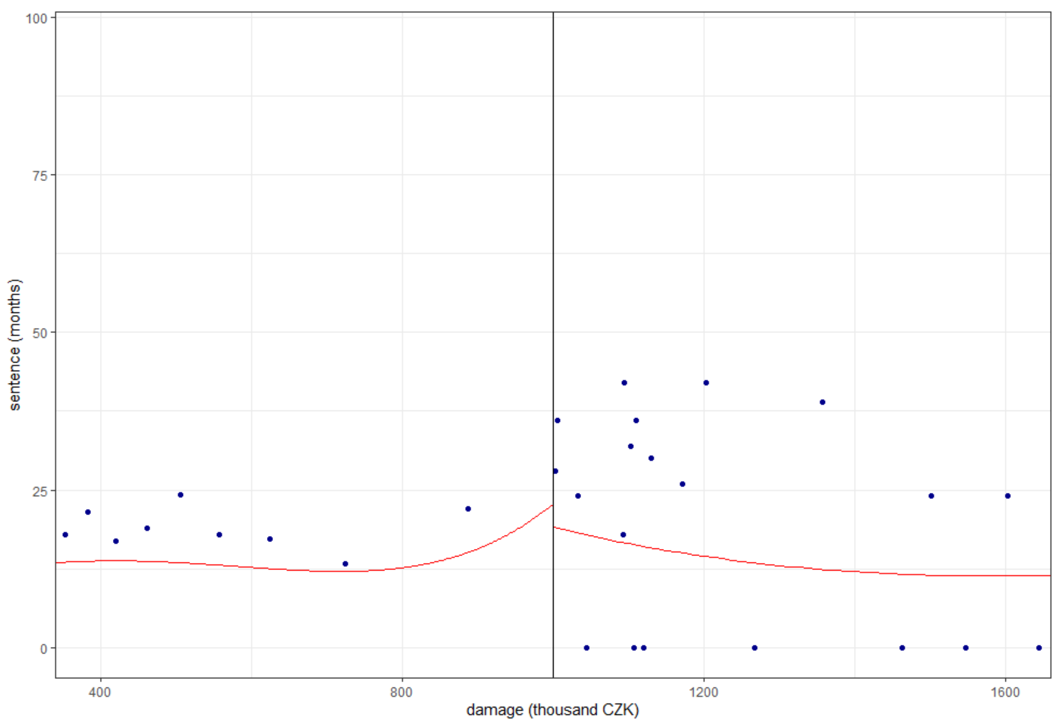
The results provide mixed evidence in terms of the presence of a discontinuity in an average sentence at the sentencing range threshold. The regression analysis does not provide convincing evidence for the after-reform period — most coefficients in Table 7 are insignificant, and most of them are negative, which would support the dominance of the reference effect.

Conversely, the before-reform coefficients (Table 8) provide much clearer evidence. Coefficients for the 50k threshold are all positive and mostly significant. That suggests that there is a positive jump in sentences upon crossing the sentencing range threshold. This jump ranges approximately from 17 to 27 months (focusing on the significant coefficients only). For the 500k thresholds, the coefficients are all positive but lose their significance.

In terms of the mechanisms summarized in Table 3, the positive coefficients may signal the presence of a severity effect. On the other hand, a negative jump could speak towards a reference effect. Section 6 discusses the possible interpretations of the results in more detail.

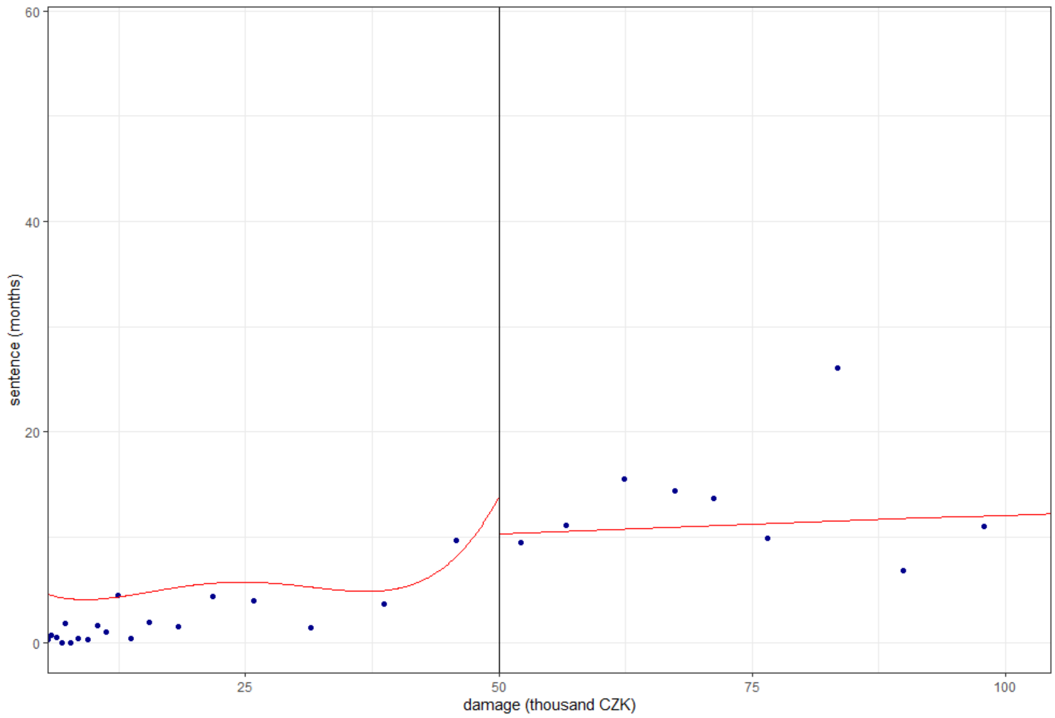


(a) 100k CZK threshold

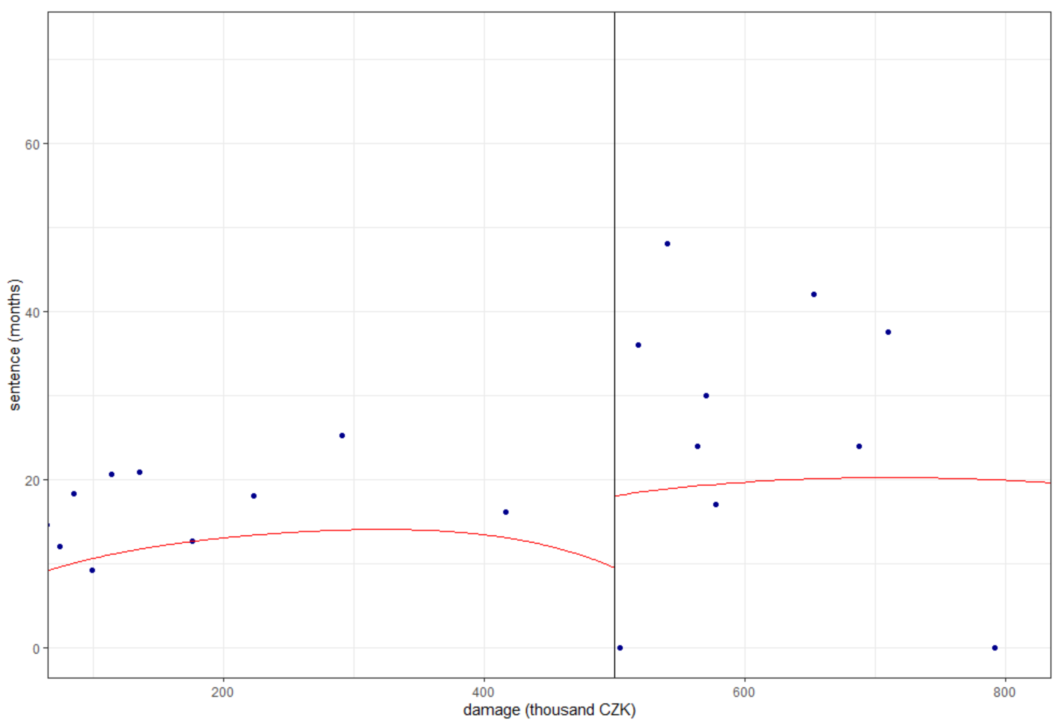


(b) 1m CZK threshold

Figure 10: Sentence as a function of damage caused around two different sentencing ranges thresholds. After-reform cases were used for estimation. The regression line was fitted using non-parametric local linear regression with controls. Controls include the age and gender of the offender, a recidivist dummy and a number of previous convictions, a dummy for concurrence, and the number of different types of punishment imposed for the given offense.



(a) 50k CZK threshold



(b) 500k CZK threshold

Figure 11: Sentence as a function of damage caused around two different sentencing ranges thresholds. Before-reform cases were used for estimation. The regression line was fitted using non-parametric local linear regression with controls. Controls include the age and gender of the offender, a recidivist dummy and a number of previous convictions, a dummy for concurrence, and the number of different types of punishment imposed for that offense.

Table 7: The discontinuity in sentence upon crossing a sentencing range threshold for after-reform cases. In all cases, local linear regression with a triangular kernel was used to non-parametrically estimate the model. The controls include a recidivism dummy, the gender and age of the offender, a concurrence dummy, and a number of different punishments imposed for the given crime. The bandwidth is expressed in thousands CZK.

	Dependent variable: sentence							
	Threshold 100k				Threshold 1m			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Conventional	-5.195	-5.308**	-2.660	-3.641**	-0.991	-7.648	-7.534	-6.151*
	(4.280)	(2.689)	(3.340)	(1.668)	(6.919)	(6.164)	(6.255)	(3.727)
Bias-Corrected	-4.727	-5.190*	-3.395	-4.661***	-2.135	-7.490	-2.799	-6.601*
	(4.280)	(2.689)	(3.340)	(1.668)	(6.919)	(6.164)	(6.255)	(3.727)
Robust	-4.727	-5.190*	-3.395	-4.661*	-2.135	-7.490	-2.799	-6.601
	(4.875)	(3.088)	(3.703)	(2.399)	(7.547)	(6.473)	(9.950)	(5.275)
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Bandwidth	37^+	40^+	20	100	319^+	101^+	100	300
Observations	2096	2096	2096	2096	2096	2096	2096	2096

Note: *p<0.1; **p<0.05; ***p<0.01; $^+$ denotes the optimal bandwidth

Table 8: The discontinuity in sentence upon crossing a sentencing range threshold for before reform cases. In all cases, local linear regression with a triangular kernel was used to non-parametrically estimate the model. The controls include a recidivism dummy, the gender and age of the offender, a concurrence dummy, and a number of different punishments imposed for the given crime. The bandwidth is expressed in thousands CZK.

	Dependent variable: sentence							
	Threshold 50k				Threshold 500k			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Conventional	17.239**	14.994***	17.691**	0.641	8.333	5.006	11.127**	4.527
	(7.440)	(3.944)	(7.956)	(2.058)	(18.705)	(6.838)	(4.777)	(6.324)
Bias-Corrected	19.431***	15.927***	26.778***	0.069	7.410	5.325	11.199**	10.194
	(7.440)	(3.944)	(7.956)	(2.058)	(18.705)	(6.838)	(4.777)	(6.324)
Robust	19.431**	15.927***	26.778*	0.069	7.410	5.325	11.199	10.194
	(7.749)	(4.189)	(15.377)	(2.855)	(21.088)	(7.801)	(6.831)	(8.986)
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Bandwidth	13^+	13^+	5	100	173^+	170^+	60	200
Observations	936	936	936	936	936	936	936	936

Note: *p<0.1; **p<0.05; ***p<0.01; $^+$ denotes the optimal bandwidth

6 Discussion

6.1 Reform analysis

In the main part of my empirical analysis, I split the sample of theft cases into two treatment groups based on how the 2020 reform affected them. In this subsection, I discuss the implications of the results for each treatment.

By Treatment A, I denote the cases with damage 500k-1m CZK, for which the sentencing range itself shifted downwards — from 2-8 years to 1-5 years. At the same time, less severe cases were added into the same sentencing range. The DD estimation suggests that the sentence decreased by 5 months for this subset of cases. This decline in sentences is also confirmed by the event study plot. The event study plots also show that the sentence for this group decreased immediately after the reform. Nevertheless, this plot should be interpreted with caution, as in the before-reform period, there seems to be a slight pattern — some coefficients turn out to be significantly negative.

The decrease in sentences can be interpreted as a sign of the severity effect — when the sentencing range decreases, the cases are suddenly perceived as less severe. This result may seem straightforward; however, it still represents an important piece of evidence that the sentencing rule is not independent of the policy and that the judges actually respond to sentencing range design change. Since the new and the old sentencing ranges overlap, the judges could, in principle, still impose the same sentence for a given value of damage. However, that seems not to be the case. Overall, the main takeaway from this finding is that judges are quite responsive to sentencing ranges change, which highlights the importance of their reasonable and substantiated design.

Second, I also analyzed the cases with damage 100k-500k (Treatment B), for which the sentencing range remained unchanged (1-5 years), but there were some more severe cases added to the same sentencing range. The DD estimate suggests that after the reform, the average sentence dropped by 1 month. This pattern is observable mostly when using a binary after-treatment indicator in the DD analysis. The event-study plot shows a mild decrease in most after-reform periods. The drop in sentences seems to be quite gradual and gains its significance only one year after the reform. In this case, there seem to be no significant trends in the before-treatment period, which increases the credibility of these results.

The drop in the sentence may be explained through the presence of a reference effect — the cases seem to be less severe compared to the cases that were added, which leads the judge to lower their sentence. This could be interpreted as important evidence that the judges actually compare the cases to others in the same sentencing range.

In the Appendix, I show that these results are mostly robust for different treatment sample choices (Section A.3). I also examine how the estimated treatment effects change when I adopt two alternative control groups. Unfortunately, with alternative control groups, the event study plots seem not to replicate exactly. Nevertheless, I still document a convincing drop in average sentence using the binary treatment indicator (Appendix Section A.2).

6.2 Around-threshold cases

I also examined the sentences imposed for cases closely around a sentencing range threshold. This part of my research is closely related to Drápal and Šoltés (2023), who studied a similar question using an experiment with Czech prosecutors. They report that the recommended sentence increased by 54 % (10 months) upon crossing the sentencing range threshold switching from 0-2 years to 1-5 years and by 12 % (5 months) when switching the sentencing range from 1-5 to 2-8 years.

First, I examine the around-threshold cases using the standard RDD approach for two thresholds after reform (100k and 1m) and two thresholds before reform (50k and 500k). Interestingly, my findings differ for these two time periods.

For the before-reform period, I find that the sentence increases by approximately 14 to 27 months at the threshold that shifts the sentencing range from 0-2 years to 1-5 years and by 5 to 12 months at the threshold that changes the range from 1-5 years to 2-8 years. These estimates even exceed the results of Drápal and Šoltés (2023). This could be driven by the fact that they use an experiment with prosecutors, whereas I analyze observational data of sentences imposed by judges. There might be some differences not only between the different subject pools but also between the behavior in an experiment and in a real-world setting. Theoretically, the presence of an upward jump would speak towards the dominance of the severity effect. The judges seem to increase their sentence instantly when it falls into a more severe sentencing range. This induces a disparity in sentences at the sentencing range threshold leading to the cases below the threshold being significantly less punished than the cases above the thresholds.

Conversely, for the after-reform period, I observe a negative jump in sentences at both thresholds. However, most coefficients turn out to be insignificant. That suggests that there is either no or negative jump in sentences at the sentencing range thresholds. Linking this to the underlying mechanisms, the negative jump can be driven by the reference effect, where the cases above the threshold are compared to a more severe reference group and thus appear to be less severe. However, this finding is contrary to Drápal and Šoltés (2023).

The difference in the around-threshold behavior of sentences between the two time periods is difficult to interpret. The explanation through a small sample size after the reform is not very plausible since the sample is, in fact, approximately two times larger.

In Appendix Section A.4, I explore this difference further using a difference in discontinuities approach. This exercise could help incorporate both before- and after-reform samples into one regression estimating a causal effect of a sentencing range threshold presence. However, the coefficients turn out to be mostly insignificant.

Alternatively, the difference in the RDD results for the before- and after-reform sample might be driven by the judges slowly adjusting to the reform. I try to address this concern by analyzing a sample of after-reform cases that were decided after 2023 when the sentencing rules should already have been quite established (Appendix Section A.5). The coefficients are closer to zero and become mostly insignificant, which could suggest the judges slowly diverge back to the dominance of the severity effect. However, an analysis of more recent data (which is not available at the moment) would be required to confirm this interpretation.

6.3 Limitations and possible extensions

Several important limitations need to be borne in mind when interpreting the results presented in this paper. First, in the Czech Republic, many criminal cases are punished with alternative forms of punishment, which complicates the analysis of sentencing trends and necessitates a credible strategy for handling these cases. Therefore, in this paper, I focus on more severe theft cases where the punishment by imprisonment rate is higher. For cases punished with alternative sentences, I assume that the sentence is zero. This approach reflects the idea that judges may occasionally opt for a milder punishment by using alternative methods. I support this definition of variables by carefully examining the extensive margin of sentencing before and after the reform.

Second, my research was impaired by a quite low amount of data in the pre-treatment period. That is mostly driven by the fact that the damage caused started to be reported only after 2019. Nevertheless, I still describe the main sentencing patterns. A fruitful extension of this research might be to examine other sentencing range reforms where the amount of data is larger.

Finally, my research focuses only on the judges's side and does not capture any other actors in the legal environment. That requires adopting the assumption that other actors do not respond to a sentencing ranges reform. Such an assumption may be justifiable by a recent study with Czech prisoners (Chen et al., 2024), where the authors show that the inmates have very little knowledge about the legal provisions. Alternatively, one could model and analyze how the offenders responded to the reform by examining the crime rates and composition before and after the reform. In addition, one could also define a social welfare function and study the general equilibrium effects of different sentencing ranges designs. That could lead to answering the question of which sentencing range design is optimal in terms of welfare and optimal criminal policy. These results could be of great interest to policymakers and help them to improve the current sentencing ranges design.

Conclusion

This paper focuses on the impact of sentencing ranges design on sentences. In particular, I describe how sentencing ranges shape judicial decisions. This question is relevant not only to the criminology and behavioral economics literature but also to an optimal criminal policy design.

Building on a previous study in this field (Drápal & Šoltés, 2023), I explain the decision of the judge through a severity and reference effect. The severity effect occurs when the judge interprets sentencing ranges as distinct categories, signaling that certain crimes are more severe than others. The reference effect arises when the judge compares a case to others within the same sentencing range. Both effects can significantly influence the judge's sentencing decisions and may lead to sentencing disparities.

To identify these effects, I analyze a dataset of Czech criminal cases. In particular, I focus on theft, which represents the most frequent offense and offers a straightforward measure of case severity — the damage caused. I take advantage of a 2020 reform that shifted the sentencing ranges for theft towards a milder scheme. I split the sample into the cases where the sentencing range decreased (from 2-8 years to 1-5 years) and the cases for which the sentencing range remained constant (1-5 years), but more severe cases were added into that sentencing range. I examine the change in sentences for each treatment group using difference in differences, taking the sentences for obstruction of justice and obstruction of a sentence of banishment as a control group.

My findings indicate that when a sentencing range is shifted downwards, the average sentence decreases by 5 months. This may be due to the severity effect, where a case is viewed as less severe when it falls into a lower sentencing category. When more severe cases are added to a sentencing range, the average sentence decreases by 1 month. That speaks towards the existence of a reference effect, where the judge compares cases to more severe ones within the same range, making them appear less severe by comparison. The key takeaway from the reform analysis is that judges adjust their decisions in response to changes in sentencing ranges and actively compare cases when determining the appropriate punishment.

Additionally, I also study the cases around sentencing range thresholds using a regression discontinuity design. I estimate the discontinuity in sentences at sentencing range thresholds separately for the before- and after-reform period. Interestingly, the results for these two time periods contradict. For the before-reform cases, I find a significant upward jump in sentences upon crossing a sentencing range threshold. This finding is in line with a previous study that studied the same question experimentally (Drápal & Šoltés, 2023) and speaks towards the severity effect of sentencing ranges, where cases in the more severe sentencing range are subject to harsher punishment. Nevertheless, the evidence for the after-reform period is mixed, and most discontinuity estimates are negative. A further investigation of this change in sentencing patterns may be a fruitful path for future research.

To conclude, my paper confirms that sentencing ranges design shapes sentencing decisions. In particular, the results suggest that sentencing ranges are associated with both, severity and reference effect on the decision of the judge. My results can contribute to the general understanding of the impact of sentencing ranges on sentences and may represent one of the first important steps towards a debate about an optimal sentencing ranges design.

References

Anderson, J. M., Kling, J. R., & Stith, K. (1999). Measuring interjudge sentencing disparity: Before and after the federal sentencing guidelines. *The Journal of Law and Economics*, *42(S1)*, 271–308.

Bjerk, D. (2017). Mandatory minimums and the sentencing of federal drug crimes. *The Journal of legal studies*, *46*(1), 93-128.

Butts, K. (2023). Geographic difference-in-discontinuities. *Applied Economics Letters*, *30*(5), 615–619. Retrieved from `https://doi.org/10.1080/13504851.2021.2005236` doi: 10.1080/13504851.2021.2005236

Chen, D. L., Cingl, L., Philippe, A., & Šoltés, M. (2024). Exploring inmates' perceptions, attitudes, and behavior: Implications for theories of crime. *CERGE-EI working paper*, *779*.

Cohen, A., & Yang, C. S. (2019). Judicial politics and sentencing decisions. *American economic journal. Economic policy*, *11*(1), 160-191.

The Czech Criminal Code, Act No. 40/2009 Coll. (2009).

Drápal, J. (2023). Punitive by negligence? The myths and reality of penal nationalism in the Czech Republic. *European Journal of Criminology*, *20*(4), 1549-1567. doi: 10.1177/14773708211063753

Drápal, J., & Šoltés, M. (2023). Sentencing decisions around quantity thresholds: theory and experiment. *Journal of experimental criminology*, *1*, 1-54.

Frankel, M. E. (1972). Lawlessness in sentencing. *University of Cincinnati law review*, *41*, 1-54.

Grembi, V., Nannicini, T., & Troiano, U. (2016). Do fiscal rules matter? *American Economic Journal: Applied Economics*, *8*(3), 1–30. Retrieved 2024-08-19, from `http://www.jstor.org/stable/24739127`

Hofer, P. (2019, 02). Federal sentencing after Booker. *Crime and Justice*, *48*, 000-000. doi: 10.1086/701712

Leibovitch, A. (2017). Punishing on a curve. *Northwestern University law review*, *111*(5), 1205-1280.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698-714.

Skugarevskiy, D. (2017). *Essays in law and economics of enforcement* (Dissertation thesis). Graduate Institute of International and Development Studies in Geneva.

Sporer, S. L., & Goodman-Delahunty, J. (2009). Disparities in sentencing decisions. *Disparities in Sentencing Decisions. Social Psychology of Punishment of Crime*, 379-401.

Travova, E. (2023). Under pressure? performance evaluation of police officers as an incentive to

cheat. *Journal of Economic Behavior Organization*, *212*, 1143-1172.
Tuttle, C. (2019). Racial disparities in federal sentencing: Evidence from drug mandatory minimums.

A Appendix

A.1 Damage report rate

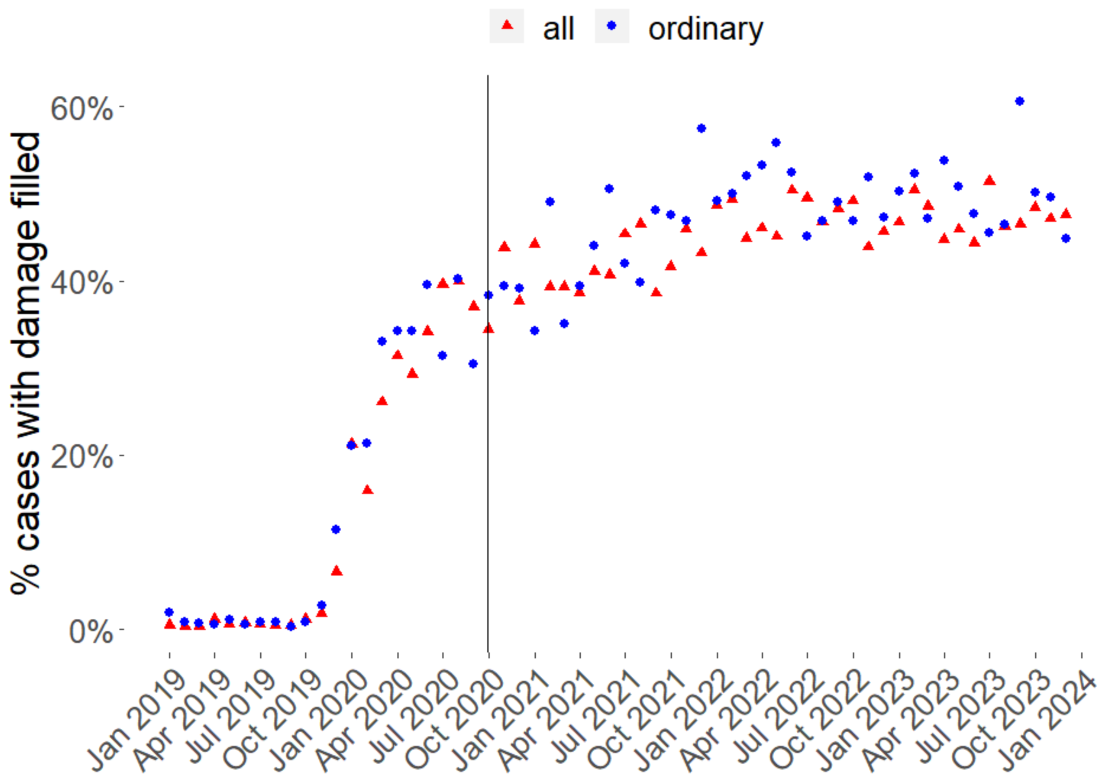


Figure A.1: The rate of cases with damage filled before and after the reform. Red triangles represent all theft cases; the blue dots represent ordinary theft cases (cases where damage is the criterion determining the sentencing range; see Section 2 for definition). The black line denotes the 2020 reform. The date relates to the sentence coming into legal power, which determines the use of the pre-/post- reform legal norm.

A.2 Alternative control groups

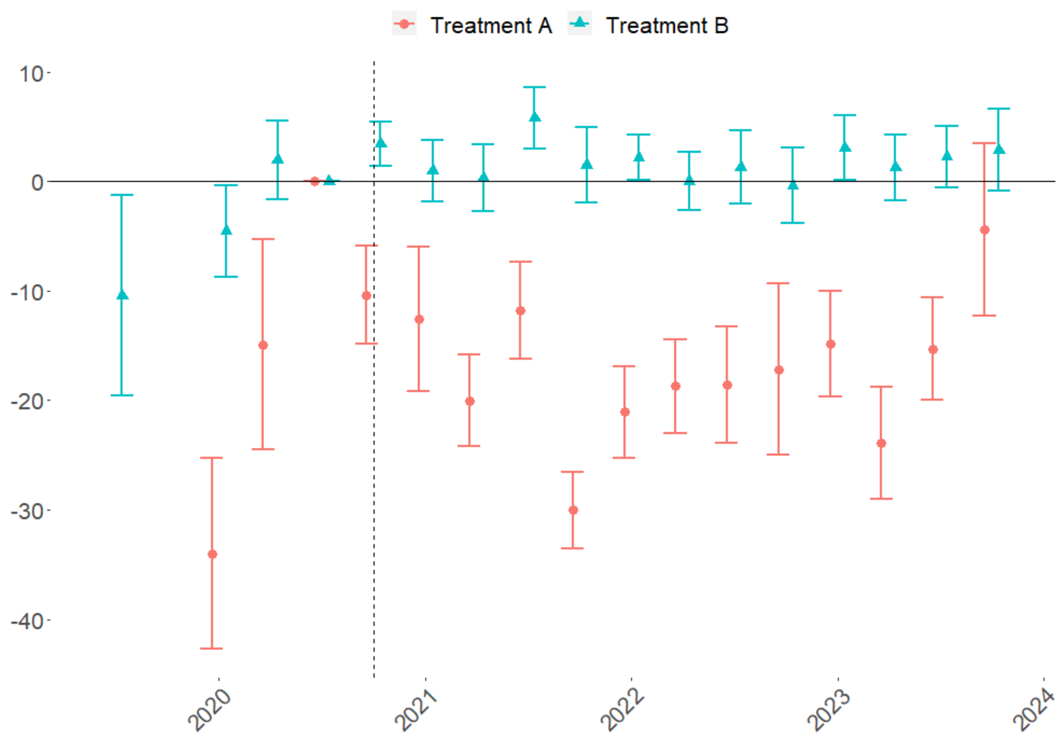
To address the concerns that my findings may be driven by an arbitrary choice of control group, I replicate the results using two alternative control groups.

It should be noted that given the scope of the 2020 reform, it is extremely difficult to come up with a set of control cases that were absolutely unchanged by this reform. For instance, all crimes against property were at least partially affected, which unfortunately disables them from becoming a control group in my analysis.

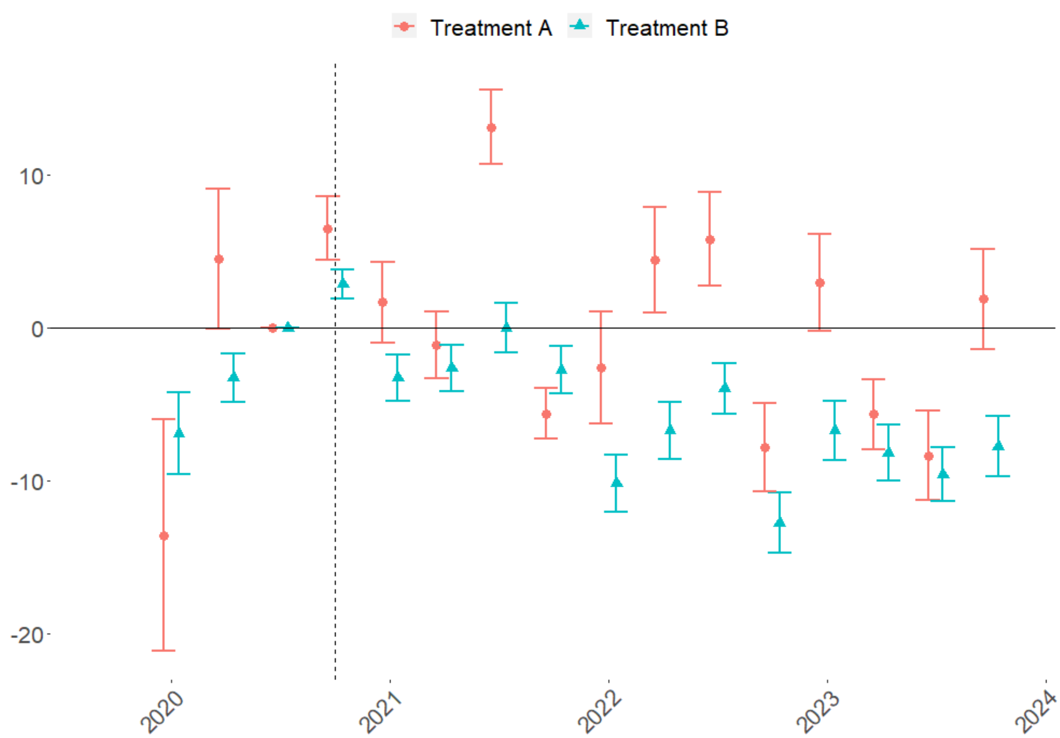
Nevertheless, I introduce two more alternative control groups. First, I use *breaking and entering*. This crime is committed upon wrongfully entering another's dwelling. This choice of control group seems to be convenient as it is quite similar to theft. In 45 % of cases, these two crimes are committed together. However, this may induce some concerns about the potential spillover effects of the reform to the sentences for this crime, whose direction and magnitude are not straightforward to interpret. Since we cannot easily disentangle these effects, we need to bear this important limitation in mind when interpreting the results obtained using this control group.

Second, I consider the negligence of mandatory support. The main advantage of this control group is that the offense is related to the offender's property; thus, the sentencing trends could approximate the counterfactual sentencing for theft well. The main disadvantage is that this offense has quite low sentencing ranges (0-2 years, 0-1 year) and is only rarely punished by imprisonment.

I plot the results of the event study model to examine both pre-tends and the actual effect of the treatment. The event study plots were produced using exactly the same method as in the previous analysis. Figure A.2 shows the results for two alternative control groups. Different confidence interval widths are driven by different numbers of cases in each quarter.



(a) breaking and entering (§ 178)



(b) negligence of mandatory support (§ 196)

Figure A.2: The quarterly effects on sentence for two alternative control groups. The baseline rate corresponds to Q3 2020. The dashed vertical line represents the 2020 reform. The regressions control for judge fixed effects, number of previous convictions, age of the offender, concurrence dummy, and the number of different punishments for the given crime. 95 percent confidence intervals are plotted.

For both alternative control groups, the estimates become quite noisy and differ significantly from zero also in the before-reform period, which impairs the interpretation of the results. Nevertheless, with breaking and entering, I at least replicate the drop in sentences for Treatment A. For the negligence of mandatory support, the pattern in Treatment A is not that clear; however, I at least get a drop in sentences in the Treatment B sample.

The results of this exercise show that the event plots may be control group sensitive. This limitation may be, however, caused by data availability in the before-reform period. To partially overcome the lack of data, I also re-ran the regressions with a dummy after-treatment period indicator. Tables A.1 and A.2 report the results. These estimates are in line with the main analysis. For Treatment A, both alternative control groups confirm a reduction in sentences around 5-8 months. For Treatment B, the estimated drop in sentences is around 1.8-2.8 months.

Table A.1: Estimates of the treatment effect obtained using DD approach and breaking and entering (§ 178) as a control group. Controls include judge fixed effects, age, number of previous convictions, damage, number of different punishments for the given crime and concurrence, recidivism, juvenile and gender dummies of the offender.

	Dependent variable:	
	sentence	
Panel A: Treatment A		
	(1)	(2)
After:Treatment	-7.587^{***}	-6.256^{***}
	(1.877)	(1.768)
Intercept	Yes	Yes
Controls	No	Yes
Observations	7,399	7,399
R^2	0.167	0.363
Adjusted R^2	0.167	0.318
Residual Std. Error	6.564 (df = 7395)	5.939 (df = 6906)
Panel B: Treatment B		
	(1)	(2)
After:Treatment	-2.702^{***}	-1.831^{***}
	(0.657)	(0.600)
Intercept	Yes	Yes
Controls	No	Yes
Observations	8,792	8,792
R^2	0.292	0.498
Adjusted R^2	0.292	0.468
Residual Std. Error	8.671 (df = 8788)	7.516 (df = 8292)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table A.2: Estimates of the treatment effect obtained using DD approach and negligence of mandatory support as a control group. Controls include judge fixed effects, age, number of previous convictions, damage, number of different punishments and concurrence, recidivism, juvenile and gender dummies of the offender.

	Dependent variable:	
	sentence	
Panel A: Treatment A		
	(1)	(2)
After:Treatment	-7.291^{***}	-5.909^{***}
	(1.079)	(1.057)
Intercept	Yes	Yes
Controls	No	Yes
Observations	27,346	27,346
R^2	0.156	0.408
Adjusted R^2	0.156	0.397
Residual Std. Error	3.785 (df = 27342)	3.200 (df = 26845)
Panel B: Treatment B		
	(1)	(2)
After:Treatment	-2.406^{***}	-2.269^{***}
	(0.367)	(0.346)
Intercept	Yes	Yes
Controls	No	Yes
Observations	28,739	28,739
R^2	0.326	0.532
Adjusted R^2	0.326	0.523
Residual Std. Error	5.054 (df = 28735)	4.250 (df = 28234)
Note:		*p<0.1; **p<0.05; ***p<0.01

A.3 Cases pooling

In the main analysis, I use a narrow range of damage as a Treatment A and Treatment B sample (damage 500k-1m and 100k-500k, respectively). Nevertheless, in principle, Table 2 implies that the range of cases is wider for each treatment group. In this section, I analyze different subgroups of each treatment. I first pooled cases with damage 50k-100k, 500k-1m, and 5m-10m for the Treatment A sample and cases with damage 10k-50k, 100k-500k, and 1m-5m for the Treatment B sample. Then, I focused only on cases 50k-100k to re-estimate the effect of treatment A and on cases 10k-50k to re-estimate the effect of treatment B. Other sample choices were not possible because of the low number of cases with very high damage (over 1m CZK).

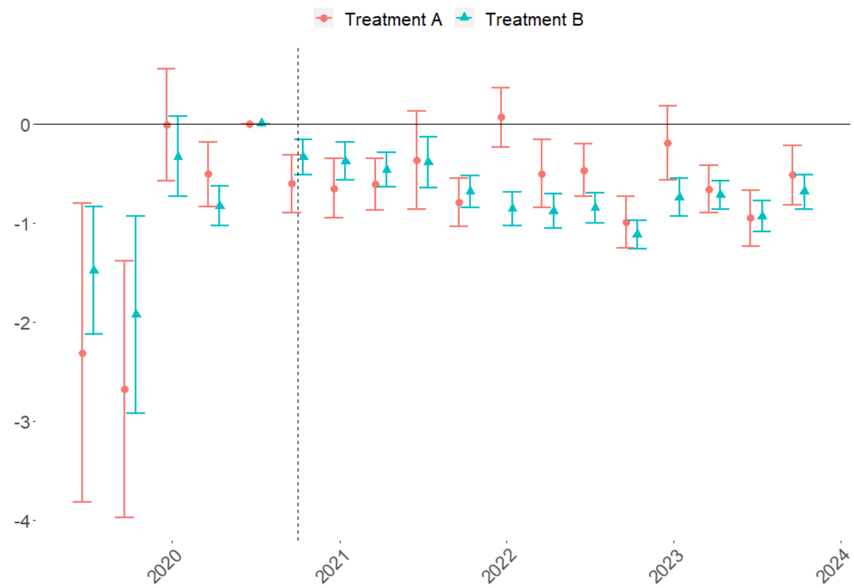
To overcome the differences in the sentencing ranges for different treatment groups, I standardized the sentences using the before-treatment control group mean and standard deviation (z-score normalization). The coefficients show how much the sentence changes in terms of the standard deviations relative to the control group. Table A.3 presents the results.

I document a clear and significant drop in sentences for all treatment group definitions. The drop ranges from 0.31 to 1.39 multiples of standard deviations in the Treatment A group, and 0.17 to 0.48 standard deviations in the Treatment B group. That means that the presence of the drop in average sentence after the reform is not dependent on the choice of sample.

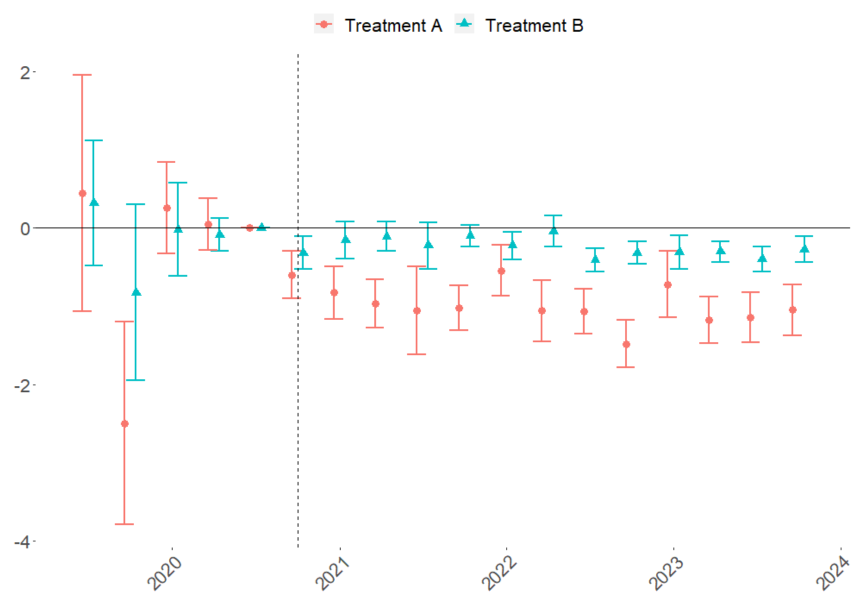
I conclude this section with event study plots for standardized outcome variables. The pooled sample elicits not quite clear patterns in the pre-treatment period. Conversely, the alternative sample shows almost all pre-treatment coefficients to be zero and confirms the patterns observed for the main sample used in the analysis.

Table A.3: Estimates of the treatment effect for different samples of cases with standardized outcome variable. Controls include judge fixed effects, age, number of previous convictions, damage, number of different punishments for the given crime and concurrence, recidivism, juvenile and gender dummies of the offender.

	Dependent variable:		
	z-score standardized sentence		
Panel A: Treatment A			
	(1)	(2)	(3)
After:Treatment	-0.903^{***}	-0.313^{***}	-1.385^{***}
	(0.086)	(0.083)	(0.269)
Damage range	50k-100k	pooled	500k-1m
Intercept	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	45,160	45,415	44,815
R^2	0.229	0.233	0.219
Adjusted R^2	0.220	0.224	0.210
Panel B: Treatment B			
	(1)	(2)	(3)
After:Treatment	-0.169^{***}	-0.349^{***}	-0.474^{***}
	(0.063)	(0.054)	(0.076)
Damage range	10k-50k	pooled	100k-500k
Intercept	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	46,543	48,241	46,208
R^2	0.191	0.288	0.327
Adjusted R^2	0.182	0.280	0.319
Note:			*p<0.1; **p<0.05; ***p<0.01



(a) pooled treatment groups



(b) alternative definitions of treatment groups (A: 50k-100k, B: 10k-50k)

Figure A.3: The quarterly effects on normalized sentence for different treatment group definitions. The baseline rate corresponds to Q3 2020. The last column represents the original sample choice. The dashed vertical line represents the 2020 reform. The regressions control for judge fixed effects, number of previous convictions, age of the offender, concurrence dummy, and the number of different punishments for the given crime. 95 percent confidence intervals are plotted.

A.4 Difference in discontinuities

Given that the standard RDD estimation provides different results for the before- and after-treatment period, I extend the basic RDD analysis with an estimation of a difference in discontinuities (Grembi, Nannicini, & Troiano, 2016) before and after reform. I estimate the difference between the pre-treatment and post-treatment discontinuity at the threshold of 50k, 100k, 500k, and 1m (denoted as D_c henceforth). The rationale is that I restrict the sample to cases in the interval $D_i \in [D_c - h; D_c + h]$ and run a difference in discontinuities local linear regression for cross-sectional data proposed by Butts (2023)

$$\begin{aligned} S_i = \delta_0 + \delta_1(D_i - D_c) + \mathcal{I}_{D_i \geq D_c}\left[\gamma_0 + \gamma_1(D_i - D_c)\right] + \\ +T_i\left\{\alpha_0 + \alpha_1(D_i - D_c) + \mathcal{I}_{D_i \geq D_c}\left[\beta_0 + \beta_1(D_i - D_c)\right]\right\} + \\ + \sum_{j=1}^{k} \lambda_j X_{ij} + e_i, \end{aligned} \tag{A.1}$$

where S_i is the sentence, T_i a dummy indicating whether the given D_c is or is not a sentencing range threshold. That means that for thresholds 100k and 1m, T_i is identical to the after-reform indicator; for thresholds 50k and 500k, it is the negation of the after-reform indicator. X_{ij} is a set of covariates.

Coefficient β_0 is then the difference in discontinuities estimator. In this case, it can be interpreted as a change in discontinuity once the given value of damage becomes a sentencing range threshold. Table A.4 reports the results of this exercise.

The results of this exercise are quite difficult to interpret — most coefficients end up being insignificant, and they are both positive and negative. It should be noted that the variable T_i is equal to 1 when the value of damage is a sentencing range threshold. The only significant coefficient turns out to be strongly negative, which would mean that once 1m CZK starts being the sentencing range threshold, a large negative discontinuity occurs. That is, however, not entirely supported by other results obtained with RDD.

Table A.4: Difference-in-discontinuities estimator of the change in discontinuity when the given value of damage starts being the sentencing range threshold. The controls include a recidivism dummy, the gender and age of the offender, a concurrence dummy, and a number of different punishments imposed for the given crime. The bandwidths correspond to the optimal bandwidths computed in the RDD analysis.

	Dependent variable: sentence			
Panel A: After-reform sentencing range thresholds				
	(1)	(2)	(3)	(4)
$\mathcal{I}_{D_i \geq D_c} : T_i$	-6.055	-7.204	-78.500^{***}	-88.353
	(4.575)	(4.998)	(27.801)	(221.434)
Threshold	100k	100k	1m	1m
Intercept	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	695	773	97	38
R^2	0.324	0.066	0.448	0.091
Adjusted R^2	0.312	0.058	0.377	-0.085
Panel B: Before-reform sentencing range thresholds				
	(1)	(2)	(3)	(4)
$\mathcal{I}_{D_i \geq D_c} : T_i$	-1.489	-0.340	7.378	16.542
	(4.725)	(5.386)	(13.364)	(17.987)
Threshold	50k	50k	500k	500k
Intercept	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	412	452	235	179
R^2	0.344	0.057	0.462	0.078
Adjusted R^2	0.324	0.042	0.433	0.040

Note: *p<0.1; **p<0.05; ***p<0.01

A.5 RDD subsample analysis

Previous analyses persistently show that the sign of discontinuity in sentences at the sentencing range threshold significantly differs for the before- and after-treatment period. I hypothesize that this might be due to a slow response of the judges to the reform. To examine this hypothesis, I run the RDD regression for the after-reform period on a restricted sample of cases starting in 2023.

Table A.5: The jump in sentence upon crossing a placebo threshold for cases after 01/01/2023. In all cases, local linear regression with a triangular kernel was used to non-parametrically estimate the model. The controls include a recidivism dummy, the gender and age of the offender, a concurrence dummy, and a number of different punishments imposed for the given crime. The bandwidth is expressed in thousands CZK.

	Dependent variable: sentence							
	Threshold 100k				Threshold 1m			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Conventional	-1.201	-2.686	0.048	-3.729	10.509	-4.749	-3.261	1.066
	(7.364)	(4.996)	(5.880)	(3.120)	(8.938)	(6.081)	(5.765)	(3.632)
Bias-Corrected	-0.167	-1.515	0.945	-4.296	13.388	10.523*	5.261	-0.640
	(7.364)	(4.996)	(5.880)	(3.120)	(8.938)	(6.081)	(5.765)	(3.632)
Robust	-0.167	-1.515	0.945	-4.296	13.388	10.523	5.261	-0.640
	(8.069)	(5.534)	(7.513)	(4.297)	(9.767)	(26.360)	(25.882)	(4.586)
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Bandwidth	32^{+}	30^{+}	10	100	269^{+}	104^{+}	100	300
Observations	1338	1338	1338	1338	335	335	335	335
Note:	*p<0.1; **p<0.05; ***p<0.01; $^{+}$ denotes the optimal bandwidth							

This exercise shows that if we restrict the sample to cases judged long after the adoption of the reform, the discontinuity at the sentencing range thresholds slightly shifts toward zero or even toward positive values. That may be because the judges may need some time to understand and internalize the new sentencing range design. However, most coefficients remain insignificant, and the pattern is still far from what I found for the before-reform period.