The Impact of Sentencing Ranges on Judicial Decisions: Evidence from a Czech Reform[*][†]

April 8, 2025

Tereza Burýšková[‡] , CERGE-EI[§]

Abstract

Sentencing ranges represent a common policy tool to reduce disparities in judicial decisions by providing structured sentencing guidelines. However, existing research suggests that their effects on sentences may arise through several different mechanisms. This paper leverages a 2020 Czech reform of the sentencing ranges system and estimates its effects using difference-in-differences (DD). I find that judges respond to lowering the sentencing ranges by reducing sentences on both extensive and intensive margins. Specifically, sentences decrease not only for cases directly affected by a lower sentencing range but also for cases in unchanged ranges where more severe offenses were added. These findings provide empirical support for two key mechanisms: the severity effect, where sentencing ranges serve as signals of case severity, and the reference effect, where judges compare cases within the same range and adjust sentences accordingly. The results highlight the unintended consequences of sentencing range reforms and suggest that the reference effect should be considered when designing sentencing policies and conducting empirical research on judicial behavior.

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.

[†]I am grateful to Michal Šoltés for many important insights that helped me to develop this paper.

[‡]née Hofrichterová, Email: tereza.buryskova@cerge-ei.cz

[§]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. In the civil law legal system, policymakers often introduce a system of sentencing ranges to overcome such sentencing disparities. The sentencing range determines the lower and the upper limit for the years of imprisonment that the judge can impose. Typically, the sentencing ranges are related to the severity of the particular case (e.g., the damage caused, characteristics of the victim, or the amount of drug possessed). While these ranges are intended to unify the sentencing decisions for similar cases, their design may introduce new complexities into the consideration of the judge.

My paper builds mainly on the notion of severity and reference effect of sentencing ranges introduced by Drápal and Šoltés (2024). The authors develop a simple behavioral model of the sentencing process, introducing two main effects driving the decision of the judge - reference and severity effect. The severity effect builds on the idea that the sentencing range the case belongs to offers a rough signal about its severity. Therefore, cases that belong to a lower sentencing range should, all else being equal, be associated with lower sentences. The reference effect of sentencing ranges operates through the judges comparing cases within the same sentencing range. The relative position of the case in the corresponding sentencing range in terms of severity then shapes the final sentences - the cases that are less severe compared to their reference group get lower sentences than similar cases that are more severe in their reference group.

My research develops on these effects and tests them using Czech court data on theft cases. I exploit a 2020 reform that reclassified certain types of theft, resulting in a quasi-exogenous change in sentencing ranges. Particularly, for some damage ranges, the sentencing ranges lowered; for others, the ranges remained the same, but more severe cases were added into the same sentencing range. I examine the extensive margin (represented by the proportion of cases punished by unconditional imprisonment) and the intensive margin (represented by the average sentence) of sentencing for each group separately.

Using the DD approach, I find that judges respond to the reform of sentencing ranges by decreasing the sentences on both extensive and intensive margins. This decrease occurs for cases whose sentencing range has been lowered, as well as for cases whose sentencing range remained the same, but more severe cases were added to the same sentencing range. The results for the first group speak in favor of the severity effect; the results for the second group support the reference effect, indicating that sentencing ranges play the role of a reference group to which the cases are compared. This finding implies that even cases that were not directly affected by the reform could be influenced through this indirect mechanism. My results may also have methodological implications for empirical legal research. Given that sentencing ranges reforms may affect sentences indirectly, it is necessary to choose the control group carefully to perform any comparisons.

My research brings empirical evidence for the already introduced hypotheses regarding the impact of sentencing ranges on sentences using court data. I show that reference considerations significantly shape the sentence imposed and that this pattern should be considered when designing or reforming the sentencing ranges system or performing empirical research.

The rest of this paper is organized as follows: Section 2 summarizes the related literature; Section 3 briefly describes the legal context; Section 4 explains the intuition about the underlying mechanisms; Section 5 introduces the dataset used in the empirical analysis; Section 6 presents the results of the empirical analysis, and Section 7 discusses their implications. The main contribution is summarised in the Conclusion.

2 Literature Review

Many legal scholars and economists have studied the impact of sentencing policies on sentences. The majority of this literature focuses on the US context and mandatory minimum imprisonment lengths, which were introduced to unify sentences for certain drug offenses. Bjerk (2017a, 2017b) examines the impact of mandatory minimum sentencing reforms, showing how they constrain judicial discretion and alter sentencing outcomes, particularly for drug-related offenses. Anderson et al. (1999) analyze interjudge sentencing disparity before and after the implementation of federal sentencing guidelines, finding that while the guidelines reduced variation in sentences, disparities persisted due to prosecutorial discretion. Tuttle (2023) highlights racial disparities in federal sentencing, providing empirical evidence on the disproportionate impact on Black defendants. Fischman and Schanzenbach (2012) explore judicial behavior under structured sentencing regimes, demonstrating that while guidelines reduce variability, judges often adapt their decision-making in ways that can sustain or even exacerbate disparities.

Beyond mandatory minimums, recent research has explored the broader effects of sentencing guidelines, judicial discretion, and external factors influencing sentencing decisions. Cohen and Yang (2019) analyze how judicial ideology impacts sentencing decisions, showing that partisan alignment with the executive branch can lead to systematic differences in sentencing severity. Similarly, Leibovitch (2017) examines how judges adjust sentencing decisions in response to implicit benchmarking created by sentencing distributions, finding that judges are influenced by the broader distribution of sentences within their jurisdiction. Moreover, Schulhofer (1992) highlights the unintended consequences of sentencing guidelines, demonstrating that efforts to standardize sentencing can inadvertently shift discretion from judges to prosecutors, altering case outcomes in unintended ways.

More recent evidence suggests that changes in sentencing guidelines can have long-term and subtle impacts on judicial behavior. Witwer (2023) uses longitudinal data to show that departures from guidelines are sensitive to whether the guidelines are presumptive or advisory, indicating that even moderate shifts in structure can lead to meaningful changes in sentencing patterns. Similarly, Beckett, Evans, and Harris (2023) highlight how policies that aim to tailor sanctions to a defendant's ability to pay — such as in the context of monetary fines — can inadvertently maintain or exacerbate racial disparities. They show that judicial discretion, even when guided by reformist policies, remains a key channel through which bias can persist, reinforcing the notion that structural reforms alone may be insufficient to fully mitigate inequality in sentencing outcomes. Teichman, Zamir, and Ritov (2023) further supports this view by showing that biases in legal decision-making are present across prosecutors, defense attorneys, law students, and laypersons alike, suggesting that institutional design must account for persistent psychological and cognitive factors.

In addition to disparities, researchers have examined the relationship between sentence length and future criminal behavior. Weswasi, Sivertsson, Bäckman, and Estrada (2023) present a quasi-experimental study leveraging three policy reforms in Sweden to assess whether longer incarceration reduces recidivism. Their results indicate that extended sentences have limited effects on reoffending, challenging the deterrence rationale for harsher punishments and supporting more rehabilitative or individualized approaches.

A common approach to identifying the impact of sentencing ranges empirically is to focus on sentences for around-threshold cases (Bjerk, 2017b; Skugarevskiy, 2017). The rationale is that the cases that are close to a sentencing range threshold (and are similar in observable characteristics) differ only in their sentencing range. Therefore, by comparing the sentences for below and above-threshold cases, the authors assess the impact of the sentencing range. Using this approach, Skugarevskiy (2017) examines Russian cases of drug possession, reporting a significant increase in sentences when crossing the sentencing range threshold. Nevertheless, the interpretation of these results may be strongly dependent on the legal context.

A similar approach is adopted in the paper most relevant to mine, where Drápal and Šoltés (2024) focus on the around-threshold cases of drug possession and theft. They ran an online experiment with 200 Czech prosecutors where 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. They asked the prosecutors to recommend a sentence for each scenario. The results show 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 confirming the mechanisms described in their paper using court data.

2.1 The 2020 reform

In October 2020, the provision of the Criminal Code was modified significantly. The definition of the terms determining the extent of damage shifted towards higher values of actual damage. The change in term definitions implies different legal classifications for cases before and after the reform. For example, a case with damage of CZK 75k 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 *moderate* 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.

Table 1: Sentencing ranges for ordinary theft cases in the Criminal Code

	Sentencing Range	
Damage (CZK)	Before September 2020	After September 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

Note: The sentencing ranges for different types of theft as prescribed by the Criminal Code before and after the 2020 reform. The reform came into power on October 1, 2020.

Table 1 shows that the effects of the 2020 reform were twofold. For some cases, the sentencing range shifted (e.g., cases with damage CZK 500k-1m 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 CZK 100k-500k face the same sentencing range of 1-5; however, the same sentencing range now relates also to cases with damage CZK 500k-1m, 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. Table 2 highlights this distinction. Analyzing changes in sentencing patterns for these two groups separately will help identify the mechanisms through which the reform operates 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 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 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

Note: A distinction between Treatment A and Treatment B type cases based on the impact of the 2020 reform. Own summary based on the Criminal Code. For simplicity, I omit cases with damage > CZK 10m.

3 Mechanisms in sentencing: Severity and reference effect

As for the mechanisms behind sentencing decisions, I build on Drápal and Šoltés (2024), who introduce severity and reference effects and demonstrate how these effects shape the sentences for cases close to a sentencing range threshold. In light of their intuition, I form predictions for the effect of the reform under each mechanism.

Drápal and Šoltés (2024) 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 A has damage of 99k CZK, and Case B has damage of 101 CZK. Assume that apart from the damage, these cases are identical in all other characteristics. The provision of the Criminal Code assigns A to the sentencing range of 0-2 years and B to the sentencing range of 1-5 years. When determining the punishment, the higher sentencing range for case B signals its increased severity, which increases the sentence imposed (severity effect). However, case B is compared to more severe cases falling into the same sentencing range, which decreases the sentence (reference effect). Here, these two effects work against each other. Potentially, one can determine which effect prevails by comparing the sentences for cases A and B. If the sentence for case A is higher than the sentence for case B, we could conclude that the reference effect dominates. Conversely, if the sentence for case B exceeds the one for case A, it demonstrates the dominance of the severity effect. The authors leverage this intuition when comparing the sentences recommended for around-threshold cases. In this paper, I introduce yet another approach to identify the severity and reference effect tailored to the 2020 reform.

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 (if existent) 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 3 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 3. Since these cases are now classified as less severe, we expect 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 work 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 the impact of severity and reference effect on the sentences.

	Severity Effect	**Reference Effect**
	Impact on Mean Sentence	
Treatment A *(Sentencing Range Downward Shift)*	↓	↑
Treatment B *(Addition of More Severe Cases)*	0	↓

Note: A summary of the impact of severity and reference effect (as introduced by Drápal and Šoltés (2024)) on the Treatment A and Treatment B cases analyzed in this paper. The severity effect decreases the sentence for Treatment A cases and does not influence Treatment B cases. The reference effect increases sentences for Treatment A and decreases them for Treatment B. This intuition provides interpretation for the changes in average sentence estimated empirically.

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 the circumstances of the case (different from damage). In the experimental setting leveraged by Drápal and Šoltés (2024), it was possible to set these circumstances to be absolutely identical for the cases considered, whereas in my empirical analysis, I deal with cases differing in multiple dimensions. I overcome this issue by introducing controls (to deal with observable variables) and employing the DD design (to capture the general trends in sentencing).

4 Court data

In the Czech Republic, all criminal cases are well-documented, and the case-level data are available for research purposes. In particular, three main sets of variables are available for each case. First, there is data about the criminal procedure, including the court and the judge 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 data available capture the period between 2006 and 2023. However, the damage caused, which is central to my analysis, has been reported only since 2019, a year before the reform. Appendix Figure A.1 shows that a stable report rate of around 40 % emerged by the beginning of 2020.

Appendix Table A.1 presents the descriptive statistics of the dataset. It turns out that theft is often punished by alternative means of punishment, including a conditional sentence. Moreover, in line with the previous literature (Drápal, 2023), I find that the conditional sentences are systematically lower than the standard conditional sentences, suggesting that judges may perceive these as distinct punishment types.

5 Results

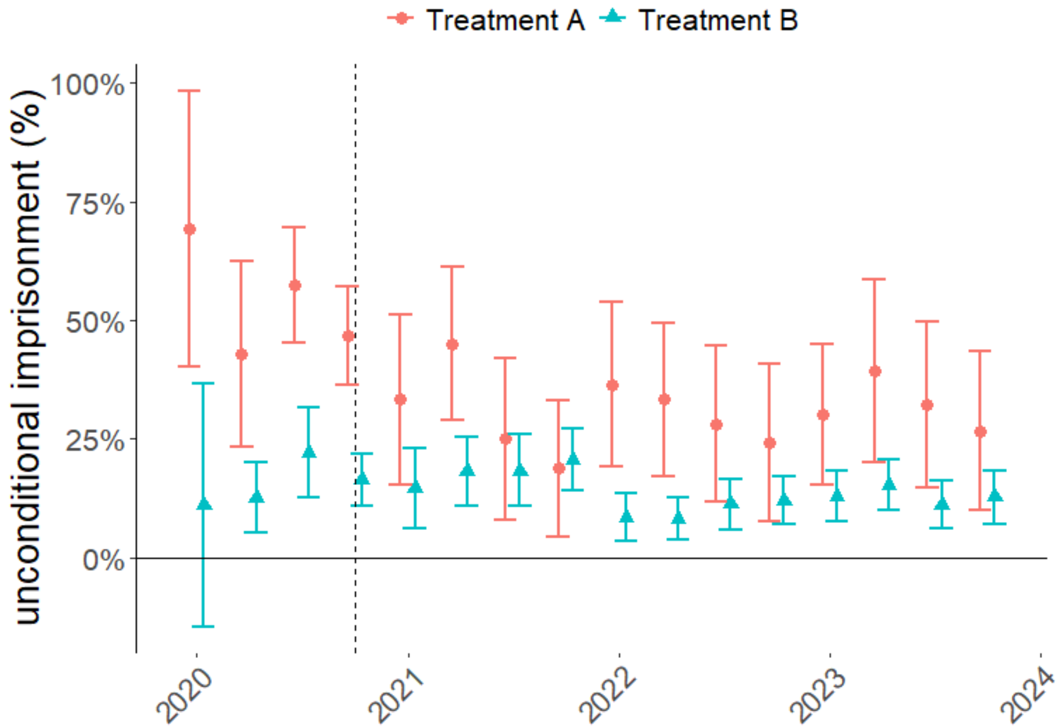
5.1 Sentencing before and after the reform

For the purpose of this analysis, I take cases with damage between CZK 50k and 100k as the Treatment A sample and cases with 10k and 50k as the Treatment B sample. Nevertheless, in the Appendix section A.3, I show that the general patterns do not depend on the choice of the sample.

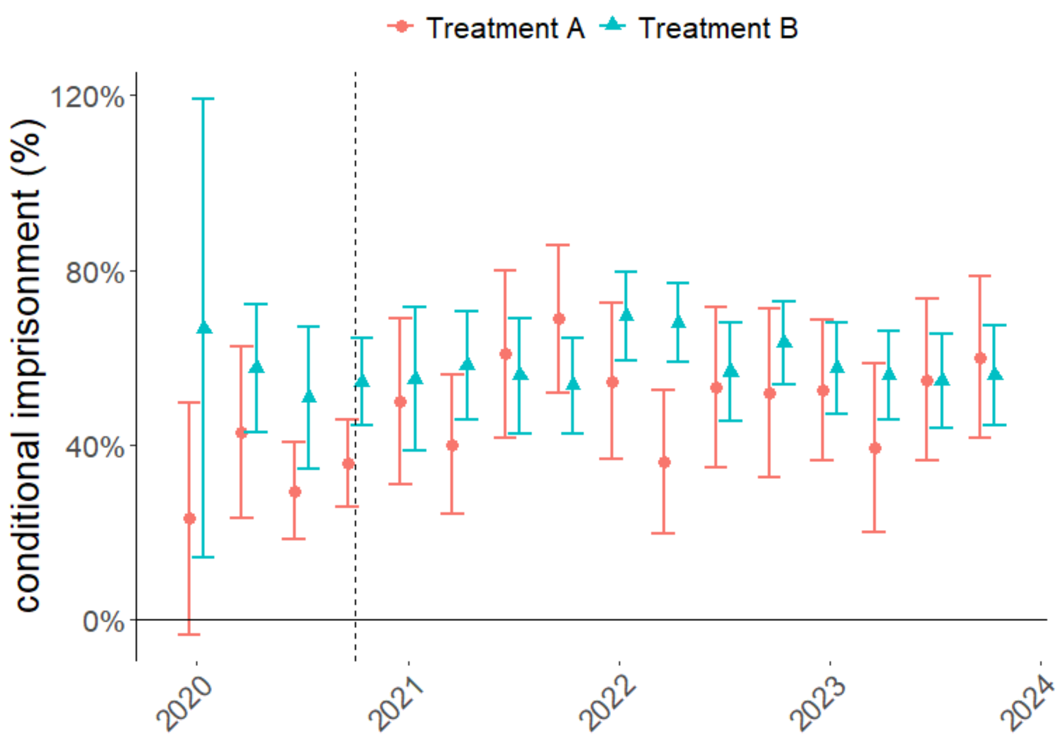
I start by treating the reform as exogenous — I simply compare sentences for the before-and-after reform period. I first explore the evolution of the extensive margin of sentencing - the proportion of cases punished by conditional and unconditional imprisonment. Figure 1 shows that for the Treatment A sample, the unconditional imprisonment rate decreases, whereas the conditional imprisonment rate increases. The decrease occurs immediately with the adoption of the reform. That suggests as the sentencing range decreases, the judges, in general, opt for milder types of punishment. That is consistent with the severity effect of sentencing ranges.

The pattern is less significant for the Treatment B sample. The decrease in unconditional imprisonment and increase in conditional imprisonment rates only became apparent one year after the reform. This result could suggest the presence of a reference effect - as more serious cases are added into the same sentencing range, the judges again seem to respond by imposing milder types of punishment. The decrease comes with a 1-year delay which could signal that the judges first need to gather some experience and gather a sufficient number of cases to form a new reference group.

I examine the intensive margin of sentencing by plotting the evolution of imprisonment length. However, it would not be reasonable to plot the simple average sentence for cases punished by unconditional imprisonment, as one part of the judge's response seems to be a switch to milder types of punishment. Therefore, when studying the intensive margin of sentencing, I recode the variable so that if the case was punished by unconditional imprisonment, I keep its length, and



(a)



(b)

Figure 1: The evolution of types of punishment (extensive margin of sentencing)

Note: Panel (a) shows the unconditional imprisonment rate, panel (b) the conditional imprisonment rate. The dots denote the treatment group quarter averages. 95 percent confidence intervals are shown. The black dashed line denotes the adoption of the reform.

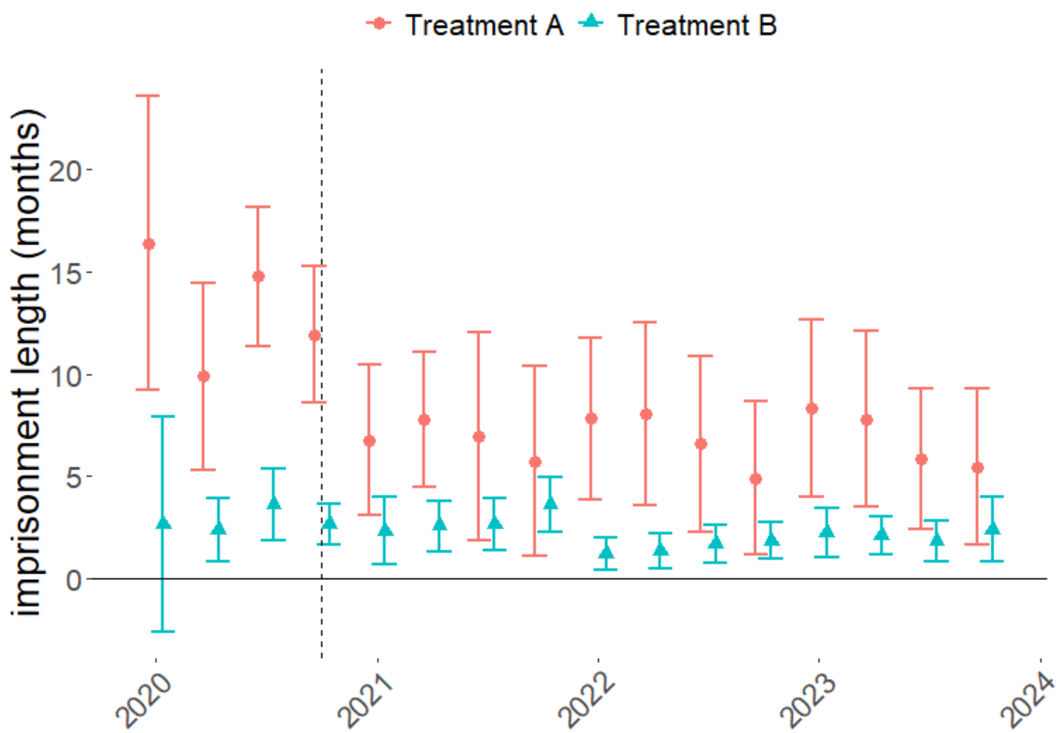


Figure 2: The evolution of unconditional imprisonment length (intensive margin of sentencing)

Note: For all cases punished by alternative means of punishment, I assume that the sentence is equal to 0. The dots denote the treatment group quarter averages. 95 percent confidence intervals are shown. The black dashed line denotes the adoption of the reform.

when it was punished with an alternative punishment (including conditional imprisonment), I plug in 0 for the imprisonment length. Figure 2 shows a decreasing trend for an average sentence defined in this way. Again, the drop for the Treatment A sample is immediate, whereas, for Treatment B, it kicks in later after the reform.

The results of this exercise suggest that the sentence decreases on both the intensive and extensive margins for both treatment groups. However, for causal interpretation, I would need to assume that the introduction of reform was exogenous and that there were no other time-specific impacts around the introduction of the reform. Such assumptions may not be fully substantiated in the setting that I focus on. Therefore, in the following subsection, I estimate the impacts of the reform using the DD approach with controls, which should at least partially overcome this issue.

5.2 DD

The DD method relies on the comparison of our treatment groups of interest to a control group that comprises cases that were unaffected by the reform. Unfortunately, the 2020 reform influenced the sentencing ranges for theft and all other crimes against property, sharply limiting the potential candidates for a control group. Given this limitation, I adopt the obstruction of justice and obstruction of a sentence of banishment (§ 337 of the Criminal Code) as a control group. Similar to theft, considering this offense belongs to the daily routine of most judges. 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 unrelated to any monetary variables; thus, arguably, the sentences should not be influenced by the 2020 reform.

First, I estimate the treatment effects using the standard DD regression with a dummy after-treatment period indicator (Table 4).

$$Y_i = \beta_0 + \beta_1 P_i \cdot T_i + \beta_2 T_i + \beta_3 P_i + \beta_j X_{ij} + e_i, \tag{1}$$

where Y_i is the outcome of interest, P_i is the dummy indicating the after-treatment period and T_i is the treatment indicator, X_i represents a set of covariates.

Table 4 presents the results for three outcomes of interest - unconditional imprisonment, conditional imprisonment, and the length of conditional imprisonment (defined as 0 for the cases punished by other means). The DD estimates confirm that the unconditional imprisonment rate decreases, conditional imprisonment increases, and imprisonment length decreases for both treatment groups.

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. The estimates are statistically significant with and without controls. The first result is perhaps not that

Table 4: DD estimates of the reform effects

Panel A: Treatment A (sentencing range downward shift)

	Dependent variable:					
	unconditional imprisonment		conditional imprisonment		imprisonment length (months)	
	(1)	(2)	(3)	(4)	(5)	(6)
After:Treatment	-0.211^{***}	-0.129^{***}	0.173^{***}	0.129^{***}	-5.758^{***}	-4.506^{***}
	(0.041)	(0.037)	(0.041)	(0.039)	(0.492)	(0.463)
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	45,161	45,160	45,161	45,160	45,161	45,160
Treated observations	594	594	594	594	594	594
R^2	0.001	0.226	0.003	0.164	0.023	0.163
Adjusted R^2	0.001	0.217	0.003	0.154	0.023	0.154

Panel B: Treatment B (addition of more severe cases)

	Dependent variable:					
	unconditional imprisonment		conditional imprisonment		imprisonment length (months)	
	(1)	(2)	(3)	(4)	(5)	(6)
After:Treatment	-0.047	-0.048^{*}	0.067^{**}	0.067^{**}	-0.995^{***}	-0.912^{***}
	(0.030)	(0.027)	(0.031)	(0.029)	(0.356)	(0.335)
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	46,544	46,543	46,544	46,543	46,544	46,543
Treated observations	1977	1977	1977	1977	1977	1977
R^2	0.004	0.225	0.017	0.174	0.0002	0.142
Adjusted R^2	0.004	0.216	0.017	0.164	0.0001	0.132

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: Controls include judge fixed effects, age, number of previous convictions, number of different punishments for the given crime, concurrence, recidivism, juvenile, and gender dummies of the offender. The imprisonment length is set to 0 for cases with alternative punishment (including conditional suspension of punishment).

surprising — the judges respond to a sentencing range decrease by lowering the sentence. The latter result could be interpreted as evidence for a reference effect — when more severe cases were added, the average sentence decreased.

To understand the time evolution of the reform effects, I leverage an event study in a DD framework. In particular, I divide the cases according to the quarter when the sentence was passed.[1]. Then, for each quarter q, I run the following specification

$$Y_i = \alpha_q + \beta_q T_i + \gamma_{jq} X_{ij} + e_i, \tag{2}$$

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.

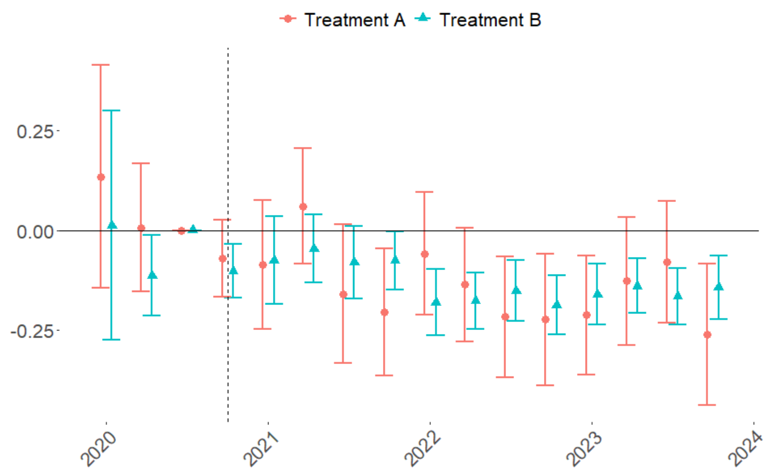
My interest falls on the coefficient β_q, which shows the difference between the control and treatment groups in a period q. I normalize these coefficients by taking the coefficient one period before the reform β_{-1} as the baseline level. Then, the coefficients show how the difference between the control and treatment groups evolved over time. Before the treatment, most coefficients should be close to zero, signaling that in the before-treatment period, there has not been any difference in trends for the control and treatment groups.

Figure 3 shows the main event study plot for both treatment groups. For Treatment A, the coefficients before treatment are insignificant. For Treatment B, however, some coefficients are significantly negative. However, these are periods with quite little data, which may influence the estimates. In the after-reform period, both treatment groups elicit a clear decreasing trend in unconditional imprisonment rate and the length of unconditional imprisonment. Conversely, the conditional imprisonment rate increased for both treatment groups.

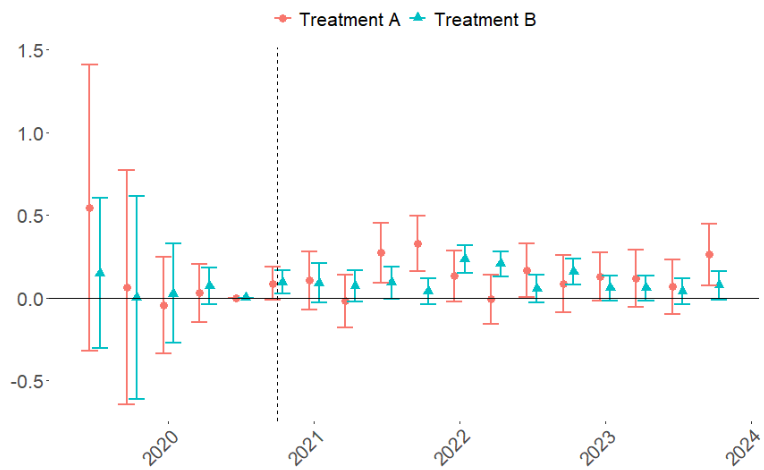
The results of this exercise are in line with the estimates obtained with the binary treatment indicator and the preliminary analysis of the trends in sentencing. In light of the mechanisms described in Section 4, I bring evidence for both severity and reference effect. The severity effect emerges for the Treatment A cases (where the sentencing range was lowered) and lowers the sentence. The reference effect arises when Treatment B cases get compared to their new, more severe control group and, therefore, are punished mildly. This pattern is applied both on the intensive (lowering the length of the sentence) and on the extensive margin (switching to less severe types of punishment).

In the Appendix, I perform several robustness checks of my DD results. In Section A.3, I replicate the DD analysis with a different control group. Even though the event study plots become quite noisy, I still reconstruct the main pattern using the binary treatment indicator. Additionally,

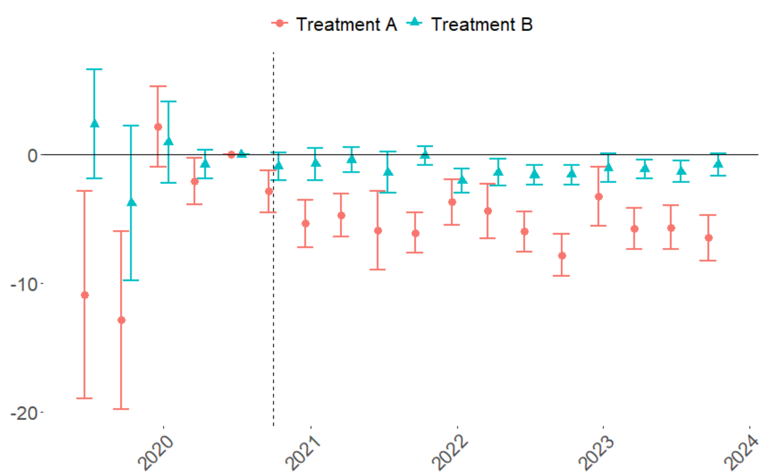
[1] Unfortunately, a finer partition was not possible due to the amount of data before reform.



(a) unconditional sentence rate



(b) conditional sentence rate



(c) unconditional imprisonment length

Figure 3: The quarterly effects on the outcomes of interest

Note: The Figure shows the quarterly coefficients on unconditional imprisonment, conditional imprisonment, and imprisonment length. 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.

in Section A.4, I examine different samples of treated cases, which seems to confirm the results obtained with the original sample.

6 Discussion

In 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 group separately

By Treatment A, I denote the cases with damage CZK 50k-100k, for which the sentencing range itself shifted downwards - from 1-5 years to 0-2 years. The DD estimation suggests that the sentence decreased by 5 months, and at the same time, it was less often punished by unconditional imprisonment. This decline occurred immediately after the reform, suggesting that judges immediately noticed and applied the new legal guidelines. Nevertheless, for this sample, we observe that the evolution of outcomes in the before-treatment period was quite turbulent. That could be explained by a lack of data from the before-treatment period, where the damage caused was often not reported. 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 respond to sentencing range design change, which highlights the importance of their reasonable and substantiated design.

For Treatment B cases, the sentencing range remained unchanged (0-2 years), but more severe cases were added to the same sentencing range. Here, the unconditional imprisonment rate decreases slightly, whereas the conditional imprisonment rate increases. The average sentence drops by 1 month. This pattern is observable mostly with 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 gradual and gains its significance only one year after the reform. The drop could be explained through the reference effect - after the reform, the cases seem to be less severe compared to the new reference group, 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.

Moreover, my results imply that, under the presence of a reference effect, even cases with unchanged sentencing ranges may be affected by reference group shifts. As a result, the judges may respond by adjusting the sentences to a new reference group. This highlights the importance of a substantiated control group choice when performing any analysis of sentencing guidelines reforms.

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 (including conditional imprisonment), which impairs the analysis of sentencing trends and necessitates a credible strategy for handling these cases. Therefore, when addressing the intensive margin of sentencing, for cases punished with alternative punishment, I assume that the imprisonment length is zero. This approach reflects the idea that judges may opt for a milder punishment by using alternative methods. I support this choice by carefully examining the extensive margin of sentencing before and after the reform.

Furthermore, 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. 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 range 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 study the general equilibrium effects of different sentencing range 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 range design.

Conclusion

This paper focuses on the impact of sentencing ranges design on sentences in the context of a 2020 Czech reform. Building on Drápal and Šoltés (2024), I model 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.

I demonstrate these effects using 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 1-5 years to 0-2 years) and the cases for which the sentencing range remained constant (0-2 years), but more severe cases were added into that sentencing range. I examine the change in sentences for each treatment group using DD, 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 unconditional imprisonment rate decreases, the conditional imprisonment rate increases, and the average sentence decreases by 5 months. I attribute this result 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, followed by a mild increase in the conditional imprisonment rate. That speaks towards the reference effect, where the judge compares cases to more severe ones within the same range.

The key takeaway is that sentencing ranges set up reference categories that may shape sentences. Additionally, I offer a methodological remark that since sentencing ranges, reforms may also indirectly affect the reference groups by reshuffling them. This could be particularly important when choosing a control group to evaluate the reform's effect. My results also 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.

Beckett, K., Evans, H., & Harris, A. (2023). Who benefits from criminal legal reform? a natural experiment to assess racial disparities in a policy targeting monetary sanctions. *Journal of Experimental Criminology*. doi: 10.1007/s11292-023-09597-3

Bjerk, D. (2017a). Mandatory minimum policy reform and the sentencing of crack cocaine defendants: An analysis of the fair sentencing act. *Journal of Empirical Legal Studies*, *14*(2), 370-396. doi: https://doi.org/10.1111/jels.12150

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

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. doi: 10.1257/pol.20170079

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. (2024). Sentencing decisions around quantity thresholds: theory and experiment. *Journal of Experimental Criminology*, *20*(4), 1323–1367. doi: 10.1007/s11292-023-09568-8

Fischman, J. B., & Schanzenbach, M. M. (2012). Racial disparities under the federal sentencing guidelines: The role of judicial discretion and mandatory minimums. *Journal of Empirical Legal Studies*, *9*(4), 729–764. doi: 10.1111/j.1740-1461.2012.01266.x

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

Schulhofer, S. J. (1992). Plea bargaining as disaster. *The Yale Law Journal*, *101*(8), 1979–2009. doi: 10.2307/796861

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

Teichman, D., Zamir, E., & Ritov, I. (2023). Biases in legal decision-making: Comparing prosecutors, defense attorneys, law students, and laypersons. *Journal of Empirical Legal Studies*, *20*(4), 827–856. doi: 10.1111/jels.12379

Tuttle, C. (2023). Racial disparities in federal sentencing: Evidence from drug mandatory minimums.

Weswasi, E. A., Sivertsson, F., Bäckman, O., & Estrada, F. (2023). Does sentence length affect the risk for criminal recidivism? a quasi-experimental study of three policy reforms in sweden. *Journal of Quantitative Criminology*, *39*, 787–823. doi: 10.1007/s11292-022-09513-1

Witwer, A. R. (2023). Discretion and disagreement: A longitudinal study of departures under presumptive and advisory sentencing guidelines. *Crime & Delinquency*. doi: 10.1177/00111287231218701

A Appendix

A.1 Damage report rate

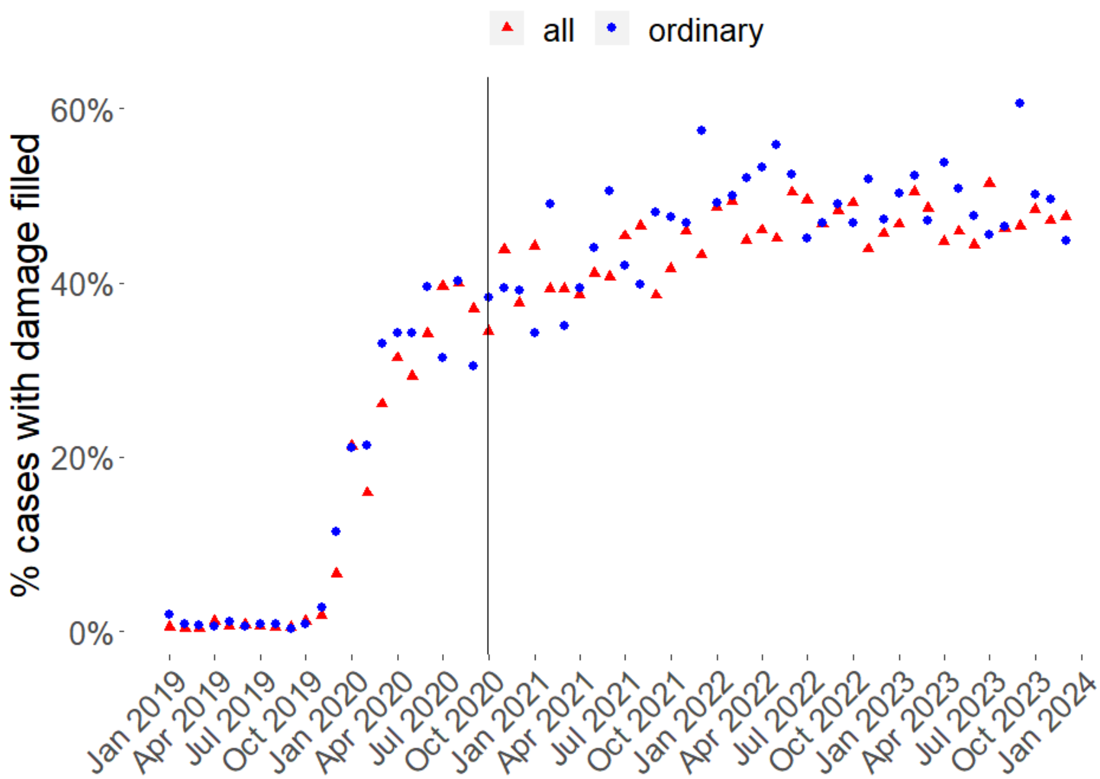


Figure A.1: The rate of cases with damage filled before and after the reform

Note: Red triangles represent all theft cases; the blue dots represent ordinary theft cases (cases where damage is the criterion determining the sentencing range that were used as the main sample). 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 Decriptive characteristics of the dataset

Table A.1: Descriptive statistics of the main dataset

	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 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
conditional 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

Note: The year range is limited to 2019-2023. Ordinary cases are defined as cases where the criterion determining the sentencing range was the damage caused. By recidivist, I denote the offenders where the court counted the offender's previous convictions as an aggravating circumstance (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.3 Alternative control group

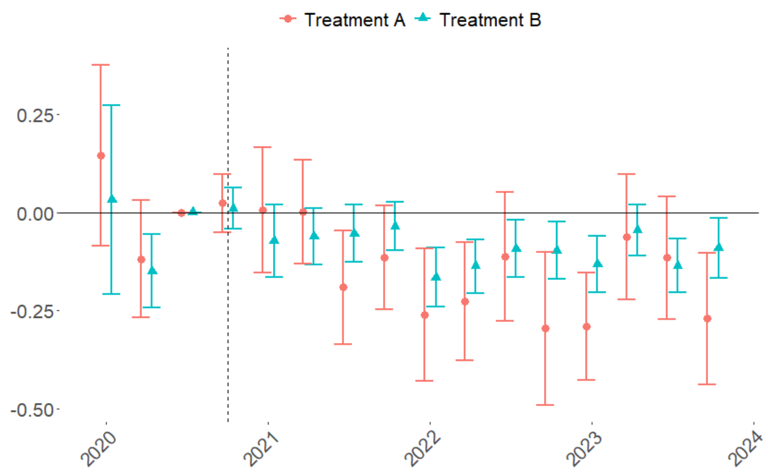
To address the concerns that my findings may be driven by an arbitrary choice of control group, I replicate the results using an alternative control group.

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.

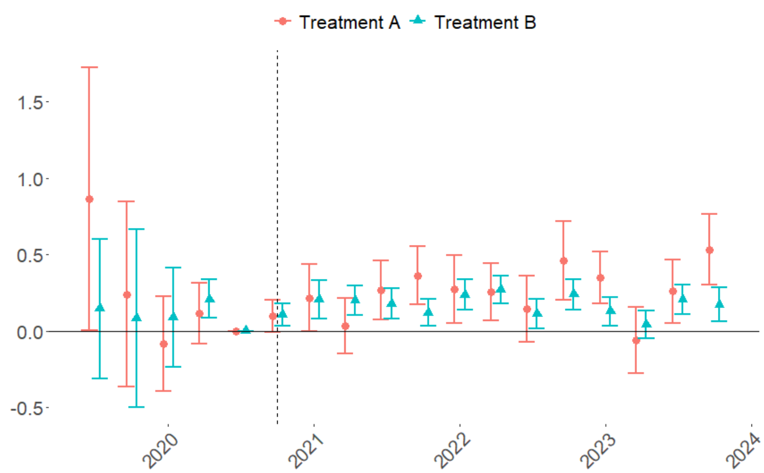
Therefore, apart from the control group used in the main analysis, I introduce the cases of negligence of mandatory support, another common crime that the courts consider regularly. Moreover, this crime is at least vaguely related to property — the offender profits from not paying the support.

I examine both pre-tends and the effect of the treatment through a DD event study for unconditional imprisonment, conditional imprisonment, and length of unconditional imprisonment (Figure A.2). Additionally, I also estimate the reform effects using the binary treatment indicator (Table A.2).

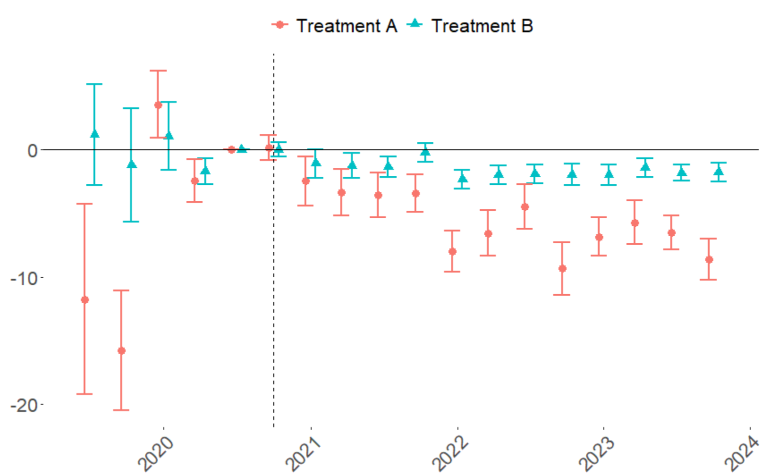
Overall, I replicate the results obtained in the main analysis -—- for both treatment groups, the unconditional sentence rate decreases, the conditional sentence rate increases, and the length of the sentence decreases by around 5 months for Treatment A and around 1 month for Treatment B. Nevertheless, even with this control group, I still obtain quite noisy coefficients in the before-reform period (especially for Treatment A). That could support the explanation that these are driven by a low amount of treated observations with damage reported in the sample.



(a) unconditional imprisonment rate



(b) conditional imprisonment rate



(c) unconditional imprisonment length

Figure A.2: The quarterly effects on sentence for the alternative control group

Note: The Figure shows the quarterly coefficients on unconditional imprisonment, conditional imprisonment, and imprisonment length for the alternative control group (negligence of mandatory support). 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.

Table A.2: DD estimates of the reform effects, negligence of mandatory support as a control group

Panel A: Treatment A (sentencing range downward shift)

	Dependent variable:					
	unconditional imprisonment		conditional imprisonment		imprisonment length (months)	
	(1)	(2)	(3)	(4)	(5)	(6)
After:Treatment	-0.203^{***}	-0.139^{***}	0.146^{***}	0.119^{***}	-5.673^{***}	-4.785^{***}
	(0.030)	(0.029)	(0.045)	(0.039)	(0.354)	(0.343)
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	27,691	27,691	27,691	27,691	27,691	27,691
Treated observations	594	594	594	594	594	594
R^2	0.014	0.136	0.002	0.303	0.088	0.208
Adjusted R^2	0.014	0.120	0.002	0.290	0.088	0.193

Panel B: Treatment B (addition of more severe cases)

	Dependent variable:					
	unconditional imprisonment		conditional imprisonment		imprisonment length (months)	
	(1)	(2)	(3)	(4)	(5)	(6)
After:Treatment	-0.038^{*}	-0.041^{*}	0.040	0.060^{**}	-0.911^{***}	-0.892^{***}
	(0.022)	(0.021)	(0.033)	(0.029)	(0.253)	(0.242)
Intercept	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
Observations	29,074	29,074	29,074	29,074	29,074	29,074
Treated observations	1977	1977	1977	1977	1977	1977
R^2	0.0003	0.124	0.0001	0.289	0.007	0.137
Adjusted R^2	0.0002	0.109	−0.00005	0.277	0.007	0.122

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: Reform effects estimated with an alternative control group (breaking and entering). Controls include judge fixed effects, age, number of previous convictions, number of different punishments for the given crime, concurrence, recidivism, juvenile, and gender dummies of the offender. The imprisonment length is set to 0 for cases with alternative punishment (including conditional suspension of punishment).

A.4 Alternative samples

In the main analysis, I use a narrow range of damage as a Treatment A and Treatment B sample. Nevertheless, in principle, Table 2 offers several damage ranges for each treatment group. In this section, I analyze different samples for each treatment. First, I focus 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. Second, I pool 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. Other sample choices are not available due to the low number of cases with very high damage (over 1m).

To overcome the differences in the scale for different treatment groups, I standardize the sentences using the before-treatment control group mean and standard deviation (z-score normalization). The coefficients show how the sentence changes in terms of the standard deviations relative to the control group. For comparison, in Table A.3, I also present the normalized estimates for the original sample (10k-50k 50k-100k).

The fundamental intuition seems robust against the choice of sample — most coefficients are significant and in the same direction as in the main analysis, supporting the interpretation of the judges moving to less severe types of punishment. However, the precise values of the estimates differ slightly.

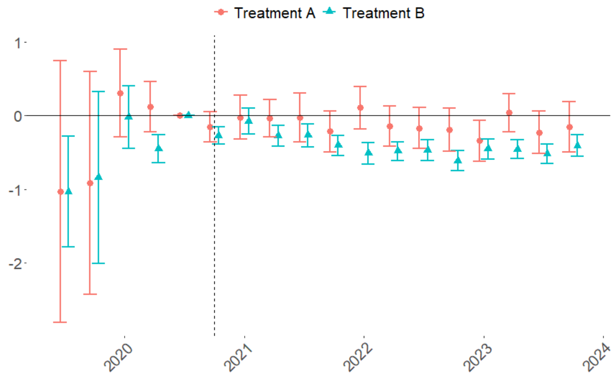
I conclude this analysis with event study DD plots for the pooled sample and the alternative sample for each treatment group (Figure A.3). The results for the pooled sample seem to reflect the results obtained with the basic sample; however, they lose some significance. Conversely, the alternative sample does not entirely confirm the patterns obtained with the main sample. For unconditional sentence rate, the coefficients end up being mostly insignificant and some even positive, for the imprisonment length, these are quite noisy both before and after treatment. This could be either driven by the low number of observations in this sample or a different nature of these cases compared to those that are less severe.

Table A.3: Estimates of the treatment effect for different samples

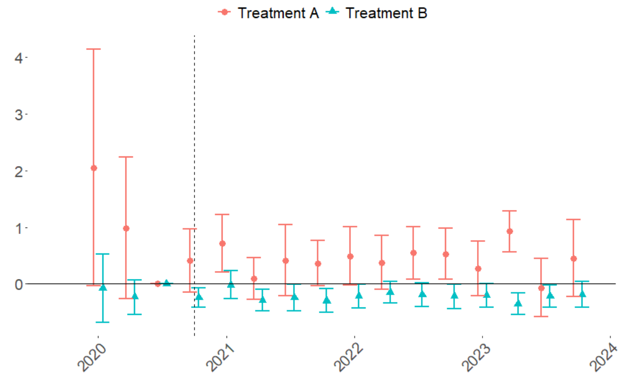
	Dependent variable (z-score standardized):								
	unconditional imprisonment			conditional imprisonment			imprisonment length (months)		
Panel A: Treatment A (sentencing range downward shift)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
After:Treatment	−0.289***	−0.071	−0.145*	0.280***	0.249	0.181**	−0.903***	−1.385***	−0.313***
	(0.083)	(0.256)	(0.076)	(0.084)	(0.259)	(0.077)	(0.086)	(0.269)	(0.083)
Damage range	50k-100k	500k-1m	pooled	50k-100k	500k-1m	pooled	50k-100k	500k-1m	pooled
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	45,160	44,815	45,415	45,160	44,815	45,415	45,160	44,815	45,415
Treated observations	594	249	849	594	249	849	594	249	849
R^2	0.226	0.225	0.227	0.164	0.164	0.163	0.194	0.180	0.190
Adjusted R^2	0.217	0.215	0.218	0.154	0.154	0.153	0.185	0.170	0.181
Panel B: Treatment B (addition of more severe cases)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
After:Treatment	−0.107*	−0.076	−0.168***	0.145**	0.115*	0.149***	−0.169***	−0.474***	−0.349***
	(0.060)	(0.065)	(0.045)	(0.062)	(0.065)	(0.046)	(0.063)	(0.076)	(0.054)
Damage range	10k-50k	100k-500k	pooled	10k-50k	100k-500k	pooled	10k-50k	100k-500k	pooled
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	46,543	46,208	48,241	46,543	46,208	48,241	46,543	46,208	48,241
Treated observations	1977	1642	3675	1977	1642	3675	1977	1642	3675
R^2	0.225	0.227	0.229	0.174	0.169	0.168	0.163	0.286	0.238
Adjusted R^2	0.216	0.218	0.220	0.164	0.160	0.158	0.153	0.277	0.229

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

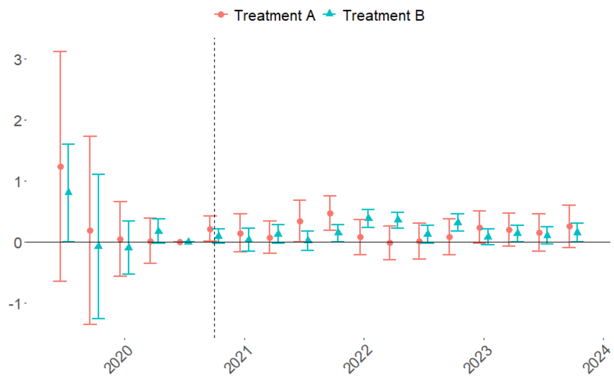
Note: The estimated impact of the reform on the variables of interest. All outcomes are z-score normalized. The first column for each outcome variable denotes the sample used in the main analysis. Controls include judge fixed effects, age, number of previous convictions, damage, number of different punishments for the given crime and concurrence, recidivism, juvenile and gender dummies of the offender.



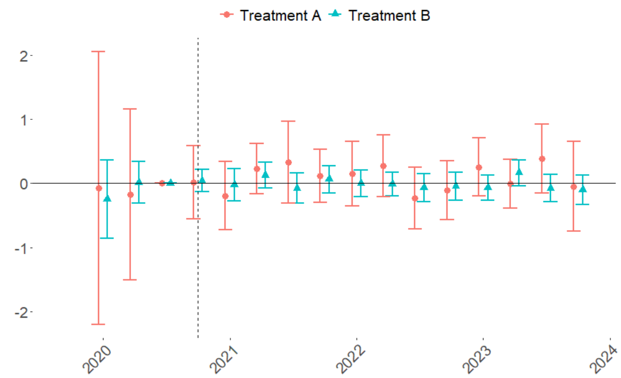
(a) unconditional sentence rate, pooled sample



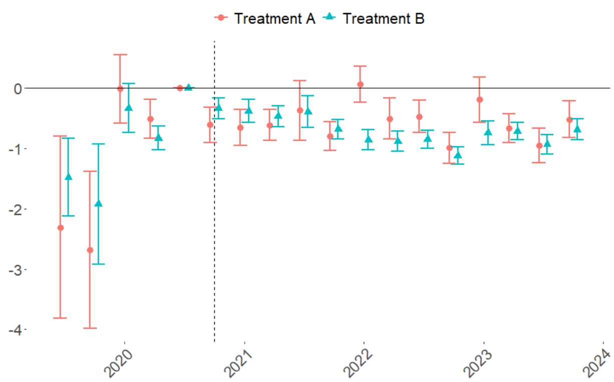
(b) unconditional sentence rate, alternative sample



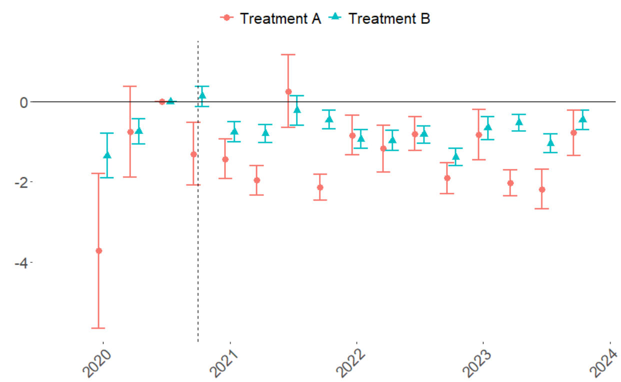
(c) conditional sentence rate, pooled sample



(d) conditional sentence rate, alternative sample



(e) imprisonment length, pooled sample



(f) imprisonment length, alternative sample

Figure A.3: The quarterly effects on the outcomes of interest for the pooled sample

Note: The Figure shows the quarterly coefficients on unconditional imprisonment, conditional imprisonment, and imprisonment length. The Treatment A pooled sample captures cases between 50k-100k, 500k-1m, and 5m-10m. The Treatment B pooled sample captures cases between 10k-50k, 100k-500k, and 10m-50m. The Treatment A alternative sample captures cases between 500k-1m The Treatment B pooled sample captures cases between 100k-500k. 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.