

# Reference and Severity Considerations in Sentencing: Evidence from a Czech Criminal Law Reform

January 30, 2026

Tereza Burýšková<sup>12</sup>

## Abstract

Sentencing ranges can influence judicial decision-making through multiple behavioral channels. This paper studies a reform of the nominal damage thresholds classifying theft offenses into statutory categories. As a result of this reform, some offenses were moved to less serious categories, facing reduced sentencing ranges (e.g., from 2–8 to 1–5 years of imprisonment), while for others the formal range remained unchanged (e.g., 1–5 years) but the category was expanded to include more severe cases. I find that judges impose significantly shorter prison sentences both for offenses with explicitly lowered sentencing ranges and for offenses whose statutory ranges remained unchanged yet came to encompass more severe conduct. The findings provide evidence for two distinct mechanisms: a severity channel, whereby sentencing ranges signal the legislator’s assessment of offense seriousness, and a reference channel, whereby judges evaluate cases relative to other offenses within the same statutory range. The findings demonstrate that sentencing range reforms can have broader and more nuanced effects than those implied by formal changes in statutory minimums and maximums alone. More broadly, the paper highlights the importance of reference-based judicial reasoning in the development and assessment of sentencing policy.

JEL classification: K14, K42

Keywords: sentencing, sentencing ranges, court data, sentencing disparities

---

<sup>1</sup>Faculty of Law, Charles University, náměstí Curieových 901/7 116 40 Prague 1, Czech Republic, tereza.buryskova164@student.cuni.cz

<sup>2</sup>I thank Keren Weinshall, Michal Šoltés, and Jakub Drápal for their helpful comments. The project has received funding from the European Regional Development Fund (Center for Inequality and Open Society, Project No. CZ.02.01.01/00/23\_025/0008690)

# 1 Introduction

Sentencing ranges — statutory minimum and maximum penalties for specific offenses — are a widely used policy tool in criminal law, promoting consistency and reducing judicial disparities (Klein, 2002; Ulmer & Johnson, 2004). These ranges usually reflect selected case characteristics such as offense type or monetary harm. While designed as legal constraints, sentencing ranges also shape judicial reasoning and may generate behavioral spillovers beyond their formal reach.

This paper examines the impact of changes in sentencing ranges on judicial decisions, leveraging a 2020 reform of theft sentencing guidelines in the Czech Republic as a natural experiment. The reform reclassified certain theft offenses by adjusting monetary thresholds, thereby changing which offenses fell into which sentencing categories. Consistent with the mechanisms introduced by Drápal and Šoltés (2024), I show that judicial responses arise both from direct changes in the statutory range (a severity channel) and from shifts in how judges perceive a case’s seriousness relative to others in the same range (a reference channel).

The severity channel reflects that the sentencing range provides a coarse signal of case severity. Accordingly, cases assigned to a lower range should, all else equal, receive lower sentences. For instance, with the 2020 reform, a case involving damage of CZK 60k (approximately 2,500 EUR) fell within the sentencing range of 1-5 years before the reform, and 0-2 years after the reform. The reference channel arises because judges compare cases within the same legal category, implying a similar sentencing range. The relative position of a case influences the final sentence — cases that are less severe relative to their reference group receive lower sentences than otherwise similar cases that are more severe within their group. For instance, a case with damage of CZK 20k (approximately 800 EUR) still faces the same sentencing range of 0-2 years; however, after the reform, the same sentencing range is also applied to a case of CZK 60k. With new, more serious cases added to a category, cases within the original range may be sentenced more leniently than they would have been otherwise, because now these seem relatively less severe in their category. This reference dependence of sentencing has been described in the law literature (Rachlinski, Johnson, Wistrich, & Guthrie, 2009; Englich et al., 2006; Leibovitch, 2017) and has its foundations in behavioral economics and psychology (Kőszegi & Rabin, 2006; Bordalo et al., 2012; Fischhoff, 1975). Consequently, the reform may also affect sentencing patterns for damage ranges, whose sentencing ranges remained unchanged. This finding raises concerns about using such cases as a control group in sentencing research.

Leveraging court data from 2019 to 2023, I examine both the extensive margin (the proportion of cases punished by custodial imprisonment) and the intensive margin (the average sentence length, conditional on imprisonment) of sentencing for each group separately. Using a difference-in-differences (DD) approach, I find that judges respond to the reform primarily on the intensive

margin, reducing the length of sentences. These reductions occur both for cases with lowered sentencing ranges and for cases whose ranges remained unchanged but were expanded to include more severe offenses. The results for the first group support the severity channel, indicating that a lower sentencing range is associated with more lenient sentences. The results for the second group support the reference channel: when more severe cases are added to a range, previously typical cases appear less serious by comparison, resulting in lighter sentences.

Conversely, the reform had little effect on the extensive margin of sentencing. The distribution of punishment types elicits no statistically significant changes. This finding is crucial for understanding the overall impact of the reform. It suggests that judges did not respond by shifting between types of sanctions, but rather adjusted sentence lengths within the existing framework of custodial punishment. Without this stability on the extensive margin, the observed reductions in sentence lengths could reflect a broader change in sentencing philosophy or composition of punished cases, making the intensive margin results harder to interpret.

These results also carry important methodological implications. Sentencing ranges that remain formally unchanged by a reform can still be indirectly affected through shifts in the composition of cases they encompass. For example, if one ignores the reference effect and assumes that such cases are unaffected by the reform, they may mistakenly use them as a control group. In my setting, this would lead to an overestimation of the reform's impact, as these cases also experience reductions in sentence length due to their new relative position within the sentencing range. Researchers using difference-in-differences or similar quasi-experimental designs should verify whether the control group has undergone a change in its reference group. For property crimes, where the main criterion is the damage caused, the reference group is easily assessable. However, for other crimes, the relevant control group may be challenging to describe. Neglecting reference considerations in the control group may lead to biased estimates and misinterpretation of the true effects of sentencing reforms. In this paper, I overcome this concern by adopting cases of obstruction of justice and obstruction of a sentence of banishment as a control group. Its sentencing ranges are not based on any monetary thresholds; therefore, it is reasonable to assume that it was not touched by the reform, and it can be used as a benchmark to identify the reference and severity channels in theft sentences.

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 data and estimation strategies used in the empirical analysis; Section 6 presents the results of the empirical analysis, and Section 7 discusses their implications. The main contribution is summarized in the Conclusion.

## 2 Literature Review

A growing interdisciplinary literature emphasizes that legal rules influence behavior not only through formal enforcement and material sanctions, but also by shaping social norms, expectations, and reference points. In this view, the law serves an expressive function, conveying information about socially acceptable behavior, and coordinating beliefs, even in settings where legal rules leave substantial discretion or are weakly enforced.

Lane et al. (2023) provide empirical evidence that legal rules affect behavior by clarifying normative expectations and shaping social norms. Using experimental and field data, they show that individuals respond to legal rules even in the absence of strong enforcement, suggesting that the law operates as a signal of social meaning and appropriateness.

Sentencing law provides a particularly salient setting in which these expressive and normative functions of law may operate. Sentencing rules typically define broad statutory ranges rather than precise prescriptions, leaving judges with substantial discretion to translate legal categories into individualized punishment decisions. As a result, sentencing ranges may serve not only as legal constraints, but also as signals of offense severity and as reference points for evaluating cases relative to others.

The impact of sentencing policies on judicial decision-making has been a topic of interest in the empirical legal literature. Much of this work focuses on the United States, studying the effects of mandatory minimums and sentencing guidelines as formal tools designed to constrain judicial discretion and reduce unwarranted disparities (Klein, 2002; Ulmer & Johnson, 2004). Mandatory minimum reforms have been shown to significantly affect sentencing outcomes, particularly in drug cases, by limiting judicial discretion (Bjerk, 2017, 2007). Anderson et al. (1999) document that the adoption of federal sentencing guidelines reduced overall sentence variability but did not eliminate inter-judge disparities, in part due to prosecutorial discretion.

Subsequent research highlights that substantial disparities persist even under formal sentencing regimes. A prominent strand of this literature documents racial inequalities in sentencing outcomes. Tuttle (2023) shows that disparities disproportionately burden Black defendants; Fischman and Schanzenbach (2012) finds that mandatory minimums may intensify racial disparities even when broader discretion could mitigate them; and Burch (2015) demonstrates that intra-racial disparities also arise, with Black defendants with darker skin tones receiving significantly longer sentences. More recent empirical work using large administrative datasets confirms that guideline-based systems continue to produce substantial disparities. Analyzing roughly 380,000 federal cases, Goldrosen et al. (2023) and Topaz et al. (2023) find that White defendants receive shorter sentences than comparable Black and Hispanic defendants, with considerable heterogeneity across judges. Examining the origins of these gaps, McConnell (2022) shows that differential sorting at

earlier stages—such as arrest, charging, and plea bargaining—does not account for the disparities; instead, sentencing decisions themselves appear to drive much of the observed inequality.

A growing comparative literature extends these insights beyond the U.S. context. Drápal and Plesničar (2025) analyze sentencing discretion in former socialist European jurisdictions, arguing that high individualisation, limited institutional guidance, and weak procedural constraints contribute to inconsistent sentencing practices. Using recent Czech data, Drápal and Dušek (2025) demonstrate that non-binding recommendations from penal elites can significantly influence judicial behavior, leading to an increase in the imposition of fines. These findings underscore that even in the absence of binding rules or formal guidelines, informal institutional signals and legal framing can shape sentencing outcomes.

Recent work has increasingly focused on the behavioral mechanisms through which sentencing rules affect judicial decisions (Rachlinski et al., 2009; Englich et al., 2006; Leibovitch, 2017). Leibovitch and Teichman (2025) argue that under conditions of normative ambiguity—where the law does not prescribe a uniquely correct action—legal rules exert influence primarily through their expressive content, anchoring social norms and reference points rather than dictating precise outcomes. Empirical studies leveraging threshold-based identification strategies exploit the fact that cases just above or below a statutory cutoff are similar in observable characteristics but subject to different sentencing ranges. Using this approach, Skugarevskiy (2017) studies Russian drug-possession cases and finds a sharp increase in sentence severity immediately above the statutory threshold. Drápal and Šoltés (2024) conduct an experiment with Czech prosecutors, varying case attributes around sentencing thresholds and manipulating hypothetical sentencing ranges. They identify two distinct mechanisms: a severity channel, whereby sentencing ranges signal the legislator’s assessment of the seriousness of the offense, and a reference channel, whereby cases are evaluated relative to others within the same statutory category. Their findings support both mechanisms.

This paper contributes to the existing literature by providing observational evidence on the operation of both the severity and reference channels in a real-world institutional setting. In contrast to experimental or threshold-based designs, it examines a comprehensive sentencing reform that reclassified offenses without uniformly changing statutory sentencing ranges. By doing so, it highlights how sentencing reforms can influence judicial decisions through expressive and comparative mechanisms, even for cases whose formal legal constraints remain unchanged.

## 3 Institutional Context

### 3.1 Czech Criminal Law

In the Czech Republic, every offense defined in the special part of the Criminal Code is assigned a statutory sentencing range, which consists of a lower and upper bound on imprisonment. These ranges reflect the characteristics of a civil law system. Judges are required to impose a sentence within this range in nearly all cases. In my sample of theft cases, only 6 % of treated cases are punished with a sentence outside the statutory range (conditional on custodial imprisonment). Within the boundaries of the range, however, judges have broad discretion and are not bound by formal sentencing guidelines. There is no centralized system of sentencing recommendations, and judicial practice is largely shaped by individual judgment and informal norms.

The Czech Criminal Code provides for several types of punishment. The most common are custodial imprisonment (served in prison) and suspended imprisonment (where the sentence is not executed unless the offender reoffends). There are several other alternatives, such as house arrest or monetary sanctions. Judges are required to consider both the seriousness of the offense and the individual circumstances of the offender when selecting the appropriate punishment. Factors that guide this assessment include the nature of the offense, the offender's personal background, behavior during and after the crime (such as efforts to mitigate harm or cooperate with authorities), and the anticipated effect of the punishment on the offender's future life. The principle of subsidiarity also applies: if a milder penalty would suffice, a harsher one should not be imposed.

When it comes to crimes against property, such as theft, the amount of damage plays a central role in assessing the severity of the offense. Damage is measured by the value of the property at the time it was stolen. This provides a basis for approximating the severity of the offense—and, by extension, the appropriate punishment—based on monetary thresholds. However, cases with special circumstances — such as burglary or pickpocketing —reflect additional aggravating factors that elevate the severity of the offense beyond what would be implied by damage alone. In my analysis, I focus solely on cases without any special characteristics, where the sentencing range is determined solely by the amount of damage.

### 3.2 The 2020 Reform

In October 2020, the Criminal Code was significantly modified. The definition of the terms determining the extent of damage shifted towards higher nominal values. The rationale of this reform was to align the damage definitions, which were established in 2002, with the current price level<sup>1</sup>.

---

<sup>1</sup>Although this intention is only hypothesized, since this was an amendment without any explanatory report

This change in term definitions implies different legal classifications for cases before and after the reform. For example, a case with damage of CZK 60k (approximately 2,500 EUR) would be classified as a case with *larger damage* before the reform and would be punished with 1-5 years of imprisonment. However, after the reform, the damage would be classified only as *moderate* and would be punished by imprisonment for 0-2 years.

Table 1: Sentencing Ranges for Theft Cases in the Criminal Code

Damage (CZK)	Sentencing Range	
	Before October 2020	Starting 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	
more than 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 summarises the sentencing ranges before and after the 2020 reform. It 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), and at the same time, they were pooled with a less severe group of cases (e.g., 100k-1m). For others, the sentencing range itself did not change; however, the composition of cases within the range shifted towards harsher offenses. For instance, cases with damage CZK 100k-500k face the same sentencing range of 1-5 years; however, the same sentencing range now captures a damage range of CZK 100k-1m, which is relatively more severe compared to the original damage range of CZK 50k-500k.

Table 2 highlights the distinction between the two groups of cases, which I further denote as Treatment A and Treatment B. Next, I form predictions regarding the change in sentences for each group in the presence of severity and/or reference channel.

Table 2: Two Different Types of Theft Cases in Terms of Reform Effects

Damage (CZK)	Sentencing Range		Sentencing Range Thresholds	Damage Range Thresholds
	Before Reform	After Reform		
<b>Treatment A</b>				
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	↓	↑
<b>Treatment B</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	–	↑
more than 10 m	5-10 years	5-10 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.

## 4 Mechanisms in Sentencing: Severity and Reference

As for the mechanisms behind sentencing decisions, I build on Drápal and Šoltés (2024), who introduce severity and reference channels and demonstrate how these effects shape the sentences for cases close to a sentencing range threshold. The authors interpret the sentencing ranges as both categorical indicators of the approximate severity of the crime and as reference groups within which the cases are compared to one another. They denote these different roles of sentencing ranges as the severity channel and the reference channel, respectively.

The authors illustrate the severity and the reference channel 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 101k 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 channel). However, case B is compared to more severe cases falling into the same sentencing range, which decreases the sentence (reference channel). Here, these two effects work against each other. Presumably, 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 channel dominates. Conversely, if the sentence for case B exceeds the one for case A, it demonstrates the dominance of the severity channel. 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 channel tailored to the 2020 reform. First, consider a group of cases whose legal classification (and the corresponding sentencing range) remains unchanged; however, the composition shifted towards higher damage values. Then, since the legal classification is the same, the impact of the severity channel remains constant; however, the original cases now seem to be less severe within their group. Thus, the reference channel (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 the group was pooled with less severe cases. This situation aligns with the Treatment A cases discussed in Section 3. Since these cases are now classified as less severe, we expect lower sentences. However, because these cases now represent the most severe within their new range, the reference channel could drive sentences upward. Thus, in this case, the severity and reference channels work in opposite directions, and the change in sentences depends on their relative importance. Table 3 summarizes the discussed impact of each channel on the mean sentence for both treatment groups. Furthermore, Table 4 provides an interpretation of all possible mean sentence shifts after the reform, which I estimate in my empirical analysis.

Table 3: A Summary of the Impact of Severity and Reference channel on the Sentences

	<b>Severity</b>	<b>Reference</b>
	<b>Impact on Mean Sentence</b>	
<b>Treatment A</b> ( <i>Sentencing Range Downward Shift</i> )	↓	↑
<b>Treatment B</b> ( <i>Damage Range Upward Shift</i> )	0	↓

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

Table 4: Interpretation of Mean Sentence Changes after the Reform.

Observed Mean Sentence	Implication
<b>Treatment A</b> ( <i>Sentencing Range Downward Shift</i> )	
↓	Severity/Severity and Reference
↑	Reference/Severity and Reference
<b>Treatment B</b> ( <i>Damage Range Upward Shift</i> )	
↓	Reference only
↑	Rules out Reference

Note: This table summarizes the interpretation of observed sentence changes under Treatments A and B. Decrease in the mean sentence in both groups indicates the presence of both severity and reference channels.

## 5 Data and Empirical Strategy

### 5.1 Data

In the Czech Republic, criminal cases are comprehensively documented, and detailed case-level data are available for research. In particular, three main sets of variables are available for each case. First, there is data about the criminal procedure, including the court and judge identifier; 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 and free from systematic bias.

The data span the period from 2006 to 2023. However, the damage caused, which is central to my analysis, has been reported only for cases decided after October 2019, 12 months before the reform. Appendix Figure A.1 shows that a stable report rate of around 50 % emerged by the beginning of 2020, as the court staff started to adjust to the newly introduced variables.

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 suspended sentence. Moreover, in line with the previous literature (Drápal, 2023), I find that the average length of a suspended sentence is systematically lower than that of a custodial sentence, suggesting that judges may perceive these as distinct punishment types.

## 5.2 Estimation

To ensure the comparability of treated and control units and reduce concerns about treatment effect heterogeneity, I restrict the analysis to cases with damage between CZK 50,000 and 100,000 (Treatment A sample) and between CZK 10,000 and 50,000 (Treatment B sample). These thresholds are chosen based on the location of the reform-induced threshold changes. In Appendix A.6, I show that the overall patterns are robust to alternative sample definitions.

### 5.2.1 DD

The DD approach compares changes in sentencing outcomes over time between treated and untreated cases. A key identifying assumption is that, absent the reform, trends in sentencing would have evolved similarly for both groups — the parallel trends assumption. In the next subsection, I partially assess this assumption using the dynamic DD model.

A central challenge in this setting is selecting a suitable control group. The 2020 reform simultaneously altered sentencing thresholds for theft and most other crimes against property, sharply reducing the availability of unaffected comparison groups. Therefore, I use cases of obstruction of justice and obstruction of a sentence of banishment (§ 337 of the Criminal Code) as the control group. These offenses are handled routinely by judges, often involve similar procedural dynamics, and were not subject to any legal changes in the period of interest. Moreover, unlike theft, they are not tied to monetary values, making them unlikely to have been indirectly affected by the reform.

I estimate the average treatment effect using the following DD specification for each treatment group separately:

$$Y_i = \beta_0 + \beta_1 \cdot P_i \cdot T_i + \beta_2 T_i + \beta_3 P_i + X_i' \delta + \varepsilon_i, \quad (1)$$

where:

$Y_i$  is the sentencing outcome of interest (length of unconditional imprisonment for the intensive margin, unconditional imprisonment indicator for the extensive margin),

$T_i$  is a treatment group indicator,

$P_i$  is an after-reform indicator,

$X_i$  is a vector of covariates including judge fixed effects, number of previous convictions, offender's age, gender, and a concurrence dummy.

$\varepsilon_i$  is the error term.

The coefficient  $\beta_1$  captures the average effect of the reform on treated cases relative to controls, while adjusting for time-invariant differences across groups and common shocks that affect both groups over time.

## 5.2.2 Dynamic DD

To analyze how the reform's effects evolve over time, I estimate a dynamic difference-in-differences (dynamic DD) model. This approach allows me to assess both the presence of pre-trends and the temporal structure of the treatment effect.

Let  $\tau(i)$  denote the time relative to the reform for individual  $i$ , measured in quarters (event time), such that  $\tau = 0$  corresponds to the quarter of the reform,  $\tau = -1$  is the last pre-reform period, and  $\tau = k$  denotes  $k$  quarters after the reform. I define a series of event-time indicators  $D_{\tau(i)}$  and interact them with the treatment indicator  $T_i$ , which equals 1 for individuals in the treatment group and 0 otherwise.

The dynamic DD specification is:

$$Y_i = \sum_{\tau \neq -1} \beta_{\tau} \cdot D_{\tau(i)} \cdot T_i + \alpha_{q(i)} + \gamma_j + X_i' \delta + \varepsilon_i, \quad (2)$$

$Y_i$  is the outcome of interest,

$D_{\tau(i)}$  is a dummy for event time  $\tau$ ,

$T_i$  is the treatment group indicator,

$\alpha_{q(i)}$  are calendar-quarter fixed effects,

$\gamma_j$  are judge fixed effects,

$X_i$  is a vector of covariates including judge fixed effects, number of previous convictions, offender's age, gender, and a concurrence dummy.  $\varepsilon_i$  is the error term.

The coefficients  $\beta_{\tau}$  capture the treatment effect in each period  $\tau$  relative to the reference period  $\tau = -1$ , which is omitted from the regression. Under the identifying assumption of parallel trends, the  $\beta_{\tau}$  coefficients for  $\tau < 0$  should be close to zero. Post-reform coefficients ( $\tau \geq 0$ ) trace out the dynamic impact of the reform over time.

This specification enables a flexible analysis of treatment effect dynamics and allows for a visual inspection of pre-treatment trends. While not a formal test, the coefficients in the pre-reform periods provide suggestive evidence regarding the plausibility of the parallel trends assumption.

## 6 Results

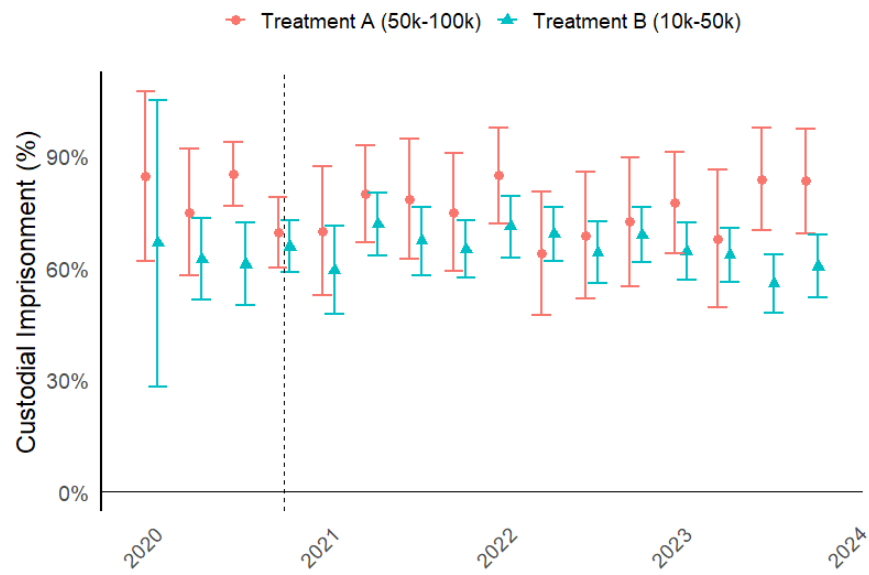
In this section, I present the effects of the reform on custodial sentences, examining both intensive and extensive margins. In the main analysis, I restrict my sample only to cases punished by a custodial sentence, which may raise some doubts about selection bias. In Appendix Section A.3, I address these concerns by running a dynamic DD for punishment type indicators and comparing characteristics of the cases punished with custodial punishment before and after the reform. There,

I find no systematic differences.

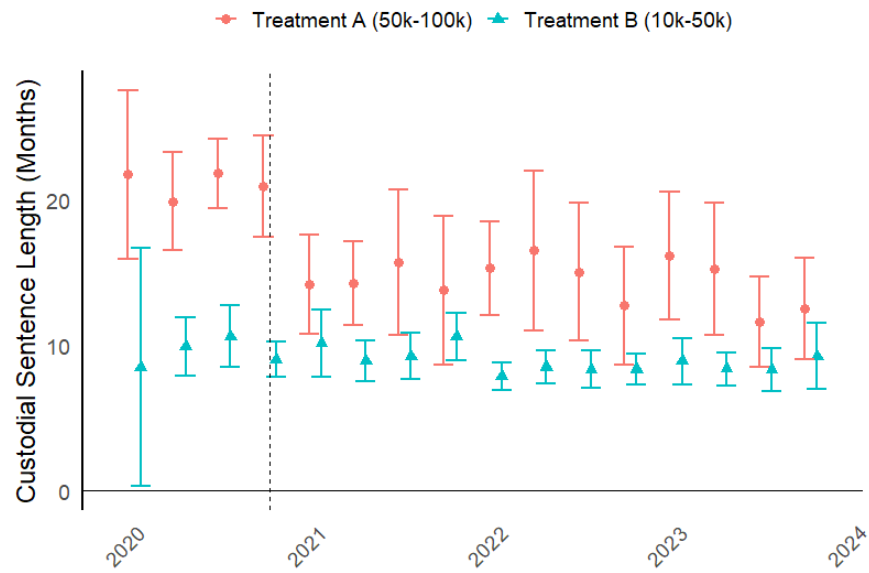
## 6.1 General Trends in Sentencing

As a first step, I plot the evolution of the outcomes of interest. I examine how sentencing outcomes change across the reform date, focusing separately on the extensive margin (the probability of suspended or custodial imprisonment) and the intensive margin (the length of custodial imprisonment conditional on being imposed). These descriptive trends offer a preliminary assessment of the reform's impact.

I compute the custodial sentence rate and average custodial sentence length for each quarter before and after the reform and plot their evolution. Figure 1 (a) shows no significant pattern for the custodial imprisonment rate in both treatment groups. Figure 1 (b) indicates that for the Treatment A sample, the custodial imprisonment length drops immediately with the adoption of the reform. This suggests that as the sentencing range decreases, judges tend to opt for milder punishments, which may be related to the severity channel. The pattern is less pronounced for the Treatment B sample. Minor decreases in imprisonment length only become apparent one year after the reform. This result could be supported by the reference channel, as more serious cases are added to the same sentencing range, the judges again seem to respond by imposing milder punishments. The decrease comes with a one-year delay, which could signal that the judges first need to gather some experience and a sufficient number of cases to form a new reference group.



(a) Extensive margin



(b) Intensive margin

Figure 1: The time evolution of custodial sentence probability and length

Note: Panel (a) shows the custodial imprisonment rate, panel (b) the custodial imprisonment length. 95 percent confidence intervals are shown. The black dashed line denotes the adoption of the reform.

## 6.2 DD

Table 5 presents the results of the DD regressions for custodial sentence. The DD estimates confirm that the imprisonment length significantly decreases by around 5-6 months for Treatment A and 1 month for Treatment B. Conversely, the custodial imprisonment rate remains unchanged. This result aligns with the trends we observed when plotting the general sentencing trends. To show that the decrease in sentences is not just mechanical (for instance, that the after-reform drop in sentences is not driven solely by the threshold cases), I plot the custodial imprisonment length as a function of damage in Appendix Figure A.2.

## 6.3 Dynamic DD

The results of this exercise align with the estimates obtained using the binary treatment indicator and the preliminary analysis of sentencing trends. The custodial sentence rate does not elicit any dominant pattern. In contrast, the length of custodial sentences decreases after the reform for both treatment groups.

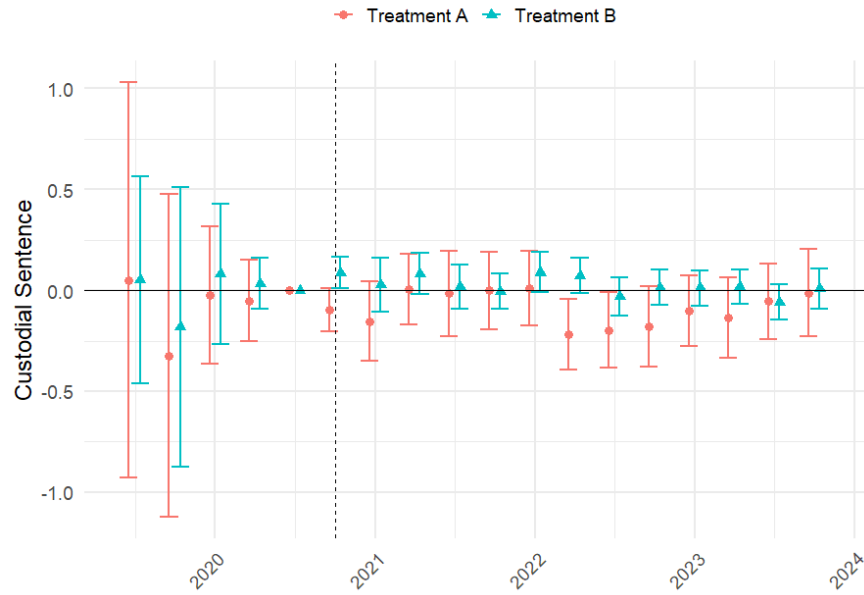
In the Appendix, I perform several robustness checks of my DD results. In Section A.3, I examine other types of punishment to address potential selection bias. Section A.5 replicates the DD analysis with a different control group. Even though the event study plots become quite noisy, I still reconstruct the main pattern using the binary treatment indicator. Additionally, in Section A.6, I construct an alternative treated sample by pooling the original one with other cases affected through the same mechanism. Again, I confirm the results obtained with the main sample.

Table 5: DD Estimates of the Reform Effects on Custodial Sentences

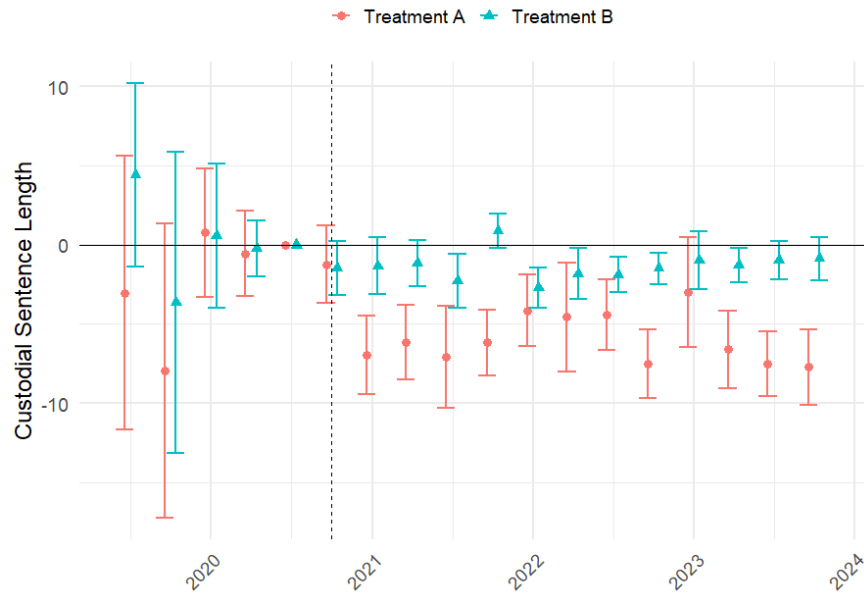
<b>Panel A: Treatment A (sentencing range downward shift)</b>				
	<i>Dependent variable:</i>			
	custodial imprisonment		imprisonment length (months)	
	(1)	(2)	(3)	(4)
Intercept	-0.044 (0.046)	-0.008 (0.044)	-6.196*** (0.621)	-4.753*** (0.596)
Controls	Yes	Yes	Yes	Yes
	No	Yes	No	Yes
Observations	45,161	45,160	24,219	24,219
Treatment group observations	594	594	450	450
R <sup>2</sup>	0.003	0.087	0.049	0.175
Adjusted R <sup>2</sup>	0.003	0.077	0.049	0.157
<b>Panel B: Treatment B (damage range upward shift)</b>				
	<i>Dependent variable:</i>			
	custodial imprisonment		imprisonment length (months)	
	(1)	(2)	(3)	(4)
After:Treatment	0.042 (0.034)	0.033 (0.033)	-1.294** (0.505)	-0.949** (0.481)
Intercept	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	46,544	46,543	25,053	25,053
Treatment group observations	1977	1977	1284	1284
R <sup>2</sup>	0.002	0.085	0.005	0.139
Adjusted R <sup>2</sup>	0.002	0.074	0.005	0.121

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Note: Controls include judge fixed effects, age, number of previous convictions, concurrence, recidivism, juvenile, and gender dummies of the offender.



(a) quarterly effects on custodial sentence rate (extensive margin)



(b) quarterly effects on custodial sentence length (intensive margin)

Figure 2: Dynamic DD for custodial sentence

Note: The Figure shows the quarterly coefficients on custodial imprisonment and custodial imprisonment length. The baseline rate corresponds to Q3 2020. The dashed vertical line represents the 2020 reform. The regressions control for judge fixed effects. 95 percent confidence intervals are plotted.

## 7 Discussion

Table 6 summarizes the results of my empirical analysis: the imprisonment length decreases for both treatment groups.

Table 6: Estimated Impact of the Reform

	<b>Estimated Impact on Mean Sentence</b>
<b>Treatment A</b> <i>(Sentencing Range Downward Shift)</i>	↓
<b>Treatment B</b> <i>(Damage Range Upward Shift)</i>	↓

Note: A summary of the results of the empirical analysis. For both treatment groups, we observe a decrease in sentences at the intensive margin. This could be supported only by the presence of both severity and reference mechanisms.

To interpret the results, I refer to Tables 3 and 4. The decrease in Treatment A could indicate the existence of a severity consideration and prevalence of the severity channel over the reference channel. Nevertheless, a decrease in the mean sentence in the reference group indicates the existence of a reference channel (the severity channel could not support a decrease in sentences for this group). Given that, my empirical analysis brings evidence for the existence of both mechanisms.

Moreover, the decrease under the severity mechanism occurs immediately after the reform, whereas the decrease under the reference effect kicks in with a delay. The delay in the sentencing response for Treatment B may reflect the time judges need to adjust their internal reference points. Although the judges were aware of the reform in advance, they may have reacted only after it was implemented. Unlike the immediate effect of a mechanical change in the sentencing range (as in Treatment A), the formation of new reference groups likely requires repeated exposure to the redefined case mix. This gradual adjustment is consistent with the literature on reference-dependent decision-making.

Several limitations need to be considered when interpreting the results presented in this paper. First, in the Czech Republic, many criminal cases are punished with alternative forms of punishment (including suspended imprisonment), which impairs the interpretation of results related to custodial imprisonment. I address this concern in Appendix Section A.3, where I find that the composition of punishment types has not changed with 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 with a richer dataset.

## 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 channel. The severity channel operates through the judge interpreting sentencing ranges as distinct categories, signaling that certain crimes are more severe than others. The reference channel 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 two groups: cases where the sentencing range decreased (from 1-5 years to 0-2 years) and cases for which the sentencing range remained constant (0-2 years), but with more severe cases added to that 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 average sentence decreases by five months. I attribute this result to the severity channel, where a case is viewed as less severe when it falls into a lower sentencing category. Similarly, when more severe cases are added to a sentencing range, the average sentence decreases by one month. That speaks towards the reference channel, where the judge compares cases to more severe ones within the same range.

Conversely, for custodial imprisonment rates, I do not find any major patterns. The reform seems to have only affected the intensive margin of sentencing, not the extensive margin.

The central contribution of my paper is demonstrating that sentencing ranges establish reference categories that may influence sentencing outcomes. Additionally, I offer a methodological remark that sentencing ranges reforms may also indirectly affect the reference groups by reshuffling them. This is particularly important when selecting a control group to study trends in sentencing. My results also contribute to the general understanding of the impact of sentencing ranges on sentences by identifying both reference and severity channels and their magnitudes. These findings may represent one of the 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. (2007). Racial profiling, statistical discrimination, and the effect of a colorblind policy on the crime rate. *Journal of Public Economic Theory*, 9(3), 521-545. doi: <https://doi.org/10.1111/j.1467-9779.2007.00318.x>
- Bjerk, D. (2017). 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>
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, 127(3), 1243–1285.
- Burch, T. (2015). Skin color and the criminal justice system: Beyond black-white disparities in sentencing. *Journal of Empirical Legal Studies*, 12(3), 395–420. doi: 10.1111/jels.12077
- 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.
- 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., & Dušek, L. (2025). Law or authority: How czech penal elites shaped sentencing policy by non-binding interventions. *Journal of Quantitative Criminology*. doi: 10.1007/s10940-025-09614-w
- Drápal, J., & Plesničar, M. M. (2025). Sentencing elsewhere: Structuring sentencing discretion in post-communist Europe. *European Journal on Criminal Policy and Research*, 31(2), 251–270. doi: 10.1007/s10610-023-09568-4
- 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
- Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, 32(2), 188–200. doi: 10.1177/0146167205282152
- Fischhoff, B. (1975). Hindsight  $\neq$  foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1(3), 288–299.

- 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
- Goldrosen, N., Smith, C. M., Ciocanel, M.-V., Santorella, R., Sen, S., Bushway, S. D., & Topaz, C. M. (2023). Racial disparities in criminal sentencing vary considerably across federal judges. *Journal of Institutional and Theoretical Economics*, 179(1), 92–113. doi: 10.1628/jite-2023-0005
- Klein, D. (2002). *Sentencing guidelines and judicial discretion*. Santa Monica, CA: RAND Corporation.
- Kószegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4), 1133–1165.
- Lane, T., Nosenzo, D., & Sonderegger, S. (2023, May). Law and norms: Empirical evidence. *American Economic Review*, 113(5), 1255–93. doi: 10.1257/aer.20210970
- Leibovitch, A. (2017). Punishing on a curve. *Northwestern University law review*, 111(5), 1205–1280.
- Leibovitch, A., & Teichman, D. (2025). Normative ambiguity, social norms, and the expressive power of law. *Journal of Empirical Legal Studies*, 22(4), 475–501. doi: <https://doi.org/10.1111/jels.70005>
- McConnell, B. (2022). Racial sentencing disparities and differential progression through the criminal justice system: Evidence from linked federal and state court data. *arXiv*.
- Rachlinski, J. J., Johnson, A. J., Wistrich, A. J., & Guthrie, C. (2009). Anchoring in sentencing. *Cornell Law Review*, 54, 1–45.
- Rachlinski, J. J., Wistrich, A. J., & Guthrie, C. (2015). Can judges make reliable numeric judgments? distorted damages and skewed sentences. *Indiana Law Journal*, 90(2), 695–739. Retrieved from <https://scholarship.law.cornell.edu/facpub/1417>
- Skugarevskiy, D. (2017). *Essays in law and economics of enforcement* (Dissertation thesis). Graduate Institute of International and Development Studies in Geneva.
- Topaz, C. M., Ning, S., Ciocanel, M.-V., & Bushway, S. (2023). Federal criminal sentencing: race-based disparate impact and differential treatment in judicial districts. *Humanities and Social Sciences Communications*, 10(1), 366. doi: 10.1057/s41599-023-01879-5
- Tuttle, C. (2023). Racial disparities in federal sentencing: Evidence from drug mandatory minimums.
- Ulmer, J. T., & Johnson, B. (2004). Sentencing in context: A multilevel analysis. *Law & Society Review*, 38(4), 777–803. doi: 10.1111/j.0023-9216.2004.00063.x

# A Appendix

## A.1 Damage report rate

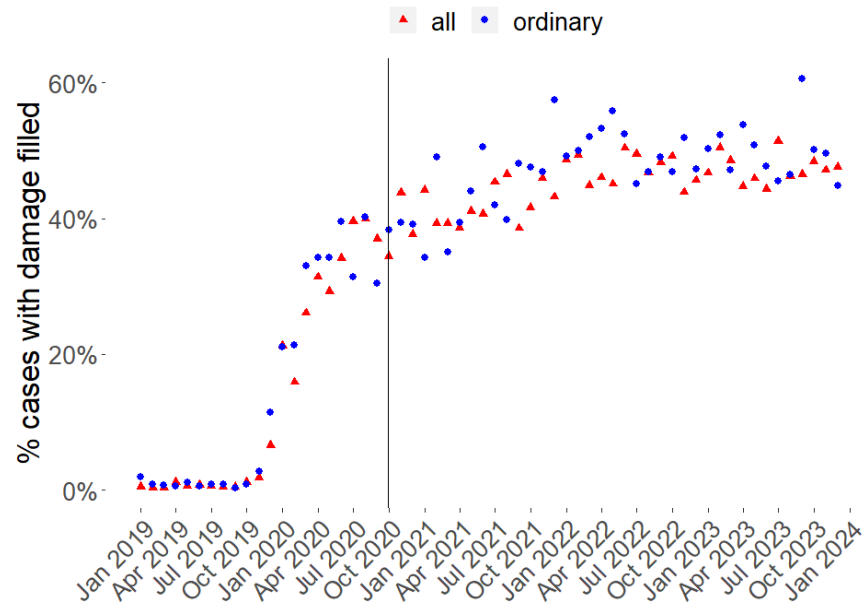


Figure A.1: The rate of cases with damage filled before and after the reform

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 refers to the sentence coming into legal effect, which determines the application of the pre- or post-reform legal norm.

## A.2 Descriptive characteristics of the dataset

Table A.1: Descriptive Statistics of the Main Dataset

	All		Ordinary Cases	
	Before	After	Before	After
Cases	22,383	35,677	936	4,095
Custodial Sentences	14,209	22,885	680	2,954
Suspended Sentences	1,666	2,139	77	323
Mean Custodial Length (m)	14.33	14.12	16.64	16.89
Mean Suspended Length (m)	12.94	12.88	15.26	15.12
Mean Damage (thousand CZK)	68.11	70.38	122.44	159.53
Recidivist Rate (%)	11	12	13	11
Mean Age	32.47	33.42	33.12	33.64
Male (%)	83	85	81	83

Note: The year range is limited to 2019-2023. Ordinary cases are defined as cases with no special circumstances, where the criterion determining the sentencing range is the damage caused. By recidivist, I refer to offenders for whom the court counts their previous convictions as an aggravating circumstance (under the Criminal Code, it is the court's discretion to consider the previous convictions as aggravating circumstances).

## A.3 Other types of punishment

To investigate whether the reform affected the use of different types of punishment, I begin by estimating a multinomial logit model where the dependent variable is the type of punishment imposed (custodial sentence, other punishment, or suspended imprisonment).

Table A.2 presents multinomial regression results comparing the likelihood of receiving alternative forms of punishment (relative to custodial sentences) before and after the reform. In Treatment A, the coefficients are positive but statistically insignificant. This suggests no clear evidence that the reform altered the probability of receiving alternative punishments in this subgroup.

In contrast, for Treatment B, both coefficients are negative and again not statistically significant, indicating that the reform did not result in systematic shifts in punishment type relative to custodial sentences in this group either.

While Table A.2 examines the relative use of punishment types, it does not reveal whether the reform affected the severity of cases within each sentencing category. It is possible that the overall proportions of punishment types remained unchanged, while the reform subtly altered the composition of cases assigned to each category — shifting more or less severe cases between them

Table A.2: Multinomial Regression Results for Punishment Type

<b>Panel A: Treatment A</b>		
<i>Dependent variable: Punishment Type</i>		
	Other Punishment	Suspended Imprisonment
After reform	0.405 (0.263)	0.164 (0.352)
Observations	594	
<b>Panel B: Treatment B</b>		
<i>Dependent variable: Punishment Type</i>		
	Other Punishment	Suspended Imprisonment
After reform	-0.067 (0.151)	-0.308 (0.243)
Observations	1977	

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Note: Multinomial logit model of the impact of the reform on the type of punishment. Custodial sentence cases serve as a reference category.

without affecting their relative frequencies.

Table A.3 presents the average amount of damage (in CZK) associated with each punishment type, before and after the reform. This helps identify whether the composition of offenses within each sentencing category shifted after the reform.

A significant increase in the average damage associated with suspended imprisonment after the reform may suggest a shift toward assigning this punishment to more severe cases. However, since all other estimates remain statistically insignificant, this isolated finding alone is insufficient to undermine the overall conclusions of the main analysis. Moreover, this increase could partially reflect broader economic factors such as inflation or general price level increases, rather than changes in judicial behavior.

In Treatment B, there is no statistically significant shift in average damages for any punishment type. The largest observed difference is a decrease of over CZK 5,000 for suspended imprisonment, but this result is not significant. This indicates that for this group the reform did not disrupt the relation between damage and punishment type.

Table A.3: Mean Damage Before and After Reform

<b>Panel A: Treatment A</b>			
	<i>Average damage (CZK)</i>		
	Custodial	Other Punishment	Suspended Imprisonment
Before reform	72,305 (1,225)	72,752 (2,731)	69,968 (4,284)
After reform	71,492 (814)	71,792 (1,533)	73,065 (2,228)
Difference	-813	-960	3,097**
Observations	450	99	45
<b>Panel B: Treatment B</b>			
	<i>Average damage (CZK)</i>		
	Custodial	Other Punishment	Suspended Imprisonment
Before reform	22,524 (934)	20,973 (1,254)	26,021 (2,350)
After reform	22,664 (319)	21,954 (473)	20,993 (919)
Difference	140	981	-5027
Observations	1284	553	140

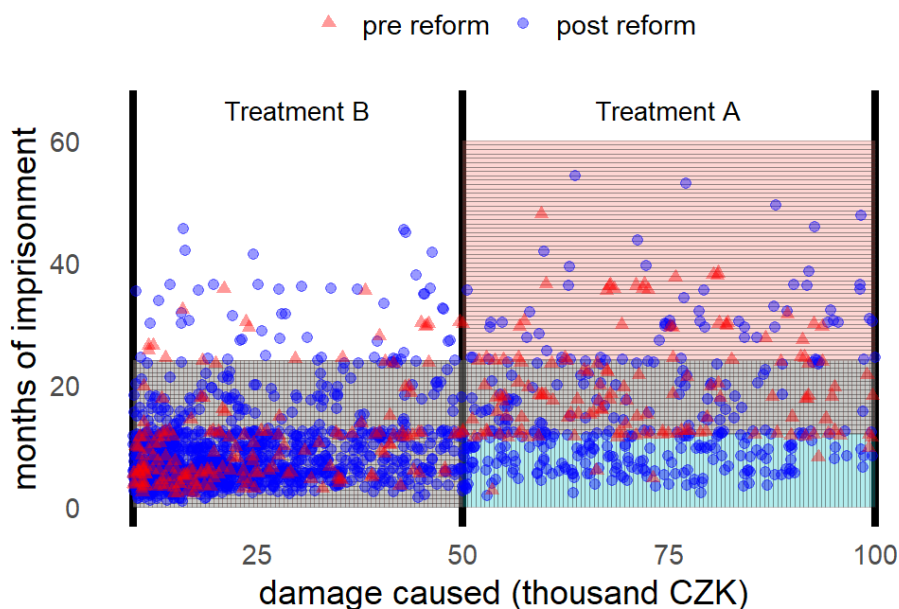
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Average damage for each punishment type before and after the reform. The differences are computed as After-Before.

## A.4 Custodial imprisonment length as a function of damage

Figure A.2 shows that the decrease in the average sentence is not driven solely by the threshold or outlying cases. Instead, after the reform, there is indeed a general downward trend in sentence length.

Figure A.2: Custodial imprisonment length as a function of damage for Treatment A and Treatment B samples



Note: Red triangles represent before reform cases, blue dots represent after reform cases. Red rectangles indicate sentencing ranges before reform, while blue rectangles indicate sentencing ranges after reform. The points falling outside the ranges are cases with some important aggravating or mitigating circumstances, where the judge is allowed to impose a sentence below or above the statutory sentencing range.

## A.5 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 challenging to identify 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.

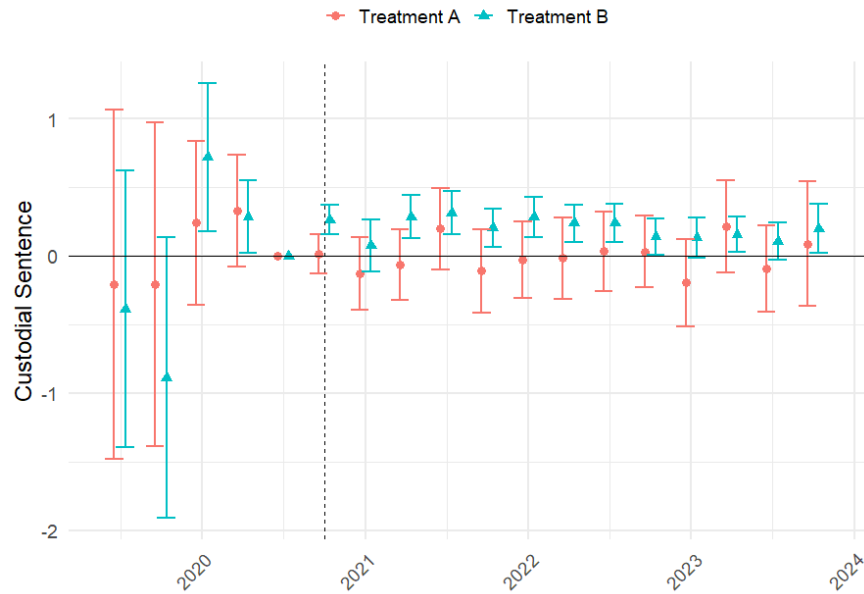
Therefore, apart from the control group used in the main analysis, I introduce the cases of negligence of mandatory support, another common crime that the courts regularly consider. Moreover,

this crime is at least vaguely related to property — the offender profits from not paying the support.

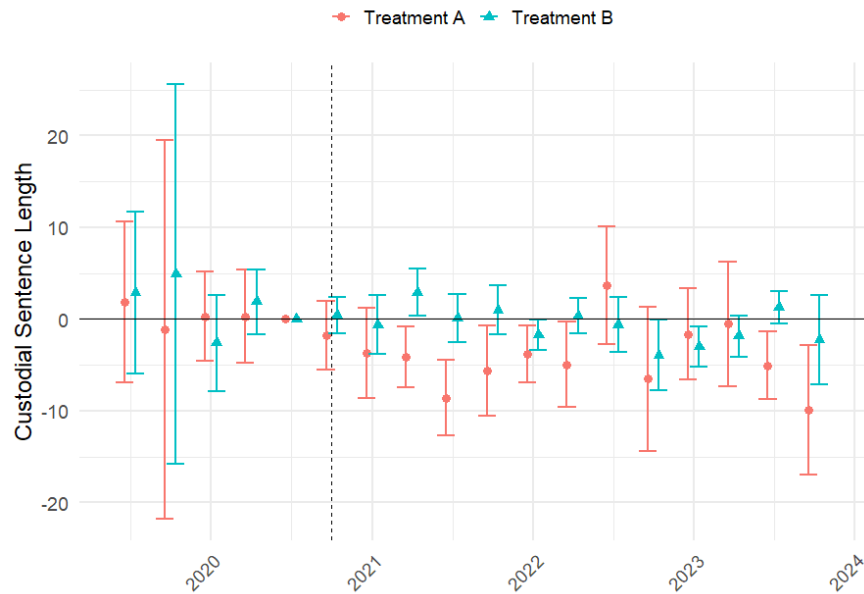
Figure A.3 presents quarterly effects for both Treatment A (sentencing range downward shift) and Treatment B (damage range upward shift), using the alternative control group: For Treatment A, the effects on the custodial sentence rate generally fluctuate around zero both before and after the 2020 reform (marked by the dashed vertical line). This suggests that for Treatment A, the reform did not significantly alter the probability of receiving a custodial sentence when using this alternative control group. As for the sentence length, there's a significant downward trend starting around the time of the reform. This strongly supports the hypothesis that the reform led to shorter custodial sentences for cases under Treatment A.

For Treatment B, similarly, the effects on the custodial sentence rate mostly hover around zero. Some of them turn out to be significantly positive. However, for this group, the trend seems to be suspicious even before the reform; therefore, I do not derive any systematic conclusions from that. A slight downward trend in custodial sentence length is observable after the reform; however, it starts only after 2023. This again supports my main intuition regarding the reference channel.

Overall, even with the alternative control group, I replicate my main results. For completeness, Table A.4 performs the DD regressions with binary treatment indicators. The estimates closely match those obtained in the main analysis.



(a) quarterly effects on custodial sentence rate (extensive margin)



(b) quarterly effects on custodial sentence length (intensive margin)

Figure A.3: Dynamic DD for an alternative control group

Note: The Figure shows the quarterly coefficients on custodial imprisonment and custodial imprisonment length for an alternative control group of 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. 95 percent confidence intervals are plotted.

Table A.4: DD Estimates of the Reform Effects, Negligence of Mandatory Support as a Control Group

<b>Panel A: Treatment A (sentencing range downward shift)</b>				
	<i>Dependent variable:</i>			
	custodial imprisonment		imprisonment length (months)	
	(1)	(2)	(3)	(4)
After:Treatment	-0.203*** (0.030)	-0.139*** (0.029)	-5.673*** (0.354)	-4.785*** (0.343)
Intercept	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	7,744	7,744	4,376	4,376
Treatment group observations	594	594	450	450
R <sup>2</sup>	0.014	0.136	0.088	0.208
Adjusted R <sup>2</sup>	0.014	0.120	0.088	0.193

<b>Panel B: Treatment B (damage range upward shift)</b>				
	<i>Dependent variable:</i>			
	custodial imprisonment		imprisonment length (months)	
	(1)	(2)	(3)	(4)
After:Treatment	0.016 (0.035)	0.026 (0.035)	-1.539*** (0.564)	-1.355*** (0.517)
Intercept	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	9,127	9,127	5,210	5,210
Treatment group observations	1977	1977	1284	1284
R <sup>2</sup>	0.013	0.170	0.104	0.405
Adjusted R <sup>2</sup>	0.012	0.114	0.103	0.333

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Note: Reform effects estimated with an alternative control group (negligence of mandatory support). Controls include judge fixed effects, age, number of previous convictions, concurrence, recidivism, juvenile, and gender dummies of the offender.

## **A.6 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 pool all cases affected in a similar way to my alternative treatment sample. However, since now we are comparing cases with different severity, I standardize the outcome of interest using z-score normalization. Thus, the regression coefficients now represent multiples of the standard deviation in the control group. For easier comparison, I standardize the variables and re-estimate the coefficients also for the original sample. Table A.5 presents the results. The estimates are robust against the choice of sample — for the original and pooled samples, they match in sign, approximate value, and significance.

Table A.5: DD Estimates of the Reform Effects, Pooled and Original Sample

<b>Panel A: Treatment A (sentencing range downward shift)</b>								
	<i>Dependent variable:</i>							
	standardized custodial imprisonment				standardized imprisonment length (months)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
After:Treatment	-0.034 (0.084)	0.014 (0.082)	-0.088 (0.091)	-0.015 (0.089)	-0.329*** (0.096)	-0.184** (0.092)	-1.022*** (0.102)	-0.784*** (0.098)
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Sample	Pooled	Pooled	Original	Original	Pooled	Pooled	Original	Original
Observations	45,416	45,415	45,161	45,160	24,434	24,434	24,219	24,219
R <sup>2</sup>	0.005	0.089	0.003	0.087	0.110	0.232	0.049	0.175
Adjusted R <sup>2</sup>	0.005	0.078	0.003	0.077	0.110	0.215	0.049	0.157
<b>Panel B: Treatment B (damage range upward shift)</b>								
	<i>Dependent variable:</i>							
	standardized custodial imprisonment				standardized imprisonment length (months)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
After:Treatment	0.019 (0.049)	-0.003 (0.048)	0.083 (0.067)	0.067 (0.066)	-0.367*** (0.065)	-0.302*** (0.061)	-0.213** (0.083)	-0.157** (0.079)
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Sample	Pooled	Pooled	Original	Original	Pooled	Pooled	Original	Original
Observations	48,242	48,241	46,544	46,543	26,439	26,439	25,053	25,053
R <sup>2</sup>	0.011	0.093	0.002	0.085	0.157	0.313	0.005	0.139
Adjusted R <sup>2</sup>	0.011	0.083	0.002	0.074	0.157	0.299	0.005	0.121

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Note: Reform effects estimated with the original sample and with a pooled sample for Treatment A and Treatment B. Controls include judge fixed effects, age, number of previous convictions, concurrence, recidivism, juvenile, and gender dummies of the offender. All outcomes are z-score standardized.