



UNIVERSITY OF LEEDS

This is a repository copy of *What is the Value of Judicial Experience? Exploring Judge Trajectories Using Longitudinal Data*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/182204/>

Version: Accepted Version

---

**Article:**

Drápal, J and Pina Sanchez, J orcid.org/0000-0002-9416-6022 (2022) What is the Value of Judicial Experience? Exploring Judge Trajectories Using Longitudinal Data. *Justice Quarterly*. ISSN 0741-8825

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# **What is the Value of Judicial Experience?**

## **Exploring Judge Trajectories Using Longitudinal Data**

### **Abstract**

Judicial experience is considered essential for the proper functioning of the sentencing system. We investigate how it influences judicial decisions and its role in reducing sentencing disparity. To do so, we analyze all Czech criminal decisions imposed in 2007-2017 using data that includes judge identifiers. This unique feature of our data enables us to measure judges' experience directly, as the number of criminal cases processed, and to assess patterns in between-judge disparities longitudinally over the course of judges' careers. We find that experienced judges impose more prison sentences, decide fewer cases via shortened procedure and find fewer defendants guilty. In addition, as judges become more experienced, between-judge disparities reduce across all the outcomes considered. Experience is thus an instrumental factor affecting judicial decisions throughout the criminal process, and one that contributes to greater consistency.

**Key words:** sentencing, disparities, longitudinal, judicial experience, Czech Republic

## **Introduction**

The notion of just sentencing has been challenged repeatedly by empirical studies demonstrating the presence of widespread unwarranted disparities. Sentencing differences between similar cases have been associated with defendant, prosecutor, judge, and court characteristics, both in the U.S. and in Europe, (see e.g. Ulmer & Johnson, 2004; Johnson, 2005, 2006; Johnson et al., 2008; Wu & Spohn, 2010; Ulmer et al., 2011; Pina-Sánchez & Linacre, 2013; Kim et al., 2015). Extensive research efforts have been dedicated to the exploration of sentencing disparities; Ulmer & Bradley (2018) and Ulmer (2012) both offer good summaries of the relevant literature. Identifying the specific factors behind these disparities has long been seen as a key task in sentencing scholarship – arguably as the field’s most burning question (Spohn, 2015; Tonry, 2016; Ulmer, 2012).

Many have suggested that the experience of those who decide on sentencing is an important factor capable of limiting unwarranted disparities; scholars have referred to it as an invaluable (Ashworth, 1995), central force (Hester, 2017). Experience is of especial importance to those defending the judicial-defensive tradition (Tata, 2020, p. 18), especially when sentencing is viewed as a craft, a qualitative skill that is learned and improved through practice (Tata, 2007).

Since sentencing is governed primarily by principles rather than rules, the decision-making process requires numerous elements to be weighted, and slight differences of judgment during this process can lead to substantial differences in the resulting sentence. Psychological research has suggested that experience may be beneficial in recognizing similarities and differences: The more judges explain their decisions (reasoning being a cornerstone of principled sentencing), the more they should understand the subtle differences between cases (see Lombrozo, 2012 for literature summary); this should lead them to impose more principled sentences and thus likely reduces the level of unwarranted disparities.

Judicial systems are aware of the importance of experience, and they value it. This is evidenced both in the Anglo-Saxon tradition, which relies on judges having prior careers as proficient legal practitioners, and in the continental European systems that operate mandatory clerkships for would-be judges. Yet even with such academic and practical criminal law knowledge, novice judges embark on a journey of discovery, exploring the realm of sentencing with, at best, a smudged map of general principles and some observations from the past. Does gaining more experience enable them to better understand the subtle differences between cases? Does it reduce unwarranted disparities? Do judges become more punitive as their careers progress? How does their experience influence the other decisions they make as judges, apart from sentencing? Do they become bolder in exerting their full autonomy or more considerate for the individual features of each case processed – or vice versa? And do all judges follow similar decision-making trajectories?

In spite of the relevance of these questions, how sentencing practices develop over time is not a subject that has been directly examined to date. Multiple studies using U.S. Federal and States Sentencing Commissions data have considered the effect of proxies for experience, such as time on bench or age, on sentence severity (e.g. Hauser, 2012; Johnson, 2006; Spohn, 1991 or Steffensmeier & Britt, 2001). Most of that literature has, however, focused on the effect on severity, leaving unexplored other important aspects of sentencing where experience could play a role. Furthermore, all previous studies on the topic are based on cross-sectional designs, thus neglecting the longitudinal dimension of the judicial .

Here, we employ a longitudinal design that enables us to shed new light on the effect of judicial experience on sentencing and on other features of the judicial decision-making process. To do so, we examine the trajectories of a large sample of judges across time; this is the first study of this kind undertaken in sentencing research. We rely on a unique dataset that records the decisions taken by Czech judges from their very first case onwards. By observing

the total number of cases decided by each judge in the dataset, we can generate a direct measure of judicial experience, avoiding the use of proxies such as years on bench. Furthermore, by modeling individual judge trajectories, we can correctly account for between-judge variability and assess how that variability changes across time.

Lastly, we seek to differentiate our analytical strategy from the archetypal quantitative sentencing study by moving beyond the modeling of sentence severity and broadening our attention to cover additional aspects of judicial decision-making practice that can be distinctively measured. These include guilt adjudication, signaling (lack of) confidence in opposing the prosecution, and the use of shortened procedures, as a measure of (un)willingness to fully investigate the case. This broader approach provides a more holistic assessment of the effects of judicial experience.

We proceed by discussing the theoretical lessons that can be drawn from the psychological literature on expertise and from the literature that has examined the effects of various proxies for experience on sentencing. We then present the background to our study, the Czech sentencing system and the data we use for the analysis. Subsequently, we present our methods and results, first examining the changes in our observed variables with increasing experience and then discussing the impact of experience on sentencing disparities. The article concludes with a discussion of our results and of their implications for further research on judicial experience and sentencing, together with a series of policy recommendations to enable criminal justice systems to benefit maximally from judicial experience.

### **The impact of experience on sentencing: psychological considerations**

To become an expert in any discipline or task, one needs to practice the skills it involves, deliberately and attentively (Winegard et al., 2018). Feedback, evaluation, training and

opportunities for repetition and gradual refinement are crucial in achieving expertise (Feltovich et al., 2018). In what is of relevance to the sentencing practice, the psychological literature indicates that mere exposure to examples is insufficient for learning to differentiate between cases. To fully profit from the experience, it is necessary to identify cases' distinguishing features explicitly and receive corrective feedback (Mosier et al., 2018). Research on expertise shows us that certain learning methods help to elicit and refine tacit knowledge, such as discussing issues and weighing arguments for and against various solutions (Cianciolo & Sternberg, 2018). Becoming an expert in a given profession requires one to perform it, yet benefits of experience are greatly limited when it is hard to observe others and imitate them (Billett et al., 2018).

Generally, as judges gain experience they will likely reach a certain stable, average level of performance of their duties (Ericsson, 2018) and they will become better able to orientate themselves in the judicial environment by identifying patterns they have witnessed in previous cases. Once a judge has processed a substantial number of cases, s/he will have a better understanding of the requirements of sentencing, including offenders' varying circumstances and how the establishment of baselines and unofficial starting points can lead to fewer gross mistakes and disparate sentences (Ericsson, 2018). Yet it is not likely to be the judges themselves who will profit the most from their experience: No regular corrective feedback is provided to them and detailed discussions between judges over particular decisions are uncommon. We may speculate that as their experience increases, judges will likely decide in a more automatic and patterned way (Albonetti, 1991; Ericsson, 2018; Feltovich et al., 2018). Experience need not be conceived as achieving psychological maturity, but rather as becoming familiar with the "court technology" (Eisenstein et al., 1988). These insights help us to theorize below about the hypothetical effects of judicial experience on sentencing.

## **Empirical sentencing research based on proxies of experience**

Time on bench – usually measured in years – has been used as a proxy for experience in multiple studies (Frazier & Bock, 1982; Hauser, 2012; Johnson, 2006; Kritzer, 1978; Lim et al., 2016; Myers, 1988; Spohn, 1991; Steffensmeier & Britt, 2001; Steffensmeier & Hebert, 1999; Welch et al., 1988), yet the findings from this body of work appear contradictory. While some studies have found that more experienced judges are more punitive (Hauser, 2012; Welch et al., 1988), others have found that they are more lenient (Kritzer, 1978; Spohn, 1991; Steffensmeier & Hebert, 1999), and a third group of studies has reported non-significant effects (Johnson, 2006; Lim et al., 2016; Steffensmeier & Britt, 2001).

Time on bench represents “general judicial experience”, and is a useful aspect to consider in study topics linked to “being a judge” e.g. judicial culture or judges’ views of themselves. Measuring experience indirectly as time on bench is, however, problematic for several reasons. Most importantly, a judge’s time on bench does not necessarily reflect their experience with criminal cases. In many systems, judges are assigned to either civil, administrative or criminal cases and they can be re-assigned at any time. A judge might try her/his first criminal cases several or even dozens of years after becoming a judge, having previously tried other case types. Similarly, time on bench fails to consider that, over the same period of time, some judges sentence many more cases than others. We show to what extent these events can distort the time on bench measure in the Measurement Appendix. A wide range of life-course events might further distort the time on bench measure, such as parental leave, long-term illness or temporary delegation to the ministry of justice (a possibility in some systems). An “issue-specific judicial experience” measure, which considers the quantity and types of processed cases – such as the number of criminal cases in

which the given judge imposed a sentence – is thus better suited to study topics linked to “performing judicial tasks”, including sentencing.

Another possible proxy for experience could be the judge’s age, but this is probably an even less accurate proxy than time on bench, since it has no link to judicial practice. The majority of studies that have examined the impact of age on sentencing have found that older judges are more punitive (Cook, 1973; Hauser, 2012; Kritzer, 1978; Myers, 1988; Steffensmeier & Hebert, 1999), although a few studies have concluded either that older judges are more lenient (Johnson, 2006) or that age does not influence sentencing (Steffensmeier & Britt, 2001). Similarly, inconsistent effects have been reported in studies exploring both age and time on bench simultaneously (Hauser, 2012; Kritzer, 1978; Steffensmeier & Hebert, 1999; Johnson, 2006; Steffensmeier & Britt, 2001).

Various studies have examined specific types of experience, such as the impact of previously serving as a prosecutor (Frazier & Bock, 1982; Myers, 1988; Spohn, 1991; Steffensmeier & Hebert, 1999; Welch et al., 1988), which is generally found to lead to a more punitive approach, although Myers (1988) found that prosecutorial experience has contradictory effects on different sentencing outcomes. Encounters with differently severe cases in the very early stages of a judicial career have been found to influence sentencing in the following months (Leibovitch, 2016). The findings reported in this literature on specific types of experience appear, however, too contradictory to prove any specific effects and moreover these studies only explored the effect on severity, without considering how experience affects other aspects of judicial decision-making.

### **Research questions and hypotheses**

In our study we ask two key questions. First, does greater experience contribute to reducing disparities between court decisions? We hypothesize that novice judges without adequate



sentencing experience might not be able to distinguish properly between similar and different cases, and that this might lead to inconsistent or unprincipled sentencing outcomes. In this scenario, greater experience would likely reduce between-judge disparities. Second, do judges, on average, decide differently as their experience increases? Even though the potential impact of sentencing experience on various aspects of judicial decision-making remains largely unknown, empirical psychological literature indicates that, at the very least, we may expect the rate of blatant mistakes to decrease with greater experience and that more experienced judges should act more confidently. We discuss these general principles in relation to each of our variables of interest.

Most empirical studies on sentencing have so far explored variability in the imposition of *prison sentences* and/or the *length* of those sentences, aiming to measure the severity of the judge's decisions. We follow in their footsteps. Sentencing scholars hold divergent views as to what impact increasing judicial experience ought to have on the severity of imposed sentences (as yet, scholars have not considered other aspects of judicial decision-making). Some have argued that greater experience might harden judges (Hauser, 2012; Welch et al., 1988), while others have contended that the repeated experience of imposing prison sentences that do not seem to “work” and have various collateral consequences for the offenders might lead judges to perceive imprisonment as futile (Spohn, 1991) and so impose fewer prison sentences. As existing theoretical and empirical scholarship examining the influence of judicial experience on sentencing has reached contradictory conclusions as to whether judges become more or less punitive with greater experience, we do not formulate a directional hypothesis for this variable: Our study is primarily exploratory.

We further need to look beyond severity: While sentencing is considered the judge's greatest moment of decision in common law jurisdictions, continental judges consider the more important decision to be that of whether the defendant is guilty or not. Sentence

imposition, which is usually announced at the same time as the *decision on guilt*, plays a lesser role for all trial participants. Despite this, decisions on guilt have remained under-researched to date. In this study, we consider whether judges arrive at verdicts of guilt. Following Heumann (1978/2020, chpt. 5) and Wright & Levine (2014) we hypothesize that more experienced judges will find fewer people guilty as they will approach the evidence provided by the police and the prosecution more cautiously.

Criminal justice systems are under pressure to handle cases as quickly as possible, while respecting the requirements of a fair trial. Different systems offer different ways of achieving this. While common law systems employ guilty pleas and plea bargains, continental systems traditionally rely either on prosecutors deciding less serious cases themselves, with the consent of the accused, or on judges issuing simplified judgments (penal orders) without seeing the accused, against which the defendant can file an appeal that results in normal proceedings taking place. Continental prosecutors and judges are both usually limited in terms of the sanctions that can be imposed in these shorter procedures. Deciding cases by penal order saves judges time, but they must decide on the basis of very basic written information about the offense and the offender, leading to a less individualized, less adversarial procedure, and less thorough decision. We examine judges' decisions to choose the more speedy procedure (deciding by *penal order*) over a thorough consideration of guilt and sentence. We hypothesize that judges will act in an economic and efficient way and make greater use of simplified procedures as their experience increases: As the number of cases they have processed increases, judges may become more confident in deciding that a particular case is sufficiently clear-cut to merit the time-saving option.

### **The Czech judicial and criminal justice system**

The Czech Republic is a mid-sized European country (10.5 million inhabitants) with a democratic rule-of-law political and legal system. Following the 1989 Velvet revolution, which ended a 40-year-long communist authoritarian regime, Czechoslovakia peacefully divided into the Czech and Slovak Republics; both countries then joined NATO and the European Union. It is the most developed country behind the former Iron Curtain, partly due to the democratic, legal, industrial and bureaucratic legacy of the Austrian-Hungarian Empire of which it was formerly part, and of the first Czechoslovak Republic in the mid-war years. Its population is rather homogeneous with the largest groups of foreign residents being Ukrainians (1.4%) and Vietnamese (0.6%) and the largest ethnic minority group the Roma population (2.4%).

Like other continental legal systems, the Czech judicial system operates with career judges. After graduating from law school, which typically consists of five years of university study, the first step towards a judicial career is to become an assistant to a judge, prosecutor, attorney or notary. After at least three years of practice, a judicial candidate must then pass a professional exam to qualify as a judge, after which s/he may apply for relevant vacancies. To be appointed as a judge the candidate must also be at least 30 years old. While historically only those who had served as judges' assistants were appointed as judges, this practice has changed and since 2006 a substantial minority of judges in office previously worked as attorneys or prosecutors. The judiciary is attractive for its high income and long-term social prestige.

Judges in the Czech Republic are neither elected by the public nor nominated by political parties. Instead, the selection process takes place at the level of the regional courts, whose presidents decide who to recommend for vacant positions, although the chosen judges are formally nominated by the Ministry of Justice and appointed by the President of the Republic. Once appointed, Czech judges are normally expected to serve until the end of their

careers; retirement is obligatory at 70 years of age and in 2016 the average age at which judges left the judiciary was 65 years.

Czech judges are independent in every respect. They are bound only by the laws enacted by Parliament. In the decades since the 1989 Velvet Revolution, the judiciary has successfully defended itself against numerous intrusions by the executive and legislative powers. This broadly conceived judicial independence, however, makes it impossible for judges to be removed from their function unless they commit blatant mistakes or consistently seriously underperform. Cases are assigned to judges quasi-randomly within certain specializations, so a novice judge's first case could be one of the most serious cases tried at the court in years.<sup>1</sup> This strong judicial independence is combined with a high level of sentencing discretion, a common feature of Central European and other continental legal systems (for a discussion of the German system see e.g. Kaspar, 2020). Overarching sentencing rules are enumerated in the general part of the Penal Code and wide sentencing ranges are provided for individual offenses and their subsections. The judge has ample discretion to impose various alternative sanctions (fines, community work, home detention, various prohibitions), suspended prison sentences (with or without supervision) or non-suspended prison sentences and to decide on the extent of those sanctions. Judges are limited by legislative provisions that require them not to impose non-suspended prison sentences

<sup>1</sup>Specializations might include e.g. traffic offenses or whether the case began via a shortened procedure or not. Cases are usually successively assigned to individual judges using a wheel (one-by-one in a fixed order; algorithm-like random case assignment is rare). In many ways, the differences between specializations should be captured by the variables we are using, such as the type or seriousness of the offense and the type of criminal proceedings initially chosen. This is further discussed in the Analytical Appendix, where we show that, on where the observable characteristics are concerned, the cases dealt with by novice and experienced judges did not differ substantially.

unless necessary, and are also procedurally incentivized not to impose non-suspended prison sentences. This approach leads to low numbers of new prisoners in comparison to other European countries; however, Czech prisoners serve prison sentences that are many times longer than those served by their counterparts in Western Europe, and this results in a relatively larger prison population (Dünkel, 2017).

Speed of proceedings is the only characteristic according to which Czech judges are currently evaluated. Judges are reproached if their workload appears to be lagging behind others, and especially if the entire court where they operate is processing cases more slowly than other courts in the region. If judges hold proceedings up with unwarranted delays, they may be subject to disciplinary proceedings, which can lead to their removal from the bench. The primary instrument at the judges' disposal to accelerate their case load is to issue penal orders – decisions that do not include any reasoning and are imposed without an oral hearing. The defendant may either accept the decision or file a protest, which quashes the penal order and leads to the usual criminal proceedings. The majority (50-60%) of criminal cases are decided via penal order, and protests to these are filed in approximately a quarter of cases of these cases (Drápal & Vávra, 2021). However, because speeding up judicial decisions can affect their quality and lead to miscarriages of justice, the legislator established in 2001 that non-suspended prison sentences cannot be imposed by penal orders.

There are several other important features of the Czech criminal justice system. It does not make use of juries, albeit lay judges participate in the process. The appellate procedure prohibits *reformatio in peius* (worsening the defendant's position). Hence, if only the defendant appeals, the appellate court cannot impose a harsher sanction; if the prosecutor also appeals, the appellate court has the discretion to impose the sanction it considers most appropriate. Unlike in common law systems, the prosecution does not play an essential role in sentencing: While sentence recommendations have anchoring and priming effects, the judge

is not bound by either the prosecution's legal classification of the offense or their sentence recommendation. As judges and prosecutors informally admit, all this leads to the decision on sentencing being viewed as less important than the decision on guilt.

These features mean that it is advantageous to examine the influence of experience on judicial decision-making in a continental setting such as the Czech system, rather than in a common law context. In common law systems a judge may sentence several cases in a row, which has led researchers to hypothesize that the sentencing of each individual case is influenced by the cases that are sentenced immediately before them (for a thorough discussion of judges' incentives in common law systems see Emerson, 1983; for a discussion of "hearing days" and time on bench see Leibovitch, 2016). Since judges in continental systems decide on guilt as well as on sanctions, researchers studying continental systems are well placed to examine the effects of judicial experience on their decisions. Between-case dependency is less likely in continental setting. A continental judge decides at most a handful of cases each day (there are no separate sentencing sessions) and those sentenced later in the day do not come into contact with those sentenced earlier. While continental judges may still be influenced by the cases they have decided recently (as juries in the English system, see Bindler & Hjalmarsson, 2019), which may serve as their primary reference point, being the most vivid in their memory, they are not incentivized to decide cases in relation to those recently decided.

### **Analytical strategy and data**

To study the effect of experience on sentencing, we use growth curve models, a technique frequently employed in subfields of criminology such as life-course criminology (see for example Caufmann et al., 2017, or McLean et al., 2019). In essence, these growth curve models are nothing more than special variants of the multilevel models that are so commonly

used in sentencing research (Fearn, 2005; Ulmer & Johnson, 2004), with the key difference that the hierarchical structure of the dataset is defined longitudinally. To achieve this, we specify dependencies for sentences clustered within judges and include an explanatory variable indicating the order in which the sentences were imposed. Further details about our modelling strategy are provided in the Analytical Appendix.

The data we use for this research were provided by the Czech Ministry of Justice and consist of case-level sentences imposed from 1995 onwards. They include information about the defendant (sex, age at offense date, number of previous convictions), the offense (the classification of the offense by section and subsection, which determines the sentencing range), the criminal proceedings (dates of the proceedings and information on pre-trial detention), the decision and sanctions (the outcome of the criminal proceedings, including types and amounts of any sanctions imposed).

This dataset, however, does not include the necessary judge identifiers to recreate the longitudinal format needed to explore the effect of experience. To access that information, we carried out a linkage process in collaboration with the Ministry of Justice and the district courts, in which the name of the judge who decided each case was extracted from the court databases. Judge names were provided for virtually all cases in 2007-2015; in 2016 and 2017, names are missing for just under 4% of cases, largely because three of the 86 district courts (Ostrava, Příbram and Uherské Hradiště) failed to provide that data, likely due to a technical error. The total number of cases processed in the Czech courts is reported in the Measurement Appendix, as is the percentage of cases dropped due to failure to identify the deciding judge. Once cases had been linked to named judges, further data on the demographic and professional characteristics of the judges were obtained from the Ministry of Justice. These included each judge's occupation prior becoming a judge, the date of their qualifying exam, the date of taking their oath (i.e. becoming a judge), their year of birth and information about

their studies (which faculty of law they attended and when they completed their studies); the judges' sex was deduced from their names since the judges' names in our data are sex-specific.<sup>2</sup>

### ***Response variables***

We operationalize the decision to impose a *prison sentence* as whether the offender was directly incarcerated or not. When the court imposed a suspended prison sentence, which is similar to a probation in the U.S., we did not consider this to be a prison sentence. Similarly, when considering the *length of prison sentences* only non-suspended prison sentences were counted. The decision on *guilt* was operationalized as whether the court decision was recorded by the Criminal Records Office, diversions were thus considered guilty verdicts. The propensity to use a *shortened procedure* was operationalized as whether the case was concluded by a penal order. Unfortunately, the data did not identify cases that were originally decided by penal order but subsequently returned to normal proceedings following a protest; the numbers of penal orders will be thus underestimated. Descriptive statistics for all variables and the sample sizes of these samples are presented in the Measurement Appendix.

### ***Explanatory variables***

We consider the following explanatory variables in our models. Relating to defendants and offenses, these are the defendant's *sex* and *age at offense date*; their *previous convictions*, capped at 20 and demeaned (31.2% of the examined defendants were first-time offenders); *squared previous convictions*, used in line with previous literature to capture the expected

<sup>2</sup> The vast majority of female Czech last names end with the suffix “-ová” or “-á”. For those few judges whose sex could not be identified from their last names (such as those ending with -ů or -í), their sex was deduced from their first names (the vast majority of which are also sex-specific). Judges' sex was coded manually.



positive, but marginally decaying effect of previous convictions; offense seriousness defined as one third of the *sentencing range* (e.g. the value at 0.33 where 0 is the statutory minimum and 1 the statutory maximum prison sentence; 59.3% of the examined cases had no sentencing minimum and 93.6% had a sentencing minimum of 1 year or less). To reflect the varying nature of the committed offenses, the most common *offenses* are coded: Three of the most common offenses were theft, frustrating execution of an official order (primarily driving after a ban and failing to report to prison) and non-payment of alimony. It is important to note that offense type is not interchangeable with sentencing range since offenses are usually divided into several subsections, each with its own specific sentencing range. We also include a dummy variable for whether the offense was sentenced according to the *new penal code* enacted in 2010, which enabled the imposition of new sanctions such as home detention, emphasized the use of fines, increased the possibility of suspending prison sentences and raised some sentencing ranges, especially for more serious offenses (Scheinost et al. 2015). We further identify *multiple offending*, i.e. whether the offender was sentenced for more than one offense, which was true in three-quarters of cases. In such cases, the Czech criminal code prescribes that the sentence imposed should be within the sentencing range for the most serious offense committed, with the other offenses influencing the judge's choice within that sentencing range.

Pertaining to the criminal proceedings, we identify whether the case was commenced via a *simplified procedure*, which signals a more clear-cut case (and was true of half of the examined cases), or by a standard indictment. We code *pre-trial detention (remand)* as a binary variable (4.5% of all studied defendants were placed in pre-trial detention). We also consider the *presence of an attorney*, whether contracted or appointed by the state in cases

where the law prescribes this.<sup>3</sup> *Year of the decision* is also included to control for changes in the Czech criminal justice system, which evolved significantly during the 2007-2017 period, with new sanctions enacted, shifts in public and judiciary attitudes and in the crime structure, and a broad amnesty issued in 2013. In this amnesty, prison sentences of 1 year and less were pardoned unconditionally, most prison sentences of 1 to 2 years were pardoned conditionally (the condition was not to re-offend in a given period) and suspended prison sentences of up to two years, community sentences and home detentions were also pardoned. Relevant convictions were formally expunged, so that although judges could still consider them when later sentencing for new offenses, their importance was lessened.

Our final group of explanatory variables relates to judge characteristics. The key explanatory variable, *experience*, is measured as the number of criminal cases processed by each judge since taking their oath. Further details are provided in the Measurement Appendix. We also include the judge's *occupation* prior to joining the judiciary in our models. While most Czech judges began their careers as judicial clerks, many were also employed as prosecutors, attorneys, civil servants or in other legal professions. Last, we control for the judge's *age* at the time of the decision and for their *sex* (female judges decided slightly fewer than half of the studied cases).

### ***Samples and descriptive statistics***

We restrict our analysis to cases that were decided in 2007-2017 by district judges<sup>4</sup> who joined the bench (took their judicial oath) during the same period. We exclude cases related to

<sup>38</sup> For young offenders, offenders in pre-trial detention, offense categories with a sentencing maximum of at least 5 years and other relatively infrequent cases; s. 36-36b of the Code of Criminal Procedure.

<sup>4</sup> District judges deal with the vast majority (98 %) of criminal cases; only the most serious cases are tried by regional judges the most serious ones.

youth defendants (less than 2% of our original sample) since they are treated differently than adults: Restorative elements are more commonly considered and incarceration is rare. We further limit the sample to cases decided by judges with sufficient experience, which we define as at least 200 processed cases. This yields 149 judges. To capture general trends of experience and not only those of a few exceptional judges, we also exclude cases processed by any judge with experience of more than 1800 prior cases: At around this level of experience the number of judges in our sample drops below 20. These two exclusions result in the loss of 8,136 cases from our sample, out of 129,920 cases in total.

We partition the remaining cases into three samples, to serve different analytical purposes. Decisions in which a guilty verdict was reached (Sample 1) are selected to analyze the imposition of non-suspended prison sentences and the use of shortened procedures. To examine sentence length, we use a second sample composed only of cases in which non-suspended prison sentences were imposed (Sample 2). Finally, to investigate decisions on guilt we use the full sample (Sample 3).

## **Results**

All our model estimates are reported in Table 1. The models examining severity – studying the imposition of non-suspended prison sentences and their lengths – are in line with existing sentencing scholarship. The defendant’s number of previous convictions substantially influences judges’ decisions to incarcerate and affect the length of the resulting prison sentence. As the number of previous convictions increases, their effect marginally decreases; this can be seen from the negative coefficient for the squared number of previous convictions. More serious offenses and multiple offenses lead to harsher sentences. Different types of offenses are sentenced differently, and the 2013 amnesty had a noticeable impact, since it erased many offenders’ past criminal records and so led to fewer prison sentences being

imposed subsequently. Not surprisingly, we find that pre-trial detention correlates with the imposition of non-suspended prison sentences, although it appears to have only a slight influence on the length of those sentences. When the defendant was represented by an attorney, this was associated both with an increased probability of incarceration, with a lower probability of being found guilty and fewer penal orders, as might be expected. Our models, therefore, confirm several observations common in the existing sentencing scholarship.

[Table 1 near here]

Our estimates indicate that greater judicial experience influences many of the decisions judges take in criminal cases. The odds of a judge imposing a non-suspended prison sentence are positively associated with the number of cases the judge has processed (Model 1). After 1000 cases processed (mean 7.64 years, sd 1.74 years), the odds of imposing a non-suspended prison sentence are 1.27 times as high as prior to it. Experience is, however, not correlated with length of sentence (Model 2).

Increased experience has a substantial impact on decisions on guilt: after processing 1000 cases, the odds of finding the accused not guilty (or transferring their case to the domain of administrative penal law or discontinuing the proceedings) are 1.4 times as great than prior to processing them (Model 3). The strongest effect our models suggest that experience has on judges' decision-making has to do with the use of the shorter, penal order procedure: The odds of *not* employing penal order after processing 1000 cases are 1.77 higher than prior to processing them (Model 4).

To compare the effect sizes and to set these results in the context of other variables, processing 1000 cases increases the incarceration rate by approximately as much as widening the sentencing range by one year (e.g. from 0-2 years to 0-3 years). A judge with 1000 cases'

experience finds defendants guilty less often, and this effect is equivalent in magnitude to that of being represented by an attorney. The influence of processing 1000 cases to employ penal orders is of equivalent effect size to a change of 15% of the range of previous convictions.

We find the age of the judges is also correlated with their decision-making. With increasing age, judges become less likely to issue prison sentences, while those that they do impose are shorter. Interestingly, these findings are in the opposite direction compared to the effects we found to be associated with experience: Experience of 1000 cases has an effect of approximately similar magnitude on the decision as to whether to impose a non-suspended prison sentence – but in the opposite direction – as an additional 10 years of age (Model 1).

None of our models suggest the presence of any linear trend, but they do suggest that judges reacted to specific events during the examined period and that those changes were neither gradual nor linear. Changes in the imposition of non-suspended prison sentences over time (Model 1) provide evidence of the 2013 amnesty affecting decisions in 2013-2015, but there is no change observed either before or after these years. The length of non-suspended prison sentences (Model 2) does not seem to have changed over time. The probability of finding the defendant guilty (Model 3) and of deciding by penal order (Model 4) suddenly decreased in 2011, but the effect sizes remained very similar for all yearly dummies until 2017, similarly not suggesting a linear trend. To control for a possible linear trend, we replaced year dummies with the number of days between the date of the decision and January 1<sup>st</sup> 2007. The results were very similar, with the exception of the results for penal orders, for which the effect of experience increased from -0.57 to -0.8 after including linear day-trends.

Further, we examined models in which we excluded the judges' ages; the effects and their sizes remained very similar. We checked for multicollinearity by examining the variance inflation factors in Model 1: All values were below 5 and for the judge decision

number the value was 1.41. None of our robustness checks suggest that the results of our main models are unreliable.

We then considered the possibility that the effect of experience might not be stable over time (a learning curve, rather than a learning line), by examining models in which we added squared and cubed representations of number of processed cases to the linear one. The message these models communicated about the effects of judges' experience was not different, only more nuanced: While approximately the first 600 cases have a more substantial effect than cases above that number, increasing experience even beyond 600 cases does influence the decisions judges take in the same direction as we observed in our linear representation.

As extralegal disparities are a frequent focus in sentencing scholarship and since considerations of sentencing factors might change over time, we also ran models in which we included an interaction between the number of processed cases and (i) the defendant's sex or (ii) the defendant's number of previous convictions. These interactions were not statistically significant. While additional research would be necessary to investigate this issue in detail, our limited results suggest that judges may consider several factors similarly when sentencing and making other decisions, even with increasing experience.

Finally, to provide methodological insight into different measures of experience, we standardized both "issue-specific judicial experience" (the number of processed cases) and "general judicial experience" (time on bench measured in days) and re-calculated models 1-4. The effect of differently measured judicial experience was virtually the same for sentencing outcomes, yet the effect of issue-specific experience was 1.7-1.8 times higher in models studying the use of penal orders and guilty decisions. This suggests that when judicial experience is not the primary studied variable, general judicial experience can provide some

relevant insight. However, more detailed studies and researchers investigating the influence of experience on judicial decision-making should use the issue-specific measure.

We now turn to consider the random effects parts of our models, which indicate changes in the level of between-judge disparities. For all our outcomes, judges' trajectories converge with increasing experience, as the results reported in Table 1 show. All random slopes terms are statistically significant, as is the covariance between the random intercepts and random slopes. Since those covariance are negative, we can conclude that there is evidence of convergence in between judge disparities across judicial careers. However, the strength of such effect varies. While decisions about whether to incarcerate offenders and on guilt and by penal order converge substantially across judges, there is no strong evidence of convergence for decisions on the lengths of non-suspended prison sentences.

To assess the effect size of this convergence process visually we plot individual judge trajectories in Figure 1, which displays 30 randomly selected judges' trajectories and the average trajectory for the whole sample of judges included in each model. We restrict the predicted random slopes to the first 500 decided cases – this is a compromise between only displaying trajectories for judges with the most experience and displaying trajectories for sufficiently long periods. On average, judges sentenced 500 cases in 3.1 years (sd 1.64 years). The displayed trajectories suggest that judges have distinct approaches when they first begin trying criminal cases, and that there is evidence of outliers following distinct patterns, but on average with increasing experience, their approaches converge, across all outcomes considered.

[Figure 1 near here]

## **Discussion**

We provide empirical evidence of the importance of experience, which had previously only been backed up by theoretical considerations. Experience plays a crucial role in making judicial decision-making more consistent. As the number of cases a judge has processed in their career increases, between-judge disparities decrease for every measure we have investigated. While these findings do not contradict the need for official guidance structuring sentencing discretion, grounded in the legal traditions of individual countries (Council of Europe, 1992), they emphasize the role that experience plays in the judicial system, which is often overlooked. Together with other recent research, especially on the homogenizing effects of judicial rotation in the U.S. (Hester, 2017; Abrams et al., 2021) and on the impact of judges' meetings within smaller jurisdictions (O'Malley, 2013; Lappi-Seppälä, 2001), our research suggests that more attention should be paid to traditional and informal means of increasing sentencing consistency. Our findings are particularly encouraging, since examined Czech judges were given only limited learning opportunities to refine their sentencing practices. There was no training and no textbook on sentencing, the level of reasoning judges gave for the sentences imposed was much lower than desirable (Tomšů & Drápal, 2019), and neither appellate courts nor sentencing councils supported judges by providing (sufficient) guidance; just as in many other Central and Eastern European countries.

When moving beyond the analysis of disparities to that of average effects, we have also identified that as judges gain sentencing experience, they find fewer defendants guilty, but impose more non-suspended prison sentences, and decide fewer cases by shortened procedure. Judges' experience does not, however, impact the length of the non-suspended prison sentences they impose. Some of these results are not fully in line with previous theoretical research nor, in some cases, with the hypotheses we formulated. The higher propensity to impose non-suspended prison sentences might support those who have argued



that experience hardens judges (Hauser, 2012; Welch et al., 1988). Yet this is only partially corroborated, since we found no effect on length of prison sentences.

The fact that more experienced judges are less likely to find the accused guilty and make less use of simplified procedures seems to confirm that judges become more confident as they gain experience, and that they become more familiar with the court techniques that can help them cope with their case load. We interpret their apparently increased willingness to oppose the prosecution as evidence of increased confidence. The reason for it is that it involves disagreeing with professional repeat-players in the process, who usually have much more experience with the system than any novice judge (Heumann, 1978/2020, chpt. 6). While we hypothesized that more experienced judges would use the simplified penal order procedure more frequently, as it is the easiest and quickest way to likely close most cases, this hypothesis was refuted. More experienced judges deal with their cases in more detail and decide fewer cases by penal order, which will likely allow them to better individualize the sanction, achieving one of sentencers' goals (Tata, 2019; Plesničar, 2013), while providing them with the pleasure of judging (Posner, 1993).

Overall, we interpret the results as indicating that as judges gain experience, they become more prone to (or better at) differentiating the cases they deal with into three categories: Cases that should not be decided by penal order, since they require thorough consideration and a non-suspended prison sentence might be appropriate. Second, cases that should not be processed by the criminal justice system at all, due to lack of evidence or seriousness. And third, cases suited to the faster, simplified penal order procedure. We show that with increasing experience the numbers of defendants in the third category reduce in favor of the other two categories as more experienced judges impose more non-suspended prison sentences, find fewer defendants guilty and decide fewer cases via penal order. This differentiation parallels findings from the United States, suggesting that more experienced

prosecutors do not always seek severe outcomes, but rather try to distinguish between cases requiring lenient and severe outcomes, are willing not to apply the most severe charges or sanctions and question the police account of the crime (Wright & Levine, 2014).

Our analysis of procedural response variables enables us to reconcile the previous inconsistent associations found between judges' experience and the severity of the sanctions they impose. Interpreting an increase in the probability of incarceration as simple evidence of judges becoming more punitive fails to grasp the compromises that judges are forced to make on a daily basis in order to process the vast numbers of cases on their desks, as is required of them, while delivering just decisions (Tata, 2019). Judges are under pressure to review cases as quickly as possible; the easiest way to do so is to use the simplified penal order procedure. The difficulty is that they are not permitted to directly incarcerate any offender when using the penal order procedure. Early in their careers, when they have not yet mastered the court techniques, judges seem to be willing to trade speed for other types of sanctions. That approach (and the prosecution's limited resistance to it) is in line both with the observation that sentencing is considered less important than the decision on guilt in continental legal systems and with previous research, which has shown that Czech judges are generally willing not to incarcerate offenders if it enables them to process cases more quickly (Drápal, 2017). Once judges become more familiar with court techniques, they partially desist from such widespread use of penal orders; we believe the reason for this is that the full procedure enables them to deal with cases more properly. Unfortunately, the relationship between the use of penal orders and non-suspended prison sentences cannot be directly tested.

The effect of judicial experience on sentence severity, measured by the rate of imposition of non-suspended prison sentences, must be interpreted within the context of the specific criminal justice system in question, as countries differ substantially in their use of prison sentences. The Czech Republic sends far fewer offenders to prison than many other

liberal European countries; its high prison population is caused by prisoners serving long sentences (Dünkel, 2017). The length of those sentences is caused, among other factors, by the accumulation of sentences imposed one after another, of which the last is a non-suspended prison sentence (Drápal, 2021; continental systems usually do not know the institute of concurrently run prison sentences). Earlier imposition of non-suspended prison sentences might, in this context, not be a signal of any increase in punitiveness, but of a recognition that imposing further suspended prison sentences would only lead to future accumulation of sentences and thus to disproportionately long series of prison sentences being served. This hypothesis would need to be confirmed by further research, since other interpretations are also imaginable, such as that judges with greater experience tend to use fewer alternative sanctions after seeing that offenders often breach their conditions, and so resort to imposing more non-suspended prison sentences.

In applying these conclusions to the common law – and especially U.S. – context, the equivalent to overused penal orders are guilty pleas that are subject to virtually no review (Rakoff, 2021, chpt. 2). While judges have the opportunity to accept or decline plea argument, they seldom do so due to limited information (Johnson, 2019). We would expect this to change with increasing experience. Novice judges have little knowledge about sentencing and guilty pleas remove this burden from their shoulders, enabling them to deal with other problems. Yet with increasing experience, judges are better able to distinguish between cases that are worthy of trial and those not, hence some of them participate in the negotiations and set rough guidelines for prosecutors (Heumann, 1978/2020, chpt. 6), especially in states that allow this (Henderson et al., 2021). Overall, with increasing experience we would expect U.S. judges to be more willing to search for more information, leading to increased judicial oversight over one of the most contentious and problematic issues in the present U.S. criminal justice system, if they were allowed to (Henderson et al.,

2021). This would be in line with the role a judge should play in plea-bargaining, specifically that judges should not be pressured by their crowded docket to urge defendants to enter guilty pleas and rather they should ensure defendants are treated fairly (Lippke 2011, p. 205).

### **Implications for further research and policy**

Judicial experience does not only matter at sentencing but also in other aspects of penal decision-making. The unique contribution of our study is the demonstration that judicial experience matters at sentencing because it reduces the extent of between judge disparities, while it further leads judges to find fewer defendants guilty and to decide fewer cases by the quickest route possible. We show the relevance of considering procedural measures, as opposed to focus simply on sentencing, since we still know rather little about the impact of procedural choices and restraints on sentencing.

The window of observation considered in our study encompassed a maximum of 11 years with the average trajectory captured being 5.2 years. We do not know, therefore, what the impact of experience might be after 15 or 30 years of judging. The interaction of experience and age is a matter of particular interest, since our study suggests the impacts of experience and age might operate in opposite directions but to a similar extent. We expect further research to study longer periods of time, which would enable more nuanced studies capable of exploring potential non-linear trajectories. Future research should also focus on the development of trajectories in within-judge disparities, a question that we have not covered here.

Studies replicating our approach will be useful to establish whether our findings are limited to the continental judicial and sentencing culture or whether such trends also exist in common-law countries, where judges hold different positions and in which the sentencing process differs. Firstly, replications would be needed regarding whether the level of

unwarranted sentencing disparities improves with experience. Future studies could also consider whether novice judges are more likely to converge towards the national practice or towards the practice at their court. Moreover, it would be of key relevance to test the interpretations from some of our results, namely that more experienced judges might try to do their jobs as thoroughly as possible under the available circumstances and pressures (which might be differently operationalized in different contexts, as we suggested above vis-à-vis plea bargaining in the U.S. context). In the decades to come, we expect sentencing scholars to carry out research comparing the penal decisions made by differently experienced judges in civil and common law systems, which might also have implications for a discussion of what type of judicial appointment system a country should choose to adopt.

Lastly, while most legal systems currently enable judges to gain expertise in some areas of their work, many criminal justice systems are not designed to make the most of the learning opportunities that judges and other court actors could profit from (as outlined above by the psychological scholarship). Too frequently judges are neither thoroughly trained to learn from their experience, nor is the observation of other judges' sentencing processes normally encouraged. We hypothesize that if judges were provided with additional training, if they were procedurally incentivized to thoroughly think about their cases (e.g. via a requirement for more detailed reasoning) and if they received more detailed feedback on their decisions from their peers (e.g. via conversations with colleagues or via appellate courts' corrections), their decision-making would become even more consistent. The evidence supports this: When a judge is regularly in contact with other judges' sentencing practices, e.g. via judicial rotation, that experience seems to limit between-judge disparities (Hester, 2017; Abrams et al., 2021).

## **Bibliography:**

- Abrams, D., Galbiati, R., Henry, E. & Philippe, A. (2021) When in Rome ... on Local Norms and Sentencing Decisions. *Journal of the European Economic Association*. Advance online publication. <https://doi.org/10.1093/jeea/jvab038>
- Albonetti, C. A. (1991). An Integration of Theories to Explain Judicial Discretion. *Social Problems*, 38(2), 247–266. <https://doi.org/10.2307/800532>
- Anderson, A. L., & Spohn, C. (2010). Lawlessness in the federal sentencing process: A test for uniformity and consistency in sentence outcomes. *Justice Quarterly*, 27(3), 362-393. <https://doi.org/10.1080/07418820902936683>
- Ashworth, A. (1995). Reflections on the Role of the Sentencing Scholar. In Clarkson, C. & Morgan, R. (Eds.). *The Politics of Sentencing Reform* (pp. 251-256). Clarendon Press.
- Billett, S., Harteis, C., & Gruber, H. (2018). Developing Occupational Expertise through Everyday Work Activities and Interactions. In Ericsson, K.A., Hoffman, R.R. Kozbelt, A. & Williams, A. M. (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (2<sup>nd</sup> ed.; pp. 105-126). Cambridge University Press.
- Bindler, A. & Hjalmarsson, R. (2019). Path Dependency in Jury Decision Making. *Journal of the European Economic Association*, 17(6), 1971–2017. <https://doi.org/10.1093/jeea/jvy046>
- Cauffman, E., Fine, A., Thomas, A. G., & Monahan, K. C. (2017). Trajectories of Violent Behavior among Females and Males. *Child Development*, 88(1), 41-54. <https://doi.org/10.1111/cdev.12678>
- Cianciolo, A. T., & Sternberg, R. J. (2018). Practical Intelligence and Tacit Knowledge: An Ecological View of Expertise. In Ericsson, K.A., Hoffman, R.R. Kozbelt, A. & Williams,

- A. M. (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (2<sup>nd</sup> ed.; pp. 770-762). Cambridge University Press.
- Cook, B. B. (1973). Sentencing Behavior of Federal Judges: Draft Cases-1972. *University of Cincinnati Law Review*, 42(4), 597–633.
- Dhami, M.K., & Belton, I. (2016). Statistical Analyses of Court Decisions: An Example of Multilevel Models of Sentencing. *Law and Method*, 10, 247–266. <https://doi.org/10.5553/rem/.000019>
- Drápal, J. (2017). Vede větší užití trestních příkazů k ukládání méně nepodmíněných trestů odnětí svobody? *Jurisprudence*, (5), 5–17.
- Drápal, J. (2021). Sentencing multiple conviction offenders. *European Journal of Criminology*. Advance online publication. <https://doi.org/10.1177/1477370821996903>
- Drápal, J. & Vávra, J. (2021) Databáze InfoSoud a možnosti jejího využití. *Česká kriminologie*, 6(1-2), p. 1-25.
- Düinkel, F. (2017). European Penology: The Rise and Fall of Prison Population Rates in Europe in Times of Migrant Crises and Terrorism. *European Journal of Criminology*, 14(6), 629–653. <https://doi.org/10.1177/1477370817733961>
- Eisenstein J., Flemming, R. B. & Nardulli, P. F. (1988). *Contours of Justice: Communities and Their Courts*. Little, Brown and Co.
- Emerson, R. M. (1983). Holistic Effects in Social Control Decision-Making. *Law & Society Review*, 17(3), 425-455.
- Ericsson, K. A. (2018). The Differential Influence of Experience, Practice, and Deliberate Practice on the Development of Superior Individual Performance of Experts. In Ericsson, K.A., Hoffman, R.R. Kozbelt, A. & Williams, A. M. (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (2<sup>nd</sup> ed.; pp. 745-769). Cambridge University Press.

- Fearn, N. E. (2005). A Multilevel Analysis of Community Effects on Criminal Sentencing. *Justice Quarterly*, 22(4), 452–487. <https://doi.org/10.1080/07418820500364668>
- Feldmeyer, B. & Ulmer, J. T. (2011). Racial/ethnic threat and federal sentencing. *Journal of Research in Crime and Delinquency*, 48(2), 238-270. <https://doi.org/10.1177/0022427810391538>
- Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2018). Studies of Expertise from Psychological Perspectives: Historical Foundations and Recurrent Themes. In Ericsson, K.A., Hoffman, R.R. Kozbelt, A. & Williams, A. M. (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (2<sup>nd</sup> ed.; 59-83). Cambridge University Press.
- Fine, A., & Cauffman, E. (2015). Race and Justice System Attitude Formation During the Transition to Adulthood. *Journal of Developmental Life Course Criminology*, 1, 325-349. <https://doi.org/10.1007/s40865-015-0021-2>
- Frazier, C. E., & Bock, E. W. (1982). Effects of Court Officials on Sentence Severity: Do Judges Make a Difference? *Criminology*, 20(2), 257–272. <https://doi.org/10.1111/j.1745-9125.1982.tb00460.x>
- Hauser, W. I. (2012). *Do Judges' Experiences and Indelible Traits Influence Sentencing Decisions? New Evidence from Florida*. [Doctoral dissertation, Florida State University]. FSU Digital Repository. <https://diginole.lib.fsu.edu/islandora/object/fsu:183251/datastream/PDF/download/citation.pdf>.
- Henderson, K. S., Fountain, E. N., Redlich, A. D., & Cantone, J. A. (2021). Judicial involvement in plea-bargaining. *Psychology, Public Policy, and Law*. Advance online publication. <https://doi.org/10.1037/law0000327>



- Hester, R. (2017). Judicial rotation as centripetal force: Sentencing in the court communities of South Carolina. *Criminology*, 55(1), 205-235. <https://doi.org/10.1111/1745-9125.12132>
- Heumann, M. (2020). *Plea Bargaining*. The University of Chicago Press. (Original work published 1978)
- Johnson, B. D. (2005). Contextual Disparities in Guidelines Departures: Courtroom Social Contexts, Guidelines Compliance and Extralegal Disparities in Criminal Sentencing. *Criminology*, 43(3), 761-796. <https://doi.org/10.1111/j.0011-1348.2005.00023.x>
- Johnson, B. D. (2006). The Multilevel Context of Criminal Sentencing: Integrating Judge- and Country-level influences. *Criminology*, 44(2), 259–298. <https://doi.org/10.1111/j.1745-9125.2006.00049.x>
- Johnson, B. D. (2019). Trials and Tribulations: The Trial Tax and the Process of Punishment. *Crime and Justice*, 48, 313–363. <https://doi.org/10.1086/701713>
- Johnson, B. D. & Lee, J. G. (2008). Racial disparity under sentencing guidelines: A survey of recent research and emerging perspectives. *Sociology Compass*, 7(7), 503-514. <https://doi.org/10.1111/soc4.12046>
- Kaspar, J. (2020). Sentencing Guidelines vs. Free Judicial Discretion – Is German Sentencing Law in Need of Reform? In Ambos, K (Ed.). *Sentencing. Anglo-American and German Insights* (pp. 337-352). Göttingen University Press.
- Kim, B., Spohn, C. & Hedberg, E. C. (2015). Federal sentencing as a complex collaborative process: Judges, prosecutors, judge-prosecutor dyads, and disparity in sentencing. *Criminology*, 53(4), 597-623. <https://doi.org/10.1111/1745-9125.12090>
- Kritzer, H. M. (1978). Political Correlates of the Behavior of Federal District Judges: A “Best Case” Analysis. *The Journal of Politics*, 40(1), 25–58. <https://doi.org/10.2307/2129975>

- Leibovitch, A. (2016). Relative Judgments. *Journal of Legal Studies*, 45(2), 281-330.  
<https://doi.org/10.1086/687376>
- Lappi-Seppälä, T. (2001). Sentencing and Punishment in Finland: The Decline of the Repressive Ideal. In Tonry, M. & Frase, R.S. (Eds.). *Sentencing and Sanctions in Western Countries* (pp. 92-150). Oxford University Press.
- Lim, C. S. H., Silveira, B. S., & Snyder, J. M. (2016). Do Judges' Characteristics Matter? Ethnicity, Gender, and Partisanship in Texas State Trial Courts. *American Law and Economics Review*, 18(2), 302–357. <https://doi.org/10.1093/aler/ahw006>
- Lippke, R. L. (2011). *The Ethics of Plea Bargaining*. Oxford University Press.
- Lombrozo, T. (2012). Explanation and Abductive Inference. In K. J. Holyoak & R. G. Morrison (Eds.). *The Oxford Handbook of Thinking and Reasoning* (2<sup>nd</sup> ed., pp. 260-276). Oxford University Press.
- McLean, K., Wolfe, S. E., & Pratt, T. C. (2019). Legitimacy and the Life Course: An Age-graded Examination of Changes in Legitimacy Attitudes over Time. *Journal of Research in Crime and Delinquency*, 56(1), 42-83. <https://doi.org/10.1177/0022427818793934>
- Merrall, E. L. C., Dhimi, M. K., & Bird, S. M. (2010). Exploring Methods to Investigate Sentencing Decisions. *Evaluation Review*, 34(3), 185–219.
- Mosier, K., Fischer, U., Hoffman, R. R., & Klein, G. (2018). Expert Professional Judgments and “Naturalistic Decision Making.” In Ericsson, K.A., Hoffman, R.R. Kozbelt, A. & Williams, A. M. (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (2<sup>nd</sup> ed.; pp. 453-475). Cambridge University Press.
- Myers, M. A. (1988). Social Background and the Sentencing Behavior of the Judges. *Criminology*, 26(4), 649–676. <https://doi.org/10.1111/j.1745-9125.1988.tb00858.x>

- O'Malley, T. (2013). Living without Guidelines. In Ashworth, A. & Roberts, J. V. (Eds.) *Sentencing Guidelines: Exploring the English Model* (pp. 218-236). Oxford University Press.
- Penn, J., & Silverstein, M. (2012). Gender Role Ideology and Life Course Transitions of Baby-boom Women. *Advances in Life Course Research*, 17(4), 191-198. <https://doi.org/10.1016/j.alcr.2012.04.001>
- Pina-Sánchez, J., & Linacre, R. (2013). Sentence consistency in England and Wales: Evidence from the Crown Court sentencing survey. *British Journal of Criminology*, 53(6), 1118-1138. <https://doi.org/10.1093/bjc/azt040>
- Pina-Sánchez, J., & Linacre, R. (2014). Enhancing Consistency in Sentencing: Exploring the Effects of Guidelines in England and Wales. *Journal of Quantitative Criminology*, 30(4), 731-748. <https://doi.org/10.1007/s10940-014-9221-x>
- Pina-Sánchez, J., Grech, D., Brunton-Smith, I. & Sferopoulos, D. (2019). Exploring the origin of sentencing disparities in the Crown Court: Using text mining techniques to differentiate between court and judges disparities. *Social Science Research*, 84(1), 102343. <https://doi.org/10.1016/j.ssresearch.2019.102343>
- Plesničar, M. M. (2013). The individualization of punishment: Sentencing in Slovenia. *European Journal of Criminology*, 10(4), 462-478. <https://doi.org/10.1177/1477370812469858>
- Posner, R. A. (1993). What Do Judges and Justices Maximize? (The Same Thing Everybody Else Does). *Supreme Court Economic Review*, 28(3), 1-41. <https://doi.org/10.1086/scer.3.1147064>
- Rakoff, J. S. (2021). *Why the Innocent Plead Guilty and the Guilty Go Free*. Farrar, Straus and Giroux.

- Scheinost, M., Válková, H., Háková, L., Hulmáková, J., Kotulan, P., Rozum, J., Šámal, P., Škvain, P., Tomášek, J. & Vlach, J. (2015). *Sankční politika a její uplatňování*. IKSP.
- Spohn, C. (1991). Decision Making in Sexual Assault Cases: Do Black and Female Judges Make a Difference? *Women and Criminal Justice*, 2(1), 83–105.  
[https://doi.org/10.1300/J012v02n01\\_06](https://doi.org/10.1300/J012v02n01_06)
- Spohn, C. (2015). Evolution of Sentencing Research. *Criminology & Public Policy*, 14(2), 225–232.
- Steffensmeier, D., & Britt, C. L. (2001). Judges' Race and Judicial Decision Making: Do Black Judges Sentence Differently? *Social Science Quarterly*, 82(4), 749–764.  
<https://doi.org/10.1111/0038-4941.00057>
- Steffensmeier, D., & Hebert, C. (1999). Women and Men Policymakers: Does the Judge's Gender Affect the Sentencing of Criminal Defendants? *Social Forces*, 77(3), 1163–1196.  
<https://doi.org/10.1093/sf/77.3.1163>
- Tata, C. (2007). Sentencing as Craftwork and the Binary Epistemologies of the Discretionary Decision Process. *Socio-Legal Studies*, 16(3), 425–447.  
<https://doi.org/10.1177/0964663907079767>
- Tata, C. (2019). 'Ritual Individualization': Creative Genius at Sentencing, Mitigation, and Conviction. *Journal of Law and Society*, 46(1), 112–140.  
<https://doi.org/10.1111/jols.12144>
- Tata, C. (2020). *Sentencing: A Social Process*. Palgrave Pivot.
- Tomšů, K., & Drápal, J. (2019). Odůvodňování trestů: Empirická studie rozhodnutí okresních soudů. *Státní Zastupitelství*, 17(6), 9–21.
- Tonry, M. (2016). Equality and Human Dignity: The Missing Ingredients in American Sentencing. *Crime and Justice*, 45(1), 459–496. <https://doi.org/10.1086/686256>

- Ulmer, J. T. (2012). Recent Developments and New Directions in Sentencing Research. *Justice Quarterly*, 29(1), 1-40. <https://doi.org/10.1080/07418825.2011.624115>
- Ulmer, J. T., & Bradley, M. S. (Eds.). (2018). *Handbook on Punishment Decisions: Locations of Disparity*. Routledge.
- Ulmer, J. T., & Johnson, B. D. (2004). Sentencing in Context: A Multilevel Analysis. *Criminology*, 42(1), 137–178. <https://doi.org/10.1111/j.1745-9125.2004.tb00516.x>
- Ulmer, J. T., Light, M. T. & Kramer, J. (2011). The “liberation” of federal judges’ discretion in the wake of the Booker/Fanfan decision: Is there increased disparity and divergence between courts? *Justice Quarterly*, 28(6), 799-837. <https://doi.org/10.1080/07418825.2011.553726>
- Welch, S., Combs, M., & Gruhl, J. (1988). Do Black Judges Make a Difference? *American Journal of Political Science*, 32(1), 126-136.
- Winegard, B., Winegard, B., & Geary, D.C. (2018). The Evolution of Expertise. In Ericsson, K.A., Hoffman, R.R. Kozbelt, A. & Williams, A. M. (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (2<sup>nd</sup> ed.; pp. 40-48). Cambridge University Press.
- Wright, R. F. & Levine, K. L. (2014). The Cure for Young Prosecutors’ Syndrome. *Arizona Law Review*, 56(4), 1065-1128.
- Wu, J. & Spohn, C. (2010). Interdistrict disparity in sentencing in three US district courts. *Crime & Delinquency*, 56(2), 290-322. <https://doi.org/10.1177/0011128707309997>

## Tables

Table 1: Results of growth curve models

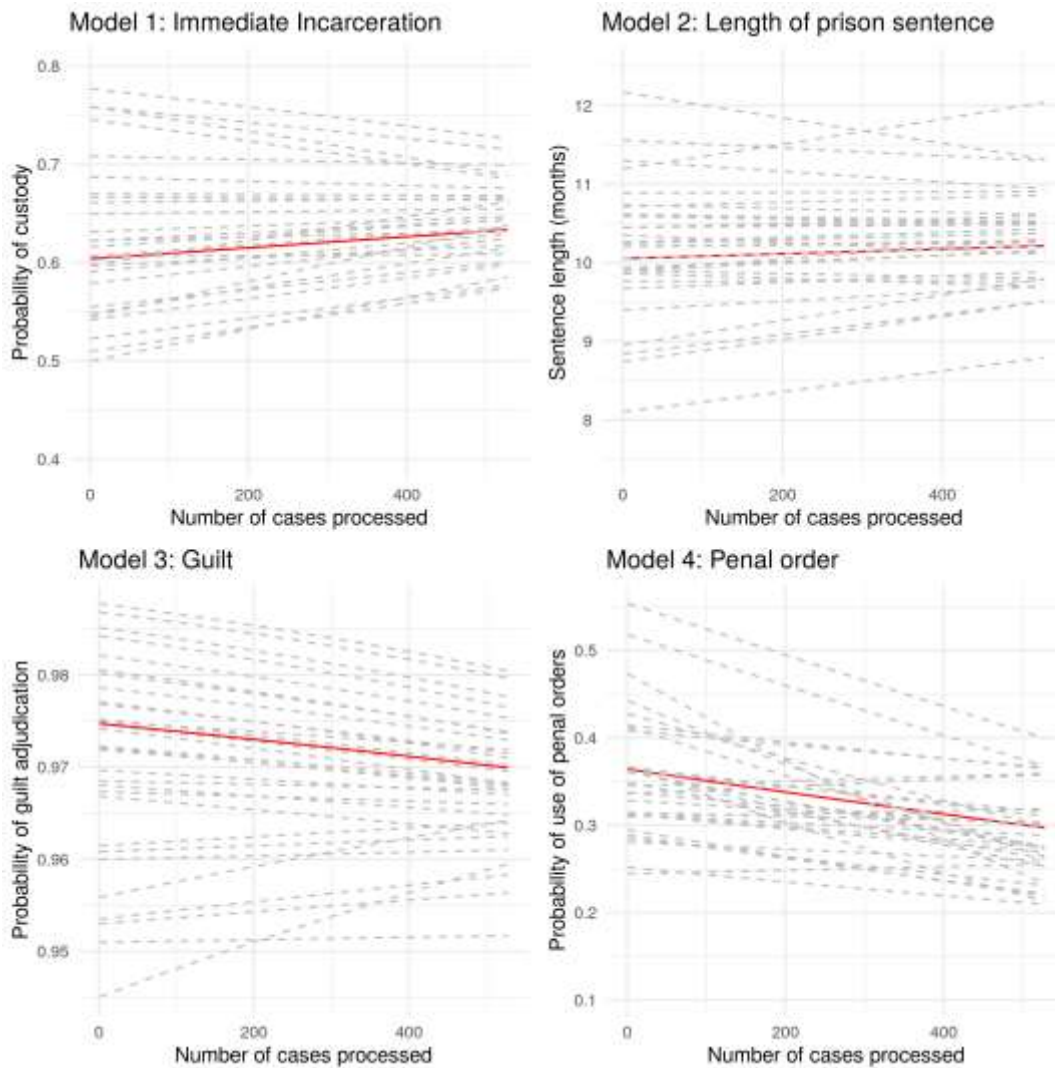
	Model 1: Imposition of non-suspended prison sentence			Model 2: Length of non-suspended prison sentence			Model 3: Guilt			Model 4: Penal order		
	Coeff.	S.E.		Coeff.	S.E.		Coeff.	S.E.		Coeff.	S.E.	
Intercept	0.42	0.3		2.63	0.09	***	3.65	0.27	***	-0.56	0.25	*
<b>Judge decision number</b> (divided by 1000)	0.24	0.07	**	0.03	0.02		-0.34	0.08	***	-0.57	0.09	***
<b>Sentence range</b>	0.47	0.02	***	0.4	0.01	***	-0.14	0.01	***	-0.75	0.01	***
<b>Previous convictions</b> (demeaned)	4.89	0.06	***	0.45	0.02	***	-0.11	0.04	**	-2.05	0.04	***
<b>Previous convictions</b> (demeaned) <sup>2</sup>	-2.27	0.05	***	-0.16	0.02	***	-0.07	0.04		0.86	0.04	***
<b>Most serious offense in the case</b> (ref: Theft): Bodily harm	-1.5	0.1	***	0.11	0.04	**	-0.23	0.06	***	0.42	0.05	***
Bodily harm (negligence)	-2.11	0.72	**	-0.45	0.35		0.99	0.18	***	-0.79	0.11	***
Credit fraud	-1.74	0.12	***	0.1	0.05		1.28	0.08	***	0.88	0.05	***
Damage to a thing	-0.96	0.17	***	-0.3	0.08	***	0.05	0.09		0.03	0.07	
Dangerous stalking	-1.1	0.18	***	-0.02	0.07		-0.4	0.11	***	0.01	0.1	
Disorderly conduct	-1.6	0.12	***	-0.08	0.05		-0.22	0.06	***	0.15	0.05	**
Driving under influence	-1.49	0.09	***	-0.09	0.04	*	1.4	0.06	***	0.36	0.03	***
Drug production	-1.19	0.08	***	0	0.03		0.74	0.08	***	0.86	0.06	***
Embezzlement	-0.95	0.11	***	0.2	0.05	***	0.43	0.07	***	0.53	0.06	***
Endangering a child's upbringing	-1.62	0.22	***	-0.11	0.1		1.15	0.15	***	1	0.08	***
Extortion	-1.73	0.14	***	0.05	0.04		-0.48	0.08	***	-0.02	0.11	
Fraud	-0.73	0.08	***	0.33	0.03	***	0.4	0.05	***	0.34	0.05	***
Frustrating execution of an official order	0.27	0.04	***	-0.38	0.01	***	0.51	0.04	***	-0.09	0.03	**
Great bodily harm (negligence)	-0.19	0.16		0.33	0.07	***	0.28	0.1	**	-0.53	0.08	***
Insurance fraud	-1.64	0.32	***	0.39	0.15	**	-0.02	0.12		-0.7	0.14	***
Money laundering	-0.97	0.16	***	-0.05	0.06		-0.15	0.11		0.09	0.1	
Non-payment of alimony	-0.39	0.05	***	0.03	0.02		-0.28	0.04	***	0.43	0.03	***
Other	-1.48	0.07	***	-0.05	0.02	*	-0.1	0.04	**	0.06	0.04	
Robbery	-1.31	0.11	***	-0.13	0.03	***	0.26	0.09	**	-1.63	0.38	***
Unauthorized possession of a credit card	-1.36	0.19	***	-0.11	0.07		0.32	0.13	**	0.47	0.09	***
Unauthorized use of another's property	-0.96	0.24	***	0.02	0.1		-0.11	0.17		0.38	0.14	**
Violence against a public official	-0.56	0.16	***	0.06	0.06		0.26	0.14		0.27	0.1	**
Violation of a domestic freedom	-0.55	0.06	***	-0.02	0.02		-0.12	0.06	*	0.37	0.05	***
<b>Offender's age at offense date</b>	-0.04	0	***	0	0	***	-0.01	0	***	0.01	0	***
<b>Sex of offender</b> (ref. male)	-0.1	0.05	*	-0.11	0.02	***	0.12	0.03	***	0.1	0.02	***
<b>Year of the decision</b> (ref. 2007): 2008	0.08	0.17		0.04	0.06		-0.19	0.17		0.13	0.11	
2009	0.14	0.17		0.04	0.06		-0.27	0.17		-0.02	0.11	
2010	0.1	0.17		0.06	0.06		-1.06	0.17	***	-0.67	0.12	***
2011	-0.06	0.18		0.05	0.06		-1.32	0.18	***	-0.83	0.12	***
2012	-0.28	0.19		0.03	0.06		-1.47	0.18	***	-0.85	0.13	***
2013	-1.18	0.19	***	0.03	0.06		-1.47	0.18	***	-0.45	0.13	***
2014	-0.79	0.19	***	0.01	0.06		-1.48	0.18	***	-0.51	0.13	***
2015	-0.43	0.19	*	0.01	0.06		-1.38	0.19	***	-0.67	0.14	***
2016	-0.29	0.2		0.02	0.06		-1.39	0.19	***	-0.64	0.14	***
2017	-0.31	0.2		-0.01	0.06		-1.41	0.19	***	-0.61	0.15	***
<b>Sentenced under old Penal Code</b>	0.2	0.07	**	-0.01	0.02		1.18	0.04	***	1.02	0.04	***
<b>Concurrence</b> (ref. yes)	-0.49	0.03	***	-0.27	0.01	***	-0.52	0.03	***	0.26	0.02	***
<b>Pre-trial detention</b> (ref. no)	2.94	0.05	***	0.2	0.01	***	1.12	0.07	***	-3.31	0.14	***

<b>Manner of beginning procedure</b> (ref. full): Other	0.34	0.38	0.26	0.11	*	-0.42	0.18	*	-3.19	0.81	***	
Shortened	-0.02	0.03	-0.07	0.01	***	0.24	0.03	***	0.46	0.02	***	
<b>Represented by attorney</b> (ref. no)	1.24	0.03	***	-0.06	0.01	***	-0.43	0.03	***	-1.83	0.03	***
<b>Judge's age at decision</b>	-0.02	0.01	**	-0.01	0	**	-0.01	0.01		0	0.01	
<b>Sex of the judge</b> (ref. male)	-0.09	0.07		-0.01	0.02		0.12	0.06	*	0.08	0.06	
<b>Previous occupation</b> (ref. jud. assistant): Attorney	0.01	0.09		0.02	0.03		0.12	0.08		0.03	0.09	
Civil servant	-0.31	0.19		-0.04	0.07		-0.28	0.18		0.43	0.22	*
Other	0.52	0.49		-0.01	0.14		0.29	0.49		-0.06	0.39	
Prosecutor	0.14	0.09		0.03	0.03		0.03	0.08		-0.03	0.09	
Random part estimates	Coef.	S.E.		Coef.	S.E.		Coef.	S.E.		Coef.	S.E.	
Court level: Variance intercept	0.2	0.05		0.005	0.002		0.07	0.02		0.09	0.03	
Judge level: Variance intercept	0.21	0.04		0.013	0.003		0.22	0.03		0.13	0.02	
Judge level: Covariance of intercept and experience	-0.2	0.04		-0.011	0.004		-0.3	0.05		-0.15	0.04	
Judge level: Variance experience	0.23	0.05		0.019	0.005		0.57	0.1		0.54	0.09	

Note: \*: p < 0.05, \*\*: p < 0.01, \*\*\*: p < 0.001

## Figures

Figure 1: Judge trajectories



NOTE: Each dashed gray line represents the linear trajectory of an individual judge over their first 500 cases; the red line symbolizes the average judge trajectory.

### Figure captions:

Figure 1: Judge trajectories



## **Methodological Appendix**

### *Measurement Appendix*

In this Measurement Appendix we provide further insights into the characteristics of the data used in our analysis and expand our justification for using the number of cases processed as a measure of issue-specific judicial experience.

### *Descriptive statistics*

The descriptive statistics for the variables used in our analysis are presented in Table A.2. This table is divided in two to distinguish between categorical and continuous variables. In each part, the descriptive statistics are presented separately for each sample.

Table A.1: Descriptive statistics

**Categorical variables**

	<b>Sample 1: Guilty verdicts (%)</b>	<b>Sample 2: Nonsuspended prison sentence imposed (%)</b>	<b>Sample 3: All cases (%)</b>
<b>Most serious offense in the case:</b> Theft	19.5	37.79	21.26
Bodily harm	2.61	1.21	2.63
Bodily harm (negligence)	0.47	0.02	0.59
Credit fraud	3.93	0.76	3.86
Damage to a thing	1.4	0.34	1.42
Dangerous stalking	0.69	0.43	0.66
Disorderly conduct	2.85	0.77	2.83
Driving under influence	13.83	1.31	11.35
Drug production	2.78	4.37	2.78
Embezzlement	2.54	0.91	2.6
Endangering a child's upbringing	1.18	0.19	1.04
Extortion	0.89	1.3	1.07
Fraud	4.09	2.62	4.53
Frustrating execution of an official order	12.41	21.53	11.49
Great bodily harm (negligence)	0.94	0.43	1.06
Insurance fraud	0.41	0.09	0.47
Money laundering	0.61	0.55	0.71
Non-payment of alimony	13.59	8.34	12.91
Other	8.44	5.77	9.14
Robbery	1.29	4.73	1.6
Unauthorized possession of a credit card	0.85	0.35	0.84
Unauthorized use of another's property	0.35	0.2	0.36
Violence against a public official	0.64	0.61	0.64
Violation of domestic freedom	3.71	5.39	4.16
<b>Year of the decision:</b> 2007	0.7	0.74	0.68
2008	2.85	2.92	2.78
2009	5.61	6.07	5.49
2010	7.81	9.09	7.84
2011	8.95	10.57	8.98
2012	9.82	10.76	9.87
2013	12.87	9.06	12.4
2014	13.05	11.08	13.2
2015	12.41	12.2	12.55
2016	13.7	14.24	13.85
2017	12.22	13.26	12.37
<b>Procedure:</b> Full	44.1	49.89	47.86
Other	0.08	0.16	0.17
Shortened	55.81	49.96	51.97
<b>Previous occupation:</b> Judicial assistant	64.96	60.14	64.46
Attorney	16.09	17.3	16.36
Civil servant	2.78	2.25	2.81
Other occupation	0.2	0.31	0.19
Prosecutor	15.96	20.01	16.18
<b>Female offender</b>	15.56	7.96	14.91
<b>Sentenced under old Penal Code</b>	14.29	14.83	15.71
<b>Concurrence</b>	79.01	63.63	76.73
<b>Pre-trial detention</b>	4.39	25.34	4.5

<b>Represented by an attorney</b>	12.09	34.7	14.35
<b>Female judge</b>	47.42	43.42	46.91
<b>N</b>	94,200	12,693	121,784

### Continuous variables

	Min.	Max.	Mean	SD
<b>Guilty verdicts (Sample 1)</b>				
Average sentence range (de-meaned)	-1.17	10.5	0	0.98
Number of previous convictions (de-meaned)	-0.28	1.72	0	0.36
Offender's age at offense date	16	112	34.04	11
Number of previous cases decided by the judge	1	1800	598	463
Judge's age at decision	21.61	61.7	36.98	5.19
<b>Non-suspended prison sentence imposed (Sample 2)</b>				
Average sentence range (de-meaned)	-1	10.5	0.49	1.24
Number of previous convictions (de-meaned)	-0.28	1.72	0.37	0.44
Offender's age at offense date	17	83	33.26	9.57
Number of previous cases decided by the judge	1	1800	596	458
Judge's age at decision	23.84	59.71	36.91	5.01
<b>All cases (Sample 3)</b>				
Average sentence range (de-meaned)	-1.23	11.43	0	1.03
Defendant's number of previous convictions (de-meaned)	-0.31	1.69	0	0.38
Defendant's age at offense date	15	112	33.83	10.92
Number of previous cases decided by the judge	1	1800	604	461
Judge's age at decision	21.61	61.7	37.01	5.18

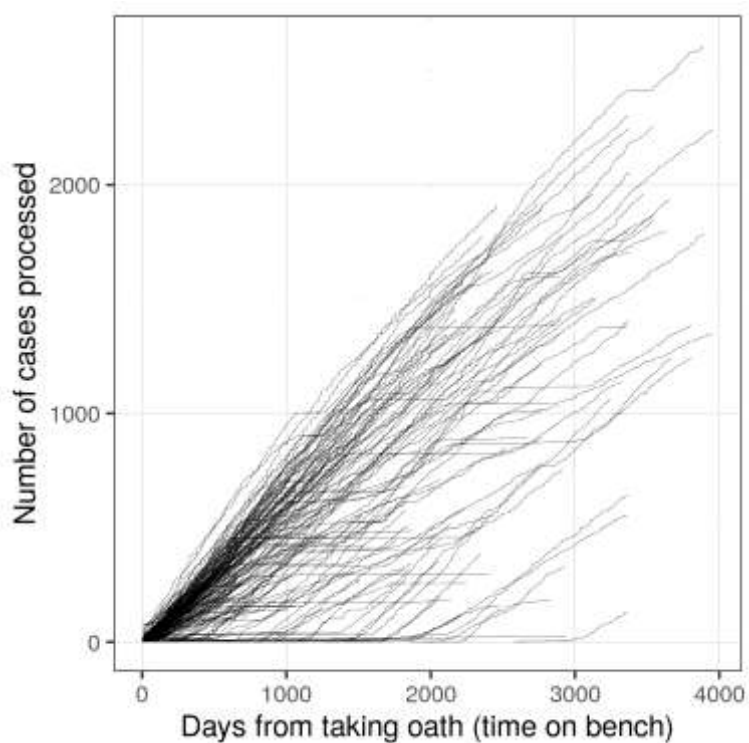
### *Measuring judicial experience*

It is important to elaborate on the distinction between issue-specific judicial experience (represented by the number of processed criminal cases) and general judicial experience (time on bench), as this represents a key contribution of our study. The main advantage of issue-specific judicial experience is shown in Figure A.1, which displays the relationship between time on bench and the number of processed cases for the judges captured in our sample across our window of observations.<sup>5</sup> Each line represents an individual judge; the figure reveals that many judges began deciding criminal cases years after taking their oath, and thus that their experience would not be accurately reflected by the time on bench measure. Time on bench would also not reflect the fact that some judges sentence more cases over a given period of time than others. There are various reasons behind these disparities in the frequency

<sup>5</sup> We only consider district court cases. Cases the judges may have been involved in when on secondment to higher courts are not counted.

of cases processed, such as work on civil or administrative cases (not criminal ones), career gaps as a result of parental leave, and so on. Using the total number of criminal cases tried overcomes both problems from these factors which mean that time on bench is not a representative measure of experience with criminal cases, although it is still not a perfect measure of experience. In particular, the number of cases processed places too much emphasis on the quantity of the experienced gained from processing cases, and fails to consider the quality of that experience. Certain judges might be given fewer, yet harder and longer-lasting cases or might specialize in a particular type of criminality, which might shape their decision-making differently.

Figure A.1: Experience measured as time on bench and as number of cases processed (Czech judges who took their oaths in 2007-2017)



NOTE: Each line captures the trajectory of an individual judge over time on bench.  $N = 99,514$ ,  $N_{judges} = 149$ .

Moreover, we are aware that our measure of issue-specific experience might be affected by missing judge identifiers. In Table A.2 we therefore disclose the number of cases in which we were not able to acquire the name of the judge deciding the case. We do not consider this missing data to be a substantial issue for our study as only a (very) small share of cases is missing. As we explain in the main text, there are several reasons behind this missing data, including that data from three district courts is missing for 2016 and 2017, likely due to a technical error.

Table A.2: Share of cases for which we lack information about the deciding judge, 2007-2017

<b>Year of decision</b>	<b>Total number of cases decided</b>	<b>% of cases with judge data missing</b>
2007	100,525	0.3
2008	98,250	3.3
2009	96,389	0.3
2010	90,253	0.3
2011	91,048	0.1
2012	91,678	0.1
2013	95,305	0.1
2014	91,979	0.2
2015	82,325	0.4
2016	75,811	1.2
2017	69,175	5.5

### *Analytical Appendix*

In this appendix we provide further details of our modelling strategy and additional robustness checks. Between-judge variability has often been ignored in sentencing research, due to the common absence of judge identifiers in official sentencing data (Pina-Sánchez et al., 2019). When accounted for, it is always assumed to be uniform across time, despite overwhelming evidence from life-course literature that it is safer to assume that individual differences do change across time (Fine & Cauffman, 2015; Penn & Silverstein, 2012). Again, we probably owe this assumption to the limitations of previously available sentencing data. Methodologically, failing to account for this variability affects a model's measures of uncertainty (Dhami & Belton, 2016; Johnson, 2006). Importantly, there are also substantive implications to be considered. If we ignore the longitudinal dimension in sentencing data, we cannot reliably estimate the true extent of between-judge disparities or examine how these change as judges gain experience.

We specify the unique longitudinal component available in our sentencing dataset using growth curve models. The various elements that make up a growth curve model can be visualized by considering a typical linear regression model where sentence length, denoted as  $Y$ , is regressed upon a list of  $k$  case characteristics,  $X_k$ :

$$Y_{ti} = \beta_0 + \beta_k X_{kti} + \epsilon_{ti}$$

with  $\beta_0$  representing the model's intercept, the average sentence length after controlling for each of the  $k$  case characteristics included in the model;  $\beta_k$  represents the effect of each of those case characteristics on sentence length; and  $\epsilon_{ti}$  represents the residual term, capturing differences in sentence length that cannot be explained by the case characteristics included in the model, which are assumed to be independent and normally distributed, with mean 0 and constant variance  $\epsilon_{ti} \sim N(0, \sigma_\epsilon^2)$ . If the model is well specified in the sense that most of the

legally relevant characteristics of the case are accounted for, then this residual term can be taken as an estimate of the extent of unwarranted sentencing disparities present in the system (Pina-Sánchez & Linacre, 2014). Notice as well that two subscripts  $t$  and  $i$  are included in each of the variables and the residual term. This is to denote that cases,  $t$ , are clustered within judges,  $i$ .

This standard model can be extended by including a covariate capturing the order in which the sentences were imposed by each judge, denoted here as  $exp$ . The regression coefficient associated with that covariate,  $\beta_1$ , can then be used to estimate the average association between  $exp$ , and  $Y$ . For example, if  $\beta_1 > 0$  and statistically significant, this indicates that judges increase the severity of their sentences as they become more experienced. To consider not just the average effect of experience, but also each judge's individual trajectory, the model can be further extended by including a random intercept term,  $u_0$ , and a random slopes term,  $u_1$ . These two random terms allow the intercept and the effect of experience to vary by judge as follows  $\beta_{0i} = \beta_0 + u_{0i}$  and  $\beta_{1i} = \beta_1 + u_{1i}$ , which is why they are now presented with their own subscript  $i$ . This linear growth curve model can be presented in the following form, which is similar to the random slopes models commonly used in sentencing research to examine disparities between judges or courts in the use of certain case characteristics (Anderson & Spohn, 2010; Johnson, 2005):

$$Y_{ti} = \beta_{0i} + \beta_{1i}exp_{ti} + \beta_k X_{kti} + \varepsilon_{ti}$$

The addition of these two random effects implies invoking further assumptions, namely that each of these judge-level residuals is distributed independently as a normal variable with mean 0 and a given variance,  $u_{0i}N(0, \sigma_{u0}^2)$ ,  $u_{1i}N(0, \sigma_{u1}^2)$ , and is independent from the case level residuals, represented as  $\varepsilon_{ti}$ . However, we do not assume that  $u_{0i}$  and  $u_{1i}$  are independent of each other; indeed, the covariance between the two random effects,  $\sigma_{u0,u1}$ ,

will be used to examine any potential changes in unobserved between-judge disparities as judges progress in their careers. For example, if  $\sigma_{u0,u1} > 0$  are statistically significant, this may indicate that between-judge disparities in sentence length increase as judges become more experienced.

The growth curve model represented above is of a linear type, as is the model we use to examine differences in the lengths of the custodial sentences imposed – we call this Model 2 (sentence length is log-transformed to adjust for the right-skewness observed in its distribution). For Model 1 and Models 3 and 4, where we explore the probabilities of incarceration, guilt adjudication and penal order procedures, we extend the model defined above using a binary logit growth curve specification. All our models are estimated using MLwiN (v. 3.0) and R (v. 3.6.3) connected by the package R2MLwiN (v. 0.8-6).

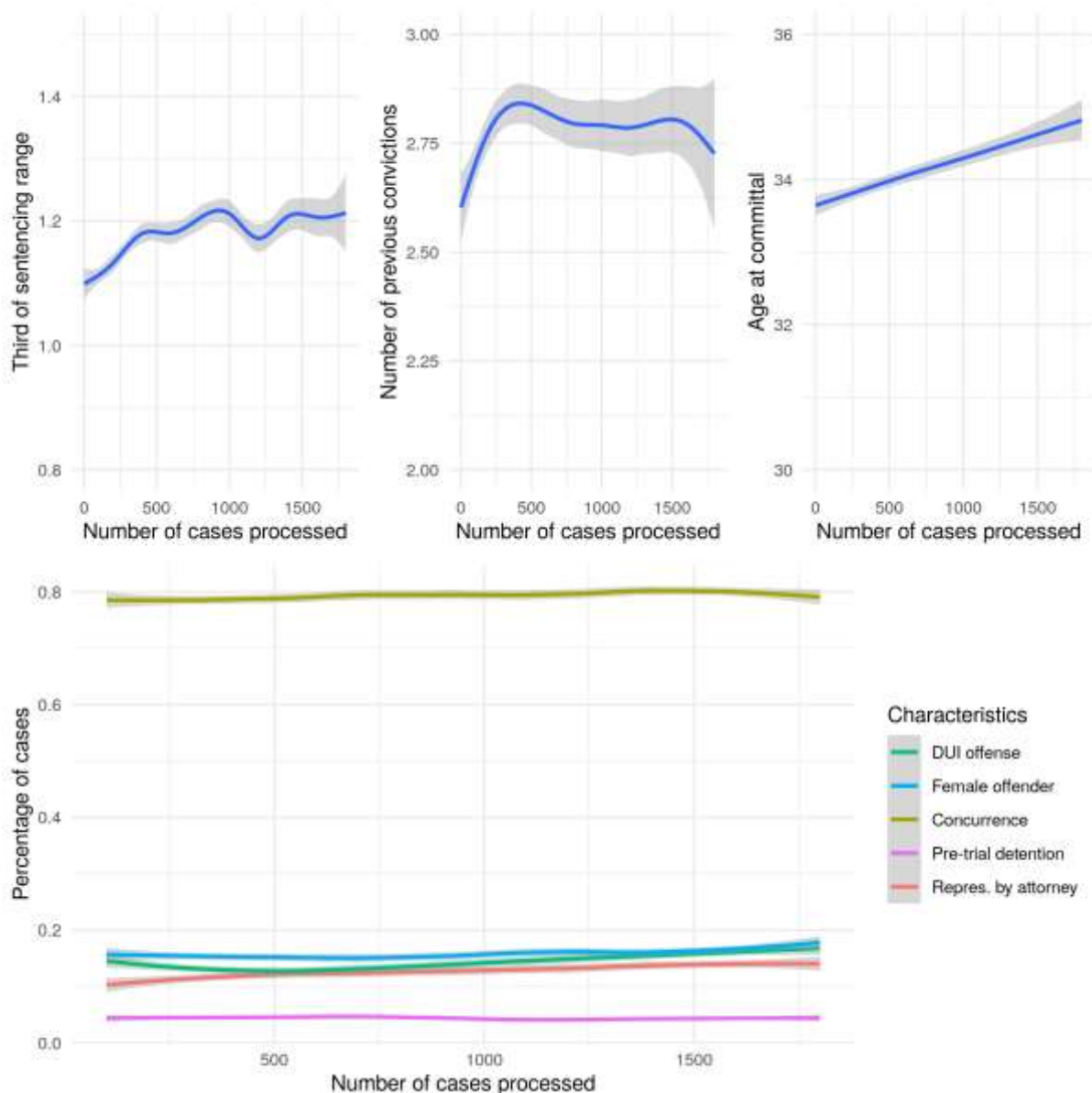
### *Robustness checks*

We have undertaken a variety of robustness checks to assess the sensitivity of our findings to a series of assumptions. We tested whether the number of cases processed is associated with changes in the Czech criminal justice system or with the cases judges are assigned, which might bias the effect that we have attributed to judicial experience. There is little evidence of such effects. The correlation between issue-specific experience (the number of cases processed by a given judge) and time (measured as the number of days elapsed from January 1<sup>st</sup> 2007, the start of the window of observation considered in our study) is relatively weak (Pearson's  $r$  of 0.36, Sample 1), as is the correlation between the number of cases processed and the age of the judge (Pearson's  $R$  of 0.37, Sample 1). Given that we observe that the judges in our sample joined the judiciary gradually, any trend would need to be stable across the entire examined period (there is no unusually large wave of judges entering the system at



any point, which could give rise to a linear effect starting from such a point). To control for such a possible trend, we also estimate models 1-4 substituting the yearly dummy variables with a continuous variable capturing the time since the onset of our window of observation. As we discussed in the Results section, this change did not affect the fixed or random estimates associated with experience. As Figure A.2 shows, our explanatory variables do not change substantially over time, which suggests an absence of changes in the Czech criminal justice system or in the assignment of cases correlated with judicial experience (the only exception is an increase in the average age of defendants, by one year over 1800 decisions; the defendant's age is one of the variables we control for in our models).

Figure A.2: Relationship between number of cases processed and case characteristics (Sample 1)



NOTE: The presented characteristics refer to average values across cases processed at specific points in judicial careers. The average value is calculated across cases decided by all judges at given numbers of cases processed. Non-continuous variables were rounded to hundreds of processed cases; the geom\_smooth function in R was used to capture the trend.

To test the quasi random allocation of cases to judges within the Czech criminal justice system, we employ a model in which the response variable is the judge decision number and the explanatory variables are all the case characteristics recorded in our dataset (sentencing range, previous convictions, offense type, offender age and sex, type of criminal code,

concurrency, pre-trial detention, manner of beginning procedure and representation by an attorney). The results are reported in Table A.3. We explicitly removed year dummies as year is inherently correlated with the judge decision number. While approximately half of the explanatory variables are statistically significant, there is no substantively significant pattern of case-composition across cases assigned to novice or experienced judges (e.g. fewer DUI offenses, more serious offenses, more pre-trial detentions). The standardized coefficients are very small with the largest (after the criminal code, which is time-dependent and inherently correlated with the judge decision number) being representation by an attorney pushing standard deviation by 3.3%; this is likely explained by the increasing wealth of the Czech population between 2007 and 2017. Hence, we can conclude, that cases are not assigned in a particular pattern across time, at least not based on the case characteristics that we were able to observe.

Table A.3: Results of model with judge decision number as response variable (Sample 1)

	Coeff.	Standardized coeff.	S.E.	
Intercept	0.238	0	0.007	***
<b>Sentence range</b>	0.006	0.013	0.002	**
<b>Previous convictions</b> (demeaned)	0.009	0.007	0.005	*
<b>Most serious offense in the case</b> (ref: Theft):				
Bodily harm	-0.003	-0.001	0.01	
Bodily harm (negligence)	0.119	0.017	0.022	***
Credit fraud	0.072	0.03	0.009	***
Damage to a thing	0.014	0.004	0.013	
Dangerous stalking	-0.016	-0.003	0.018	
Disorderly conduct	-0.009	-0.003	0.01	
Driving under influence	0.03	0.022	0.006	***
Drug production	-0.005	-0.002	0.01	
Embezzlement	0.035	0.012	0.01	***
Endangering the child's upbringing	0.036	0.008	0.015	*
Extortion	0.034	0.007	0.017	*
Fraud	0.05	0.021	0.008	***
Frustrating execution of an official order	0	-0.008	0.006	
Great bodily harm (negligence)	0.023	0.005	0.016	
Insurance fraud	0.121	0.017	0.024	***
Money laundering	0.015	0.002	0.02	
Non-payment of alimony	0.012	0.009	0.006	*

Other	0.036	0.021	0.007	***
Robbery	0.017	0.004	0.016	
Unauthorized possession of credit card	0.01	0.002	0.017	
Unauthorized use of another property	0.022	0.003	0.026	
Violence against a public official	0.007	0.001	0.02	
Violation of a domestic freedom	0.015	0.006	0.009	
<b>Age at offense date</b>	0	0.01	0	**
<b>Sex of offender</b> (ref. male)	0.003	0.002	0.004	
<b>Sentenced under old Penal Code</b>	0.392	0.296	0.004	***
<b>Concurrence</b> (ref. yes)	0.001	0.001	0.004	
<b>Pre-trial detention</b>	-0.023	-0.01	0.008	**
<b>Manner of beginning procedure</b> (ref. full):				
Other	0.148	0.009	0.051	**
Shortened	-0.011	-0.012	0.004	**
<b>Represented by attorney</b> (ref. no)	0.048	0.033	0.005	***

Note: \*: p < 0.05, \*\*: p < 0.01, \*\*\*: p < 0.001; R<sup>2</sup>: 0.088