Prosecutors, judges and sentencing disparities: Evidence from traffic offenses in France

Alessandro Melcarne, Benjamin Monnery, François-Charles Wolff

Prosecutors, judges and sentencing disparities: Evidence from traffic offenses in France[#]

Alessandro Melcarne[*] **Benjamin Monnery[**]** **François-Charles Wolff[***]**

May 2022

Paper accepted for publication at
International Review of Law and Economics

Published paper : https://doi.org/10.1016/j.irle.2022.106077

Abstract: While there is widespread evidence that sentences for similar cases tend to differ across courts, the production of sentencing disparities by prosecutors *versus* judges has received very limited attention to date. In this paper, we focus on this issue using traffic offenses data from neighboring courts in South-East France. First, we measure disparities for observably similar cases both at the extensive margin (type of sentences) and intensive margin (*quantum*) and find large differences in sentencing across courts. Second, we decompose those disparities between the influence of prosecutors through their procedural choices (simplified versus classical criminal procedures) and that of judges who always have the final word on sentences. While there is heterogeneity in the role of prosecutors between courts, we find that most sentencing disparities cannot be explained by the sole decisions of prosecutors.

Keywords : courts ; judicial disparities ; sentencing ; prosecutors ; mediation analysis

JEL codes : K14, K41, K42

[#] We are indebted to two anonymous reviewers and the editor, Eric Helland, for their very helpful comments and suggestions on a previous draft. We thank Yannick Joseph-Ratineau for providing access to the data and discussing institutional features of French courts. We also received useful feedback from participants at the Judicial Disparities Workshop (University Paris Nanterre) and the 2020 AFED annual conference. Financial support from ComUE University Paris Lumières is gratefully acknowledged.
[*] EconomiX (CNRS), University Paris Nanterre, alessandro.melcarne@parisnanterre.fr
[**] Corresponding author. EconomiX (CNRS), University Paris Nanterre, bmonnery@parisnanterre.fr
[***] LEMNA, University of Nantes and TEPP, francois.wolff@univ-nantes.fr

1. Introduction

Criminal justice officials face a trilemma in their daily work. They have to deliver equal justice for all citizens, tailor individualized decisions that fit the crime being judged, and process efficiently large caseloads of offenses. Given the equivocal and heterogeneous nature of such goals, there is room for judicial disparities both across and within courts in how cases are handled from offending to sentencing. Disparities can be explained by the preferences and constraints of two types of agents - prosecutors and judges - who essentially decide on: *i*) which criminal procedure to use and *ii*) which sentence to impose.

In France, as in most judicial systems, prosecutors are responsible for upstream decisions in criminal cases. They choose whether to prosecute or dismiss new incoming offenses, and then select one of several criminal procedures to handle them. Judges then receive cases and can either choose autonomously a sanction type and a *quantum*, or just validate the prosecutor's proposal. Thus, sentencing disparities might result from both the individual decisions and/or mutual interactions between prosecutors and judges. However, to date, most studies have analyzed the decisions of those actors separately. For example, many papers have estimated the impact of judges' characteristics like gender or ethnicity on sentencing, whereas other papers have focused on the impact of political affiliation as well as electoral incentives on prosecutors' judicial behavior (Epstein and Weinshall, 2021).

In this paper, we study the production of judicial disparities across courts located in South-East France using a sample of 3,400 cases of traffic offenses augmented with data on judges' characteristics. First, we measure sentencing disparities for observably similar cases across neighboring courts, providing the first econometric estimates on such differences in France to the best of our knowledge. Second, and more originally, we decompose those disparities between the influence of judges and that of prosecutors using a mediation analysis approach (Imai et al., 2010; Pearl, 2012). This approach allows us to capture the indirect impact of prosecutors on sentencing disparities across courts through their choices of criminal procedures. Several important features make traffic offending the most interesting type of crime to analyze for our purpose.

First, traffic offenses are massively widespread and represent 42% of all convictions in France, representing about 258,000 convictions in 2018. Second, traffic offenses receive a highly intertwined criminal treatment by prosecutors and judges, through the frequent use of "simplified criminal procedures" instead of classical procedures leading to a trial. These simplified procedures grant prosecutors the lead in sentencing, allowing them to make a sentence proposal that is later validated by judges. In practice, several prosecutors from the same court often work successively on the same case from arrival to conviction, limiting the role of prosecutor-level characteristics. Also, such simplified procedures limit judges' discretion by capping the maximum penalty or by excluding certain sentence types like prison. Thus, in such cases, disparities can originate from the interplay between two decisional sources, upstream (prosecutors) and downstream (judges).

Third, traffic offenses are very homogenous, unspecific, and often victimless, making case-heterogeneity less of a concern as a potential confounder[1]. In practice, French magistrates often make decisions relying on the same set of basic information as available in our dataset (socio-demographic information about offenders, type of offense, alcohol intake and criminal background). Fourth, the majority of traffic offenders are easily identified during police stops and guilt is often implied. As a consequence, investigation costs are usually close to zero and lead to a very high prosecution rate, which means a low sample selection of cases. Fifth, traffic offenses receive highly standardized judicial treatment using rules and guidelines that are often very explicit, although not made public. This usually entails a limited number of criminal procedures and sanctions. Yet, we observe significant and sizeable variations in decisions even for such standardized cases among neighboring courts.

According to our empirical analysis, sentencing disparities for observably similar cases are large from one court to the next, both at the extensive margin (type of sanction) and intensive margin (*quantum*). There are also large differences in the use of probation sentences as opposed to fines or in the amount of such fines across courts. These cross-court disparities prevail when we control for the characteristics of judges, in terms of gender and experience, and are robust to selection on unobservables. Then, we provide for the first time a decomposition of cross-court disparities between the role played by prosecutors choosing procedures (indirect effect) and the role played by judges making the final calls (direct effect) using a mediation analysis. There is substantial heterogeneity between courts. At the extensive margin, the indirect effect is negative and very low in three courts, but the court effect is never fully explained by the decisions made by prosecutors. We conclude that prosecutors have a rather limited independent impact on disparities when choosing between criminal procedures.

[1] As an example, a DUI (driving under the influence) with an alcohol intake of 0.9mg/L is very similar to any other DUI of 0.9mg/L.

The remainder of our paper is organized as follows. Section 2 reviews the existing literature on sentencing disparities along with their underlying explanations. Section 3 presents the institutional context of traffic offenses in France and Section 4 describes the dataset. In Section 5, we study differences in sentences between courts both at the extensive and intensive margins. In Section 6, we investigate the role of prosecutors when explaining the courts' disparities. Finally, Section 7 concludes with a discussion of our findings.

2. Literature review

The traditional model of legal formalism considers judges as some sort of robots that limit themselves to applying the law to the facts under their scrutiny without external influence. *Ceteris paribus*, this model predicts little if any disparity in sentencing across judges. However, thanks to improvements in data collection, a growing body of evidence on such disparities has emerged since the 1980s and 1990s. There is now widespread evidence that judicial disparities exist and are often significantly large, even after accounting for differences in case characteristics. As a consequence, legal scholars, political scientists and economists have gradually proposed more "realistic" models to explain such disparities (Epstein and Weinshall, 2021).

To date, existing literature has mainly focused on the comparison of the decisions made by individual judges. The behavioral model emphasizes the role of judges' social-background and personal attitudes (Rachlinksi et al., 2009; Heise, 2002; Bourreau-Dubois et al., 2020). Conversely, the attitudinal model supports the idea that judges implement their policy preferences in their decisional process (Epstein and Knight, 1997; Schauer, 2000; Fischman and Law, 2009; Fałkowski and Lewkowicz, 2021; de Castro, 2021). Since Posner (1993), judges are widely viewed also as economic agents who maximize some utility function based on their preferences and institutional constraints. According to this rational-choice model, factors related to judges' tastes, leisure or career concerns are expected to influence decisions and lead to sentencing disparities across judges (Cohen, 1991; Taha, 2004; Melcarne, 2017).

Other factors have been shown to affect judges' decisions like panel composition (Helland and Tabarrok, 2000), local economic conditions (Ichino et al., 2003), but also more mundane issues like sports results (Eren and Mocan, 2018) or even breakfast eating habits (Danziger et al., 2011). In the context of the United States where many judges are elected, the preferences of local voters and media coverage have also been shown to impact sentencing (Huber and Gordon, 2004; Berdejo and Yuchtman, 2013; Anwar et al., 2019). Such maximization processes

will be “constrained” by the overall set of institutional arrangements commonly known as judicial independence (Melcarne and Ramello, 2015) shielding judges from external incentives, but at the same time also granting them a certain degree of discretion over their decisions.
Prosecutors are also subject to similar interactions when they decide which cases to prosecute. They were shown to balance social welfare with other concerns like reputation, reelection, or private-sector job opportunities (Glaesar et al., 2000; Dyke, 2007; Rasmussen et al., 2009; Bandyopadhyay and McCannon, 2014; Nelson, 2014). What emerges from this literature is that judges and prosecutors do not simply respond as machines to the application of law, they are on the contrary affected by external factors. However, these studies are mostly concerned with explaining sentencing disparities across magistrates, and not among courts where judges and prosecutors interact. Even when interactions among magistrates are taken into consideration (Epstein et al., 2011; Berdejó and Chen, 2017), this is limited to the interplay of judges in multiple-justices deciding panels.
In this paper, we investigate the production of cross-court disparities by prosecutors and judges. Our work is most closely related to the stream of literature attempting to decompose judges and prosecutors’ role in sentencing. Examining 3,000 cases from three U.S. states, Kim et al. (2015) study the individual influence of judges and prosecutors on the length of prison sentences, as well as their interactions in judge-prosecutor dyads. They show that both influences exist and have a large impact on the severity of sentencing, with judge-prosecutor dyads playing a particularly large role. However, they also show that the influence of prosecutors and judges varies across local contexts, with significant effects in some courts and insignificant in others.
In the context of federal criminal cases in the United States, Rehavi and Starr (2014) show that, conditional on the arrest charge, prosecutors’ initial choice of court charge drives disparities in sentencing between black and white defendants. This occurs because different court charges carry different mandatory minimum sentences (if any), leading to sentencing disparities across cases and courts. Studying four southern U.S. states, Feigenberg and Miller (2021) also find large disparities in sentencing across neighboring courts. Defendants judged in a top-quartile county, in terms of punishment severity, are 2 to 4 times more likely to be incarcerated than comparable defendants in a bottom-quartile county. Also, the disparities are partly

explained by racial heterogeneity in the population and follow an inverted U-shape[2]. Such pattern is consistent with in-group bias where the presence of minorities increases voters' desire for harsher sentencing.

With the present work, we contribute to this existing literature with a focus on France. Compared to the United States, which attracted most scholarly interest, France is a civil law country where judges and prosecutors are appointed civil servants. Also, the law being applied to criminal offenses, procedures and sentences is the same across courts. Still, we observe large disparities in sentencing across neighboring courts for similar or fairly similar cases. We also contribute to the scarce literature examining the co-production of sentencing by prosecutors and judges.

3. Institutional context

Our analysis focuses on traffic offenses judged in French district, first-instance courts ("*Tribunaux de Grande Instance*"). This excludes the mildest traffic violations, like parking violations or excessive speed, which are handled by police courts. Traffic offenses are very widespread in the French population. In 2018, there were approximately 420,000 traffic-offense cases (17%) handled by courts over a total of 2,5 million criminal cases (Cocuau, 2021). 258,000 of these cases led to a conviction, corresponding to a proportion of 42% of all convictions. Traffic offenses usually include four broad categories: *i*) driving offenses like drunk-driving, drug-driving, speeding (40% of traffic-related convictions), *ii*) administrative offenses related to drivers' license or car insurance (39%), *iii*) stop-and-control offenses such as refusal to stop or comply (12%), and *iv*) involuntary injuries in car accidents (8%). Those figures have been stable over time (Chabanne and Timbart, 2017; Timbert and Minne, 2013). In our analysis, we use data on the two main categories: driving offenses and administrative offenses[3].

Criminal procedures

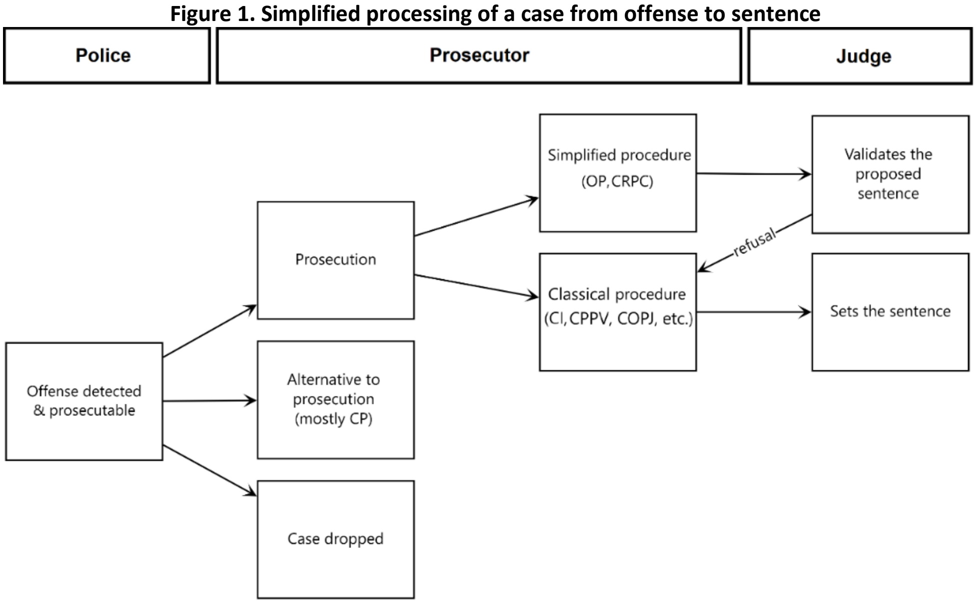
In France, criminal procedure rests on two key principles defining the roles of prosecutors and judges. First, prosecutors always make decisions about criminal procedures, a principle known as "*principe d'opportunité des poursuites*". Second, judges make decisions about conviction

[2] When local racial heterogeneity is low (mostly white or mostly black counties), punishment is relatively lenient. However, when there is large racial heterogeneity, elected prosecutors and judges exert harsher enforcement using more prison sentences.

[3] The two other types of traffic offenses are very rare in our sample.

and sentence, a principle known as "*principe d'individualisation des peines*". In case of a traffic-related offense, the criminal procedure follows a clear sequence of decisions by prosecutors and judges, from offending to sentencing, which is summarized in Figure 1. Usually, prosecutors make upstream decisions (whether to prosecute or not, which procedure to use) and judges make downstream decisions (conviction and sentence, or acquittal[4]). However, such principles stemming from classical criminal procedural law are somewhat distorted in the case of modern criminal procedures, in particular with the simplified procedures described below.

Figure 1. Simplified processing of a case from offense to sentence



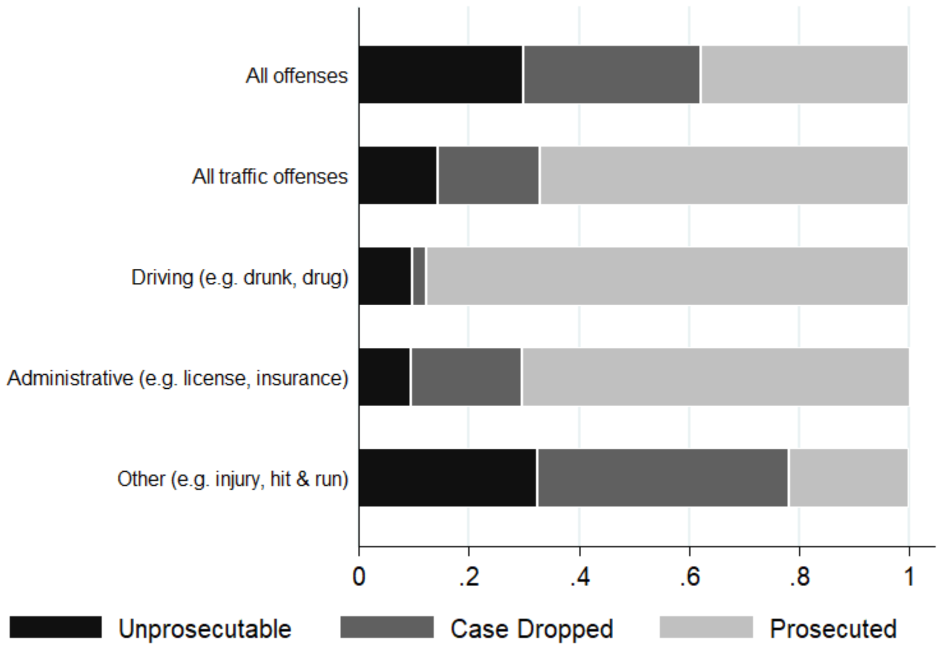
Source: figure from authors.

The first decision to make is whether the offense is prosecuted or not. Traffic offenses are characterized by a high rate of prosecution in France. In 2018, 86% of detected traffic offenses corresponded to simple cases sufficiently meritorious for prosecution (the offender was most often identified, under arrest and with sufficient evidence of guilt), so only 14% were deemed "non prosecutable" by the prosecutor[5]. For the types of traffic offenses included in our empirical analysis, only 10% of offenses could not be prosecuted and did not get processed by the judicial system. As shown in Figure 2, prosecutors effectively decided to prosecute (regardless of the procedure) in 78% of prosecutable traffic offenses, compared to less than 55% for all crimes (Cocuau, 2020). For driving offenses such as drunk-driving or drug-driving, the

[4] Acquittals are extremely rare in traffic-offense cases and represent 0.33% of our original dataset. These cases were dropped from the analysis.
[5] An offense is said "non prosecutable" in case of lack of evidence, absence of offense, or unidentified offender.

rate of prosecution reaches 97% of eligible cases. Such high prosecution rates suggest that the threat of sample selection is much lower in our context than when studying other types of crime where the prosecution stage already entails more discretion.

Figure 2. Distribution of cases in early stage in France



Source: Based on Cassiopée data from SDSE, Ministry of Justice (Cocuau, 2021)

The prosecutor's second decision is the choice between classical and simplified procedures. Historically, criminal procedures were highly focused on courtroom trials, where judges hear the defendant and have large discretion in choosing sentences. In 2000, 84% of traffic offenses leading to a conviction were adjudicated through a trial decision in the courtroom (classical procedure). In this setting, the prosecutor only recommends orally a sentence which is not binding. The judge can set sentences above or below this proposal, only limited by the maximum incurred penalty in the Criminal Code.

There are a variety of classical procedures which essentially differ in how fast the offender will appear in court and whether he risks bench warrant: the most stringent procedures are *CI* (trial on day of arrest) and *CPPV* (trial in the coming weeks), followed by more lenient procedures like *COPJ* and others[6]. However, such public courtroom hearings and deliberations are

[6] *CI* stands for immediate hearing ("*comparution immédiate*"). The offender is judged on the day of arrest or in the coming days, and often suffers a prison sentence with bench warrant. *CPPV* ("*convocation par procès verbal*") implies a trial in the next weeks or months (from 10 days to 6 months). *COPJ* ("*convocation par officier de police*

time-consuming and costly[7]. Given the importance of courts' timeliness for the overall quality of the justice sector (Melcarne et al., 2021), these procedures were progressively viewed as ineffective to handle the rapid increase in traffic offenses during the 2000s: 30% of cases' increase between 2000 and 2011 (Timbart and Minne, 2013). As a consequence, French legislators decided to introduce simpler and faster criminal procedures. There are currently three simplified procedures characterized by increasing stringency: penal composition ("composition pénale", *CP*), penal order ("ordonnance pénale", *OP*), and plea ("reconnaissance préalable de culpabilité", *CRPC*).

First, penal composition (*CP*) was introduced in 2001 as the most lenient simplified procedure. It is an alternative to prosecution, in the sense that offenders will avoid a formal criminal conviction, but still suffer a low-severity sanction such as a fine, revoking the driver's license or the obligation to follow awareness-raising courses. If the offense is of low gravity and the offender admits guilt, the prosecutor settles the sentence in agreement with the offender and the decision does not need to be validated by a judge[8]. If the offender refuses, then the prosecutor is most likely to press charges using a more stringent criminal procedure which may open the possibility of much more severe sentences (like probation or prison).

Second, penal order (*OP*) is a more stringent simplified procedure. It leads to a formal criminal conviction and opens the possibility for more serious sanctions (with the exception of prison). It is not applicable if the offender is a legal recidivist. Again, the prosecutor has the lead as he/she settles the sentence without any interaction with the offender, but in this case a judge has to validate his/her proposition (which is almost automatic in practice). In case of refusal by the judge or appeal by the defendant, the case goes through a classical procedure which is much more costly and time-consuming for judges. This simplified procedure was extended in 2003 to become applicable to much more cases (initially only the less serious offenses were concerned), among which all traffic offenses.

Third, plea (*CRPC*) was introduced in 2004. It is the most stringent of all simplified procedures. It entails criminal conviction and can lead to a prison sentence. Offenders must plead guilty on all charges and have to be defended by a lawyer during their private hearing with the

judiciaire") and other classical procedures ("*citation directe*", etc.) lead to a trial with longer delays, from months to years depending on the courts' docket.

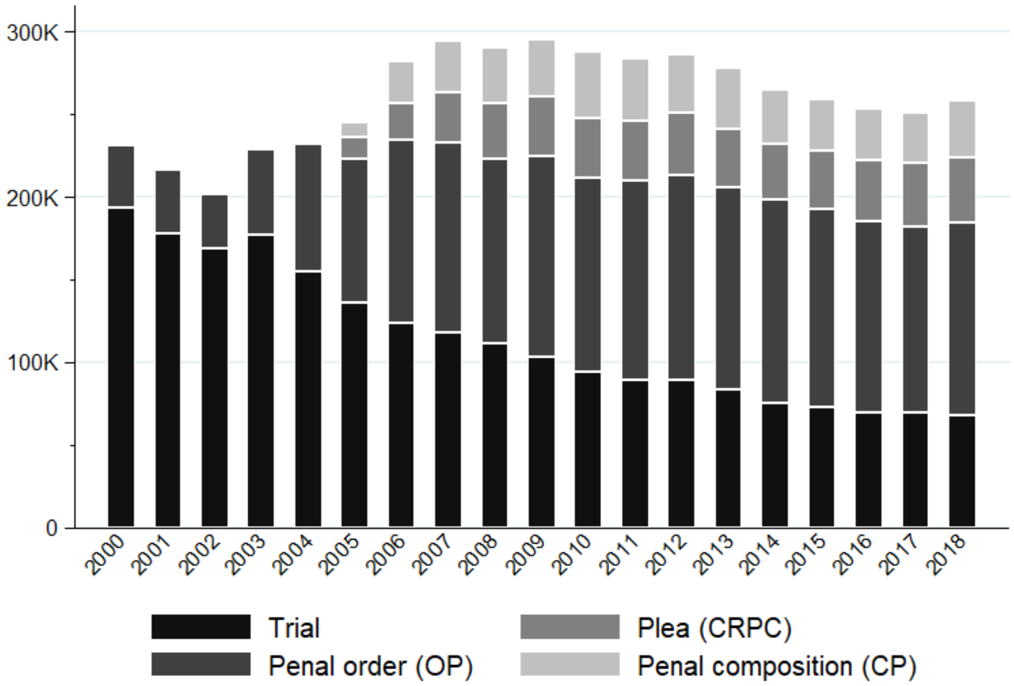
[7] Judges first have to review evidence on guilt. Defendants are often present at trial and defended by a lawyer.

[8] There are a few exceptions, corresponding to cases where the maximum incurred penalty is above 3-years of imprisonment or 3000 euros of fines.

prosecutor. The prosecutor then proposes a sentence which can go up to one-year of unsuspended prison (or even half of the incurred prison sentence if larger), with or without bench warrant. If both the prosecutor and the offender reach an agreement, the deal is proposed to a judge for validation (almost automatic in practice). If there is no agreement or the judge does not validate it, then the most likely outcome is again the opening of a time-consuming classical procedure leading to a trial in the following months.

Overall, the share of traffic-offense cases handled through classical procedures declined from 84% in 2000 to only 26% in 2018. Nowadays, as shown in Figure 3, about three quarters of the 250,000 annual convictions for traffic offending in France are dealt with simplified procedures. *OP* is the most common procedure (45% of all convictions in 2018), followed by *CRPC* (15%) and *CP* (13%).

Figure 3. Evolution of Criminal Procedures Leading to Traffic-Offense Convictions



Source: criminal records data from SDSE, Ministry of Justice (Cocuau, 2021)

The third stage of the judicial procedure concerns sentencing. French law offers a large set of sentence types ordered in three categories of increasing severity: fines and related monetary fines (day-fines), suspended prison with some form of probation, and prison sentences. While fines only entail a monetary payment to the State (often between 200 and 500 euros in traffic-offense cases), suspended prison with probation imposes supervision by judicial authorities, loss of personal autonomy and harsher sanctions in case of reoffending. As part of their probation, convicts have to comply with court-ordered requirements for a specific period (typically ranging from 12 to 36 months) such as working, treating alcohol or health-related problems, paying damages to victims, or not reoffending[9]. Finally, (unsuspended) prison sentence is the most severe type of punishment. It typically leads to imprisonment although it can also be converted by the court or later by a new judge in charge of the sentence execution into alternative sanctions such as electronic monitoring or semi-liberty, which also entail strong supervision and coercion from probation officers or prison officers (Henneguelle et al., 2016; Monnery et al., 2020).

With the increase in simplified procedures, prosecutors no longer focus only on the first decision of the judicial process (prosecute or drop charges). Instead, they now also play a leading role in sentencing through their procedural choice. By choosing whether and how to prosecute a case, they decide who has the lead in the sentencing process (either the prosecutor or the judge) and possibly which sentence (type and/or *quantum*) is incurred. For instance, no prison sentence can be issued if an *OP* procedure is chosen. In about 75% of cases involving traffic offending, prosecutors actually shape the content of sentencing decisions by making proposals that are almost always validated by judges[10]. More specifically, prosecutors first have large discretion in choosing whether a classical procedure is best-suited, or which simplified procedure to use, and then which sentence type and *quantum* to propose. The legal limitations in terms of eligibility and maximum sentences are quite lax and leave large discretion to prosecutors. Hence, prosecutors have legal leeway to impose their own views and strategies, based on their preferences, beliefs and external constraints (like inflows of cases, backlog of cases or crime trends).

Organization of Courts

In France, all judges and prosecutors are highly trained civil servants who graduated from the same national school after a very competitive entrance exam (the National School of Magistrature). They are appointed (and not elected) to courts and positions within them depending on their school rank (for the first appointment) and later depending on the annual openings and closings of positions by the Ministry of Justice. Turnover is very high since magistrates

[9] If those obligations are not met, probation officers and judges take further action and may eventually decide to revoke the suspended sentence and incarcerate.
[10] In our sample, we observed only 2 refusals over 3,000 eligible cases.

have a strong incentive to move to get promoted and are even required to move every five to seven years (often much less in practice). Also, prosecutors often switch to judging positions (and vice versa) over their career. Overall, magistrates in France form a very homogeneous and mobile body of civil servants. For these reasons, there is arguably little potential for sorting of magistrates on political grounds across courts in France. However, there tends to be sorting in terms of experience. For instance, more senior magistrates obtain positions in larger, more attractive courts.

To increase efficiency, many cases are processed successively by different prosecutors from the same prosecutorial office throughout the different steps of the criminal procedure. For instance, a first prosecutor will decide to prosecute through a plea procedure, then a second prosecutor will make later the hearing of the defendant with his lawyer. Hence, the decisions of prosecutors are collective and it is impossible in the data to single out one prosecutor who would be responsible for all the decisions throughout the handling of a traffic-offense case. Contrary to the US (Kim et al., 2015), the relationship between prosecutors and judges cannot be investigated through the prism of dyads. Conversely, only one judge is in charge of the sentencing decision and his or her characteristics may strongly influence the legal outcome.

In this institutional context, the interplay between the local team of prosecutors and individual judges could be considered as a non-cooperative game where prosecutors play first and judges second. By choosing among various different procedures, prosecutors restrict the set of options for judges. With simplified procedures, they even propose sentences that judges tend to accept almost automatically. One reason is that such validations are almost costless for judges: they are very fast to come to an end, judges essentially just sign a sheet of paper. Conversely, refusing the proposal implies that the case will later be prosecuted with a classical procedure and a formal trial, which represents extra-work for judges and the whole judicial system. Prosecutors can therefore exploit these incentives to influence sentencing, although they are formally never responsible for issuing sentences: judges always have the final word. From this perspective, the team of prosecutors may well exert strong influence on sentencing disparities.

However, a different model based on cooperation could emerge locally. If prosecutors and judges jointly consider the social welfare implications of their actions, they may prefer to settle on shared goals and rules. This could avoid conflicts between judges and prosecutors' decisions and the consequent extra-work and other costs associated with a failed simplified procedure. In such institutional arrangements, the influence of prosecutors on sentencing disparities may appear much smaller since both the choice of procedures and the final sentences are part of a larger agreement between local judges and prosecutors.

4. Data and descriptive statistics

We study differences in sentences across courts using an original dataset collected from seven courts located in the South-East of France. In each court, the database focuses on all traffic offenses that were prosecuted (not dismissed) during the six-month period from January 1st to June 30th, 2017. This includes offenses like driving under the influence of alcohol or drugs, driving license offenses, or offenses related to the lack of car insurance.

The seven sampled courts represent a small fraction of the 163 first-instance courts in France. Indeed, access to data is very difficult to obtain and most often not allowed. In our context, the data were collected in the courts by field experts (see Joseph-Ratineau, 2019). Data collection was very costly through manually collecting paper files and hand-coding decisions. As courts required confidentiality to allow data collection, we do not provide identifying information for each court or judge to maintain anonymity, but use such information in our empirical analyses. We use the labelling TGI for *"Tribunal de Grande Instance"* (from TGI1 to TGI7). Those courts belong to three neighboring courts of appeal districts labelled CA for "Cour d'Appel", from CA1 to CA3[11]. They are all located in the South-East part of France and sufficiently close from each other. The average distance between courts is 192 kilometers, dropping to 130 kilometers when excluding one court.

The sampled courts are diverse in terms of size and local characteristics. Table 1 reports some characteristics of the courts. Each court has jurisdiction over geographic areas covering between 21 and 363 cities (mostly villages) for a total population varying by a factor of almost 4 (from 271K to more than one million inhabitants). The largest court (TGI4) will later serve as the reference category in our empirical analyses. TGI4 is characterized by high unemployment and high crime rate. Courts with below average population tend to experience lower unemployment and lower crime rates. These differences in local contexts may partly explain judges and prosecutors' decisions when handling cases, although we observe too little courts to properly run statistical tests.

[11] TGI1, TGI2 and TGI5 are in CA1, TGI3, TGI6 and TGI7 are in CA2, and TGI4 is in CA3.

Table 1. Characteristics of the courts

Court	Number of cities	Population (2018)	Unemployment rate (in %, 2018)	Crime rate (x 100,000 people, 2017)
TGI1	79	284 467	8.40	2 179.87
TGI2	150	271 288	10.20	2 059.43
TGI3	282	733 423	10.98	3 710.00
TGI4	21	1 074 299	15.82	5 837.48
TGI5	139	339 705	11.30	2 118.90
TGI6	363	522 337	13.75	2 364.18
TGI7	105	281 793	11.40	1 939.37
Average	163	501 045	11.69	2 887.03

Source : data collected by authors from INSEE and Ministry of Interior.

The dataset includes three main types of variables. First, we have some offender-level characteristics: gender, age (in categories), occupational status and existence of past convictions which trigger legal recidivism. Second, we know the type of offense committed and the alcohol intake if any (usually controlled during police stops). Third, we know the type of procedure chosen by the prosecutor as well as the type of sentence and the associated *quantum*. Overall, the original sample includes 4,223 offenses. From this sample, we exclude four types of offenses that are very rare and do not show up in all seven courts[12]. We focus on the six following types of offenses: drunk-driving with low alcohol intake between 0.4 and 0.8 mg per liter of exhaled air, drunk-driving with high alcohol intake (≥ 0.8 mg), driving under the influence of narcotics, default of car insurance, default of driver's license, and multiple offenses. We exclude the few observations without any information on the procedural choice made by prosecutors (N=13) and exclude offenders who were not convicted (N=246)[13].

We end up with a sample of 3,885 offenders judged in 7 courts: 314 in TGI1 (8.2%), 333 in TGI2 (8.6%), 243 in TGI3 (6.3%), 1,588 in TGI4 (41.2%), 587 in TGI5 (15.2%), 588 in TGI6 (15.3%) and 202 in TGI7 (5.2%). When explaining court disparities, judges may have an influence on the severity of the sentence. As there were the initials of the first name and last name of each judge in the database, we decided to collect additional data on judges using public records from the *Journal Officiel de la République Française.* This includes information on gender, date of birth, year of entry as judge and year of entry in the current court, from which we deduce

[12] We exclude cases of over drunk driving (N=24), drunk driving with very low alcohol intakes (below 0.4 mg per liter of exhaled air) (N=31), hit and run (N=46), and unknown offenses (N=8).

[13] We exclude diversion measures such as penal composition (CP) since they are observed only for a subset of courts.

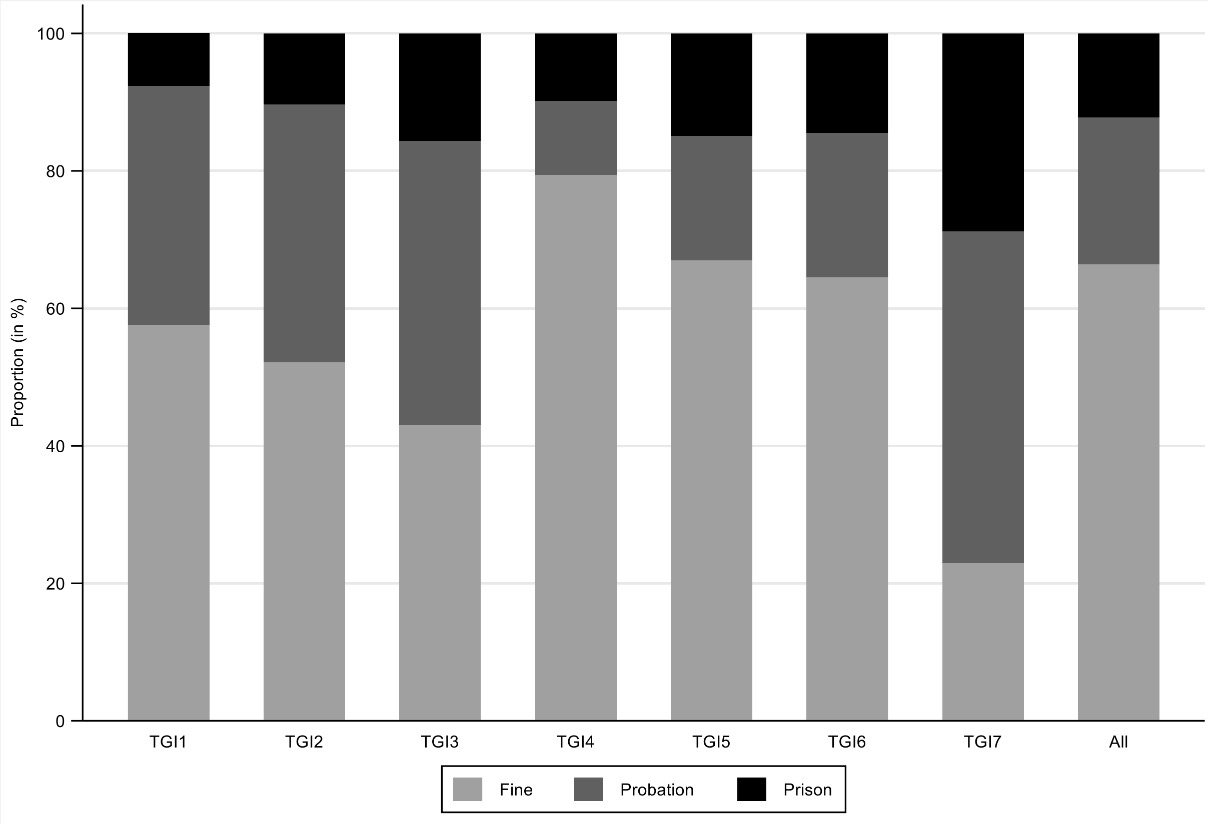
age, years of experience and court-specific years of experience. We were able to obtain those additional characteristics on judges for 3,395 offenses, corresponding to a matching rate of 88.1%. The contributions of courts to the final sample are 9.5% (TGI1), 9.4% (TGI2), 7.0% (TGI3), 38.8% (TGI4), 16.6% (TGI5), 17.1% (TGI6) and 2.6% (TGI7).

The outcome under consideration is the type of sentence. We sort the observed sentences in the sample into three categories ordered in terms of severity: fines, suspended prison under probation, and unsuspended prison[14]. Figure 4 shows the distribution of sentences across the seven courts. Fines (66.4%) are much more common than probation (21.4%) and prison (12.2%) for traffic offenses. The results show substantial differences between courts. While fines are predominant in all courts, the corresponding proportion ranges from 23.0% (TGI7) to 79.4% (TGI4). The proportion of probation ranges between 34% and 42% for TGI1, TGI2 and TGI3, but it is equal to 10.8% in TGI4. Finally, prison sentences exceed 14% in TGI3, TGI5, TGI6 and TGI7 while they are less frequent in TGI4 (9.8%).

A first explanation of these raw differences in sentencing can be due to dockets' disparities, *i.e.* differences in the types of offenses that are judged among courts. If some courts have to judge more serious offences, they are likely to be characterized by more severe sentences on average. As shown in the Appendix in Figures A1 and A2, the distribution of offense types differs markedly across courts, with different offenses receiving different types of sentences. For example, fines are used in 98.9% of car insurance cases, but in only 50.3% of cases for drunk driving with high alcohol intake.

[14] In case of multiple sentences, for instance a combination of fine and probation, we consider the most severe sentence (probation in that case).

Figure 4. Distribution of sentence types across courts



Source: data from seven courts in the South-East of France, authors' calculations.
Note: courts are anonymized (from TGI1 to TGI7) due to confidentiality issues.

These systematic differences mechanically generate disparities in the distribution of sentences observed among courts. To account for those compositional effects, we consider a normalization procedure to measure sentencing disparities net of the effect of differences in offenses. By reweighting observations to obtain caseload structures that are similar across courts in terms of offense types, we find that cross-court disparities in sentencing are substantially reduced after this correction (see Table A1 in the appendix). In TGI3 for example, the gap in the use of fines (43.0% of cases compared to a mean of 66.4% across all courts) is reduced by about one-half when we adjust for the structure of offenses (up to an adjusted fraction of 50.9% of fines).

Disparities in sentencing can also be explained by the different profiles of offenders across courts' dockets. For instance, judges may account for the economic situation of offenders as well as their criminal background when setting the sentence. In Table 2, we present some descriptive statistics by courts for the following defendants' characteristics: gender, age, occupational status, and recidivism status. Almost all offenders are male (more than 90% in all courts except TGI7) and around 7 out of 10 are less than 40 years old. The proportion of offenders having a job ranges from around 70% in TGI1, TGI2 and TGI5, 56.5% in TGI3, but only 39.4% in TGI4 (the proportion of undocumented situations being higher in that court). Finally, more than 80% of offenders have already been convicted in TGI3, while the proportion of first-time offenders is much larger in TGI4 (74.6%) and TGI6 (74.9%).

The type of sentence can also be influenced by the characteristics of the judges. In Table 2, we report the average values of the selected covariates calculated from the sample of offenses[15]. Overall, the proportion of cases judged by women is 40.1%, the average age of judges is nearly 50 and the average experience is 18 years. Again, there are large differences between courts. While more than 90% of decisions involve female judges in TGI6 and TGI7, this proportion is only 13.8% in TGI2 and even 8.2% in TGI5. Also, judges are substantially older in TGI3, TGI2 and TGI6 than in TGI1 and TGI7. The most experienced judges are found in TGI4, TGI3 and TGI2 and the highest court-specific experience is observed in TGI3. In what follows, we study whether there remains any difference between courts once both the offenders and judges' characteristics are controlled for in the regressions.

Table 2. Descriptive statistics of the sample, by courts

Variables	TGI1	TGI2	TGI3	TGI4	TGI5	TGI6	TGI7	All
Offenders' characteristics								
Gender : male	0.899	0.928	0.937	0.953	0.931	0.912	0.874	0.932
Age : ≤25	0.139	0.219	0.224	0.241	0.172	0.239	0.172	0.215
Age : 26-30	0.212	0.216	0.228	0.233	0.188	0.205	0.115	0.214
Age : 31-40	0.330	0.244	0.215	0.271	0.316	0.274	0.230	0.276
Age : 41-50	0.163	0.153	0.186	0.143	0.151	0.146	0.287	0.154
Age : >50	0.156	0.169	0.148	0.112	0.168	0.131	0.195	0.138
Occupation : Unemployed	0.167	0.219	0.257	0.277	0.151	0.256	0.264	0.236
Occupation : Employed	0.743	0.691	0.565	0.394	0.715	0.656	0.644	0.568
Occupation : Inactive (student, retiree)	0.056	0.056	0.068	0.059	0.050	0.055	0.023	0.056
Occupation : no information	0.035	0.034	0.110	0.270	0.085	0.033	0.069	0.140
Legal recidivism	0.229	0.125	0.456	0.080	0.183	0.179	0.391	0.165
Repeat offender	0.139	0.400	0.388	0.175	0.337	0.072	0.356	0.222
No recidivism	0.632	0.475	0.156	0.746	0.480	0.749	0.253	0.613
Number of offenders								
Judges' characteristics								
Gender : female	0.531	0.138	0.608	0.247	0.082	0.981	0.908	0.401
Age (average)	40.97	53.35	57.21	49.28	48.41	52.96	37.31	49.69
Years of experience as judge	15.48	19.17	19.46	24.06	8.57	15.70	11.75	18.23
Court-specific years of experience	2.99	6.29	9.61	5.67	4.53	1.78	2.43	4.83
Number of observations	288	320	237	1,318	564	581	87	3,395

Source: data from seven courts in the South-East of France, authors' calculations.
Note: the courts are anonymized (from TGI1 to TGI7) due to confidentiality issues.

[15]The minimal number of judges is 4 in TGI6 and up to 12 in TGI2.

5. Differences in sentences between courts

Extensive margin

Since we have ordered sentences on a scale of increasing severity with three categories (monetary fines, probation, and prison sentences), we turn to ordered choice models to assess differences among courts net of composition effects (Greene and Hensher, 2010)[16]. We estimate ordered Probit regressions with standard errors clustered at the court level. The corresponding estimates are presented in Table 3. We start in column 1 with only court dummies as covariates. Without any control, we find a higher severity of sentences in all courts compared to the reference TGI4 (the largest court). TGI7 is ranked first, followed by TGI3 and TGI2. A Wald test shows that the whole set of court effects is significantly different from 0, with a chi-squared statistic equal to 898.4 (p=0.000).

We introduce the type of offense as additional controls in column 2. Accounting for the pattern of offenses reduces substantially the magnitude of the court fixed effects (it is three times lower for TGI6). The variance of the court fixed effects is reduced by around one-half when adding the pattern of offenses, from 0.0828 to 0.0397. Again, the null assumption of court effects being equal to 0 is rejected (p=0.000). Controlling for offenses does not modify the order of the courts with the largest fixed effects in terms of severity (TGI7, followed by TGI3 and TGI2). Results for the various types of offenses are in line with expectations. Compared to drunk driving with a low blood-alcohol level, sentences are more severe for drunk driving with high blood-alcohol level, driving without license and multiple offenses. The reverse pattern is found for default of car insurance.

As shown in column 3, adding the offender's characteristics changes the coefficients associated with the courts, but not for all of them[17]. While the court fixed effects remain rather stable for TGI1, TGI2 and TGI7, they are no longer significant for TGI5 and significant at the 10 percent level for TGI3. Still, the null assumption of no court effect is strongly rejected (p=0.000). According to the data, the severity of sentences is not influenced by offenders' gender or age. There is a positive correlation between sentence severity and unemployment

[16] For the sake of robustness, we have also estimated ordered Probit models with a finer ordering of severity. Specifically, we consider the six following levels presented by increasing order: fines (66.4%), awareness-raising courses at the charge of the offender (3.1%), suspended prison without probation (9.3%), community service or probation (9.0%), mix of suspended prison (with or without probation) and unsuspended prison (2.0%), and unsuspended prison (10.2%). The corresponding estimates, not reported, are very similar.

[17] Further adding individual characteristics reduces by around 18% the variance compared to the case with offenses as covariates (from 0.0432 to 0.0353).

status (at the 10 percent level), and also for cases lacking information on the defendant's occupation. Finally, sentences are much more severe among repeat offenders and especially in case of legal recidivism (same offense type committed within five years).

Table 3. Ordered Probit estimates of sentences (extensive margin)

Variables	(1)		(2)		(3)		(4)	
	coef.	t-test	coef.	t-test	coef.	t-test	coef.	t-test
Court (ref : TGI4)								
TGI1	0.441***	(62.88)	0.206***	(4.80)	0.206***	(4.49)	0.302***	(4.99)
TGI2	0.568***	(47.44)	0.378***	(7.76)	0.352***	(9.66)	0.592***	(9.60)
TGI3	0.775***	(37.56)	0.535***	(14.89)	-0.118*	(-1.93)	-0.151	(-0.95)
TGI5	0.362***	(40.48)	0.176***	(3.49)	0.017	(0.66)	0.642***	(4.95)
TGI6	0.404***	(42.39)	0.137***	(4.57)	0.374***	(6.60)	0.398***	(4.77)
TGI7	1.211***	(29.97)	0.929***	(15.86)	0.604***	(14.19)	0.849***	(12.18)
Offense (ref : DUI with low intake)								
DUI with high intake			0.480**	(2.33)	0.499***	(2.59)	0.480**	(2.54)
Driving under narcotics			0.252	(1.13)	0.374*	(1.82)	0.335*	(1.72)
Default of car insurance			-1.541***	(-6.61)	-1.334***	(-8.51)	-1.298***	(-6.85)
Default of driver's license			0.490**	(2.09)	0.392**	(2.14)	0.363**	(2.10)
Multiple offenses			0.834***	(4.68)	0.728***	(4.85)	0.683***	(4.62)
Offenders' characteristics								
Gender : male					0.115	(1.24)	0.126**	(2.01)
Age : 26-30 (ref : ≤25)					-0.059	(-1.04)	-0.069	(-1.26)
Age : 31-40					0.014	(0.14)	-0.002	(-0.02)
Age : 41-50					0.071	(0.70)	0.031	(0.36)
Age : >50					0.136	(1.34)	0.113	(1.22)
Unemployed (ref : employed)					0.366*	(1.88)	0.369*	(1.82)
Inactive (student, retiree)					0.030	(0.30)	-0.013	(-0.14)
Occupation : no information					0.375***	(2.94)	0.308**	(2.53)
Legal recidivism (ref: no recidivism)					1.837***	(10.54)	1.720***	(10.42)
Repeat offender (ref: no recidivism)					1.215***	(6.14)	1.143***	(5.72)
Judges' characteristics								
Gender : female							0.531***	(2.60)
Years of experience as judge							0.031***	(4.13)
Court-specific years of experience							0.033	(1.60)
μ_1	0.753***	(25.29)	0.838***	(4.53)	1.739***	(7.05)	2.783***	(7.69)
μ_2	1.522***	(14.68)	1.711***	(8.81)	2.883***	(11.66)	3.960***	(12.71)
Observations	3,395		3,395		3,395		3,395	
Log likelihood	-2,818.5		-2,491.0		-2,022.8		-1,968.3	

Source: data from seven courts in the South-East of France, authors' calculations.
Note: the courts are anonymized (from TGI1 to TGI7) due to confidentiality issues. Estimates are obtained from ordered Probit models, with standard errors clustered at the court level. Significance levels are 1% (***), 5% (**) and 10% (*).

In column 4, we further introduce the characteristics of judges as covariates. This has a noticeable impact on the magnitude of some court fixed effects which tend to increase, in particular for TGI5 (from 0.017 to 0.642) and TGI2 (from 0.352 to 0.592). Overall, we end up with three different groups of courts in terms of severity of sentences. The more severe group includes TGI2, TGI5, and TGI7, the intermediate group includes TGI1 and TGI6, and the less severe group includes TGI3 and TGI4[18]. Interestingly, this last group corresponds to the two

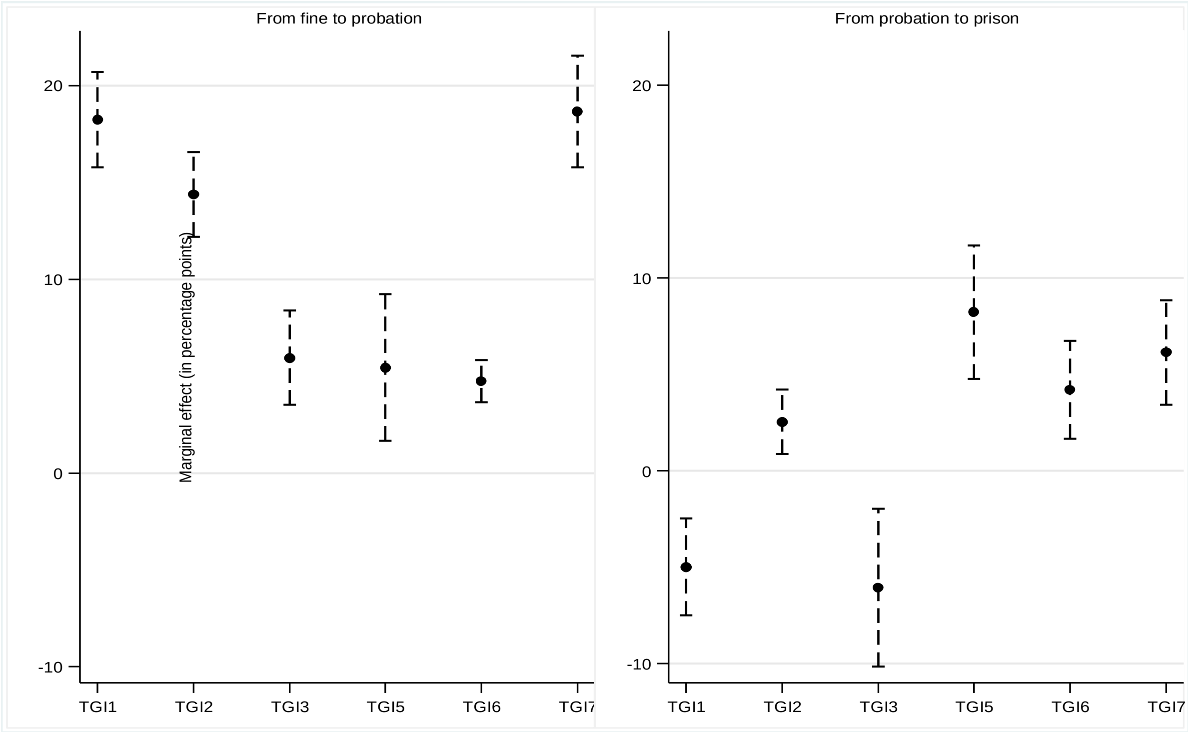
[18] However, there is no clear link between those groups and the different courts of appeal district.

largest cities of the sample, which are also those with the highest crime rate. Furthermore, results in column (4) show that characteristics of judges have an influence on judiciary outcomes. The sentence is more severe when the judge is a woman. Also, more experienced judges tend to be more severe, while years of experience in the current court does not significantly affect the outcome.

A limitation of the standard ordered model is that the cutoff values (that trigger shifting from one sentence type to the next) are the same for each offender. This means that there is no heterogeneity in the set of thresholds among courts. However, this assumption may turn invalid if each court uses its own threshold of case severity to decide between a fine and probation or between probation and prison: some courts may be eager to use probation instead of monetary sanctions, but hesitant to use prison for example, yielding a mixed picture overall in terms of sentence severity. At a more detailed level, each court may also adapt the thresholds to each type of offense, depending on local circumstances. Terza (1985), Groot and van den Brink (1999), and Boes and Winkelmann (2006) have considered some generalizations of ordered models in order to account for such threshold heterogeneity. In what follows, we relax the assumption of homogeneous cutoffs and estimate generalized ordered Probit models, in which cutoff values are allowed to differ both across courts and by offense types.

We present in Figure 5 the effect of court dummies along with confidence intervals at the 95% level obtained from the generalized ordered Probit model. The left panel shows cross-court disparities in the use of probation rather than fines and the right panel displays cross-court disparities in the use of prison as opposed to probation. Relative to the reference court TGI4, all courts display an inclination for probation over fines. The disparities are very large in TGI1, TGI2 and TGI7: defendants face a 15-18 percentage points (pp) higher probability of facing probation compared to reference court TGI4. Results are more mixed in terms of the trade-off between probation and prison sentences. Two estimates are negative (TGI1 and TGI3) and four are positive (TGI2, TGI5, TGI6, TGI7). Still, disparities are large since offenders in TGI5 face an 8.8 pp higher risk of prison instead of probation compared to if they were judged in TGI4. The gap in relative risk of incarceration is nearly 14 pp when comparing the two most extreme courts (TGI5 and TGI1, respectively).

Figure 5. Marginal effects of courts from generalized ordered Probit regressions



Source: data from seven courts in the South-East of France, authors' calculations.
Note: the courts are anonymized (from TGI1 to TGI7) due to confidentiality issues. Estimates are obtained from a generalized ordered Probit model, with standard errors clustered at the court level. The thresholds are allowed to vary according to both court and type of offenses. Characteristics of offenders and judges are also included as covariates, but do not vary according to the thresholds.

Another concern is the possibility that offenders receive multiple sentences. By construction, it is neglected in our ordered specification since we account only for the most severe sentence. Starting from the six types of sentences (fines, awareness-raising courses, suspended prison, probation, or community service, mix of suspended and unsuspended prison, unsuspended prison), we find that the proportion of multiples sentences is 8.4% (284 cases). The most frequent multiple-sentences case is a combination of probation and fine (112 cases, 39.4%), followed by a combination of unsuspended prison and fine (67 cases, 23.6%). Starting from the 11 different combinations (either single sentence or multiple sentences), we have also estimated a multinomial Logit model with the fine-only sentence as base outcome[19]. The main result is that for all outcomes we always find significant differences between courts. For instance, the probation-fine sentence is more likely in TGI1, TGI2, TGI5 and TGI7 compared to TGI4, while the court effects of TGI3 and TGI6 are not significant.

[19] The detailed MNL Logit estimates are available upon request.

Intensive margin

Next, we investigate to what extent *quantums* vary across courts. The different *quantums* investigated are the amount of fines in euros, the duration of suspended prison sentences in months, the duration of suspended prison with probation in months, and the duration of prison in months. Compared to our analysis at the extensive margin, we decompose the "probation" category between suspended prison (measured in months of prison incurred in case of reoffending) and suspended prison with probation (measured in months of probation to accomplish). For each *quantum*, we estimate linear regressions explaining the logarithm of each type of sentence. The corresponding OLS estimates with standard errors clustered at the court level are reported in Table 4.

Table 4. OLS estimates of sentences (intensive margin)

Variables	(1) Fines		(2) Suspended prison		(3) Suspended prison with probation		(4) Prison	
Court (ref : TGI4)								
TGI1	-0.422***	(-8.19)	-0.240**	(-3.18)	0.097**	(2.96)	0.099**	(2.75)
TGI2	-0.184***	(-6.93)	-0.551***	(-14.48)	0.179***	(9.00)	-0.348***	(-5.33)
TGI3	-0.213**	(-2.53)	0.084	(0.53)	0.045	(1.53)	0.280***	(6.00)
TGI5	-0.037	(-0.52)	-0.003	(-0.05)	0.147***	(5.01)	0.304**	(3.19)
TGI6	-0.319***	(-6.32)	-0.110**	(-2.06)	0.115**	(3.17)	0.430***	(6.52)
TGI7	-0.737***	(-24.01)	0.055	(0.64)	0.068	(1.49)	0.227***	(2.82)
Offense (ref : DUI with low intake)								
DUI with high intake	0.221***	(4.96)	0.182	(1.77)	0.031	(1.03)	0.191**	(2.16)
Driving under narcotics	0.147*	(1.94)	0.183**	(2.33)	0.075***	(4.73)	-0.058	(-0.68)
Default of car insurance	0.080	(0.88)	0.222	(1.55)				
Default of driver's license	0.351***	(4.22)	0.104	(1.52)	0.007	(0.20)	0.150**	(2.08)
Multiple offenses	0.525***	(7.20)	0.271**	(2.23)	0.052**	(2.29)	0.251**	(3.56)
Offenders' characteristics								
Gender : male	0.037	(0.98)	-0.016	(-0.78)	-0.028**	(-2.33)	0.169	(0.70)
Age : 26-30 (ref : ≤25)	0.016	(0.94)	-0.106	(-0.93)	0.010	(0.28)	0.075	(1.32)
Age : 31-40	0.037	(1.40)	-0.068	(-0.64)	0.031	(1.26)	0.228**	(2.10)
Age : 41-50	0.004	(0.19)	-0.175	(-1.44)	0.036	(0.78)	0.187**	(2.49)
Age : >50	0.010	(0.38)	-0.099	(-1.05)	0.037	(1.25)	0.100	(1.52)
Unemployed (ref : employed)	-0.031	(-0.90)	0.040	(0.63)	-0.000	(-0.00)	0.170*	(2.17)
Inactive (student, retiree)	-0.119***	(-4.99)	-0.036	(-0.30)	-0.042	(-1.36)	0.121	(0.79)
Occupation : no information	0.069	(1.61)	-0.085	(-0.99)	0.006	(0.28)	0.061	(1.20)
Legal recidivism (ref : no recidivism)	0.093	(0.38)	0.210***	(3.08)	0.019	(0.59)	0.399***	(3.55)
Repeat offender (ref: no recidivism)	0.027	(0.26)	0.089	(1.09)	0.015	(0.44)	0.174**	(2.09)
Judges' characteristics								
Gender : female	-0.002	(-0.02)	-0.106	(-0.69)	0.079**	(2.17)	-0.072	(-0.90)
Years of experience as judge	0.002	(0.38)	0.007	(1.44)	-0.001	(-0.64)	0.002	(0.28)
Court-specific years of experience	-0.017	(-1.01)	-0.021	(-0.95)	0.007	(1.60)	0.008	(0.81)
Constant	5.878***	(43.90)	0.768***	(5.76)	2.891***	(48.85)	0.155	(0.53)
Observations	2,268		314		300		346	
R^2	0.184		0.215		0.137		0.229	

Source: data from seven courts in the South-East of France, authors' calculations.
Note: the courts are anonymized (from TGI1 to TGI7) due to confidentiality issues. Estimates are obtained from OLS regressions, with standard errors clustered at the court level. Significance levels are 1% (***), 5% (**) and 10% (*).

Again, there is substantial heterogeneity among courts, but the ranking of courts depends on the type of sentences. We begin with fines in column 1. The average fine is 476.8 euros, with a standard deviation of 859.1 euros. Net of the influence of offenders' and judges' characteristics, the lowest averages are found in TGI7 and TGI1 while there is no difference between the other courts. Fines are strongly influenced by offenses and are much higher when driving without any valid license and when multiple offenses are reported. Conversely, both offenders' and judges' characteristics play almost no role. In particular, contrary to what was observed for sentence severity, recidivism status has no influence on the average fine.

When considering suspended prison sentences, two groups of courts emerge. The average sentence is much lower in TGI1 and TGI2 compared to the other courts. The sentence is more severe in case of driving under narcotics, multiple offenses and for recidivist offenders. Compared to TGI4, the duration of suspended prison with probation is higher in four courts out of six (TGI1, TGI2, TGI5, TGI6). Only two explanatory variables influence the duration of the sentence, with a reduced duration for male offender and an increased duration when the judge is a woman. Finally, compared to TGI4, the duration of prison is much lower in TGI2 and much higher in both TGI3, TGI5, TGI6 and to a lesser extent TGI7. Prison duration is positively correlated with age (for those above 40 and then less than 50) and legal recidivism, while none of the judges' characteristics play a role.

Robustness to unobserved heterogeneity

We now investigate whether differences in severity between courts could be due to unobserved heterogeneity. This would occur in particular if there are unobserved offender characteristics in our data (like place of birth) influencing the judiciary outcome, with different averages between courts. Over the last years, a few papers have suggested using the observables to identify the bias which can stem from the unobservables (Altonji et al., 2005, 2008, Krauth, 2016, Oster, 2019). In the case of one endogenous regressor and a continuous outcome, both Krauth (2016) and Oster (2019) propose some sensitivity analysis to calculate bias-adjusted treatment effects. The impact of the endogenous regressor is calculated for various proportions of selection on observables and unobservables. We rely on such methods and proceed in the following way with our data.

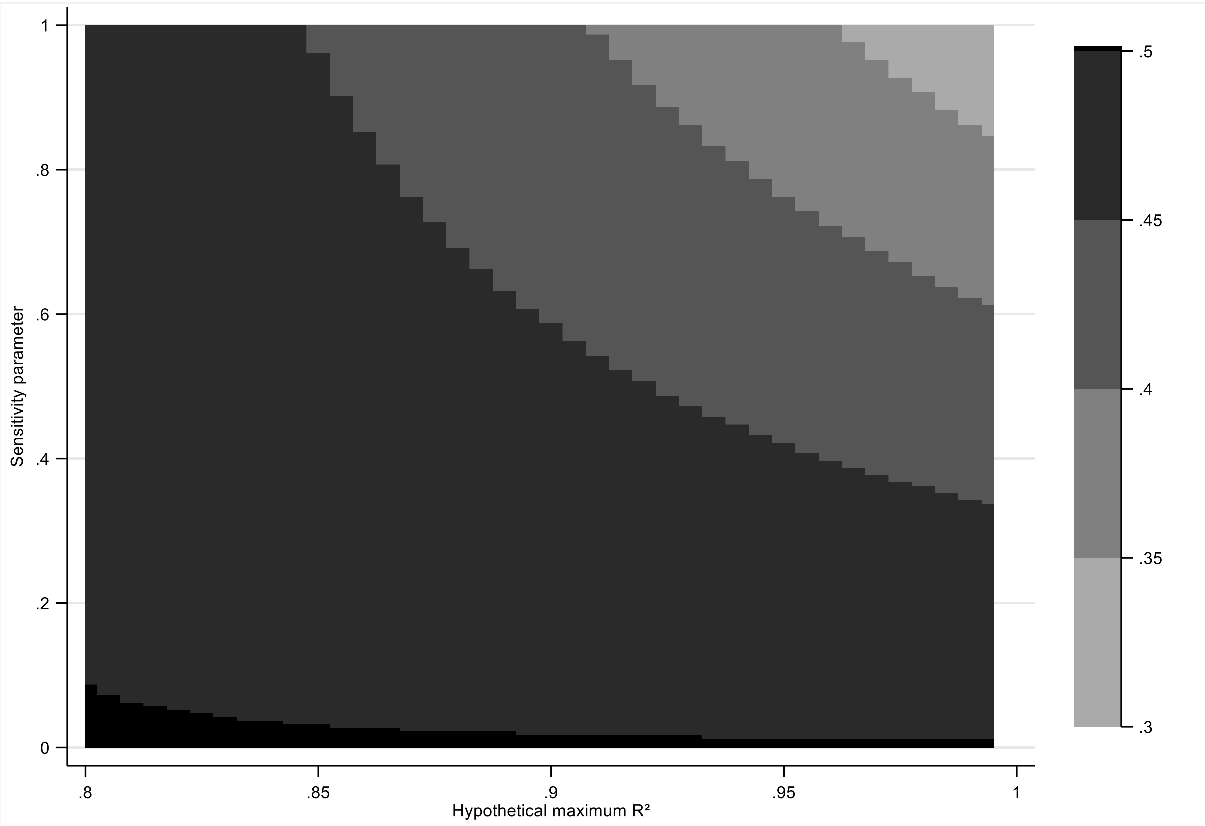
A first challenge is that we have an ordered outcome, while the estimator of Oster (2019) is based on both regression coefficients and R^2 movements. We turn to a simulated residual method to obtain a continuous measure of sentence severity (Gouriéroux et al., 1987). Let Y be an ordered indicator of severity with $Y = 1$ in case of fine, $Y = 2$ in case of probation, and $Y = 3$ in case of prison. We denote by Y^* a latent continuous outcome measuring the sentence severity such that $Y^* = X\beta + \varepsilon$, with ε a random perturbation. By definition, Y^* remains unobserved but we have $Y = 1$ when $Y^* \leq \mu_1$, $Y = 2$ when $\mu_1 < Y^* \leq \mu_2$ and $Y = 3$ when $\mu_2 < Y^*$, where μ_1 and μ_2 are threshold values estimated jointly with the coefficients β. We simulate values of the unobserved latent variable Y^* using the two following steps. First, we obtain maximum likelihood estimates $\hat{\beta}$, $\hat{\mu}_1$ and $\hat{\mu}_2$ from an ordered Probit model and get the predicted outcome $\hat{Y}^* = X\hat{\beta}$. Second, simulated residuals $\hat{\varepsilon}_s$ are drawn from the normal distribution $N(0; 1)$. The simulated outcome $\hat{Y}_s^* = X\hat{\beta} + \hat{\varepsilon}_s$ is the first value satisfying $\hat{Y}_s^* \leq \hat{\mu}_1$ when $Y = 1$, $\hat{\mu}_1 < \hat{Y}_s^* \leq \hat{\mu}_2$ when $Y = 2$, and $\hat{Y}_s^* < \hat{\mu}_2$ when $Y = 3$.

The second challenge is that we have several court effects as potentially biased due to omitted variables, while the methodology for evaluating robustness is designed for one covariate in Krauth (2016) and Oster (2019). As developing a specific methodology to account for court-specific bias with multiple courts is beyond the scope of our paper, we restrict our problem in the following way. Going back to the estimates reported in column (4) of Table 3, we decide to separate courts in two groups: TGI3 and TGI4 on the one hand (the reference group), and TGI1, TGI2, TGI5, TGI6 and TGI7 on the other hand (we call them the “treated” group). In Table A2 in appendix, we report both the ordered Probit estimates and the OLS estimates explaining the simulated latent severity. As expected, coefficients are very similar in both specifications. When explaining the simulated outcome, we find a coefficient of 0.498 for the treated group of courts. Accordingly, there is a large difference with respect to TGI3 and TGI4 since the average simulated severity is equal to 2.176 (with a standard deviation of 1.342). Another result in Table A2 is that the selected covariates provide a good explanation of the latent severity since the R^2 is equal to 0.773.

Next, we apply the estimator for omitted variable bias proposed in Oster (2019). There are two key parameters when estimating a bias-adjusting treatment effect. The first one is the value of the R^2 that would be obtained if all information on observed and unobserved covariates could be included. As the R^2 from our linear regression is high, we consider that the “worst” fit we could obtain by adding unobservables would be 0.8 and we consider a set of values ranging between 0.8 and 1 for the maximal hypothetical R^2. The second parameter, called δ, is the relative degree of selection on observables and unobservables. Following Oster

(2019), we consider a value of one as upper bound for this parameter. In the context of traffic offenses which are massive, homogenous, and processed very rapidly by courts (often without hearing), it seems unreasonable to expect that unobserved factors could explain sentence severity as much as our combination of sociodemographic and criminological control variables. For each combination (δ, R^2), we calculate the bias-adjusted treatment effect for the treated group of courts. If unobserved heterogeneity is a concern, then the null hypothesis that the treated group is not different from the control group in terms of sentence severity should be accepted.

Figure 6. Sensitivity of cross-court disparities to confounding factors



Source : data from seven courts in the South-East of France, authors' calculations.
Note: the bias-adjusted effects are obtained for various combinations (R^2, δ) using the estimator of Oster (2019).

We report our results in Figure 6, where we plot for each combination (δ, R^2) the point estimate of the bias-adjusted effect. We reach two main conclusions. First and as expected, we find that the bias-adjusted effect decreases with the hypothetical R^2 for a given proportion of unobservables and with the contribution of unobservables for a given hypothetical R^2. Second, even in the worst scenario corresponding to a hypothetical R^2 of 1 and a selection on unobservables of same magnitude as that on observables, we find a large difference between the treated group and the reference group of courts (TGI3 and TGI4). The lowest point estimate that we obtain is 0.312. For each combination (δ, R^2), we have also estimated the corresponding confidence interval. The lowest bound is 0.198, so that we can rule out the possibility that our court effects are driven by large omitted factors.

6. Assessing the contribution of prosecutors to court differences

Classical versus simplified procedure

We now investigate the duality of the sentencing process and attempt to decompose the total effect of courts on sentencing into a direct effect from judges who have the final word on sentences and an indirect effect from prosecutors who choose between simple and classical procedures.

We begin with a description of the decisions made by prosecutors across courts. Prosecutors choose a simplified procedure in nearly 8 cases over 10 (79.1%). As shown in panel A of Figure A3 in Appendix, five courts are characterized by a proportion of simplified procedures ranging between 75% and 85%, while TGI3 and TGI7 are substantially below the average (around 55%). Obviously, part of those differences may be due to the fact that the pattern of offenses varies across courts. Again, we apply a standardization method and calculate some adjusted proportions of classical versus simplified procedures by using the average pattern of offenses from all courts. As shown in panel B, differences in prosecutor's decisions do not really stem from differences in offenses since the largest gap between the raw and standardized proportions does not exceed 10 percentage points (in TGI7).

Also, we quantify the magnitude of court effects when explaining prosecutors' decisions. We turn to a Probit regression to explain the choice of the prosecutor to consider a simplified versus a classical procedure. Without any control variables, we find a positive coefficient for TGI2 (+7.5 points), TGI3 (+26.0 points) and TGI7 (+30.9 points). Conversely, there is no difference between TGI1, TGI4, TGI5 and TGI6. The estimates are more nuanced after controlling for both the type of offense and both offenders as well as judges' characteristics. The assumption of null court effects is rejected with a statistic of 47.95 (p=0.005). The probability of a classical procedure is higher in TGI2 (+8.4 points), TGI5 (+6.5 points) and TGI7 (+9.3 points), while it is lower in TGI1 (-5.9 points) and TGI6 (-7.8 points)[20]. Both the type of offenses and offenders' characteristics have a strong influence when explaining differences across courts in terms of prosecutors' decision. In particular, the simplified procedure is more likely in case of default of driver's license or multiple offenses.

The mediation analysis setting

Next, we turn to a mediation analysis technique to assess the contribution of prosecutors when explaining differences among courts (Imai et al., 2010; Pearl, 2012)[21]. A mediation model involves a treatment (exposure) T and a mediator M. Both the treatment and the mediator are expected to affect an outcome Y, conditional on a set of pre-treatment observable characteristics X. The treatment T has a direct effect on Y, but it may also have an indirect effect on Y through the mediator M. The total effect of T on Y is given by the sum of the direct and the indirect effect. Assuming that the mediator M and the outcome Y are continuous and that the treatment is either dichotomous or continuous, then both the total, direct and indirect effects can be estimated with either OLS regressions or structural equation modelling. The regression explaining the mediator is:

$$M = \alpha_M + \gamma_T T + X\beta_M + \varepsilon_M \tag{1}$$

where α_M, γ_T and β_M are coefficients to be estimated and ε_M is an error term. In (1), γ_T is the direct effect of the treatment T on the mediator M. The regression explaining the outcome is:

$$Y = \alpha_Y + \delta_T T + \delta_M M + X\beta_Y + \varepsilon_Y \tag{2}$$

where α_Y, δ_T, δ_M and β_Y are coefficients to be estimated and ε_Y is an error term. Using (1), it follows that the outcome equation (2) may be expressed as $Y = (\alpha_Y + \delta_M \alpha_M) + (\delta_T + \delta_M \gamma_T)T + X(\delta_M \beta_M + \beta_Y) + (\delta_M \varepsilon_M + \varepsilon_Y)$. Thus, the total effect of the treatment T on the outcome Y is $\delta_T + \delta_M \gamma_T$, corresponding to the sum of the direct effect δ_T and the indirect effect $\delta_M \gamma_T$ (through the mediator).

Here, we assess the contribution of prosecutors with regards to their choice of a simplified versus classical procedure when explaining differences in sentences across courts net of the influence played by the pattern of offenses, defendants' and judges' characteristics. In our setting, Y corresponds to the type of sentence with three ordered categories (fine, probation,

[20] The detailed estimates are available upon request.

[21] Identification of direct and indirect effects is possible when there is no confounding variable influencing the effect of the treatment on the mediator, the effect of the mediator on the outcome, and the effect of the treatment on the outcome.

prison), M is the prosecutor's decision to either turn to a simplified or a classical procedure, T corresponds to the court, and X is a set of pre-treatment variables including both the type of offense, offenders' and judges' characteristics. So, the sentence Y may be directly influenced by the treatment-court T and indirectly by the prosecutor's decision M, which itself may vary depending on the court C. With respect to the mediation model summarized by (1) and (2), we need to account for the two following issues.

First, the treatment is multi-categorical rather than binary (or continuous) as there are several courts. In this case, there is no single parameter representing the effect of T on M and of T on Y. As emphasized in Hayes and Preacher (2014), the appropriate strategy is to choose a reference category and to introduce the other categories as covariates. In doing so, we obtain relative effects of the treatment and the mediator on the outcome, respectively. Considering one court as reference, the mediator equation corresponding to the prosecutor's decision will be expressed as $M = \alpha_M + \sum \gamma_{\mathrm{T}} \mathbb{1}_T + \sum \gamma_o \mathbb{1}_o + X\beta_P + \varepsilon_P$ (where o refers to offenses) and the outcome equation will be expressed as $Y = \alpha_Y + \sum \delta_{\mathrm{T}} \mathbb{1}_T + \sum \delta_o \mathbb{1}_o + \delta_M M + X\beta_S + \varepsilon_S$. It follows that the relative total effect of court T on the outcome Y is given by $\delta_T + \delta_M \gamma_T$, which can be decomposed as the sum of the relative direct effect δ_T and the relative indirect effect $\delta_M \gamma_T$.

Second, we have non-linear models explaining the relationships between our variables of interest since the sentence outcome Y is ordered and the prosecutor's decision M is dichotomous. Furthermore, there are several treatment variables corresponding to the various court dummies, so that we cannot turn to a parametric framework. Instead, we rely on the inverse odds ratio-weighted approach proposed in Tchetgen Tchetgen (2013) to estimate both the direct and indirect effects. The procedure requires an estimation of the treatment-mediator conditional odds ratio function given the pre-treatment characteristics, X. The estimated weights are then used to estimate the direct effect of the treatment via a weighted regression model. The treatment and the mediator become independent when applying the weights. The indirect effect is finally obtained by subtracting the direct effect from the total effect which is calculated using an unweighted regression model.

Relying on odds ratios is very useful in our setting with multiple courts. Specifically, we use the invariance property of odds ratios, according to which the same odds ratio for the relationship between two variables, A and B, is obtained when A is the dependent variable and B is the independent variable, or when B is the dependent variable and A is the independent variable. As a consequence, the odds ratios may be estimated from a unique regression explaining the treatment (corresponding to the various courts in our framework) as a function of the mediator (the prosecutors' decision) conditional on a set of exogenous variables. When turning to the data, we choose to consider inverse odds weights (IOW) rather than inverse odds ratio weights (IORW) as the former procedure leads to more efficient estimates according to Nguyen et al. (2015). The IOW weights are stabilized by calculating inverse odds from the inverse of the predicted probability explaining the treatment as a function of the mediator and the selected covariates.

We proceed in the following way. First, we estimate a multinomial Logit explaining the effect of being judged in a given court (still with TGI4 as reference category) and introduce both the prosecutors' decision, type of offenses, offenders, and judges' characteristics as control variables. For each convicted person, we calculate the predicted probability $\hat{p}_{i(T)}$ of being judged in court T from which we derive the inverse odd ratio $\hat{r}_{i(T)}$. By construction, the inverse odd ratio is set to 1 for all offenders judged in the reference court TGI4. Then, we obtain the relative total effect $\delta_T + \delta_M \gamma_T$ for each court T by estimating an unweighted ordered Probit model which explains the sentence as a function of the court dummies, the type of offenses as well as offenders and judges' characteristics. Similarly, we obtain the relative direct effect δ_T for each court T by estimating the weighted version of the previous ordered Probit regression using the weights $\hat{r}$. The relative indirect effect $\delta_M \gamma_T$ is obtained by subtracting the relative direct effect δ_T from the total effect $\delta_T + \delta_M \gamma_T$[22].

Extensive margin

Panel 1 of Table 5 presents the mediation analysis results obtained at the extensive margin with three types of sentences (fines, probation, prison) and the prosecutors' decision of either a simplified or classical procedure. The total effect is positive and significant for the five courts (TGI1, TGI2, TGI5, TGI6 and TGI7) delivering more severe sentences compared to TGI3 and TGI4. In TGI5 and TGI7, the court effect is fully driven by a direct effect of the court on sentences and the indirect effect (mediation through prosecutors' decision) appears very low. Compared to TGI4, the more severe sentences found in those courts are not explained by the

[22] Both for the direct and indirect effects, we calculate standard errors using a bootstrap procedure with 2500 replications.

prosecutors making different choices between classical and simplified procedures. The situation is more mixed in other courts. For instance, the indirect effect is positive and larger than the negative direct effect in TGI1, while the positive direct effect is offset by the negative indirect effect of the prosecutors' decisions in TGI3. However, the various indirect effects are never statistically significant.

Table 5. Direct and indirect effects of courts on sentences : extensive margin

Court		(1) Simplified vs classical procedure			(2) 4 alternatives (OP, CRPC, CI/CPPV, COPJ/other)		
		Direct	Indirect	Total	Direct	Indirect	Total
TGI1	Coefficient	-0.202	0.503	0.302***	-0.150	0.452	0.302***
	St. error	0.411	0.391	0.108	0.479	0.462	0.108
TGI2	Coefficient	0.193	0.400	0.592***	0.427	0.165	0.592***
	St. error	0.460	0.442	0.092	0.486	0.478	0.092
TGI3	Coefficient	0.963	-1.114	-0.151	1.209	-1.360	-0.151
	St. error	0.843	0.830	0.118	0.866	0.856	0.118
TGI5	Coefficient	1.081***	-0.439	0.642***	1.598***	-0.956*	0.642***
	St. error	0.405	0.384	0.102	0.508	0.504	0.102
TGI6	Coefficient	0.584	-0.186	0.398***	0.545	-0.147	0.398***
	St. error	0.560	0.554	0.099	0.609	0.600	0.099
TGI7	Coefficient	0.903*	-0.054	0.849***	0.852	-0.003	0.849***
	St. error	0.495	0.479	0.150	0.560	0.544	0.150

Source: data from seven courts in the South-East of France, authors' calculations.
Note : the courts are anonymized (from TGI1 to TGI7) due to confidentiality issues. Estimates for the direct, indirect and total effects are obtained from a mediation analysis. The total effect is obtained from a multinomial Logit model explaining the probability of being judged in a given court, the direct effect is obtained from a weighted version of the same regression, and the indirect effect is calculated by difference. Standard errors are bootstrapped with 2500 replications. Significance levels are 1% (***), 5% (**) and 10% (*).

In panel (2), we disaggregate the simplified procedure between *OP* (54.8% of cases) and *CRPC* (24.2%), and classical procedures between most stringent and fast-track procedures (*CI* and *CCPV*, 3.7%) and less stringent and slower procedures (*COPJ* and others, 17.3%). *OP* is used much more in TGI4 (63.2%), TGI5 (62.7%) and TGI6 (58.9%) than in the other courts. Again, we turn to a multinomial Logit model to explain the probability of being judged in a given court as a function of covariates and the ordered categories for the prosecutor's decision. When implementing the mediation analysis, we find very similar results compared to the binary decision (classical versus simplified). In TGI5, the negative indirect effect is substantial and significant at the 10 percent level. In that court, prosecutors' decisions in terms of procedures are more lenient than in the reference court (TGI4), but this leniency is compensated by more severe sentences in later decisions.

Intensive margin

Next, we replicate the same decomposition of the total effect at the intensive margin by considering fines. Still with TGI4 as reference, all the other courts except TGI5 issue lower amounts of fines on average. The estimates reported in panel 1 of Table 6 show that the indirect effect is significant in two courts. In TGI3, the indirect effect is positive and strongly reduces the influence of the court on the average fine. Conversely, in TGI6, fines are lower because of differences in decisions made by prosecutors, with a large negative indirect effect which compensates the positive direct effect. In panel 2, we find that further decomposing the simplified and classical procedures does not affect the result. We have also implemented the same mediation for the duration of prison. Our estimates, not reported, show that none of the indirect effect through the prosecutors' decision is significant.

Table 6. Direct and indirect effects of courts on fines (intensive margin)

Court		(1) Simplified vs classical procedure			(2) 4 alternatives (OP, CRPC, CI/CPPV, COPJ/other)		
		Direct	Indirect	Total	Direct	Indirect	Total
TGI1	Coefficient	-0,550	0,127	-0,422***	-0,359	-0,064	-0,422***
	St. error	0,388	0,392	0,057	0,380	0,385	0,057
TGI2	Coefficient	-0,067	-0,117	-0,184***	-0,210	0,026	-0,184***
	St. error	0,440	0,436	0,058	0,407	0,404	0,058
TGI3	Coefficient	-2,952***	2,739***	-0,213***	-2,794***	2,581**	-0,213***
	St. error	1,075	1,069	0,076	1,041	1,041	0,076
TGI5	Coefficient	-0,205	0,168	-0,037	-0,219	0,182	-0,037
	St. error	0,345	0,343	0,063	0,348	0,347	0,063
TGI6	Coefficient	2,570*	-2,890**	-0,319***	2,709*	-3,028**	-0,319***
	St. error	1,433	1,415	0,048	1,427	1,409	0,048
TGI7	Coefficient	-1,003*	0,266	-0,737***	-1,954*	1,217	-0,737**
	St. error	0,549	0,561	0,121	0,783	0,793	0,121

Source: data from seven courts in the South-East of France, authors' calculations.
Note : the courts are anonymized (from TGI1 to TGI7) due to confidentiality issues. Estimates for the direct, indirect, and total effects are obtained from a mediation analysis. The total effect is obtained from a multinomial Logit model explaining the probability of being judged in a given court, the direct effect is obtained from a weighted version of the same regression, and the indirect effect is calculated by difference. Standard errors are bootstrapped with 2500 replications. Significance levels are 1% (***), 5% (**) and 10% (*).

7. Concluding comments

The purpose of our paper was to contribute to the existing literature on differences in sentences between courts, but with a fresh angle related to the influence of prosecutors on such disparities. For that purpose, we have used unique data on traffic offenses from a sample of seven courts located in South-East France. We complement this dataset with some individual characteristics of judges. The situation is different for prosecutors. As a given case is processed by different prosecutors throughout the legal procedure in French courts, we focus on the type of procedure (simplified versus classical) chosen by prosecutors. We reach two main conclusions.

First, there are sizeable disparities in sentencing between neighboring courts handling highly homogenous cases of traffic offending. Disparities are found both in terms of sentence type and *quantums* net of the influence of both offenders' and judges' characteristics. At the extensive margin, we find that the two courts delivering the less severe sentences are located in the two most populated cities which are also characterized by the highest crime rates. Robustness checks show that unobserved heterogeneity cannot explain differences in terms of sentence severity between courts. Our results are consistent with recent evidence from North Carolina according to which cross-court disparities do not disappear over time with judge rotations, since judges tend to adapt to local norms (Abrams et al., 2021). Second, results from a mediation analysis show that there is some heterogeneity in the role of prosecutors between courts. However, the indirect effect related to prosecutors' choice of procedure is low in three courts out of seven and none of the court effects is fully explained by the prosecutors' decision.

The fact that observably similar offenders face very different probabilities of serving a prison or probation sentence from one court to the next is particularly striking in a civil law country like France. Indeed, judges and prosecutors work under the constitutional principle of equal justice for all and dispose their sentences applying the same substantial and procedural criminal law, which should lead to similar sentences for comparable cases. Furthermore, those differences are not explained by the fact that judges may have different gender or experience and they are not due to some unobserved heterogeneity at the case level. Contextual elements suggest that local conditions seem to play a role, but the very small number of courts prevents us from further exploring this issue.

According to the Criminal Code, prosecutors have strong leeway to file charges and choose among several criminal procedures, which can give them a leading role in sentencing and reduce judges' discretion (through the exclusion of certain sentence's types or limitation on *quantums*). In simplified procedures, judges can only accept or refuse the prosecutor's proposed sentence, and refusals lead to the launch of classical procedures with a (long) trial and extra-work for judges. This situation could grant prosecutors more discretion to impose their preferences in terms of sentencing, which could generate disparities across courts. However, our mediation analysis shows that despite some diversity between courts the leading role played by prosecutors in the procedure is never sufficient to explain the disparities between courts. As they stand, our results are more consistent with some form of cooperation between judges and prosecutors, as was observed by Kim et al. (2015) in three U.S. district courts.

A few *caveats* have to be kept in mind when interpreting our results. First, prosecutors may influence sentencing through other channels than the criminal procedures they choose. In particular, they could affect decisions through their requisitions during traditional trials or through the cases they drop completely, but such information is not available. Second, we study a kind of collective decision resulting from all prosecutors involved in the legal procedure. As a consequence, we are not able to account for the potential influence of individual characteristics of prosecutors as we do not know the identity of all of those who are involved in each case. Also, this rules out the possibility of working on prosecutor-judge dyads as done in Kim et al. (2015). In our framework, the prosecutor's heterogeneity is picked up in the type of procedure (classical versus simplified).

Third, we investigate cases handled by seven neighboring courts in 2017. Given this proximity, one could expect more similar decisions, meaning that our estimation of court disparities is presumably a lower bound of the overall disparities that would be observed with a sample of all French courts. At the same time, having only a few courts precludes any attempt to correlate the courts effects with some local indicators like composition of the population, crime rate or available beds in prisons. Production of exhaustive data on all criminal procedures in all courts by the Ministry of Justice would be very welcome to further investigate and understand differences in sentence decisions between courts in France.

References

Abrams, D., Galbiati, R., Henry, E. and Philippe, A. (2021). When in Rome… On Local Norms and Sentencing Decisions, *Journal of the European Economic Association, 20(2),* 700-738.

Altonji, J. G., Elder, T. E. and Taber, C. R. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools, *Journal of Political Economy, 113(1)*, 151-184.

Altonji, J. G., Elder, T. E. and Taber, C. R. (2008). Using Selection on Observed Variables to Assess Bias from Unobservables when Evaluating Swan-Ganz Catheterization, *American Economic Review, 92(2)*, 345-350.

Anwar, S., Bayer, P. and Hjalmarsson, R. (2019). Politics in the courtroom: Political ideology and jury decision making, *Journal of the European Economic Association, 17(3),* 834-875.

Bandyopadhyay, S. and McCannon, B. (2014). The effect of the election of prosecutors on criminal trials, *Public Choice, 161(1).*

Berdejo, C., and Chen, D. L. (2017). Electoral cycles among U.S. courts of appeals judges. *The Journal of Law and Economics*, *60*(3), 479-496.

Berdejo, C., and Yuchtman, N. (2013). Crime, punishment and politics: an analysis of political cycles in criminal sentencing. *The Review of Economics and Statistics, 95(3),* 741-756.

Boes, S., and Winkelmann, R. (2006). Ordered response models. *Allgemeines Statistisches Archiv*, *90*(1), 167-181.

Bourreau-Dubois, C., Doriat-Duban, M., Jeandidier, B., and Ray, J. C. (2020). Does gender diversity in panels of judges matter? Evidence from French child support cases. *International Review of Law and Economics*, *63*, 105929.

Chabanne, M. and Timbart, O. (2017). La délinquance routière devant la justice, *Infostat Justice #153*.

Cocuau, V. (2020). Les auteurs d'infractions à la sécurité routière devant la justice, *Infostat Justice #180*.

Cohen, M. A. (1991). Explaining Judicial Behavior or What's "Unconstituitional" about the Sentencing Commission? *Journal Law, Economics, and Organization*, *7*(1), 183–199.

Danziger, S., Levav, J., and Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, *108*(17), 6889–6892.

de Castro, A. S. (2021). Judicial attitudes under shifting jurisprudence: Evidence from Brazil's new drug law of 2006. *International Review of Law and Economics*, *68*, 106006.

Epstein, L. and Knight, J. (1997). The New Institutionalism, Part II. Law and Courts, *7*(2), 4–9.

Epstein, L., Landes, W. M., and Posner, R. A. (2011). Why (and when) judges dissent: A theoretical and empirical analysis. *Journal of Legal Analysis*, *3*(1), 101-137.

Epstein, L. and Weinshall, K. (2021). The strategic analysis of judicial decision making, a comparative perspective, Cambridge University Press.

Eren, O., and Mocan, N. (2018). Emotional judges and unlucky juveniles. *American Economic Journal: Applied Economics*, *10*(3), 171–205.

Fałkowski, J., and Lewkowicz, J. (2021). Are Adjudication Panels Strategically Selected? The Case of Constitutional Court in Poland. *International Review of Law and Economics*, *65*, 105950.

Feigenberg, B., and Miller, C. (2021). Racial Divisions and Criminal Justice: Evidence from Southern State Courts, *American Economic Journal: Economic Policy*, 13 (2): 207-40.

Fischman, J. B., and Law, D. S. (2009). What is judicial ideology, and how should we measure it. *Washington University Journal of Law & Policy, 29*(1), 133–214.

Glaesar, E. L., Kessler, D. P. andPiehl, A. M. (2000). What Do Prosecutors Maximize? An Analysis of the Federalization of Drug Crimes, *American Law & Economics Review,* 259.

Gourieroux, C., Monfort, A., Renault, E. and Trognon, A. (1978). Generalised residuals, *Journal of Econometrics, 34(1-2),* 5-32.

Greene W. H. andHensher D. A., (2010). Modeling ordered choices. A primer, Cambridge University Press.

Groot, W., and van den Brink, H. M. (1999). Job satisfaction and preference drift. *Economics letters*, *63*(3), 363-367.

Hayes A. F. and Preacher K. J. (2014). Statistical mediation analysis with a multicategorical independent variable, *British Journal of Mathematical and Statistical Psychology*, 451-470.

Heise, M. (2002). The Past, Present and Future of Empirical Legal Scholarship: Judicial Decision Making and the New Empiricism. *University Of Illinois Law Review*, 819–850.

Helland, E., & Tabarrok, A. (2000). Runaway judges? Selection effects and the jury. *Journal of Law, Economics, and Organization*, *16*(2), 306-333.

Henneguelle, A., Monnery, B. and Kensey, A. (2016). Better at home than in prison? The effects of electronic monitoring on recidivism in France, *Journal of Law and Economics, 59(3*), 629-667.

Huber, G. A., and Gordon, S. C. (2004). Accountability and Coercion: is justice blind when it runs for office?, *American Journal of Political Science, 48(2),* 247-263.

Ichino, A., Polo, M., & Rettore, E. (2003). Are judges biased by labor market conditions?. *European Economic Review*, *47*(5), 913-944.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical science*, *25*(1), 51-71.

Joseph-Ratineau, Y. (2019). La barémisation de la justice : barémisation et droit pénal, *Report for GIP Recherche Droit et Justice*, 777 pages.

Kim, B., Spohn, C. and Hedberg, E. C. (2015). Federal sentencing as a complex collaborative process: judges, prosecutors, judge-prosecutor dyads and dispartity in sentencing. *Criminology*, *53(4),* 597-623.

Krauth, B. (2016). Bounding a Linear Causal Effect Using Relative Correlation Restrictions, *Journal of Econometric Methods, 5(1),* 117-141.

Melcarne, A. (2017), Careerism and Judicial Behavior, *European Journal of Law and Economics*, *44*, 241–264.

Melcarne, A. and Ramello, G. B. (2015). Judicial Independence, Judges' Incentives and Efficiency. *Review of Law & Economics*, 11(2),149–169.

Melcarne, A., Ramello, G. B. and Spruk R. (2021). Is justice delayed justice denied? An empirical approach. *International Review of Law and Economics*, *65*, 105953.

Monnery, B., Wolff, F-C. and Henneguelle, A. (2020). Semi-liberty, prison and recidivisim: Bounding causal effects in a survival model, *International Review of Law and Economics* 61.

Nelson, M. J. (2014). Responsive justice? Retention elections, prosecutors and public opinion, *Journal of Law and Cours, 2(1),* 117-152.

Nguyen, Q. C., Osypuk, T.L., Schmidt, N. M., Glymour, M. M. and Tchetgen Tchetgen, E. J. (2015). Practical Guidance for Conducting Mediation Analysis With Multiple Mediators Using Inverse Odds Ratio Weighting. *American Journal of Epidemiology, 181*(5), 349–356.

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence, *Journal of Business & Economic Statistics, 37(2),* 187-204.

Pearl, J. (2012). The causal mediation formula—a guide to the assessment of pathways and mechanisms. *Prevention science*, *13*(4), 426-436.

Posner, R. (1993). What do Judges Maximize? (The Same Thing Everybody Else Does). *Supreme Court Economic Review*, 3(1993),1–41.

Rachlinski, J. J., Johnson, S., Wistrich, A. J., and Guthrie, C. (2009). Does unconscious racial bias affect trial judges. *Notre Dame Law Review*, *84*(3), 1195-1246.

Rasmusen, E., Raghav, M. and Ramseyer, M. (2009). Convictions Versus Conviction Rates: The Prosecutor's Choice, *American Law & Economics Review, 11*(1).

Rehavi, M. M., and Starr, S. B. (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy*, *122*(6), 1320-1354.

Schauer, F. (2000). Incentives, Reputation, and the Inglorious determinants of judicial behavior. *University of Cincinnati Law Review*, 68, 615–636.

Taha, A. E. (2004). Publish or Paris? Evidence of How Judges Allocate Their Time. *American Law and Economics Review*, *6*(1), 1–27.

Tchetgen Tchetgen, E. J. (2013). Inverse odds ratio-weighted estimation for causal mediation analysis. *Statistical Medecine*, *32*(26), 4567–4580.

Terza, J. V. (1985). Reduced-form trinomial probit: a quantal response model without a priori restrictions. *Journal of Business & Economic Statistics*, *3*(1), 54-59.

Timbart, O. and Minne, M-D. (2013). Le traitement judiciaire de la délinquance routière, *Infostat Justice #123*.

Appendix

A1. Normalization procedure to account for offense structure

We denote each type of offense by o_i with $o \in \{o_1, \dots, o_6\}$ and define S^o as the sentence associated to each offense type. For a given court c, the average sentence S_c is $S_c = \sum w_c^o S_c^o$ where w_c^o is the proportion of offenses of type o. We neutralize the role of the weights w_c^o by calculating the normalized average $S_{cN} = \sum w_N^o S_c^o$, where the weights w_N^o correspond to normalized weights similar for all courts[23]. As shown in Table A1, in almost all cases adjusting for the composition of offense types within courts tends to reduce substantially the raw disparities in sentencing. For TGI3 for instance, fines are used in only 43.0% of cases as compared to a mean of 66.4% across courts. However, this difference is partly due to the fact that this court disproportionately deals with serious offenses (high alcohol intakes and multiple offenses) that are often treated with more severe sanctions (probation or prison). Once the pattern of offenses is controlled for, the adjusted share of fines increases to 50.9%.

Table A1. Observed and adjusted distributions of sentences across courts

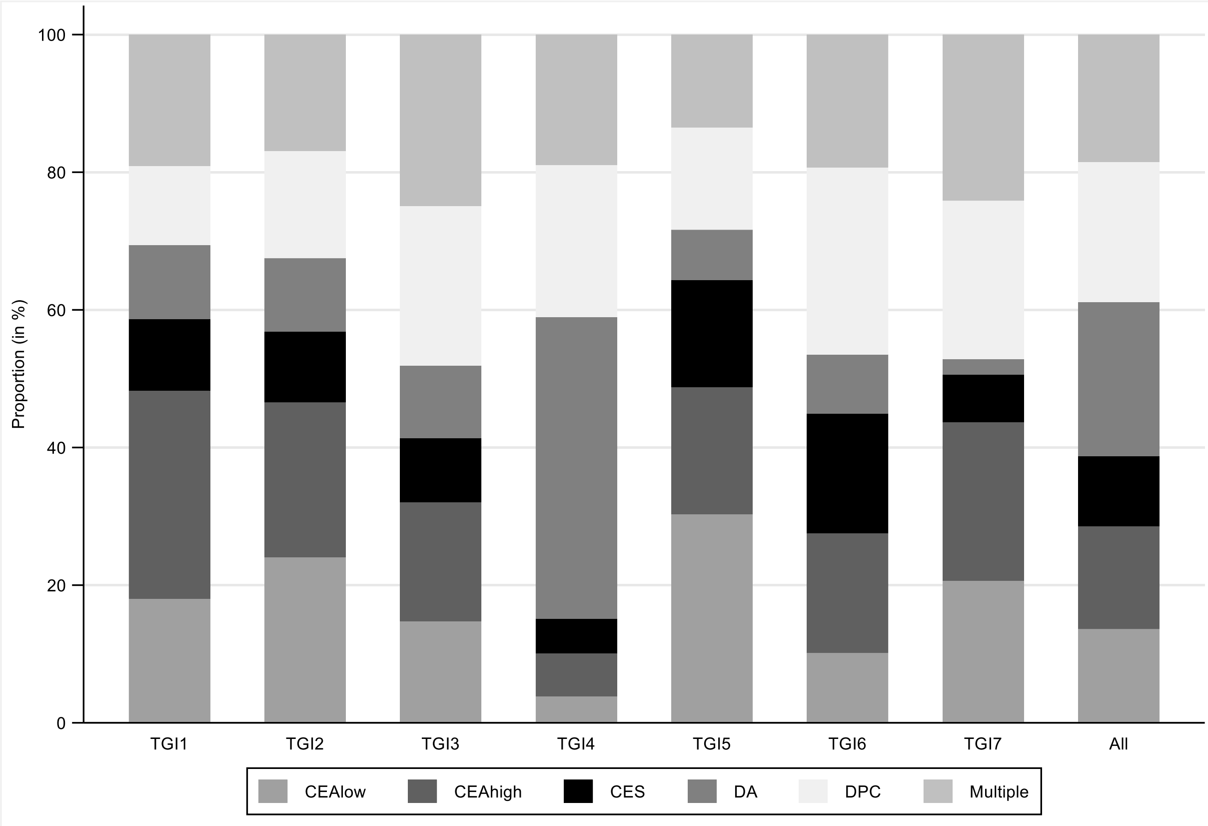
Court		Fine	Probation	Prison
TGI1	Observed	0.576	0.347	0.076
	Adjusted by offense pattern	0.611	0.312	0.077
	Ratio	1.060	0.899	1.004
TGI2	Observed	0.522	0.375	0.103
	Adjusted by offense pattern	0.564	0.330	0.107
	Ratio	1.080	0.879	1.034
TGI3	Observed	0.430	0.414	0.156
	Adjusted by offense pattern	0.509	0.363	0.128
	Ratio	1.182	0.878	0.821
TGI4	Observed	0.794	0.108	0.098
	Adjusted by offense pattern	0.726	0.157	0.117
	Ratio	0.914	1.454	1.198
TGI5	Observed	0.670	0.181	0.149
	Adjusted by offense pattern	0.668	0.175	0.158
	Ratio	0.996	0.965	1.060
TGI6	Observed	0.645	0.210	0.145
	Adjusted by offense pattern	0.672	0.198	0.130
	Ratio	1.041	0.942	0.899
TGI7	Observed	0.230	0.483	0.287
	Adjusted by offense pattern	0.405	0.363	0.232
	Ratio	1.762	0.753	0.806
All		0.664	0.214	0.122

Source: data from seven courts in the South-East of France, authors' calculations.
Note: the courts are anonymized (from TGI1 to TGI7) due to confidentiality issues. We consider the average distribution of offenses among all courts as weights to calculate the adjusted distribution of sentences.

[23] The average pattern of offenses for all courts is used to construct the normalized weights w_N^o. Detailed results are available upon request.

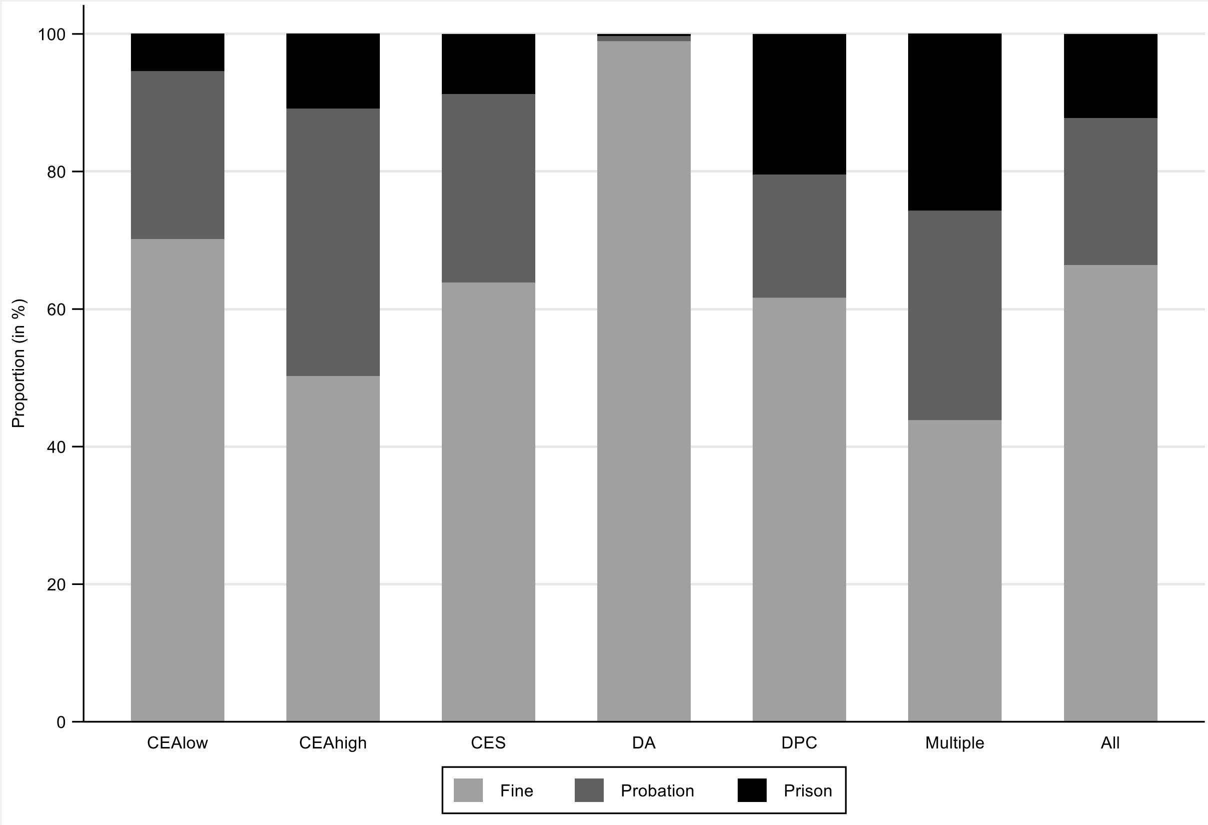
A2. Figures

Figure A1. Distribution of offense types across courts



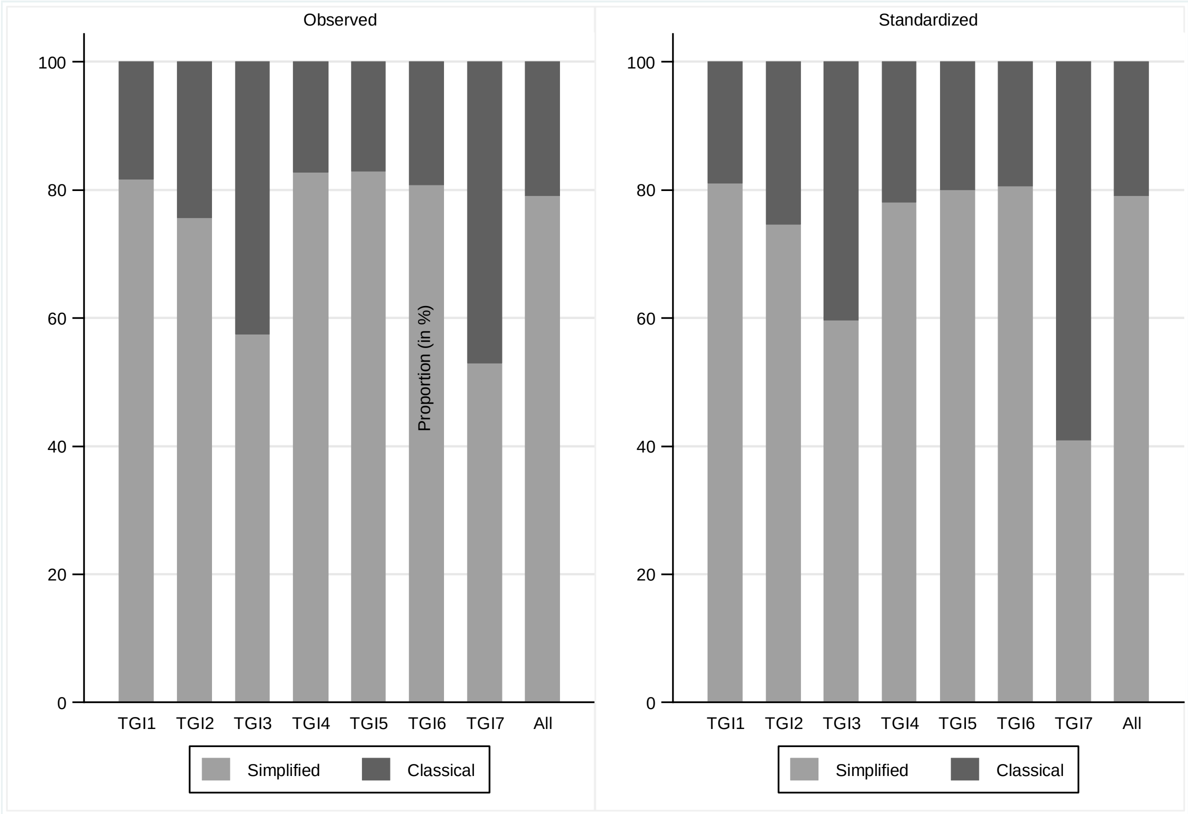
Source: data from seven courts in the South-East of France, authors' calculations.
Note: the courts are anonymized (from TGI1 to TGI7) due to confidentiality issues. CEAlow = DUI with low intake, CEAhigh = DUI with high intake, CES = driving under narcotics, DA = default of car insurance, DPC = default of driver's license, Multiple = mix of offenses.

Figure A2. Distribution of sentences by offense types



Source: data from seven courts in the South-East of France, authors' calculations.
Note: the courts are anonymized (from TGI1 to TGI7) due to confidentiality issues. CEAlow = DUI with low intake, CEAhigh = DUI with high intake, CES = driving under narcotics, DA = default of car insurance, DPC = default of driver's license, Multiple = mix of offenses.

Figure A3. Observed and adjusted decisions of prosecutors across courts



Source: data from seven French courts, authors' calculations.
Note : the courts are anonymized (from TGI1 to TGI7) due to confidentiality issues. We consider the average distribution of offenses among all courts as weights to calculate the adjusted distribution of prosecutors' decision.

A3. Estimates of latent severity of sentences

Table A2. Ordered Probit estimates and OLS estimates of latent severity of sentences

Variables	(3) Ordered Probit		(4) OLS simulated residuals	
	coef.	t-test	coef.	t-test
Courts (ref : TGI3 and TGI4)				
TGI1-TGI2-TGI5-TGI6-TGI7	0.552***	(7.42)	0.498***	(10.11)
Offense (ref : DUI with low intake)				
DUI with high intake	0.462**	(2.44)	0.486***	(6.91)
Driving under narcotics	0.325	(1.60)	0.354***	(5.38)
Default of car insurance	-1.309***	(-7.08)	-1.404***	(-16.63)
Default of driver's license	0.359*	(1.92)	0.391***	(6.83)
Multiple offenses	0.678***	(4.51)	0.648***	(11.16)
Offenders' characteristics				
Gender : male	0.125*	(1.86)	0.132***	(7.18)
Age : 26-30 (ref : ≤25)	-0.086	(-1.62)	-0.084**	(-3.06)
Age : 31-40	-0.009	(-0.10)	-0.011	(-0.32)
Age : 41-50	0.040	(0.45)	0.038	(1.07)
Age : >50	0.117	(1.25)	0.110**	(3.20)
Unemployed (ref : employed)	0.374*	(1.83)	0.362**	(2.76)
Inactive (student, retiree)	-0.018	(-0.18)	-0.041	(-1.41)
Occupation : no information	0.340***	(2.60)	0.292**	(3.66)
Legal recidivism (ref: no recidivism)	1.734***	(12.53)	1.300***	(11.41)
Repeat offender (ref: no recidivism)	1.183***	(6.91)	0.896***	(11.89)
Judges' characteristics				
Gender : female	0.401***	(3.25)	0.344***	(5.87)
Years of experience as judge	0.028***	(5.91)	0.024***	(8.83)
Court-specific years of experience	0.028	(1.64)	0.029***	(4.83)
μ_1	2.692***	(8.31)		
μ_2	3.861***	(13.66)		
Constant			0.541***	(3.92)
Observations	3,395		3,395	
Log likelihood – R^2	-1978.5		0.773	

Source: data from seven courts in the South-East of France, authors' calculations.
Note : the courts are anonymized (from TGI1 to TGI7) due to confidentiality issues. Estimates are obtained from an ordered Probit model in (1), while (2) are estimates from an OLS regression using simulated residuals. In both cases, standard errors are clustered at the court level. Significance levels are 1% (***), 5% (**) and 10% (*).