BETA
Bureau
d'économie
théorique
et appliquée

Documents de travail

«Do sentencing guidelines result in lower inter-judge disparity? Evidence from framed field experiment »

Auteurs

Cécile Bourreau-Dubois, Myriam Doriat-Duban,
Bruno Jeandidier, Jean Claude Ray

Document de Travail n° 2020 – 28

Juin 2020

Bureau d'Économie
Théorique et Appliquée
BETA

www.beta-umr7522.fr

@beta_economics

Contact :
jaoulgrammare@beta-cnrs.unistra.fr

Université de Strasbourg · cnrs · Université de Lorraine · INRAE · AgroParisTech

Do sentencing guidelines result in lower inter-judge disparity? Evidence from framed field experiment[1]

Cécile Bourreau-Dubois, Myriam Doriat-Duban, Bruno Jeandidier et Jean Claude Ray

Université de Lorraine, Université de Strasbourg, CNRS, BETA, F-54000 Nancy, France

June 2020

Abstract: We study decision-making of judges in an experimental setting resembling real world judicial decision making. We gave to 312 future judges 48 vignettes built from real data related to divorce cases involving children. We compare two different subject pools: judges who were asked to set child support awards with a guideline and judges who were asked to set child support awards without any guideline. We found that the introduction of a guideline contributes to reduce the disparity between judges (i.e. the variance for like cases is lower when the subjects have the opportunity to use the guideline) but this effect is not systematic, an increase in heterogeneity being observed for some specific cases.

Keywords: controlled experiment - field experiment - judicial sentencing - child support

JEL code: K42

[1] This work has benefited from the financial support of the Mission Droit et Justice (2016-2018).

1. Introduction

The judicial systems of developed countries have been marked over the last 30 years by the development of guidelines. In the United States, two areas are concerned since the mid-1980s: the length of prison sentences of criminal defendants and the amount of child support awards. In France, two official guidelines were implemented more recently: an advisory child support guideline in 2010 and a mandatory guideline for setting damages in labor courts in 2017. One of the major goals of the guidelines is to improve horizontal equity, through the reduction of unwarranted sentencing disparity, that is different treatment for similarly situated cases (Waldfogel, 1998)[2]. Such disparity may have different origins. First, a decision maker may impose different sentences to similar cases from the judicial point of view. This may be observed if the judge is influenced in her decision by factors such as the political, economic or social context of the case (Ichino, 2003, Marinescu 2011), qualitative characteristics of the case (ex: the quality of the lawyer, the remorse of the offender) or even the personal characteristics of the litigants (age, gender or ethnic origin). Furthermore, judges, like other economic agents, may also be considered victims of cognitive biases and errors (Guthrie et al., 2001; Wistrich et al., 2015; Spamann and Klöhn, 2016; Liu, 2018; Kahan, 2015). The relative bias of their decisions would then result from heuristics such as confirmation, anchoring or availability. As a result, the sentencing of a judge may be different for similarly situated litigants. The second origin of disparity is between-judge variation. The judges may be more or less influenced by factors that are not in the law (either legitimate or not) (Abrams et al. 2012). Individual characteristics of the judge may also impact on her decisions, like gender (Peresie 1985), ethnic origin (Abrams et al. 2012, Morin 2014), political affiliation (Cohen and Young, 2019). Even if judges were not driven by this kind of factors, different judges could make different sentencing for similar cases because they may give greater or lesser weight to legally relevant factors due to their individual preferences (Waldfogel, 1998, Woolredge, 2010). As a result of these philosophical and attitudinal differences, the sentence may rest in part on the judge who imposes it. In this paper we focus on this type of disparity.

The guidelines, in particular when they are mandatory, are expected to reduce the inter-judge disparity, resulting from the philosophy, ideology or bias of the sentencing judge, by constraining judicial discretion. A lot of studies have been looking for whether these guidelines have been successful in practice in achieving the reduction of disparity in decision-making. These studies mobilize real databases (mainly resulting from the sentencing activity of courts), allowing for a

[2] The elimination of unwarranted disparity is not the only goal of the guidelines. For instance, the US Guidelines had also objectives like to promote deterrence by increasing the sentence prison length or to reduce poverty among children by increasing the amount of child support award.

comparison of decisions before and after the introduction of the guidelines. One limitation of this approach is that they make it difficult to separate the pure effects of the guidelines from the effects related to characteristics that are observed by the judges during the judgment but not by the researcher. Thus, there is a risk of biased estimates due to unobserved heterogeneity. In most of the studies this problem is considered as solved provided that there is a random assignment of cases to judges, which ensures that judges receive the same distribution of case characteristics, both observed and unobserved. A second limitation of these studies is that they can examine overall interjudge heterogeneity in sentencing but have more difficulties to identify the combination of characteristics which conduct to a higher risk of disparity sentencing. Consequently, one may consider that the study of interjudge sentencing disparity on the basis of real data does not make it possible to assess the effectiveness of a guideline in terms of reducing disparity in a completely convincing manner.

To circumvent this limitation, we used a controlled experiment. Our subjects are people who have passed the exam to become judge and who will be, after three years spent at the Ecole Nationale de la Magistrature (ENM), appointed as judge in a French court. They had to set child support amounts for the same given cases, each one characterized by a limited amount of information: the income of the parties, the support amounts proposed by the parties, the age and mode of residence of the child and the number of children of the couple. The treatment consisted in enabling some of the subjects to set the amount of child support using the child support guideline made available by the French Ministry of Justice. With an experimental setting resembling real world, our paper is in line with the emerging framed field experiment literature (Ilomaki, 2012; Boulu-Reshef et al. 2016). Finally, our paper is original for two other aspects. We provide empirical evidence on judicial decision-making in a civil law institutional background and we study the impact of child support guidelines. This contrasts with the previous literature that mainly focused on the American case and on sentencing disparity in criminal cases.

Our main results are the following: the implementation of a guideline would contribute to reduce the disparity between judges (i.e. the variance for like cases is lower when the subjects have the opportunity to use the guideline) but this effect is not systematic, an increase in heterogeneity being observed for some specific cases.

The rest of the paper proceeds as follows. Section 2 provides an overview of prior works on the impact of guidelines on the elimination of interjudge sentencing disparity. In Section 3, we describe the experiment and the methodology. In Section 4, we report our results. Section 5 concludes.

2. Literature review

Inter-judge sentencing disparity issue has been widely discussed and researched over the past fifty years, by scholars in criminology, in political science, as in law and economics. Comparatively relatively little research has examined specifically the degree to which the sentencing guidelines reduced sentencing disparity. Three waves of studies can be identified.

The first existing studies reached somewhat mixed results. Following the introduction in the mid-1980s of mandatory guidelines in the United States, several studies have sought to measure whether this reform has made it possible to reduce disparity of decisions between judges[3]. These studies revealed either that overall sentence disparity attributable to the judge declined in the postguidelines era or increased in the postguidelines era, in some district courts. For example, Anderson, Kling, and Stith, (1999) use a definition of interjudge disparity that measured the difference in the mean prison sentence for each judge relative to the mean prison sentence for all judges in the district. They find that sentencing disparity attributable to the judge declined substantially from 1986–1987 to 1988–1989 and remained relatively stable from 1990 to 1993. The expected difference in the sentence lengths of two judges with comparable caseloads was 16–18 percent in the preguideline era but only 8–13 percent in the postguideline era. The authors state that "the Guidelines have reduced the net variation in sentence attributable to the happen stance of the identity of the sentencing judge" (Anderson et al. 1999:303). The work of Hofer, Blackwell, and Ruback (1999) leads to the same type of conclusion. Using several techniques to determine whether interjudge disparity had declined in the postguideline period, these authors conclude that the guidelines had "modest success at reducing inter-judge disparity" for some types of offenses and in some of the districts examined (Hofer, Blackwell, and Ruback, 1999: 290–1). These results observed in criminal issue areas are confirmed in civil issue fields. Using data from the National Longitudinal Survey of Labor market Experience of Youth, Argys et al. (2001) compare variation in child-support awards for divorced or separated mothers living in states after guidelines are adopted to those living in states prior to the adaptation of guidelines, controlling for sets of variables likely to account for differences. They show that the introduction of guidelines had significantly reduced the disparity of amounts. However, they also show that "the adoption of guidelines reduces the likelihood of extreme amounts in some cases, but does

[3] In criminal issue areas, the guidelines consist of proposing sentencing ranges by category of offense, according to the characteristics of the case (ie past criminal history of the offender and severity of the current offense). Judges are allowed to depart from this range (downward or upward) only under certain circumstances provided that they give relevant reasons, which may relate to the characteristics of the case or legal issues (Schanzenbach et al. 2007).

not appear to improve horizontal equity in awards for the entire distribution of families" (Argys et al. 2001, p. 246).

By contrast, other studies conclude that the variance attributable to the judge to whom the case was assigned increased in the postguideline period (Waldfogel, 1991; Payne, 1997; Lacasse and Payne, 1999). For instance, in the context of guilty pleas in the United States, Lacasse and Payne (1999) examined whether the variability of sanctions (prison sentences) attributable to judges had been eliminated by the use of guidelines. Their study shows that, once selection biases related to the characteristics of the judges' and defendants' cases were addressed, this variability would have increased. Specifically, the amount of variation attributable to the judge in trial decisions was 4% before reform and between 5 and 13% after reform, depending on the type of crime and the court. The authors conclude that the guideline is probably insufficient to eliminate any variability in sanctions between judges, since judges retain a margin of discretion even in the presence of mandatory guidelines. In the case of this study, the authors believe that judges may indeed be sensitive to the accused's remorse or may assess the evidence differently.

Some other works have showed that the implementation of guidelines did not prevent that extra-legal factors to influence the judicial sentencing, resulting in the maintenance of interjudge disparity. The sentencing Guidelines prescribed that sentences were to be set by the judge within a range determined in accordance with a computed criminal history–level score and a computed offense-level score, and judges were explicitly forbidden from considering factors such as race, gender, socioeconomic status or family circumstances. Though, empirical findings reveal that sentencing disparities by race, gender, education, and socioeconomic status are prevalent in the federal criminal justice system. Abrahms et al. (2012) find evidence of significant interjudge disparity in the racial gap in incarceration rates, which provides support for the model in which at least some judges treat defendants differently on the basis of their race. Sorensen et al (2012) show that personal circumstances do in fact figure into the determination of sentences and that racial and gender sentencing disparities remain even after conditioning on personal circumstances and the criminal history score and the severity of current offense score, judges punishing women less severely and black men more severely.

Other studies have looked at the effect on interjudge disparity of relaxing the mandatory nature of the guidelines, after the Supreme Court's decision in Booker v. United States in 2005 which greatly increased the degree of judicial discretion. Empirical work on the impact of Booker suggests increases in interjudge sentencing disparities (Scott 2010; Yang 2014). Besides researchers find that increased judicial discretion after Booker has led to large and robust increases in racial disparities (Fischman et al. 2012, Yang, 2015). It results that the loosening of

the binding nature of the guidelines in criminal matters in the United States has led to an increase in racial disparities in sentencing. Several authors (Fischman et al. 2012; Yang 2015, Rehavi and Starr, 2014) point out that the increase in racial disparities in sentencing observed after Booker is undoubtedly the result of greater discretion for judges but also the consequence of the prosecutorial choices that preceded the judicial sentencing decisions. The findings suggest that prosecutors would respond to increased judicial discretion after Booker by charging black defendants with binding mandatory minimum sentences.

Taken together, research on interjudge disparity in US state and federal courts suggest that significant and nontrivial variation in sentences across judges remains even in jurisdictions with mandatory guidelines. These results suggest that constraining judicial discretion did not eliminate this interjudge disparity.

3. Methods

We describe the design of the experiment in a first sub section (3.1.). Then we present the implementation and the sample (3.2).

3.1. Design of the experiment

Following the typology proposed by Harrison and List (2004), our controlled experiment corresponds to a framed field experiment, characterised by the selection of a panel of particular subjects (students of the ENM) and a significant contextualisation of the task to be carried out (the setting of child support, as if they were judges).

The experimental protocol consisted of placing subjects in a situation quite similar to the one they would encounter if they had to deal with divorce cases and focusing on the decision of setting the amount of child support. We gave to the subjects 48 vignettes, each one aiming at representing a frequent divorce situation involving children. These 48 situations were chosen on the basis of statistics from a representative database of divorce decisions at first instance (CEEE-TGI, 2012). We have retained six criteria to characterize each of the 48 vignettes. Four correspond to the criteria of the national child support guideline[4]: the size of the siblings for whom a child support award has be set, the type of accommodation of the child, the income of

[4] In 2010 the French Ministry of Justice produced a guideline for the determination of child support award. This guideline is advisory, judges being free not to set the amount prescribed by the guideline. The French guideline belongs to the flat Percentage of Income Model, which sets support as a percentage of only the noncustodial parent's income. The percentage is the same regardless of the paying parent's income, given the type of accommodation of the child and the size of the siblings.

the creditor parent, the income of the debtor parent. The two other criteria are the proposal of the creditor parent and the proposal of the debtor parent.

Two sizes of siblings were selected: one child and two children, representing 90% of divorce cases with children in the first instance[5]. Two accommodation situations were selected: main accommodation with the mother (every second weekend and half of the holidays) and almost exclusive accommodation with the mother (the child has very little accommodation with the father). For siblings of one or two children, these two types of accommodation are decided in just over 7 out of 10 first instance decisions.

For parental incomes, we searched in the CEEE-TGI 2012 database for the most statistically typical "father-mother" income couples. The first combination is where both parents have relatively close average incomes, but the father has a slightly higher income than the mother. For couples with one or two children living with the mother, the median incomes of this type of combination are respectively around 1,500 euros per month for the father and 1,600 euros for the mother. The second combination consists of a relatively low female income but higher than the minimum income (RSA) and a much higher male income. The analysis of the cross-distribution of parental incomes then led us to retain the following amounts: 1,000 euros per month for the mother and 1,900 euros per month for the father. In contrast, we add a third, rarer combination, in order to take into account situations where the mother earns more than the father. Here again, the precise amounts were determined by analysing income distributions to arrive at a father's income equal to 1,100 euros and a mother's income equal to 2,500 euros.

Concerning the proposals of the parties[6], the most frequent combinations of proposals are those where the proposed amounts are average and fairly close, i.e., taking the median values in such a configuration, an offer of 140 euros made by the father and a request of 200 euros made by the mother[7]. The second case is less frequent but contrasts with the previous case because of a more pronounced disagreement between the parents. The precise amounts retained correspond to the median values calculated from the cases corresponding to this type of combination of proposals, i.e. an offer equal to 100 euros and a request equal to 300 euros. Thirdly, we have also retained the fairly frequent case where the father does not want to pay a child support award, so his offer is 0, and the mother asks for a relatively standard amount of child support, i.e. a median request equal to 150 euros. Finally, we chose to retain the fairly frequent situation where the father does

[5] Whether or not the children are over the age of majority can affect the amount of support. In order to avoid this dimension being a source of heterogeneity, we have specified the ages of the children by using median ages according to the size of the siblings (5 years for siblings of one child, 6 and 10 years for siblings of two children).
[6] The parents' proposals are not taken into account by the child support guideline. Nevertheless these proposals are crucial in the judicial decision-making since, in French law, the judge has to make an *infra petita* decision.
[7] The cases where the parents agree on the child support award are not considered in the experiment.

not make a precise offer (he is willing to pay a child support, but less than the mother's expressed demand, without explicitly indicating the precise amount) and the mother asks for a relatively standard amount of support, i.e. an undefined offer and a demand equal to the median value of the standard demands, i.e. 150 euros.

In total we have therefore retained 2 x 2 x 3 x 4 = 48 test cases.

3.2. Implementation and sample

The experiment took place on 13 October 2017 with students from the *Ecole Nationale de la Magistrature* (ENM). Each of the students was asked to decide on an amount of child support for 48 different test cases. Half of them (151 students) had at its disposal the child support guidelines of the Ministry of Justice (in a simplified version, for a quicker reading), while the other (161 students) did not. None of the sub-group was aware that the experiment concerned the use of a guideline. A first session took place without the guidelines, then, a second session organized without delay (to avoid contact between the students of the two sub-groups), took place with the guideline (the second sub-group was therefore unaware that the first had not benefited from the guideline). The distribution of the subjects in the two groups was carried out on the basis of pre-existing training groups set up by the pedagogical team of ENM, according to a logic of mixity in terms of background. This led to two homogeneous sub-groups in terms of socio-demographic characteristics (see Table 1).

Each of the 48 cases was summarized in a document, which was very visual for ease of use, presenting the test cases and including a box for entering the amount of child support. In order to test whether the order of the cases could influence decisions, we produced four sets of answer sheets, each organized in a different order, and these sheets were distributed randomly to the students (see Appendix 1). The first group was asked to set an amount of child support without any comments on the existence of guidelines. The second group was asked to do the same exercise and received a simplified version of the Ministry of Justice's advisory guideline. They were simply told that they were free to use it or not. In both cases, the introduction to the exercise was brief: a simple explanation of the context and how to fill in the document. Besides, we explained to them that since the cases were very simplified, they had to consider that any information not given in the test case, but which they could think of, should be considered identical from one case to another. No oral questions were accepted during the proceedings. No chatting with a neighbour was allowed. A few students asked comprehension questions individually at the beginning of the exercise, we only answered them when they were questions of

understanding, we abstained when the question more or less amounted to asking for help on the best way to proceed. The students completed the exercise with varying degrees of speed, but none of them ran out of time.

The 312 experiment sheets (one per student) had been filled in very correctly and practically without missing data. The few missing data relate to the information sheet and not to child support amount decisions. These missing data led us to discard two sets of responses (one in each of the two groups) and to make 11 imputations for missing data (out of a total of 3,100: 10 items on the information sheet * 310 students).

Table 1: Characteristics of the subjects, by sub groups

	Subjects with « guideline » Non-weigthed (1)	**Subjects with « guideline » Weighted (2)**	**Subjects with « no guideline » (3)**
Men	23,3%	27,5%	27,5%
Age	29,2	28,6	28,6
Married couple	13,3%	12,5%	12,5%
"Pacsé" couple	16,7%	10,6%	10,6%
Non-married couple	19,3%	25,6%	25,6%
Not in couple	50,7%	51,3%	51,3%
With children	16,7%	16,3%	16,3%
Curriculum in law only	68,7%	68,1%	68,1%
Curriculum in law + other	22,0%	25,0%	25,0%
No Curriculum in law	9,3%	6,9%	6,9%
Worked in a profession before entering ENM	42,7%	33,8%	33,8%
Ever handled a divorce case	46,0%	50,0%	50,0%
Altruism scale	4,79	4,77	4,91
Risk scale	4,17	4,10	4,04
Inequality scale	7,65	7,71	7,90
N	**151**	**151**	**161**

Source : base de données Expérimentation Barème ENM 2017.

4. Results

We present the impact of the child support guideline on the level of child support award (4.1.). and on the inter-judge disparity (4.2).

4.1. The impact of the guideline on the level of child support awards

It is quite difficult to put forward a relevant hypothesis about the effect of the guidelines on the level of child support. However, we could simply point out that when the guidelines propose an amount that is far from the parties' proposals, it is likely that the judge's decision will tend to

move closer to the guidelines amount (i.e., up or down depending on whether the distance is positive or negative).

In fact, on the basis of all 48 cases of the experiment, we observe that the average child support decided by the subgroup "with guidelines" is slightly higher than the average of the decisions made by the subgroup "without guidelines": 150.3 euros *versus* 146.6 euros, and this difference is statistically significant at the threshold of 0.1%. The large size of the two sub-samples of decisions (7,676 "without guidelines" decisions and 7,200 "with guidelines" decisions) may explain why such a small difference is statistically significant. What about each of the test cases? The effect of the guidelines is significantly positive (higher average amount of child support) in 20 cases, significantly negative in 14 cases, and the effect is not significant in 14 cases. Are students who have had the opportunity to use the guidelines more sensitive to them, in the sense of setting a significantly different amount than other students, when dealing with a case with one criterion than when dealing with a case with another criterion? The econometric analysis presented in Table 2 helps to answer this question.

On average, and all other things being equal, when dealing with a case of almost exclusive accommodation with the mother, students in the sub-group "without guidelines" set an amount about 17 euros higher than in cases of main accommodation with the mother; in the same situation, students "with guidelines" increase the child support even more (23 euros), and the difference (attributable to the possibility of using the guidelines) of 6 euros is statistically significant. From the point of view of sibling size, the two sub-groups decide in a more similar way: they award lower child support (-13 euros) when dealing with a case with two children than when dealing with a case with one child, and this in almost the same proportion (the estimated difference of 2 euros is only significantly different from zero at the 5% threshold). The use of the guidelines would therefore not have a major impact from this point of view.

Compared to the fairly common situation where the parties make average and not very different proposals (140-200), the students, whoever they are, set significantly lower amounts when demand is low (150), and this undercutting behaviour is influenced by the possibility of using the guidelines, since students "with guidelines" undercut significantly less (about 5 and 8 euros respectively for the two cases with an offer equal to 150) than students "without guidelines". With a pair of proposals "unspecified offer-150", the reduction for the sub-group of students "without guidelines" is estimated at -32 euros and for the sub-group of students "with guidelines" it is estimated at -27 euros.

Table 2. Estimated amount of child support

Constant	157.81 ***
With guidelines	0.009
Without guidelines	Ref.
Woman	0.14
Male	Ref.
Age	-0.11
In a common-law relationship	0.20
Pacse(e)	-0.79
Married	-0.98
Not in couple	Ref.
With child(ren)	1.19
No children	Ref.
No law school background	-3.45 ***
Law school + other education	-5.11 ***
Only law school	Ref.
Worked before ENM	-1.16
Did not work before ENM	Ref.
Ever handled a divorce case	-0.35
Never handled a divorce case	Ref.
Altruism scale	-0.03
Inequality aversion Scale	0.35 *
Risk aversion scale	-0.02
Almost exclusive accommodation with the mother	17.39 ***
Main accommodation with the mother	Ref.
Siblings of two children	-13.31 ***
Siblings of one child	Ref.
Proposals « unspecified offer –150 »	-31.93 ***
Proposals « 0 – 150 »	-33.91 ***
Proposals « 100 – 300 »	2.55 *
Proposals « 140 – 200 »	Ref.
Incomes « 1 100- 2 500 »	-40.53 ***
Incomes « 1 900 – 1 000 »	24.15 ***
Incomes « 1 600 – 1 500 »	Ref.
Lot 1	10.15 ***
Lot 2	19.13 ***
Lot 4	13.70 ***
Lot 3	Ref.
Guidelines * Almost exclusive accommodation with the mother	5.68 ***
Guidelines * Siblings of two children	-2.11 *
Guidelines * Proposals « ? –150 »	5.30 ***
Guidelines * Proposals « 0 – 150 »	8.10 ***
Guidelines * Proposals « 100 – 300 »	-3.05 *
Guidelines * Incomes « 1 100 – 2 500 »	-10.87 ***
Guidelines * Incomes « 1 900 – 1 000 »	9.13 ***
Log-likelihood	-72 575
N	14 876

Source: 2017 Guideline Experiment at ENM database. ***: significant at the 0.1% threshold. **: significant at the 1% threshold. *Significant at the 5% threshold.

With a pair of proposals "0-150", the reduction for the sub-group of students "without guidelines" is estimated at -34 euros and for the sub-group of students "with guidelines" it is estimated at -26 euros. However, this reduction behaviour, compared to the cases with a pair of proposals "140-200", is not observed when the students deal with cases with proposals equal to "100-300".

Finally, from the point of view of the pair of income, two opposite behaviours are observed. Compared to the median situation where both parents have similar average incomes (1,600-1,500), students set higher child support amounts in cases where the father earns significantly more than the mother (1,900-1,000)[8] and lower amounts in the opposite case (1,100-2,500)[9]. But what is interesting to note here is that these two behaviours are reinforced by the possibility of using the guidelines, the differences (-11 euros and +9 euros) between the two subgroups being statistically significant.

It should also be noted that the individual characteristics of students have very little influence on decision-making behaviour, with the exception of their educational background: those who have only attended law school are significantly more generous. Likewise, it can be noted that, from the point of view of the order in which the cases are proposed, the specificity of lot n° 3 is highly significant.

So far, we have used a standard econometric methodology using simple linear regressions. However, the nature of the data, because they are experimental, suggests that this type of tool is not the most relevant. The general idea is as follows. The effects of case characteristics (the seven modalities of the five criteria: number of children, type of accommodation, income pair and proposal pair) on the amount of child support would depend in part on the characteristics of the students. Even if we control for known individual characteristics of students, it is not excluded that there is a correlation due to unobserved characteristics. Indeed, since each of the 48 responses is processed by the same student (and this for each of the 310 students), our estimates may suffer from a non-observation bias and in this case the standard deviations of the estimators would be biased[10].

Hence the need to use a multi-level model where each of the 310 sets of 48 responses is "nested" within the 310 students. From a specification point of view, we considered that each of the effects of case characteristics could vary across students partly randomly and partly determined

[8] +24 euros for the sub-group "without guidelines" and +33 euros for the sub-group "with guidelines".
[9] -41 euros for the sub-group "without guidelines" and -51 euros for the sub-group "with guidelines".
[10] They are underestimated, giving the illusion that such effect is significant when it is not.

by the ten known student characteristics (gender, age, marital life, child(ren), education, experience in the legal treatment of divorce, pre-ENM occupation, risk, inequality, altruism).

To capture the effects of the guidelines in such a cross-effect specification, the coefficients involved in the interactions and the direct effects should be summed and the significance of these sums of coefficients tested. In Table 3, we have appended the results from these types of calculation applied to the coefficients estimated using mixed regression (see Table 2 above) and to those estimated using a multi-level model. The two sets of results are very similar, with coefficients that are identical to within three or four euros at most, identical signs and essentially identical levels of significance. Therefore, we will dispense with commenting on the results from the multilevel estimation procedure, as these comments would be identical in all respects to what we have written above.

Table 3: the effect of the guidelines according to the criteria of the cases

MULTI-LEVEL MODEL	**Without guidelines**	**With guidelines**	**Differences**
Almost exclusive… *versus* Main accommodation with mother	14,91 ***	21,21 ***	+6,30 ***
Two children *versus* one child	-11,09 ***	-13,83 ***	-2,74
Proposals « 100 – 300 »…	-0,26	-2,74 *	-2,48
Proposals « 0 – 150 »…	-32,43 ***	-23,83 ***	+8,60 ***
Proposals « non explicite-150 »… *versus* Proposals « 140-200 »	-30,62 ***	-24,48 ***	+6,14 ***
Incomes « 1 100- 2 500 »…	-38,78 ***	-49,66 ***	-10,88 ***
Incomes « 1 900 – 1 000 »… *versus* Incomes « 1 600 – 1 500 »	19,72 ***	29,21 ***	+9,49 ***

MIXED REGRESSION	**Without guidelines**	**With guidelines**	**Differences**
Almost exclusive… *versus* Main accommodation with mother	17,39 ***	23,07 ***	+5,68 ***
Two children *versus* one child	-13,31 ***	-15,42 ***	-2,11 *
Proposals « 100 – 300 »…	2,55*	-0,50	-3,05 *
Proposals « 0 – 150 »…	-33,91 ***	-25,81 ***	+8,10 ***
Proposals « non explicite-150 »… *versus* Proposals « 140-200 »	-31,93 ***	-26,63 ***	+5,30 ***
Incomes « 1 100- 2 500 »…	-40,53 ***	-51,40 ***	-10,87 ***
Incomes « 1 900 – 1 000 »… *versus* Incomes « 1 600 – 1 500 »	24,15 ***	33,28 ***	+9,13 ***

Source: 2017 Guideline Experiment at ENM database. N = 14 876. ***: significant at the 0.1% threshold. **: significant at the 1% threshold. *Significant at the 5% threshold.

It is therefore legitimate to question the value of using multi-level modelling. Only the coefficient associated with the "Incomes 1,900-1,000" modality varies significantly by method (24.15 *versus* 19.72), but this does not change the conclusions about the impact of this case characteristics modality on the level of child support. The multi-level methodology, because it takes into account intra-group correlations, corrects for standard deviations of the non-independence bias in the observations, thus giving more credence to the use of significance thresholds to select the most robust effects. In this respect, it can be seen that the explanatory factors whose coefficients were insignificant (at the 5% threshold) with the mixed linear regression method are no longer significant with the multilevel method, which makes it possible to decide in favour of the absence of an effect.

This is the case for the two interaction terms "Guidelines * Proposals 100-300" and "Guidelines * Two Children". Students therefore reduce child support for siblings of two in comparison to child support for only one child. The reduction is comparable in magnitude for both students with access to the guidelines and students without access. And the absence of any significant difference between child support for children whose parents' proposals are "100-300" and those whose parents' proposals are "140-200" can be observed whether or not the students have the option of using the guidelines.

4.2. The impact of the guideline on the disparity of child support amounts

The variance calculated on all child support amounts is greater when students can use the guidelines than when they cannot (2,498 *versus* 2,272), but this positive difference in variance is due solely to the difference in variance between groups (between test cases)[11]. Conversely, the variance within groups is lower when students have the possibility of using the guidelines than when they do not (817 *versus* 1048). This negative difference makes it possible to validate the hypothesis that the use of a guidelines would reduce the disparity of legal decisions of child support amount. However, it should be pointed out that the negativity of the difference in variance "with - without guidelines" is not observed for all the test cases analyzed, but only in 36 out of 48 cases[12]. This non-systematism (similar to what we observed with regard to the impact

[11] Using statistical indicators of influence we show that the magnitude of these variances is not due to a few students who would have made decisions quite systematically and very different from those of other students (outliers). And the individual characteristics of the students would not be statistically related, *ceteris paribus*, to their individual contribution to the variance. The results of these additional analyses are available from the authors.

[12] The negative difference is significant at the 10% threshold for 30 out of 36 cases and the positive difference is significant for 10 out of 12 cases.

on the level of child support amount) leads us to investigate why, in certain types of cases, the use of the guidelines generates an increase in the heterogeneity of decisions.

To do this, we estimate the 96 variances with the characteristics of the test cases as independent variables (Table 4). The estimation without interaction (column 1) shows us mainly that the possibility of using the guidelines does have a significant negative effect (-216) on the one hand, and that the levels of variance do depend on case characteristics (for example, decisions are significantly less heterogeneous when students are dealing with cases involving two children than when they are dealing with cases involving one child), on the other hand. The specification in column 2 addresses, by incorporating interaction terms, the issue of the specific impacts of the guidelines according to the characteristics of cases. The first lesson that can be drawn from this regression is that the impact of the guidelines would not differ according to the criteria relating to children (sibling size and type of accommodation).

Table 4: estimations of the variance of the 96 cases according to their criteria

	(1)	**(2)**	**(3)**
Constant	531***	586***	721***
With guidelines	-216*	-327	-558*
Without guidelines	Ref.	Ref.	Réf.
Almost exclusive accommodation with mother	354***	206	374**
Main accommodation with the mother	Ref.	Ref.	Réf.
Siblings of two children	-272**	-206	-170
Siblings of one child	Ref.	Ref.	Réf.
Proposals « unspecified offer –150 »	220	165	257
Proposals « 0 – 150 »	298*	258	350
Proposals « 100 – 300 »	660***	948***	740***
Proposals « 140 – 200 »	Ref.	Ref.	Réf.
Incomes « 1 100- 2 500 »	126	265	223
Incomes « 1 900 – 1 000 »	438***	110	388*
Incomes « 1 600 – 1 500 »	Ref.	Ref.	Réf.
Guidelines * Almost exclusive accommodation with the mother	/	295	175
Guidelines * Siblings of two children		-132	-194
Guidelines * Proposals « unspecified offer –150 »		110	177
Guidelines * Proposals « 0 – 150 »		81	148
Guidelines * Proposals « 100 – 300 »		-576*	-332
Guidelines * Incomes « 1 100 – 2 500 »		-278	-375*
Guidelines * Incomes « 1 900 – 1 000 »		656**	430
Guidelines < offer	/	/	-431
Guidelines > demand			-814***
Guidelines > offer and < demand			Réf.
Guidelines * (Guidelines < offer)			742*
Guidelines * (Guidelines > demand)			726*
R^2 adjusted	*0.36*	*0.52*	*0.65*
N	*96*	*96*	*96*

Source: 2017 Guideline Experiment at ENM database. ***: significant at the 0.1% threshold. **: significant at the 1% threshold. *: Significant at the 5% threshold.

Table 5 summarizes the results for the other two criteria (after summing the interaction coefficients with the coefficient associated with the "with guidelines" modality for the "with guidelines" test cases) and adds the same type of results from a regression where the dependent variable is not the variance but the coefficient of variation (relative dispersion).

With regard to the parties' proposals, it can be observed that the difference in effect (compared to the reference situation "140-200") of the potential use of the guidelines on the variance is only significant for the pair of proposals "100-300". Thus, in the case of very different proposals, the potential use of the guidelines significantly reduces the variance more than it does for proposals that are close to each other[13]. But this difference is not observed if heterogeneity is measured in terms of relative dispersion.

Table 5: Effect of guidelines on Dispersal of Decisions according to proposals and incomes criteria

	Variance	**Coefficient of variation**
Proposals "Non-explicit–150" with guidelines *versus* without guidelines	-217	-0,053
Proposals "0–150" with guidelines *versus* without guidelines	-246	-0,065
Proposals "100–300" with guidelines *versus* without guidelines	-903*	-0,079
Proposals "140–200" with guidelines *versus* without guidelines (Ref.)	-327	-0,044
Incomes "1 100–2 500" with guidelines *versus* without guidelines	-605	-0,083
Incomes "1 900–1 000" with guidelines *versus* without guidelines	329**	0,020**
Incomes "1 600–1 500" with guidelines *versus* without guidelines (Ref.)	-327	-0,044

Source: 2017 Guideline Experiment at ENM database. **: significant at the 1% threshold. *Significant at the 5% threshold.

When the parental couple's income is highly unequal and in favor of the father (1,900 - 1,000), the eventual use of the guidelines leads to an increase in variance which, compared to the negative effect observed for couples with equivalent middle incomes (1,600 - 1,500), is very significantly different from the reduction in variance observed for these couples with similar incomes[14]. The same type of effect is observed by studying the coefficient of variation.

[13] Further calculations allow us to note that the difference (in the effect of the guidelines on the variance) is also statistically different (at the 1% threshold) between this pair of proposals (100-300) and respectively the other two pairs of proposals "0-150" and "unspecified offer-150", but the differences are not significant are not significant in the regression relating to the estimation of the coefficient of variation. On the other hand, there would be no significant difference between these last two pairs of proposals in terms of variance (-217 is not statistically different from -246) or in terms of coefficient of variation (-0.053 is not statistically different from -0.065).

[14] Additional calculations show that the difference is also significant between the "1,100-2,500" and the "1,900-1,000" income couples: -605 is different from +329 at the 0.1% threshold. This significant difference is also observed in terms of the coefficient of variation (-0.083 is statistically different from +0.020).

In French law, there is a procedural rule according to which a judge must decide *infra petita* (except in very special situations which must be justified). In the case of child support, this means that the judge must choose an amount between the offer and the demand expressed by the parties. To go further in explaining the variability of the effect of the guidelines on the heterogeneity of decisions, it may be interesting to study the impact of this procedural rule. Indeed, when the guidelines suggest an amount outside the range of proposals, it may be thought that it encourages some judges, but not all, to depart from this rule. It may therefore be hypothesized that a suggestion of the guidelines outside the parties' proposals is a source of heterogeneity. Estimate No. 3 in Table 4 tests this hypothesis by identifying cases where the value suggested by the guidelines is either higher than demand or lower than supply.

All other things being equal, the effect of using the guidelines in a situation where the suggestion of the guidelines is within the range of proposals is -558 points of variance, compared with +168 points (726 - 558) for cases where the guidelines suggestion is greater than demand, and +184 points (742 - 558) for cases where the guidelines suggestion is less than supply. These econometric results thus confirm that when the value suggested by the guidelines is outside the range of proposals, the possibility of using the guidelines produce more heterogeneous decisions. In both cases, the differences in variance between cases with suggestions "in the interval of proposals" and, respectively, suggestions "greater than demand" and "less than supply" are statistically significant (cf. the significance of the coefficients associated with the interaction terms). In this new specification, the positive coefficient associated with the interaction between the guidelines and the 1900-1000 income pair loses its significance. This variance-increasing effect, contrary to the general hypothesis concerning the expected effect of a guidelines, would therefore in fact be linked, at least in part, to the position of the value suggested by the guidelines within or outside the range of proposals.

We explore this idea using the regression presented in Table 6, which specifies all possible combinations of income and proposal pairs. The suggestion of the guidelines is systematically (regardless of the number of children and type of accommodation) greater than the demand when the couple has 1900-1000 income type and the demand is equal to 150. In such a configuration we therefore expect, according to our hypothesis, to observe an effect of increasing the variance associated with the possible use of the guidelines; this effect is indeed estimated to be positive and very significant (interaction coefficients equal to 1002 and 1228).

Conversely, the guideline suggestion is systematically lower than the offer when the couple has 1100-2500 income pair and the offer is equal to 140 or 100. In this case, an increase in the

variance associated with the use of the guideline is again expected to be observed, but this positive effect is not confirmed by the estimate.

Table 6: estimation of the variance of the 96 cases taking into account the combination of income and proposals criteria

Constant	665***
With guidelines	-370
Without guidelines	Réf.
Almost exclusive accommodation with the mother	206*
Main accommodation with the mother	Réf.
Siblings of two children	-206*
Siblings of one child	Réf.
Incomes « 1 600 – 1 500 » + Proposals « 100 – 300 »	1028***
Incomes « 1 600 – 1 500 » + Proposals « 0 – 150 »	80
Incomes « 1 600 – 1 500 » + Proposals « unspecified offer – 150 »	-53
Incomes « 1 900 – 1 000 » + Proposals « 100 – 300 »	1597***
Incomes « 1 900 – 1 000 » + Proposals « 0 – 150 »	36
Incomes « 1 900 – 1 000 » + Proposals « unspecified offer – 150 »	-76
Incomes « 1 900 – 1 000 » + Proposals « 140 – 200 »	-61
Incomes « 1 100 – 2 500 » + Proposals « 100 – 300 »	357
Incomes « 1 100 – 2 500 » + Proposals « 0 – 150 »	796***
Incomes « 1 100 – 2 500 » + Proposals « unspecified offer – 150 »	763***
Incomes « 1 100 – 2 500 » + Proposals « 140 – 200 »	200
Incomes « 1 600 – 1 500 » + Proposals « 140 – 200 »	Réf.
Guidelines * Almost exclusive accommodation with the mother	295*
Guidelines * Siblings of two children	-132
Guidelines * (Incomes « 1 600 – 1 500 » + Proposals « 100 – 300 »)	-512
Guidelines * (Incomes « 1 600 – 1 500 » + Proposals « 0 – 150 »)	118
Guidelines * (Incomes « 1 600 – 1 500 » + Proposals « unspecified offer – 150 »)	184
Guidelines * (Incomes « 1 900 – 1 000 » + Proposals « 100 – 300 »)	-559
Guidelines * (Incomes « 1 900 – 1 000 » + Proposals « 0 – 150 »)	1202***
Guidelines * (Incomes « 1 900 – 1 000 » + Proposals « unspecified offer – 150 »)	1228***
Guidelines * (Incomes « 1 900 – 1 000 » + Proposals « 140 – 200 »)	543
Guidelines * (Incomes « 1 100 – 2 500 » + Proposals « 100 – 300 »)	-147
Guidelines * (Incomes « 1 100 – 2 500 » + Proposals « 0 – 150 »)	-569
Guidelines * (Incomes « 1 100 – 2 500 » + Proposals « unspecified offer – 150 »)	-574
Guidelines * (Incomes « 1 100 – 2 500 » + Proposals « 140 – 200 »)	-34
R^2 adjusted	*0,79*
N	*96*

Source: 2017 Guideline Experiment at ENM database. ***: significant at the 0.1% threshold. **: significant at the 1% threshold. *: Significant at the 5% threshold.

In conclusion, the experimental protocol implemented shows two salient results. First, the impact of potential use of the guideline is most often to reduce the heterogeneity of child support awards. Then, there are specific situations where the likely use of the guideline results in an increase in the variance of decisions. These cases are characterized by the presence of a significant income gap within the couple to the benefit of the debtor and a modest demand expressed by the creditor. We interpret this result by the fact that some subjects would correct the apparent inconsistency between a high debtor's income and a low demand of child support

comparatively to what suggests the guideline, while others would validate the amount claimed by the creditor on the basis that a judge cannot judge *ultra petita*.

5. Conclusion

This research presents a framed field experiment on the effect of a child support guideline on the interjudge disparities. This experiment produced different salien results. Firstly we show that the impact of an advisory guideline is of different magnitude and of different signs (the variance with the guideline may be higher or lower than without guideline). Then, on average (i.e. for the 48 typical cases considered simultaneously), we find evidence that an advisory guideline reduce interjudge disparities, since we show that the intra-group variance is lower when the subjects have the opportunity to use the guideline. Nevertheless, we observe that this effect is not systematic, since we observe increases of the interjudge disparities for some cases. These cases are characterized by the presence of a significant income gap within the couple to the benefit of the debtor and a modest demand expressed by the creditor. We interpret this result by the fact that some subjects would correct the apparent inconsistency between a high debtor's income and a low demand of child support comparatively to what suggests the guideline, while others would validate the amount claimed by the creditor on the basis that a judge cannot judge *ultra petita*.

These results present two main limitations. Firstly, the subjects in the experiment were not judges, but students of ENM. As a result, it is likely that some of the observed effects are exaggerated, comparatively to what could be observed in a real judicial context. In particular, we may point out that a significant proportion of subjects ruled outside the range of the parties' proposals, when this situation is extremely rare in judicial decisions. Nevertheless, even if this type of behaviour is over-represented, it gives us clues to understand why the introduction of a guideline can lead to an increase in inter-judge variation. Second, for organizational reasons, we opted for an experiment with two sub-groups of subjects, one ruling without a guideline and the other with a guideline. This option constitutes a limitation because it restricted us in our statistical exploitations by binding us to do analyses of differences in average subgroup decisions. It would have been more relevant to analyse differences in individual decisions. This would have required all the subjects to rule successively without and then with a guideline on the forty-eight typical cases, which was not technically possible.

References

Anderson J. T., Kling J. R., Stith K. (1999), "Measuring Inter-judge Sentencing Disparity: Before and After the Federal Sentencing Guidelines", *Journal of Law and Economics*, 42(1), 271-307.

Argys L. M., Peters H. E. et Waldman D. M., (2001), "Can the Family Support Act Put Some Life Back Into Deadbeat Dads? An Analysis of Child-Support Guidelines, Award Rates, and Levels", *The Journal of Human Resources*, 36(2), pp. 226-252.

Boulu-Reshef B., Comeig I., Donze R., Weiss G. D (2016), "Risk aversion in prediction markets: A framed-field experiment", *Journal of Business Researc*h, Vol.69(11), pp.5071-5075.

Cohen A., Yang C. S. (2019), "Judicial politics and sentencing decisions", *American.Economic.Journal: Economic.Policy*, 11:160–91.

Fischman, J.B., Schanzenbach M. M. (2012), "Racial Disparities under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums", *Journal of Empirical Legal Studies*, 9:729–64.

Gennaioli N., Shleifer A. (2008), "Judicial Fact Discretion", *Journal of Legal Studies*, 37, pp. 1-35.

Guthrie C., Rachlinski J. J., Wistrich A. J. (2001), "Inside judicial mind", *Cornell Law Review*, 86, pp. 778-830.

Harrison G. W., List. J. A. (2004), "Field Experiments." *Journal of Economic Literature*, 42 (4): 1009-1055.

Hofer P. J., Blackwell K. R., Ruback R. B. (1999), "The effect of Sentencing Guidelines on Inter-Judge Sentencing Disparity", *The Journal of Criminal Law and Criminology*, 90(1), 239-321.

Jukka J. (2012), "Framed Field Experiment with Stock Market Professionals*", Journal of Behavioral Finance*, Vol.13(4), pp.251-258.

Lacasse C. and Payne A. A., (1999), "Federal Sentencing Guidelines and Mandatory Minimum Sentences: Do Defendants Bargain the Shadow of the Judge?", *The Journal of Law & Economics*, 42(S1), pp. 245-270

Marinescu I. (2011), "Are judges sensitive to economic conditions? Evidence from UK employment tribunals", *Industrial and Labor Relations Review*, 64(4), pp. 673–698.

Payne A. (1997), "Does Inter-Judge Disparity Really Matter? An Analysis of the Effects of Sentencing Reforms in Three Federal District Courts", *International Review of Law and Economics*, 17, pp. 337-366.

Schanzenbach M. M., Tiller E. H. (2007), "Strategic Judging under the U.S. Sentencing Guidelines: Positive Political Theory and Evidence." *Journal of Law, Economics, and Organization,* 23(1): 24–56.

Scott R. W. (2010), "Inter-judge sentencing disparity after Booker: A first look", *Stanford Law Review,* 63:1–66.

Spamann H., Klöhn L. (2016), "Justice Is Less Blind, and Less Legalistic, Than We thought: Evidence from an Experiment with Real Judges", *Journal of Legal Studies*, 45, pp. 255-280.

Waldfogel J., (1998), "Does Inter-Judge Disparity Justify Empirically Based Sentencing Guidelines?", *International Review of Law and Economics*, 18, pp. 293–304.

Wistrich A. J., Rachlinski J. J., Guthrie C. (2015), "Heart Versus Head: Do Judges Follow the Law or Follow Their Feelings", *Texas Law Review*, 93, pp. 855-923.

Wooldredge J. (2010), "Judges' Unequal Contribution to Extralegal Disparities in Imprisonment", *Criminology*, 48(2), pp.539-567.

Yang C. S. (2015), "Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing", *The Journal of Legal Studies*, 44(1), pp. 75-111.

Appendix 1 -The 48 vignettes given to the subject

		1 child, 5 years old		2 children, 6 and 10 years old	
Income	**Proposals**	**Classical custody**	**Limited custody**	**Classical custody**	**Limited custody**
Father: 1900€ Mother: 1000€	**Father: 100 Mother: 300**	**1**	**5**	**25**	**29**
	Father: 0 Mother: 150	**2**	**6**	**26**	**30**
	Father: unknown Mother: 150	**3**	**7**	**27**	**31**
	Father: 140 Mother: 200	**4**	**8**	**28**	**32**
Father: 1100€ Mother: 2500€	**Father: 100 Mother: 300**	**9**	**13**	**33**	**37**
	Father: 0 Mother: 150	**10**	**14**	**34**	**38**
	Father: unknown Mother: 150	**11**	**15**	**35**	**39**
	Father: 140 Mother: 200	**12**	**16**	**36**	**40**
Father: 1600€ Mother: 1500€	**Father: 100 Mother: 300**	**17**	**21**	**41**	**45**
	Father: 0 Mother: 150	**18**	**22**	**42**	**46**
	Father: unknown Mother: 150	**19**	**23**	**43**	**47**
	Father: 140 Mother: 200	**20**	**24**	**44**	**48**

Lot 1: 1, 2, 3, 4, 5, 6, 7, 8 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48.

Lot 2: 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 1, 2, 3, 4, 5, 6, 7, 8, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 25, 26, 27, 28, 29, 30, 31, 32.

Lot 3: 3, 2, 1, 4, 7, 6, 5, 8, 11, 10, 9, 12, 15, 14, 13, 16, 19, 18, 17, 20, 23, 22, 21, 24, 27, 26, 25, 28, 31, 30, 29, 32, 35, 34, 33, 36, 39, 38, 37, 40, 43, 42, 41, 44, 47, 46, 45, 48.

Lot 4: 11, 10, 9, 12, 15, 14, 13, 16, 19, 18, 17, 20, 23, 22, 21, 24, 3, 2, 1, 4, 7, 6, 5, 8, 35, 34, 33, 36, 39, 38, 37, 40, 43, 42, 41, 44, 47, 46, 45, 48, 27, 26, 25, 28, 31, 30, 29, 32.