"Do jurors and professional judges differ in their treatment of crime?
Evidence from French reform"

Arnaud Philippe

Toulouse
School
of Economics

Do jurors and professional judges differ in their treatment of crime?

Evidence from French reform

Arnaud Philippe[1]

Institute for Advanced Studies in Toulouse;
Toulouse School of Economics; University of Toulouse Capitole
arnaud.philippe@iast.fr

January 2017

Abstract: Do citizens and professional judges agree on the accuracy of sentences? While surveys regularly point out a demand by citizens for harsher punishment, the differences between surveys' and real decisions' conditions are large enough to cast a doubt on the results. The introduction of two jurors into a court composed of three professional judges in two French regions and for a subsample of crimes in 2012 offers a good natural experiment for documenting the question of the differences between professional judges and citizens. Difference-in-differences or triple-difference methods do not permit me to identify any change in the probability of being convicted or in sentences given by a court including jurors. If some characteristics of the reform could partly explain those null results, they clearly go against the hypothesis of a major disagreement between professional judges and citizens when they have to make real decisions in criminal cases.

Keywords: courts, sentencing, crime, judicial decision, jury members

JEL codes: D83, K4, K14

1. Introduction

Numerous countries all over the world include juries in their criminal procedures. Juries are frequently used in the judgement of the most severe cases. However, they can also be associated with judicial proceedings (in the United States) or decisions on sentence enforcement (Belgium, Italy)[2]. Regarding trial decisions, jurors can decide convictions (England, United States) or both conviction and sentence (France, Germany, Italy).

The presence of jurors is usually justified by the need to connect judicial decisions and the evolution of society's perceptions. In particular, jurors tend to avoid judicial decisions to be "captured" by professional judges only—a group that has specific socioeconomic characteristics. Additionally, jurors are supposed to increase courts' legitimacy. For this reason, before 2000 in France, judicial decisions

---

[2] Some comparative law elements on jurors can be found at: http://www.senat.fr/leg/etudes-impact/pjl10-438-ei/pjl10-438-ei.html (in French)

made by juries could not be appealed because they were supposed to be the direct reflection of citizens' will[3].

Another justification of the presence of jurors in court is to increase society's understanding of the criminal justice system. Jurors obtain a personal experience of justice and, through discussion with friends and relatives, spread some form of knowledge that legitimates the institution.

While those arguments could be viewed as convincing, their empirical foundations have rarely been documented. Two main questions could be formulated. First, do jurors and professional judges make different decisions, and if yes, in what areas? Second, is the presence of jurors changing the way citizens perceive the judicial institution?

The media frequently addresses the first of these questions, through surveys. Those surveys usually conclude that the majority of citizens strongly desire harsher sentences[4]. The demand for harshness seems to be perceived by judges, as their sentences increase when the media closely follow their decisions (Lim et al., 2015). However, the fact that people ask for harsher sentences in surveys does not necessarily mean that citizens' and professional judges' preferences diverge. Indeed, the real conditions of a trial, the court's ceremonial, the study of defendants' lives in addition to their crimes, the importance of a decision that could put a person in jail, the ideal of justice that comes with that responsibility—all of these factors probably make real decisions made by one person differ greatly from what he could express in a survey.

A natural experiment that occurred in France allows documenting more precisely the differences between decisions made by professional judges and jurors. In 2012, the jurisdictions of seven counties depending on two appeals courts (Dijon and Toulouse) experimentally modified the composition of their criminal courts in order to include jurors. In practice, two jurors were added to the three professional judges in the judgement of severe violence and severe property offenses. The instigators of the reform—the right-wing government of Nicolas Sarkozy—presented it as a way to move justice closer to the will of the citizens. On the opposite side, a large number of judges saw the law as an act of defiance towards magistrates and a way to increase sentences for the targeted offenses. The opponents clearly perceived jurors as harsher than professional judges.

The way this reform was conducted constituted a perfect natural experiment to evaluate the differences between jurors and professional judges in real decisions. Indeed, only specific place and specific crimes during a limited period of time were affected by the reform. Based on this design, it is possible to use several methodologies to measure the effect of the introduction of jurors: difference in differences among targeted crimes using non-targeted courts as a control group; difference in differences using non-targeted crimes in targeted courts as a control group; and triple differences. Those methodologies allow me to test the assertion of the evaluation report of the reform, issued in fall 2012 by the new government and based on interviews that stated: "no element allows us to say that the reform induced harsher sentences"[5].

---

[3] This was modified by a law passed in June 2000. Those decisions can now be appealed.

4 This is, for example, the case in a survey realized in France in 2013. According to this survey, 76% of the people think that sentences should be harsher. See http://www.lefigaro.fr/actualite-france/2013/03/21/01016-20130321ARTFIG00366-les-francais-pour-une-justice-plus-severe-selon-un-sondage.php (in French). See also another survey published in French newspapers:
http://www.contrepoints.org/2016/04/06/245801-sondage-les-francais-veulent-une-justice-plus-dure;
http://www.lexpress.fr/actualite/societe/justice-les-francais-veulent-plus-de-severite-et-de-moyens-selon-un-sondage_1273313.html.

[5] Report to the ministry of Justice on the introduction of jurors in criminal courts in the jurisdiction of Dijon and Toulouse: "We first noticed that the introduction of jurors did not change the precedents in criminal courts. Especially, there is no

None of the methods used in this article provide any evidence of a significant change in the judicial decisions following the reform. Neither acquittal. nor the probability of being sentenced to prison, probation or suspended prison, nor the length of those sentences is affected. Standard errors of the estimates are small and the null effects could be interpreted as precisely estimated zeros. In addition, the dispersion of the sentences does not seem to be affected. Thus, the effect of the introduction of jurors is, at best, very limited.

However, this null result is not sufficient to claim that judges' and citizens' preferences are perfectly equal. Indeed, in the experimental courts, jurors were less numerous than professional judges—two versus three. This could have restricted the capacity of the jurors to participate in the debates and to affect the decisions. This restriction could even have occurred voluntarily if, for example, jurors relied on professional judges because they did not trust their own capacities. It could also occur in a more constrained process if jurors expressed different preferences but were always outvoted by professional judges. The explanation of the null result based on difficulties in imposing their preferences is hard to rule out in the institutional setting of the reform. However, the explanation based on the systematic minority position of the juror is hardly convincing. Indeed, this could be the case only if the preferences of professional judges were perfectly homogeneous. If they were not, the expression of divergent preferences of the jurors should have changed the equilibrium that might exist when the professional judges acted alone.

The literature on determinants of judicial decisions is relatively old. A large share of it uses mock juries to test for the effects of procedural characteristics, case characteristics or court composition (see Devine et al., 2001 for a review). Others use experimental design and surveys with fictitious cases (see, e.g., Guthrie et al., 2007, or English et al., 2006). However, the extrapolation of those results to real conditions is difficult. This is a common problem for lab studies, but it could be particularly the case regarding judicial decisions. First, decisions are particularly important: they could lead to sending someone to jail for several years. Second, real crimes drive emotions—and even trauma for the most severe crime—in a way that clearly makes them different from fictitious cases and that could distort people's behaviour. Lastly, the ceremonial nature of trials is especially designed to make offenders, witnesses, and experts as well as jurors aware of the importance of the moment. Those conditions could hardly be reconstructed in lab experiments. Then, real decisions and lab experiments could diverge largely.

To overcome those problems, several recent papers studied real judicial decisions. Among them. it is possible to distinguish at least three groups. The first group focused on the cognitive bias of the judges: gambler's fallacy (Chen et al., 2016), mood (Chen, 2014) or tiredness (Dantziger et al., 2011). The second group focused on the characteristics of the actors: judges, jurors, defendants, and victims. Those papers studied the importance of race (Anwar et al., 2012a), age (Anwar et al., 2012b), sex (Schanzenbach, 2005; Philippe, 2016), political opinions (Berdejó et Yuchtman, 2013; Anwar et al., 2014), familial structure (Glym et Sen, 2015), skills (for the lawyer: Abrams and Yoon, 2007). The last group focused on the influence of exogenous events: conflicts (Shayo and Zussman, 2011), media coverage of crime and judicial errors (Philippe and Ouss, 2016), media pressure on judges (Lim et al., 2015).

Those papers documented numerous factors affecting judicial decisions. They usually focused on either professional judges or jurors, and the importance of courts' experience is rarely addressed (see,

---

evidence of harsher sentences. Interviews go in the direction of a null effect or higher leniency due to empathy for the defendants while their personal life has been presented".
http://www.presse.justice.gouv.fr/art_pix/1_1_Rapport_bilan_experimentation_citoyens_assesseurs.pdf

however, Philippe Ouss, 2016). This comes from the fact that professional judges and jurors are usually involved in very different courts and that their decisions could hardly be compared (jurors are usually involved in the most severe cases). For this reason, the natural experiment that occurred in France offered a rare occasion to document the alleged differences between professional judges and jurors.

The paper is organized as followed. Section 2 provides an overview of the institutional context. Section 3 presents the data. Section 4 present the empirical strategy and section 5 the results using the strategies presented. The last section discusses the results and conclusions.

2. Institutional context

2.1. French penal structure

The French criminal code distinguishes three different types of crime, defined by their sentences and the courts that judge them.

The less severe crimes are the "contraventions". This group includes the vast majority of road-related offenses as well as minor violence. Infractions cannot be punished by prison. A single judge decides sentences.

The second category—the "délits", translated in this paper as "misdemeanour"—includes crimes that can be punished by sentences of up to ten years. It contains the vast majority of property crimes, violent crimes, drug-related offenses, and minor sex crimes as well as some road-related offenses (driving under influence or without a license) and other less numerous types of crime (e.g., forgery, insults). This is the type of crime this paper focuses on. Apart from the experiment studied here, the court is composed of three professional judges (both for the first-instance proceeding and for appeal). Between 600,000 and 650,000 such crimes are judged each year. In 2011, the year before the reform, 25% of the crimes were punishable by one prison year, 50% were punishable by two to four years and 25% were punishable by five years or more. In practice, sentences are far below the maximum. In 2011, the average sentence was equal to 46 prison days, 28 probation days and 20 suspended prison days. Of the defendants, 80% get no prison sentences, and only 4.4% get one year or more.

The third category is the most severe crimes. They are mainly sex crimes (rape, paedophilia, etc.), homicides, very severe violence and armed robbery. Those crimes are punishable by more than 10 years. Between 2,500 and 3,000 are judged per year. The court is composed of professional judges and jurors[6]. This is the only category in which jurors are included in the judicial process in France (apart from the experiment).

A map of the territorial organization of the French criminal justice system is presented in figure 1. Regarding misdemeanours, the metropolitan territory is organized in 155 first-instance courts (164 including Corsica and overseas territories) depending on 30 appeals courts (36 including Corsica and overseas territories). The superior court, held in Paris, only makes judgements based on the statutes. Those (very rare) judgements will not be considered in this work.

---

6 Nine jurors (12 in appeals) before 2012, 6 (9 in appeals) after 2012.

2.2. The reform introducing jurors in court judging misdemeanours

A law passed on 11 August 2011 introduced jurors into courts judging misdemeanours on an experimental basis beginning on 1 January 2012. The experimental courts were the appeals courts of Dijon and Toulouse and the nine first-instance courts that are under their authority. In those jurisdictions, two jurors were added to the three professional judges who normally hear the cases. This modification occurred only for a limited set of crimes: some violent crimes, sex crimes and property crimes that are punishable by five years or more.

The government justified the experiment by the fact that "the low participation of the citizens in the functioning of justice" leads to "decisions that do not take into account the evolutions of society"[7]. According to the government:

> citizens could think that judicial decisions do not take into account the evolution of the society. In comparison with other countries, one could also claim that the limited participation of the citizens in judicial decisions does not allow a good understanding of the institution (…). The law aims at increasing citizens' participation in the judicial system. Social cohesion and the state's legitimacy could only be reinforced by this reform.

Thus, officially, the goal of the reform was to move the citizens closer to their justice system. However, this is not the way the reform was interpreted by the opposition, the judges—and more generally the lawyers—and the press. The introduction of jurors into courts for misdemeanours was perceived as a sign of the ruling political establishment's distrust of judges[8]. Moreover, the law was interpreted as a way of increasing sentences—in a context in which judges were criticized as being too lenient—one year before the presidential election. Along those lines, during the parliamentary debates, the first speaker of the opposition denounced the law:

> the only meaning I see in your bill, Mr. Minister of Justice, is the distrust of judges. (…) Indeed, you want to supervise judges—whom you perceived as excessively lenient—with citizens—whom you expect to be more severe.

In the same line, the Syndicat de la magistrature (left-wing union representing between a quarter and a third of the magistrates) stated, "moreover, the second objective of this law—and probably the only one that is important for the political power—is to fight against the alleged leniency of the judges".

After taking place in the jurisdictions under the authority of the appeals courts of Toulouse and Dijon beginning on 1 January 2012, the reform was supposed to be generalized to the entire country beginning on 1 January 2013. The new minister of justice, Christiane Taubira, appointed in May 2012 after the presidential election[9], first asked for an audit of the reform by two high-ranking judges (see Salvat and Boccon-Gibod, 2013). Their report was based on interviews of judges, lawyers, defendants and former jurors in the treated jurisdictions. It concluded that the reform improved the perception of the justice system among jurors and their relatives; slowed down and weighted courtroom sessions mainly because the judges had to explain procedures to the jurors; did not affect sentences because jurors left those decisions to the professional judges; did not affect defendants; and increased the cost

---

[7] Preamble.

[8] This lack of trust has been symbolized by Nicolas Sarkozy's sentence comparing judges to peas with same color and same taste (October 2007, 7th, TV program "vivement dimanche prochain"). Less anecdotally, the law introducing mandatory sentences passed 4 years before was also perceived as a lack of trust.

[9] The presidential election held in May 2012 led to a change of the political majority. François Hollande, a liberal, was elected to succeed Nicolas Sarkozy, a conservative.

of justice. Based on those observations, the ministry of justice cancelled the experiment and gave up the reform on 1 January 2013.

At the end of the day, only specific crimes in the courts under the authority of 2 appeals courts —over thirty—were affected for a year.

3. Data

3.1. Discharges

The first decisions of interest are convictions. Discharges are not individually recorded in France. Indeed, the judicial system cannot record information on anyone who has not been convicted. However, the number of cases as well as the number of convictions is recorded at the court level. The difference gives the number of discharges.

This article uses information on discharges per first-instance courts (155 units) and per month between January 2011 and December 2012. For each court and month, the data set contains the number of cases and the number of discharges for two types of crimes: those targeted by the law—part of violent crimes and property crimes that are punishable by five years or more—and those that are not targeted among property and violent crimes (control group). In the latter group, only violent crimes and property crimes have been conserved. In particular, road-related offenses and drug crimes have been excluded, as none of them belong to the treatment group.

On average, discharges are rare. They represent 7.25% of the cases over the period. They slightly decrease in 2012 in the treated areas for both targeted crimes (7.4% to 7%) and non-targeted crimes (6.9% to 6.5%). At the same time, they remain stable (8.1%) for targeted crimes or slightly increase (7% to 7.2%) for non-targeted crimes in the control courts.

3.2. Sentences

The main data set used here comes from criminal record compilations from the statistics service of the French Ministry of Justice (Sous Direction de la Statistique et des Etudes). These compilations contain the results and details of almost all criminal cases judged each year. Only the smallest infractions (e.g., parking offenses in the "contraventions" category defined above) are not registered. In particular, the compilations contain all misdemeanours. Procedural characteristics, date, place, precise facts and descriptions of sentences are available. Sociodemographic variables such as age, sex, and nationality are also included.

Data from 2008 to 2014 are used in this paper. The reform studied here targeted only some, but not all, violent crimes and property crimes for which offenders could be sentenced to five or more years in prison. In order to minimize the distance between targeted and non-targeted crimes used as control group in some regressions, the dataset is restricted to violence and property crimes that could be sentenced to five years and more. In this subgroup, approximately 33% of the crimes are targeted.

The final data set contains 597,000 observations (over approximately 4 million misdemeanours and 1.3 million violent crimes and property crimes). Crimes that are targeted by the reform represent 33%

of this data set. The sample sizes of the different groups over the different periods are presented in table 1. The treated sample—i.e., treated infractions in treated courts in 2012—regroups 1,352 observations.

Descriptive statistics of this sample are presented in table 2. Columns 1 (treated areas) and 2 (other areas) present the descriptive statistics for targeted crimes. Columns 3 (treated areas) and 4 (other area) present the same information for non-targeted crimes.

Defendants are mainly males—this is even truer for targeted crimes, with 93% male—and French (approximately 80%). They are on average 30 years old, and the majority have been previously convicted of a crime. Targeted crimes receive more severe sentences: 50% get a prison sentence—compared to only 35% for non-targeted crimes—and the average length is 6 months—compared to less than 3 months for non-targeted crimes.

4. Empirical strategy

4.1. Method

The introduction of jurors into courts judging misdemeanours at one moment in time in a small group of courts and for a small group of crimes allows me to evaluate the reform using three different strategies: two difference in difference and one triple difference.

The first difference in difference is based on crimes that have been targeted by the law. It uses the courts under the authority of the appeals courts of Toulouse and Dijon as the treatment groups and the rest of France as the control group. The effects are measured using regressions of the form:

$$Y_{i,ac,t} = \alpha_{ac} + \gamma_t + \beta_1 * 1_{2012} * Treat_{ac} + \delta * X_i + \varepsilon_i \ (1)$$

where $Y_{i,ca,t}$ is the outcome of interest—conviction, prison sentence—for person i judged in the area of the appeals court ac during year t; α_{ac} are fixed effects for the areas depending on the appeals courts (30 different appeals courts); γ_t are year fixed effects; 1_{2012} is a dummy equal to one in 2012; $Treat_{ac}$ is a dummy equal to one in the area depending on the appeals courts of Toulouse and Dijon; X_i are control variables (crime, age, sex, etc.); and β_1 is the parameter of interest. It measures the evolution of variable Y in 2012 in treated areas. This strategy leads to my preferred estimate as treatment and control group are very close before the reform (see table 2).

The second difference in difference uses only trial in the treated areas. It uses non-targeted crimes as the control group. As mentioned before, the data set is restricted to types of crime—violent crimes and property crimes—that are punishable by five years or more in prison. Thus control group is composed of crime that are not very different from the targeted ones.

With this difference in difference, the effects are measured using regressions of the form:

$$Y_{i,delit\ t} = \theta_{délit} + \gamma_t + \beta_2 * 1_{2012} * Treat_{delit} + \delta * X_i + \varepsilon_i \ \ (2)$$

where $Treat_{delit}$ is equal to one for the crime targeted by the reform.

Finally, it is possible to combine those two approaches by using triple difference in difference strategy. The effects are measured using regressions of the form:

$$Y_{i,ca,délit,t} = \theta_{délit} + \alpha_{ca} + \gamma_t + \beta_3 * 1_{2012} * Treat_{ca}$$

$$\beta_4 * 1_{2012} * Treat + \beta_5 * 1_{2012} * Treat_{delit} * Treat_{ca} + \delta * X_i + \varepsilon_i \quad (3)$$

The parameter of interest is now β_5 that measure the specific evolution of targeted crimes in treated areas in 2012.

4.2. Balancing checks

The identification strategies presented here are valid only if offenders in the treatment groups are not correlated with any changes in offenders' characteristics. To test for this possible bias, it is possible to run balancing checks. The idea is to run the regressions 1, 2 and 3 presented above using offenders' socioeconomic or procedural characteristics as the outcome variables. Those variables are not available when defendants are not convicted. Thus, balancing tests could be realized only for offenders who were convicted.

The effects of the reform on age, nationality (dummy equal to one if the defendant is French), gender (dummy equal to one if the defendant is a woman), past conviction (dummy equal to one if the defendant has been previously convicted in the past five years), investigation length and pretrial detention days are successively measured. The first four variables should not be affected. The last two variables could be affected if the introduction of jurors slows down the court sessions.

The results are presented in table 3. Panel A, in the upper part of the table, presents the results for the difference in difference using non-targeted courts as a control group (following equation 1). Panel B, in the central part, presents the results for the difference in difference using non-targeted crime as a control group (following equation 3). Panel C, at the bottom, presents the results for the triple difference in difference (following equation 3). Sample sizes are presented after the label of the panel. It is larger for the triple difference in difference, which uses all of the relevant information (panel C, 597,156 observations), and lower for the difference in difference using non-treated crimes as a control group because this sample is restricted to the treated area (panel B, 27,299 observations). Only the relevant coefficients (β_1, β_2 and β_5) are presented.

Coefficients presented in table 3 are mainly small and not significant. According to the triple-difference estimate, the reform decreases the proportion of female defendants, but this result is not confirmed by the difference in difference. Table 3 confirms that the introduction of jurors into criminal courts judging misdemeanours did not affect defendant characteristics or procedural variables.

5. The effect of introducing jurors on judicial decisions

5.1. Graphic evidence

9

Before going into the details of the regressions, it is possible to present the evolution over time of the main variables for the different subgroups. As previously mentioned, four groups could be distinguished: targeted crimes in treated areas, targeted crimes in non-treated areas, non-targeted crimes in treated area and non-targeted crimes in non-treated areas.

The discharge rates per year and subgroup are presented in figure 2. They increase in the treated areas for targeted crimes while they decrease for the same crimes in non-treated areas. However, the standard deviations are large, as discharges are measured at the court level. Differences do not seem to be significant.

The evolution of the probability of being convicted to prison per group is presented in figure 3. Figure 4 presents the evolution of prison sentences' length per group (set to zero if no prison). Targeted crimes in treated areas do not seem to be more or less severely punished during 2012, the year of the reform. This is true both in comparison to preceding/upcoming years and in comparison to other groups.

The same absence of effect is observable for probation time or suspended prison time (see appendix A). A more precise visualization, per month and using a synthetic control method, is presented in appendix B. It leads to the same absence of results.

5.2. Main results

As cases characteristics are not affected by the reform, it is now possible to measure the effect of the introduction of jurors on judicial decisions. The main goal is to see if, as the opponents of the reform feared, the introduction of jurors induced harsher decisions. Several outcomes are of interest here. It is first important to measure whether the proportion of discharges is affected. It is both a crucial outcome and an important thing to measure before turning to sentences. If convictions are affected, and the pool of sentences observed changes over time, this must be taken into account in the analysis. The second main outcome is sentences. Several variables are used here: the probability of being sentenced to prison, probation or a suspended prison sentence and the length of the prison sentence, probation or suspended prison sentence.

Table 4 presents the main results. They are presented in a similar way as in table 3. Panel A presents the first difference in difference using a non-targeted court as a control group and based on equation 1; panel B presents the second difference in difference using non-targeted crimes as a control group and based on equation 2; Panel C presents the triple difference in difference.

The first column of table 4 presents the effect of the reform on discharges. The unit of observation is the court, which is why sample sizes are much smaller than for columns 2 to 7. The different estimates do not provide any evidence of an effect of the reform on discharges. Coefficients go from 0.5% to 1.5%, and the most precisely estimated one (in panel B) has a small standard error of 0.2%. Thus, column 1 could be interpreted as a null effect of the introduction of jurors on discharges even if a very small positive effect cannot be fully excluded. If anything, coefficients go in the direction of an increase in the discharge rate, which contradicts the idea that jurors are more severe.

Columns 2 to 7 measure the effect of the reform on sentences for convicted people. Columns 2-4 present the results on the extensive margin. The outcome variables are dummy equal to one if defendants are sentenced to any prison time (column 2), any probation time (column 3) or any suspended prison time (column 4). Columns 5-7 present the results at the intensive margin. Outcome

variables are the length, in days, of prison sentences (from 0 to 10 years, column 5), probation (from zero to 5 years, column 6) and suspended prison sentences (from 0 to 5 years, column 7).

Only one of the heightened coefficients presented in columns 2-7 is statistically significant. Using difference in difference for targeted crime (panel A), the probability of being sentenced to some suspended prison time (column 4) increases by almost 3% when jurors are present in the court. However, this result is not confirmed by the other two strategies.

Regarding the probability of getting some prison time (column 2), point estimates of the effect go from -0.5% (panel C) to -2% (panel A) with standard errors of approximately 1.5%. Considering that the average probability of getting prison time for targeted crimes is approximately 50%, this could be interpreted as a relatively well-estimated null effect. The same is true for the probability of getting a suspended prison sentence (column 3). Coefficients go from -0.3% to -2.2% with small standard errors (approximately 1.5%) for an average of 35% for targeted crimes. Note that all coefficients in columns 2 and 3 are negative. If anything, the effect seems to be in the direction of more lenient decisions when jurors are included.

Turning to length of sentences, the results are, once again, insignificant. They are marginally less precisely estimated—point estimates are slightly higher compared to the averages—but coherent in their absence of results. Regarding prison terms (column 5), point estimates go from -4 to 6 with a standard error of approximately 8, while average sentences for targeted crimes are approximately 185 days. If anything, the effect is limited. The same is true for probation (column 6) and suspended prison sentences (column 7).

The results for discharges are similar if partial discharge (instead of full discharge) is used as the outcome variable (Appendix C, table C1, column1). The results for sentences are similar if the sample is restricted to 2011-2012 except that the probability of getting some probation time seems significantly affected when using difference in difference (Appendix C, table C1, panel A and B, column 3). Lastly, permutation tests measuring the effect of placebo reforms occurring in the jurisdictions of two random appeals courts confirm that the results presented in table 4 are small and not significant (see appendix C, figure C1, for the distribution of the coefficients when measuring the effect of placebo reforms on prison time).

All in all, the results presented in table 4 do not support the idea that introducing jurors into the court had a general effect on convictions or sentences. All three methodologies used provide precisely estimated coefficients that are not different from zero. This is consistent with what was (not) observed in the graphs. If anything, the effects of the reform are small and go in the direction of more lenient judgement.

5.3. Variability of the sentences

Lastly, it is possible that, although the reform did not change judicial decisions, the introduction of jurors increased the variability of the sentences. Indeed, as jurors changed frequently and as they could have heterogeneous preferences—especially because they did not have a common norm or training, as did professional judges, who could be more uniform—they could distort the sentences in many different ways that sum to zero mean but increase heterogeneity.

If this was true, confidence intervals presented on figure 4 should increase for targeted crimes in treated courts in 2012. This does not seem to be the case. A more precise way to measure the effect of

the reform on sentence dispersion is to follow the evolution of the distance to mean sentences. The idea is to calculate the absolute value of the difference between individual sentences and average sentences for the same crime in the same court and then to run regressions of the form presented in the previous sections.

The results are presented in table 5. Column 1 presents the effect of the reform on prison time, column 2 on probation time and column 3 on suspended prison time. None of the coefficients are significant except the one for suspended prison using the first difference in difference.

The introduction of two jury members in courts did not significantly increase the dispersion of sentences.

5.4. Subgroups

Although the reform did not have a general effect on sentences, it is still possible that some subgroups were affected. This is possible if jurors are particularly sensitive to certain crimes. For example, it is possible that professional judges—who have high socioeconomic status and usually live in relatively rich neighbourhoods—are more sensitive to property crimes and less sensitive to violent crimes than the average citizen. In this case, the null effect observed in table 4 would mask the heterogeneous effect on different subgroups.

This hypothesis could not be tested for discharges, as this outcome is not available at the individual level.

The effect of the reform on prison sentences for different subgroups is presented in table 6. For tables 3 and 4, the coefficients from the three different methodologies are presented. The effects on violent crimes or property crimes are presented in columns 1 and 2. Columns 3 and 4 present the effects for first-time offenders or recidivists. The last two columns present the results for men and women.

None of the results are significant. Standard errors increased because of smaller sample sizes, but even beyond statistical significance, no clear pattern emerges from point estimates that are alternately positive and negative and always small. Thus, the general null result does not seem to hide heterogeneity.

7. Discussion and conclusion

Introducing two jurors into criminal courts composed of three professional judges did not have any detectable effect. Neither convictions nor sentences—at the extensive or the intensive margin—were affected. This absence of a significant result is observed with all three strategies used: difference in difference or triple difference. The effects of the reform, if any, were small and pointed in the direction of a decrease in the severity of the courts.

This null result could have different origins. First, it is possible that judges and jurors shared common preferences regarding convictions and sentences. This interpretation contradicts surveys mentioned by the instigators of the law, who stated that French citizens had asked for harsher punishments. However, as mentioned in the introduction, those surveys elicited general theoretical opinions on

sentences that could differ strongly from opinions on real cases. Moreover, the people surveyed could be more extreme when they had to express their opinions without making real decisions. While several studies find that experience can affect the way people decide cases (Devine et al. 2001; Philippe and Ouss 2016), the differences in average preferences seem far smaller than what is measured in surveys.

Even if it seems to be the most plausible one, the interpretation in terms of equality in preferences is not the only one. It is also possible that preferences diverged but that jurors were always outvoted by the professional judges. This could be the case, as the reform introduced two jurors into courts already composed of three professional judges. However, this possibility seems unlikely. Indeed, it could only be true if the professional judges had homogeneous preferences and voted for exactly same decisions. If this is not the case, then the average decision of professional judges will come from a mix of harsher and more lenient decisions that average to something in the middle. In this case, if jurors had preferences that were harsher (or more lenient) than the average decision of the professional judges, then they should have modified the equilibrium upward (or downward).

Lastly, it is possible that jurors mainly did not express any preferences. This could be the case for several reasons. First, their weight in the procedure was small. Of course, they voted on sentences, but they did not have the investigation files in advance, they did not lead the debates and, as mentioned before, there were only two of them. Second, their limited knowledge of law and their inexperience in practical judicial decisions could have made them consider their presence in the court more as an interesting experience than as participation in the judicial process. For those reasons, they could have relied on the professional judges' advice and let them decide the cases. The author of the audit when asked by the ministry of justice in 2012 stated this hypothesis.

Even if it is hard to fully rule out alternative hypothesis, the results presented in this paper clearly go against the idea of a great difference in preferences between professional judges and the rest of society. However, it is possible that limited divergences did not affect the equilibrium for procedural reasons.

Bibliographie

Abadie, A., Diamond, A., & Hainmueller, J. (2012). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association*.

Abrams, D. S., & Yoon, A. H. (2007). The luck of the draw: Using random case assignment to investigate attorney ability. *The University of Chicago Law Review*, *74*(4), 1145-1177.

Anwar, S., Bayer, P., & Hjalmarsson, R. (2016). *A Jury of Her Peers: The Impact of the First Female Jurors on Criminal Convictions* (No. w21960). National Bureau of Economic Research.

Anwar, S., Bayer, P., & Hjalmarsson, R. (2012a). The impact of jury race in criminal trials. The Quarterly Journal of Economics, 127(2), 1017-1055.

Anwar, S., Bayer, P., & Hjalmarsson, R. (2012b). The Role of Age in Jury Selection and Trial Outcomes. National Bureau of Economic Research Working Paper.

Anwar, S., Bayer, P., & Hjalmarsson, R. (2014). Politics and peer effects in the courtroom. National Bureau of Economic Research Working Paper.

Berdejó, C., and N. Yuchtman. (2013). Crime, punishment, and politics: an analysis of political cycles in criminal sentencing. Review of Economics and Statistics 95.3: 741-756.

Chen, D., Moskowitz, T., Shue, K. (2016). Decision-Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires », The Quarterly Journal of Economics, vol. 131, p. 1181–1241.

Chen, D. (2014). This Morning's Breakfast, Last Night's Game: Detecting Extraneous Influences on Judging. Working paper.

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

Devine, D. J., Clayton, L. D., Dunford, B. B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. Psychology, Public Policy, and Law, 7(3), 622.

Dowler, K. (2003). Media consumption and public attitudes toward crime and justice: The relationship between fear of crime, punitive attitudes, and perceived police effectiveness. Journal of Criminal Justice and Popular Culture, 10(2), 109-126.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. Personality and Social Psychology Bulletin, 32(2), 188-200.

Gissinger-Bosse, C (2012). L'impact de la participation citoyenne sur l'institution judiciaire :

une innovation en audience correctionnelle ? Sociologos

Glynn, A. N., & Sen, M. (2015). Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?. *American Journal of Political Science*, *59*(1), 37-54.

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2007). Blinking on the bench: How judges decide cases. Cornell Law Review, 93(1).

Lim, C. S., Snyder Jr, J. M., & Strömberg, D. (2015). The Judge, the politician, and the press: Newspaper coverage and criminal sentencing across electoral systems. American Economic Journal: Applied Economics, 7(4), 103–135.

MacCoun, R. J. (1990). Experimental research on jury decision-making. Jurimetrics, 223-233. Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. Organizational behavior and human decision processes, 39(1), 84-97.

Philippe, A., Ouss, A. (2015). *"No hatred or malice fear or affection": Media and sentencing*. Working paper

Philippe, A. (2016). *Gender bias in criminal justice*. Working paper

Salvat, X., Boccon-Gibod, D. (2013). Rapport à Madame la garde des sceaux, ministre de la justice sur l'expérimentation des citoyens assesseurs dans les ressorts des cours d'appel de Dijon et Toulouse, février.

Shayo, M., & Zussman, A. (2011). Judicial ingroup bias in the shadow of terrorism. The Quarterly Journal of Economics, 126(3), 1447-1484.

Schanzenbach, M. (2005). Racial and sex disparities in prison sentences: The effect of district-level judicial demographics. *The Journal of Legal Studies*, *34*(1), 57-92.

Steblay, N. M., Besirevic, J., Fulero, S. M., & Jimenez-Lorente, B. (1999). The effects of pretrial publicity on juror verdicts: A meta-analytic review. Law and Human Behavior, 23(2), 219-235.

Figure 1: French judicial and administrative organization.
Note: red dots represent first instance courts. Small houses are appeal courts. Grey lines are administrative frontiers of the counties. Additional red lines represent additional "judicial" frontiers when a county contains several first instance courts. Adjoining counties with the same color depend on the same appeal court.
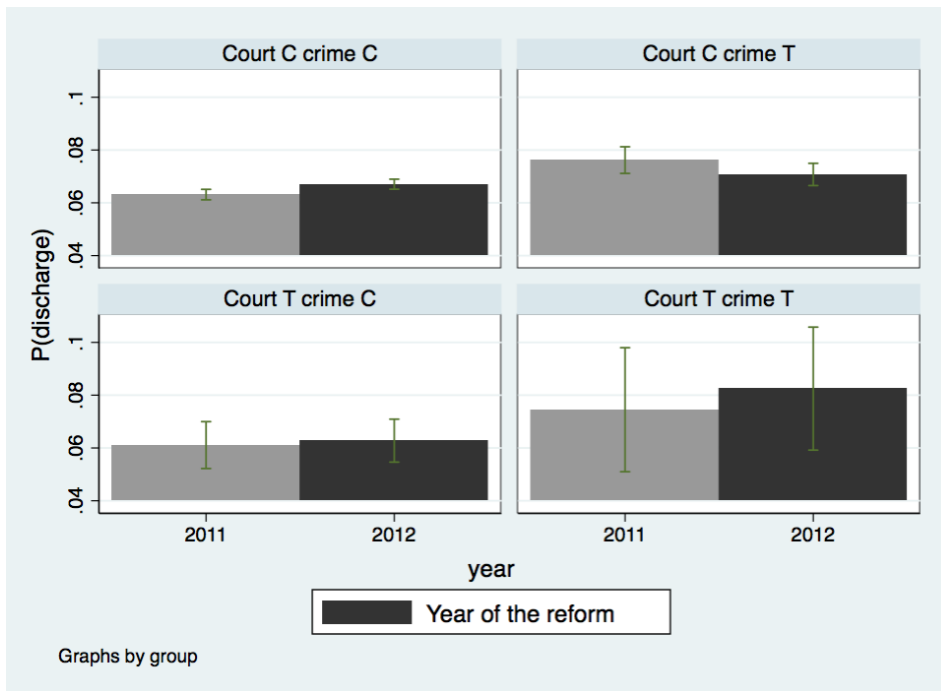
16

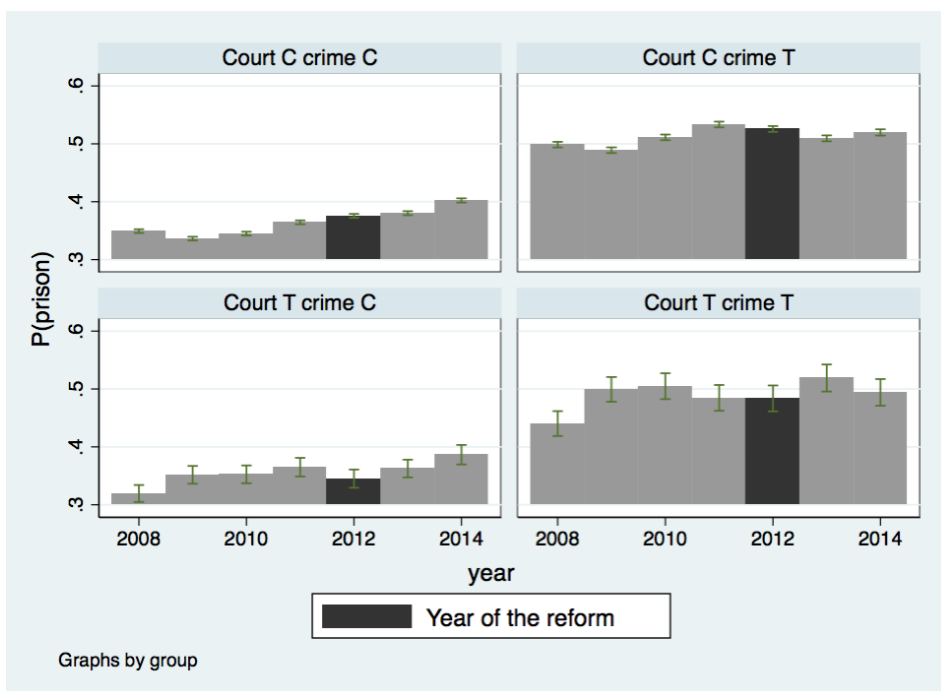Figure 2: Average discharge probability per group and per year.

Figure 3: Average probability to get a prison sentence per group and per year.
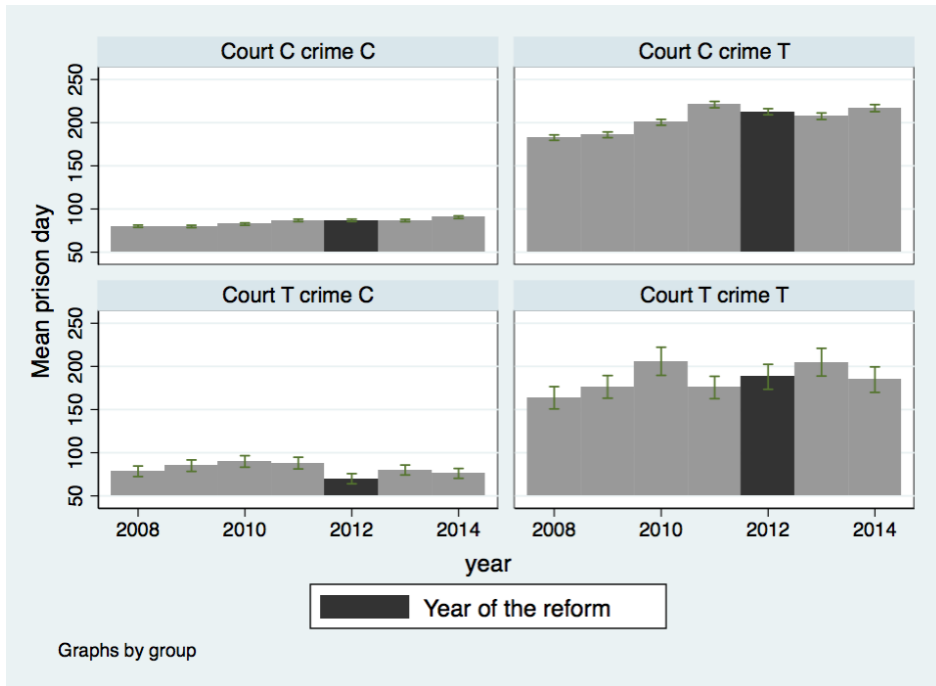
Figure 4: Average prison sentences per group and per year.

Crime	Area	Nb obs, pre 2012	Nb obs, 2012	Nb obs, post 2012
Treat	Treat	5,656	1,352	2,518
Treat	Control	112,444	27,182	48,912
Control	Treat	10,398	2,466	4,929
Control	Control	226,126	53,955	102,532

Table 1: Sample sizes in the different treatment and control groups per year.

	Targeted crimes		Non targeted crimes	
	Dijon-Toulouse	Other	Dijon-Toulouse	Other
Sex	.07	.07	.14	.14
Age	30.3	30.2	29.8	29.9
French	.86	.83	.82	.81
Former conviction	.51	.52	.51	.51
Investigation length (day)	492	425	541	494
Pre-trial detention (day)	31	30	11	9
Prison (dummy)	.49	.51	.35	.36
Probation (dummy)	.36	.35	.17	.17
Suspended prison (dummy)	.29	.3	.31	.3
Prison (day)	185	203	81	85
Probation (day)	117	116	41	43
Suspended prison (day)	60	59	36	38
N	9,526	17,793	382,613	188,538

Table 2: Descriptive statistics of the treated and control goups.

	(1)	(2)	(3)	(4)	(5)	(6)
	French	Age	Sex	Past-conviction	Investigation length	Pre-trial detention
Panel A: diff diff, non treated areas as control (N=197,615)						
Treated courts * 2012	0.00366	-0.0541	-0.00897	-0.00329	-19.28	0.628
	(0.0106)	(0.319)	(0.00739)	(0.0144)	(18.33)	(3.221)
Panel B: diff diff, non targeted crimes as control (N=27,299)						
Treated crimes * 2012	-0.00920	0.525	-0.0138	0.00154	-4.929	6.031*
	(0.0134)	(0.380)	(0.0104)	(0.0178)	(21.95)	(3.306)
Panel C: triple diff (N=597,156)						
Treated crimes * Treated courts * 2012	0.00640	0.357	-0.0200**	0.0143	4.445	4.828
	(0.0126)	(0.360)	(0.00983)	(0.0167)	(20.83)	(3.172)
Control	Crime fixed effects, court fixed effects, calendar month fixed effects					

Table 3: Balancing checks.
Coefficients presented in panel A, B and C come from separate regressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Discharge	Prison	Probation	Suspended prison	Prison	Probation	Suspended prison
	Dummy	Dummies			Quantum		
Panel A: diff diff, non treated areas as control							
Treated courts* 2012	0.0123	-0.0204	-0.0175	0.0291**	-3.535	-3.435	-1.016
	(0.0199)	(0.0128)	(0.0133)	(0.0124)	(7.912)	(5.821)	(3.278)
N	3 599	197 615	197 615	197 615	197 615	197 615	197 615
Panel A: diff diff, non targeted crimes as control							
Treated crimes * 2012	0.00537	-0.0101	-0.0220	0.0105	6.041	-8.257	-1.133
	(0.0209)	(0.0153)	(0.0152)	(0.0153)	(8.423)	(6.117)	(3.511)
N	426	22 299	22 299	22 299	22 299	22 299	22 299
Panel C: triple diff							
Treated crimes* Treated courts* 2012	0.0154	-0.00562	-0.00367	0.00575	-4.063	2.061	0.663
	(0.0148)	(0.0145)	(0.0143)	(0.0144)	(7.935)	(5.860)	(3.289)
N	7 295	597 156	597 156	597 156	597 156	597 156	597 156
Control	Court fixed effects, calendar month fixed effects	Crime fixed effects, court fixed effects, calendar month fixed effects, sex, age, French, past detention, procedural length, pre-trial detention					

Table 4: Effect of the introduction of jury members on the main characteristics of penal decisions.
Coefficients presented in panel A, B and C come from separate regressions.

	(1)	(2)	(3)
		| Quantum - mean for crime in court |	
	Prison	Probation	Suspended prison
Panel A: diff diff, non treated areas as control (N=198,064)			
Treated courts* 2012	-0.740	-0.152	-4.636**
	(6.203)	(3.740)	(2.245)
Panel B: diff diff, non tartegeted crimes as control (N=27,319)			
Treated crimes* 2012	5.167	-1.331	-3.835
	(6.589)	(4.085)	(2.405)
Panel C: triple diff (N=598,470)			
Treated crimes* Treated courts * 2012	-0.623	3.289	-0.536
	(6.199)	(3.862)	(2.267)
Control	Crime fixed effects, court fixed effects, calendar month fixed effects, sex, age, French, past detention, procedural length, pre-trial detention		

Table 5: Effect of the introduction of jury members on the absolute value of the difference between individual sentences and average sentences for the same crime in the same court.
Coefficients presented in panel A, B and C come from separate regressions.

	Quantum prison					
	Violence	Property crimes	Primo-offenders	Recidivists	Men	Women
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: diff diff, non treated areas as control						
Treated courts * 2012	-0.270	-3.375	-2.952	-2.570	-4.586	8.929
	(12.17)	(13.34)	(8.181)	(13.53)	(8.422)	(15.28)
N	55 978	85 879	95 634	101 981	183 234	14 381
Panel B: diff diff, non targeted crimes as control						
Treated crimes * 2012	6.096	13.38	3.804	14.57	5.628	7.446
	(25.32)	(13.48)	(9.008)	(14.19)	(9.078)	(16.40)
N	3 058	16 952	13 437	13 862	24 040	3 259
Panel C: triple diff						
Treated crimes * Treated courts * 2012	-9.405	2.256	-10.93	8.191	-4.114	1.611
	(23.66)	(12.96)	(8.262)	(13.54)	(8.550)	(16.39)
N	85 745	352 785	290 395	306 761	528 062	69 094

Table 6: Effect of the introduction of jury members on the prison length per sub-groups
Coefficients presented in panel A, B and C come from separate regressions.
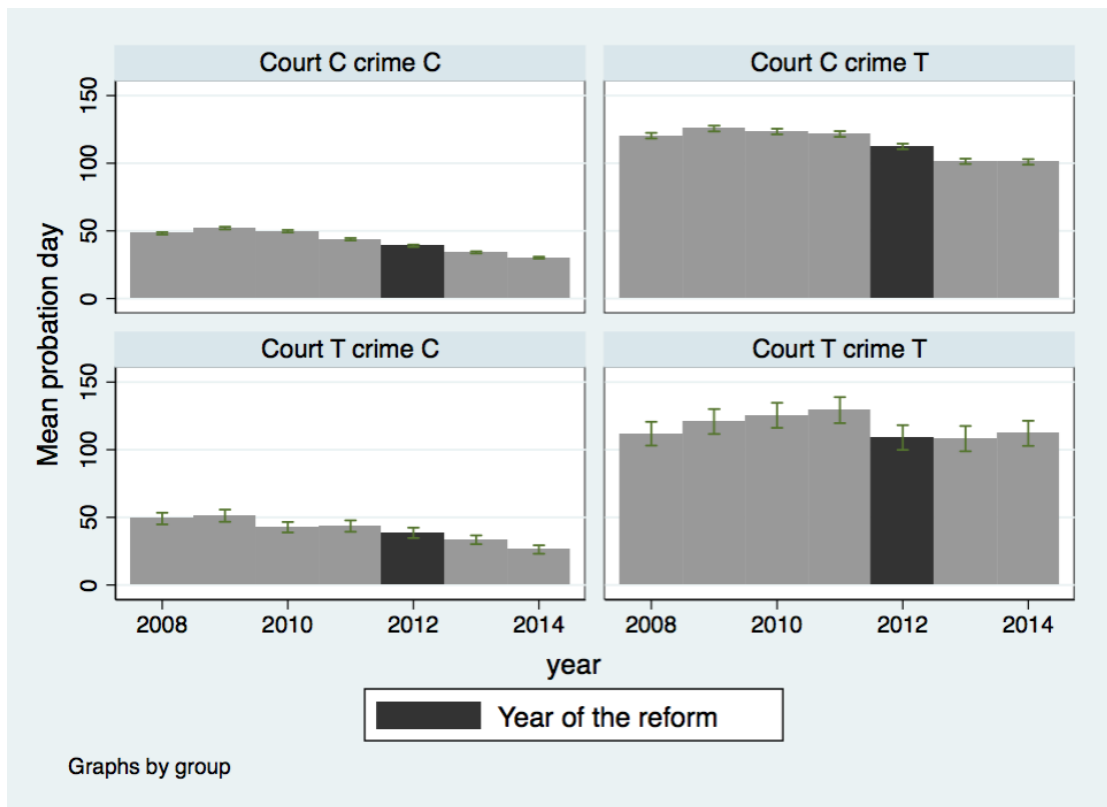
Appendix A: Additional graph.

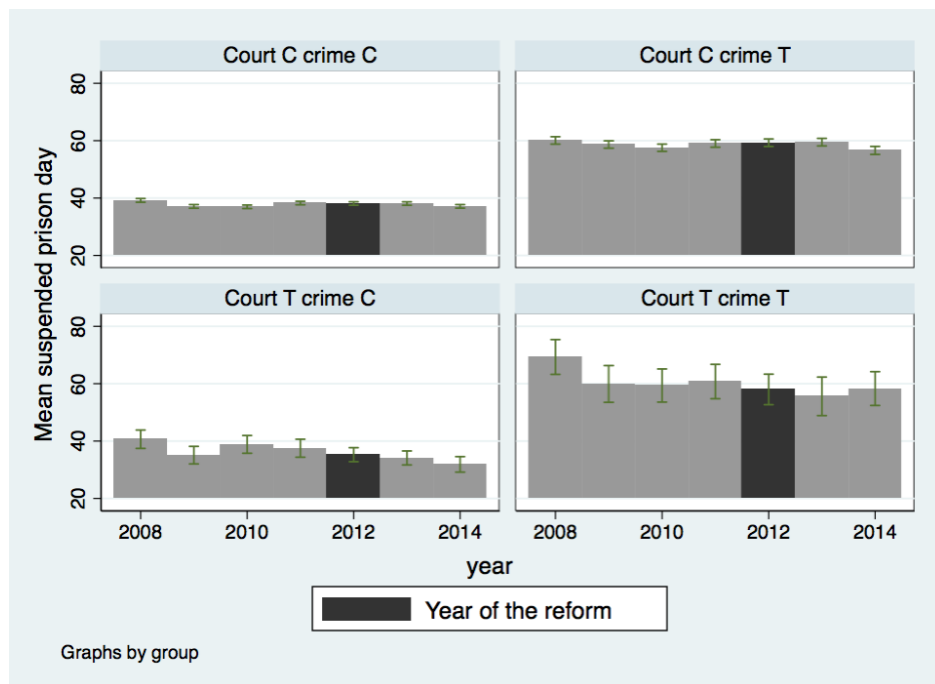Figure A1: Average probation sentences per group and per year.

Figure A2: Average suspended prison sentences per group and per year.

Appendix B: synthetic control method.

B.1. Method

In order to document the effect of the introduction of jury members in addition to professional judges it is possible to use synthetic Control method (Abadie et al, 2012). This method proceeds in two steps. The first one consists in constructing the synthetic control group for each treated groups, here the courts under the authority of Dijon and the courts under the authority of Toulouse. A Synthetic control group is the weighted combination of the control units – here the courts under the jurisdiction of the 28 non-targeted appeal courts – that approximate the treated unit before the treatment. The second step consists simply in the comparison of the evolution of the treatment area and their synthetic controls before and after the reform.

For this graphical evidence section, the data are aggregated at the appeal court level. This means that all the courts that are under the authority of one appeal court are merged together. The time unit is the month. The best weighted combinations are looked separately for Toulouse and Dijon. The other treated unit is excluded (i.e. Toulouse is not in the synthetic control of Dijon and vice versa). The synthetic controls are the best predictor for the following pre-reform characteristics:

- Proportions, among targeted crimes, of women, French, recidivist, conviction to prison, conviction to probation, conviction to suspended prison;
- Means, for targeted crimes, of age, investigation length, prison sentences, probation sentences, and suspended prison sentences.
- Means, for non-targeted crimes, of prison sentences, probation sentences, suspended prison sentences.

The results on discharges only use 2011-2012. Results on sentences use the entire 2008-2014 period.

B.2. Results

The best weighted combinations used for the main outcomes are presented in table B1. Note that the areas under the jurisdiction of Paris, Lyon or Marseille (the court is In Aix), the biggest one, are never used (their weight is zero).

Figure B1 present the evolution of the discharge probability for the treated offenses in Dijon (figure B1a) and Toulouse (figure B1b) in comparison to their respective synthetic control unit over 2011-2012. The periods during which jury members have been introduced in the court of the judicial region of Toulouse and Dijon, 2012, is indicated in grey.

Because of the limited sample size (each observation per month is the aggregation of only few courts) and the low discharge rate, the evolutions in the treated courts are volatiles and the pre-reform evolutions are not very close. No clear divergence appears after the reform. The probability of discharge does not seem to increase or decrease in Dijon and Toulouse in comparison to their synthetic control group.

The same type of results is presented in figure B2 and B3 for the probability to be sentenced to prison or the average quantum of prison for the offenses targeted by the reform. Results for Dijon are presented in figure B2a and B3a. Results for Toulouse are presented in figure B2b and B3b. Contrarily to figure B1, the evolution presented in figures B2 and B3 aggregate

numerous individual observations. This is the reason why the evolution are less volatile and the pre-reform correlation between treatment units and their synthetic controls are more similar. Note also that the period used is longer.

While figure B2 and B3 are more precise they do not provide suggestive evidence of an effect of the introduction of jury members in criminal courts. During the period of the reform, the probability to be sentenced to prison as well as the average sentences do not seem to diverge significantly in the treated courts. The same (absence of) result is observed when using probation or investigation length as the outcome (figure B4 and B5).

Jurisdictions	Discharges		Prison (dummy)		Prison (quantum)	
	Dijon	Toulouse	Dijon	Toulouse	Dijon	Toulouse
Agen	.015	.191	0	0	0	0
Aix	0	0	0	0	0	0
Amiens	0	0	0	0	0	0
Angers	0	0	.255	0	.22	0
Bastia	.167	0	.014	0	.013	0
Besançon	0	.249	0	.065	0	.085
Bordeaux	.144	.259	.145	0	.141	0
Bourges	0	0	0	0	0	0
Caen	0	0	0	0	0	0
Chambéry	0	0	0	0	0	0
Colmar	0	0	0	.341	0	.331
Douai	0	0	0	0	0	0
Grenoble	0	0	0	.537	0	.555
Limoges	0	.166	0	0	0	0
Lyon	0	0	0	0	0	0
Metz	.178	.134	0	0	0	0
Montpellier	0	0	0	0	0	0
Nancy	0	0	0	0	0	0
Nîmes	0	0	0	0	0	0
Orléans	0	0	.224	0	.222	0
Paris	0	0	0	0	0	0
Pau	0	0	0	0	0	0
Poitiers	.135	0	.003	0	0	0
Reims	.361	0	.36	0	.404	0
Rennes	0	0	0	0	0	0
Riom	0	0	0	.057	0	.029
Rouen	0	0	0	0	0	0
Versailles	0	0	0	0	0	0

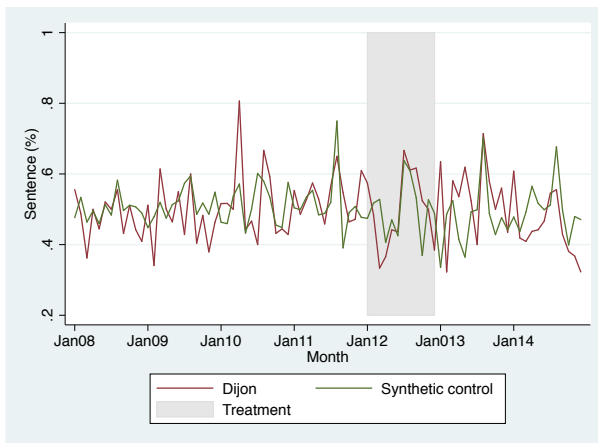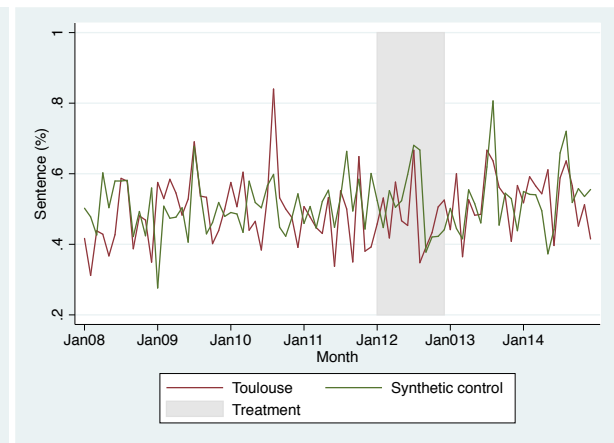Table B1: weight used in the construction of the synthetic control group.
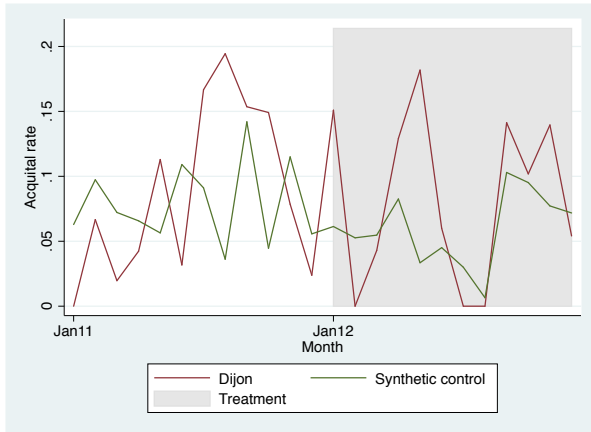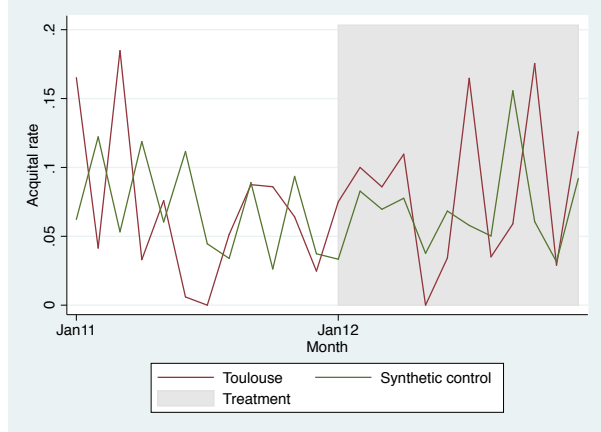
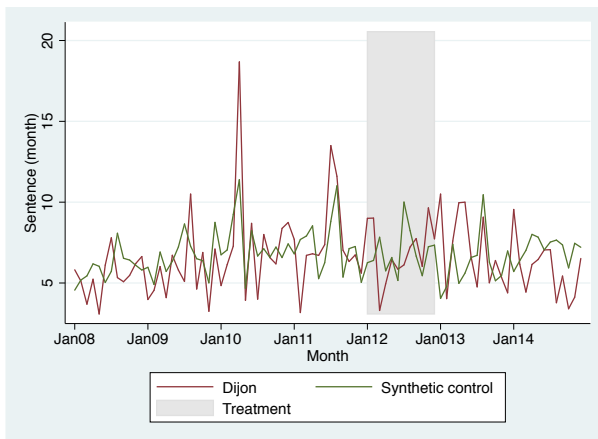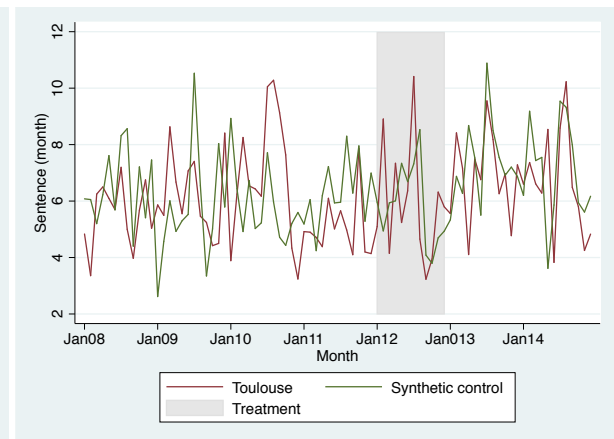Figure B1: Average discharge probability in treated courts and in their synthetic control.

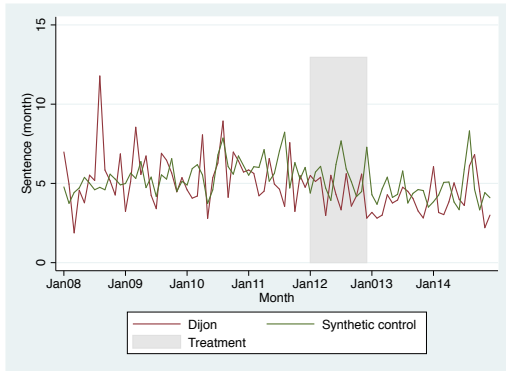Figure B2: Average probability to get a prison sentence in treated courts and in their synthetic control.
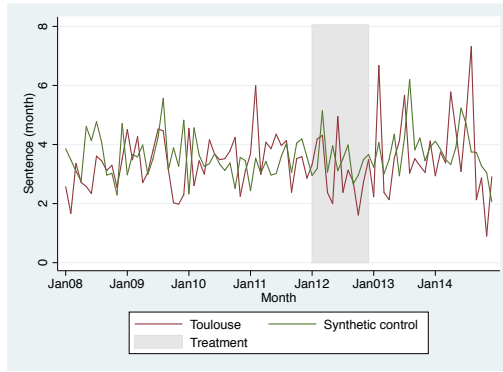

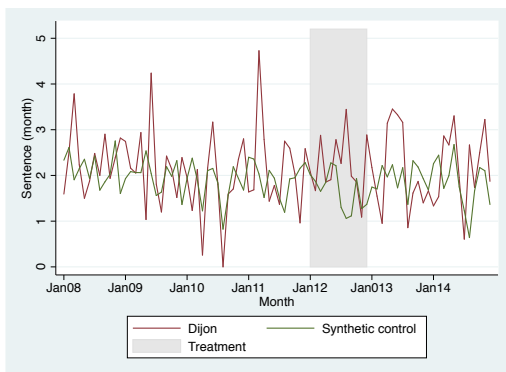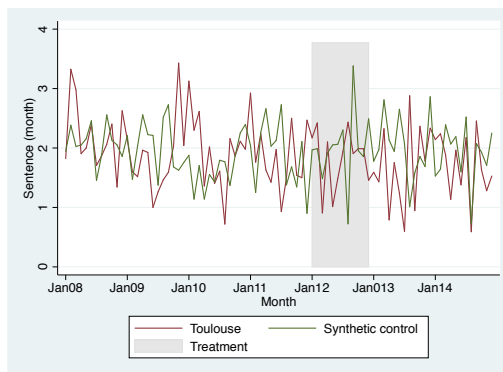Figure B3: Average prison sentences in treated courts and in their synthetic control.

Figure B4: Average probation sentences in treated courts and in their synthetic control.

Figure B5: Average suspended prison sentences in treated courts and in their synthetic control.

Appendix C

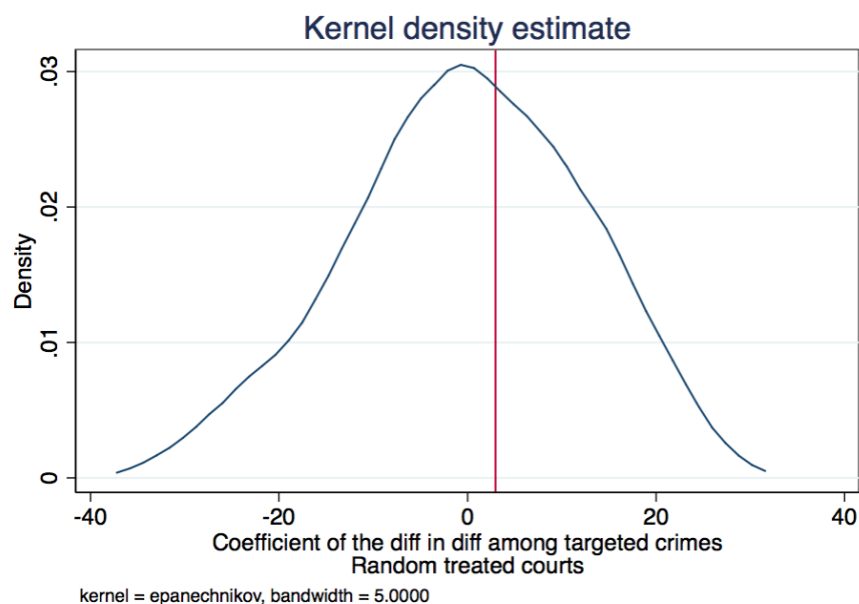kernel = epanechnikov, bandwidth = 5.0000

Figure C1: kernel distribution of the coefficients of placebo difference in difference using targeted crimes. The treated areas of the placebo are composed of the courts under the authority of two random appeal courts. The red line indicates the coefficient of the diff in diff when real treated areas are used.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Discharges	Prison	Probation	Suspended prison	Prison	Probation	Suspended prison
	Dummy	Dummies, 2011-2012			Quantum, 2011-2012		
Panel A: diff diff, non treated area as control							
Treated courts * 2012	0.0123	0.0180	-0.0351**	0.0136	21.26**	-9.869	-1.887
	(0.0199)	(0.0168)	(0.0176)	(0.0158)	(9.779)	(7.803)	(4.414)
N	3 599	57 049	57 049	57 049	57 049	57 049	57 049
Panel B: diff diff, non targeted crimes as control							
Treated crimes * 2012	0.00537	0.00467	-0.0485**	0.00616	13.75	-23.52***	-1.615
	(0.0209)	(0.0205)	(0.0203)	(0.0198)	(10.46)	(8.299)	(4.746)
N	426	7 619	7 619	7 619	7 619	7 619	7 619
Panel C: triple diff							
Treated crimes * Treated courts * 2012	0.0154	-0.00551	-0.00458	0.00732	-3.349	2.445	1.109
	(0.0148)	(0.0145)	(0.0143)	(0.0145)	(7.959)	(5.876)	(3.296)
N	7 295	170 267	170 267	170 267	170 267	170 267	170 267
Control	Court fixed effects, calendar month fixed effects	Crime fixed effects, court fixed effects, calendar month fixed effects, sex, age, French, past detention, procedural length, pre-trial detention					

Table C1: Replication of the main results using a restricted time period.
Coefficients presented in panel A, B and C come from separate regressions.