European Journal of Political Economy 88 (2025) 102678

Contents lists available at ScienceDirect

European Journal of Political Economy

journal homepage: www.elsevier.com/locate/ejpe

Full length article

Lady Justice: The impact of female judges on jury trial verdicts in North Carolina

Alessandra Foresta[1]

University of Southampton, United Kingdom

ARTICLE INFO

JEL classification:
K10
K40
J16

Keywords:
Gender
Judge
Trials behaviours

ABSTRACT

This study evaluates the impact of judges' gender on jury trial outcomes in the U.S. state of North Carolina. The identification strategy is based on conditional random assignment of judges to cases. Specifically, I take advantage of the compulsory judges' rotation imposed by the North Carolina Constitution. The results indicate that the presence of a female judge increases of 9.64–13.50 percentage points the probability of having at least one guilty verdict from the jury and of 9.5%–13.45% increase in the proportion of guilty verdicts expressed by the jury. Additionally, I perform a series of robustness and heterogeneity checks. I also investigate the potential mechanisms driving the results, exploring the influence of the jury selection process and women's attitudes toward the courts and sentencing.

1. Introduction

In many countries, the concept of *Justice* is represented by *Lady Justice*, a blindfolded woman holding a scale and a sword. The sword symbolises authority, while the scale measures the balance between support and opposition in a given case. The blindfold represents impartiality, a fundamental principle that dictates justice should be applied without regard to power, wealth, or status. Impartiality is a cornerstone in numerous justice systems, with judges often described as its "embodiment". However, in reality, judges do not always exhibit the level of neutrality and impartiality they are expected to uphold. For example, works like Spitzer and Talley (2013) and Johnson (2014) indicate that certain judges' characteristics, such as political affiliation or gender, might have an impact on the duration of sentences in bench trials within the US.[2] In jury trials, Lenehan and O'Neill (1981) and Halverson et al. (1997) find that judges non-verbal behaviours can sway juries toward one decision or another.

In this work, I explore the influence of judges' gender on the verdicts rendered in jury trials. The empirical analysis uses data on felony trials conducted in North Carolina between 2010 and 2012. The identification strategy relies on the conditional random assignment of judges to cases. Specifically, the North Carolina constitution mandates a periodic rotation (every six months) of judges across districts to prevent conflicts of interest and corruption. This rotation, combined with year and county-fixed effects,

E-mail address: a.foresta@soton.ac.uk.

[1] I thank the North Carolina Sunshine Project, Ron Wright, Francis Flanagan, Shea R. Denning for the help with the dataset and the information over the North Carolina Judicial system. I would also like to thank Andrew Pickering, Cheti Nicoletti, Emma Tomminey, Neil Cummins, Brendon Mcconnell, Matthias Flueckiger, Mariyana Zapryanova, Gabriella Conti, Mai Mahmoud, Rigissa Megalokonomou, along with the participants of the AME cluster seminar in York, SES2022, EPCS2022, seminar in Southampton, the 2nd Workshop in Gender and Economics at the University of Luxembourg, and Texas Economics of Crime Workshop 2024 for their constructive suggestions that greatly enriched this work. I would also like to thank the Editor of the *European Journal of Political Economy* and the anonymous reviewers for their comments and helpful suggestions. This article has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 897497.

[2] *Bench trials* involve judges taking on the role of fact-finder in addition to delivering sentences. The trials examined in this paper are *jury trials*, where the jury determines the facts, and the judge oversees the trial, issuing a sentence after the jury reaches a verdict. For further details, refer to Section 2.

https://doi.org/10.1016/j.ejpoleco.2025.102678
Received 29 July 2024; Received in revised form 15 January 2025; Accepted 15 April 2025
Available online 27 May 2025

ensures the random assignment of judges to different cases. The findings suggest that female judges enhance the likelihood of guilty verdicts in jury trials. More specifically, the presence of a female judge increases of 9.64-13.50 percentage points the probability of having at least one guilty verdict from the jury and of 9.5%-13.45% increase in the proportion of guilty verdicts expressed by the jury. Furthermore, I implement a battery of robustness and heterogeneity checks. The outcomes consistently align with the primary results. Finally, I investigate the possible mechanisms behind these findings. First, I evaluate the potential differences in leniency between genders (Leslie and Pope, 2017). Second, I investigate the effect of judges' gender on jury selection.

This article draws heavily from existing literature in various ways. First, it is connected with the vast research exploring gender differences in decision-making. For example, Gilligan (1982) suggests that men and women employ distinct criteria when making decisions. According to this author, women often base decisions on an "ethic of care", emphasising factors such as "correctness", responsibility and contextual considerations. In contrast, men tend to utilise an "ethics of rights", relying on abstract rules and principles such as individual liberty and hierarchical structures. These gender-based differences could result in variations in the decision-making and behaviours of judges. Fox and Van Sickel (2000) indicate that female judges are more likely to align with the prosecution, while male counterparts are more inclined to align with the defence.

Additionally, this article is linked to the literature that explores the relationship between judges' characteristics and trial outcomes. While a significant portion of prior research has concentrated on bench trials and sentencing, jury trials have garnered relatively less attention. According to the existing body of work, the impact of judges' gender on sentencing appears to be less straightforward compared to the influence of other judicial attributes, such as political affiliation (as exemplified by studies like Schanzenbach and Tiller, 2008, Fischman and Schanzenbach, 2011, and Spitzer and Talley, 2013). Johnson (2014) suggests that female judges lean toward leniency in their sentencing, while contrasting viewpoints are presented by Steffensmeier and Hebert (1999) and Spohn (1991), indicating that women judges may impose more severe sentences. Songer et al. (1994) proposes that distinctions in sentencing behaviours between male and female judges are predominantly observable in cases related to employment discrimination. Another more recent work is Lim et al. (2016), which focuses on Texas state district courts between 2004 and 2013. The authors do not find any difference in sentencing based on judges' gender. On the contrary, Boyd et al. (2010) does find some statistically significant differences. The authors evaluate the role of gender in federal appellate courts with respect to specific crimes and find that judges' gender matters, but not in all cases. According to the authors, female and male judges show differences in sentencing in sex discrimination cases, while no difference is present in the other trials. Philippe (2020) examines the influence of gender in the French judiciary system, specifically focusing on disparities in prison sentencing between men and women. The study finds that women generally receive lighter sentences than men for similar offences. However, it also highlights that an increase in the proportion of women judges reduces the gender gap in sentencing. Another study centred on continental Europe is Arantegui et al. (2024), where the authors examine the influence of gender within the Spanish judiciary system. Their research reveals that women are likely to impose longer prison sentences. Additionally, Anwar et al. (2019) explores the impact of gender on convictions, specifically highlighting the presence of female jurors. The authors assess the effects of the first female jurors in England, finding that their influence on convictions is particularly evident in cases that are salient to women, such as sexual offences.

Furthermore, this study also builds upon prior research examining the influence of judges on juries. While judges are ideally perceived as impartial figures devoid of any sway over juries, existing literature indicates that juries are often privy to the inclinations of judges regarding a case with a significant role played by judges' non-verbal behaviours (e.g., Hart, 1992, 1995, Halverson et al., 1997). Ekman and Friesen (1969) investigate the impact of judges' non-verbal behaviours on juries' verdicts, and they find that even the most composed and controlled judge can inadvertently convey biased information through non-verbal cues. Moreover, there are multiple cases in which courts rule that defendants' rights have been violated by judges' non-verbal behaviours (e.g., Blanck et al., 1985). This article tries to (partially) fill the gap and investigates the impact of judges' gender on the relationship between judges and juries.

The remainder of the paper is organised as follows: in Section 2, I focus on the legal system in North Carolina; in Section 3 on the dataset; and in Section 4 on the empirical strategy. In Section 5, I present the main results; in Section 6, the robustness checks; and in Section 7, the heterogeneity checks. Finally, in Section 8, I speculate over the possible explanations behind the results, and 9, I list my conclusions.

2. North Carolina Legal system

Jury trials in the US involve various key "participants": the defendant, the defence attorney, the prosecutor, a (seated) jury and a judge. The defendant is the individual who stands accused of the offence(s), and the defence attorney is the lawyer responsible for advising and advocating on behalf of the defendant. Prosecutors, on the other hand, act on behalf of the state and present the case against the defendant. They also substantially influence whether and when trials proceed, as suggested by Bandyopadhyay and McCannon (2014). In North Carolina, prosecutors are called *Assistant District Attorneys*. At the helm of the prosecution office is the "District Attorney", an elected official who holds a four-year term.[3]

Not all cases will result in jury trials. According to the data provided by the North Carolina Court system,[4] between 2009 and 2012 only 2.23% of cases resulted in jury trials.[5] The majority of cases are generally solved through plea bargaining, similar to the

[3] North Carolina is divided into 48 prosecutorial districts in the analysed period.

[4] Source: NCC courts statistical publications.

[5] These computations are based on the data provided by the NCC statistical reports.

rest of the US states. As suggested by Abrams and Fackler (2018), in plea bargaining, prosecutors may offer sentencing reductions (such as a lighter sentence or lesser charges) if defendants plead guilty and waive his/her right to a jury trial. If it is not possible to reach an agreement between the parties, a jury trial will take place. In Section A1 of the Online Appendix, I discuss and test the possible influence of judges' gender on pre-trial negotiations and conclude that the gender of the judges does not seem to play a role.

Juries are a key component in the Anglo-American legal system. Their role is to evaluate the evidence and reach a verdict. In layman's terms, the jury's job is to indicate if the defendant is guilty or not (while the judge's job is to decide the sentence in case of a guilty verdict).[6] Specifically, the jury overseeing a trial is referred to as the "seated jury", and in North Carolina, it is composed of 12 jurors along with alternates.[7] The jury selection process is the mechanism that transforms the jury pool, a randomly extracted group of potential jurors,[8] into the seated jury. The main idea behind this process is to remove unfit jurors (for more details, see Section 8.2).

The person with the greatest power in the courtroom is the judge. In jury trials, judges oversee the proceedings and ultimately deliver a sentence after the jury's guilty verdict.[9] The judges' role is to ensure that laws and procedures are always respected during the trials. For example, they review whether there are any illegality issues in the submitted evidence and are responsible for providing jury instructions.[10] The cornerstone of judges' conduct rests upon their impartiality, necessitating divorcing personal opinions, expectations, backgrounds, and attributes from the proceedings. The trials in the analysed sample are presided by a specific group of state judges, namely North Carolina's Superior Court Judges. The vast majority of these judges are elected officials who serve eight-year terms (around 87% of the sample). Elections for the judicial branch are held in North Carolina in even-numbered years. In the sample, there are two elections (on November 2, 2010 and on November 6, 2012) preceded by primary elections (on May 4, 2010 and on May 8, 2012).[11] Only a small part of the judges in the sample are appointed by the Governor (around 13% of the sample).

While all judges must hold legal qualifications, they are prohibited from practising law privately during their tenure on the bench. There are no term limits, though judges must be below 72 years of age.[12]

3. Data

The dataset is collected by the research team of North Carolina Jury Sunshine Project,[13] and it consists of felony trials in North Carolina between 2010 and 2012. The dataset includes judges' information such as name and surname, ethnicity, gender, political affiliation, and election details[14] and experience. The dataset also incorporates trial-specific data, such as the list of charges, verdict outcomes, sentences, defendant characteristics (age, ethnicity, and gender), and traits of both potential and seated jurors (political affiliation, gender, and ethnicity). For this article, the sample is confined to non-capital felony trials.[15] Further limitations involve focusing solely on cases with one defendant and one charge, where the verdict is known.[16] Based on the information provided by this dataset, I also generate a series of additional variables. First, I construct a set of dummies representing various charges: murder, robbery, other violent crimes, sex crimes, drug crimes, property crimes, other crimes and unknown crimes. Second, I create a variable about the judges' experience. As suggested by Johnson (2014) and Steffensmeier and Hebert (1999), bench tenure might impact sentence duration. The variable *Experience (per 100)* represents the number of years elapsed since receiving the "Juris Doctor" degree, a graduate-entry professional law degree, serving as a proxy for experience.[17] Third, I generate a variable called *Workload*. For elected judges, this variable is the ratio between the number of trials they presided over and the total number of trials in a

[6] In the U.S., under certain circumstances, judges in some states may have the power to overrule a guilty verdict through judgment notwithstanding the verdict. However, this does not apply to criminal cases in North Carolina, although it can be used in civil cases (N.C. R. Civ. P. 50). In criminal cases, defendants can instead file a Motion for Appropriate Relief (MAR) under Article 89 of G.S. Chapter 15 A, either during or after trials, primarily on legal grounds. Importantly, the success or failure of MARs does not affect the findings in this analysis, as the main outcome variables are based on jury deliberations, not the overall trial outcome. Finally, since all guilty verdicts carry a sentence in the dataset, this leads me to believe that Motions for Appropriate Relief (MARs) generally have a low rate of success in this sample. Source: UNC School of Government.

[7] Alternates are jurors designated to replace seated jurors if the latter become unable to fulfil their role. The presiding judge determines the number of alternates per trial.

[8] According to North Carolina legislation, a qualified potential juror must be a U.S. citizen, a resident of the county where the summons was issued, at least 18 years old, able to understand English and physically/mentally competent. Potential jurors must not have served as a juror during the previous two years, have served a full term as a grand juror in the last six years, and have been convicted of a felony (unless citizenship rights have been restored). Source: North Carolina Jury Service.

[9] For a more comprehensive exploration of distinct trial types, refer to note 1.

[10] Jury instructions encompass the guidelines provided by the judge to the jury after the conclusion of evidence presentation. These instructions elucidate the pertinent laws applicable to the case.

[11] Source: Ballotpedia.

[12] Source: North Carolina Judiciary branch site.

[13] Website: Sunshine Project Website. For an in-depth description of the data collection methodology, consult Flanagan (2018).

[14] In instances where election data is absent, I have leveraged sources such as Ballotpedia to fill in the gaps.

[15] Exclusions encompass murder charges classified as A1 felonies in North Carolina (38 observations).

[16] Given the structure of judicial courts in North Carolina, I remove trials with less than 12 jurors in the pool and in the seated jury (34 obs.), trials with unknown outcomes or mistrials (67 obs.), trials with more than two charges (32 obs.), if the defendant is a minor (8 obs.), and if defendant gender or race is unknown (26 obs.).

[17] To be fully authorised to practice law in North Carolina, individuals must also pass a bar examination. However, this specific information is not available in this database.

specific division in a specific year. For non-elected judges, this variable is the ratio between the number of trials they cover in a year and the total number of trials covered by non-elected judges.

Previous judge occupation might impact judges' behaviours during the trial. Following Cohen and Yang (2019), I also included a dummy for previous prosecutorial experience. I collected this information through a series of Google searches, mostly using sources from Ballotpedia, news articles, and specialist databases.[18] The main descriptive statistics are presented in Table 1.

As depicted in Table 1, within the sample, 11 % of the defendants is female, 63 % is identified as black and 34 % is identified as white and, 61 %, are over 30. On average, the composition of jury pools is roughly balanced in terms of gender, with a relatively higher percentage of individuals identifying as white (61 %) and a smaller proportion identifying as black (16 %). Moreover, the political affiliation within jury pools indicates that Republican, Democratic and other political affiliations jurors comprise approximately 26%, 31% 17%, of the composition, respectively.

Trials presided by female judges constitute 10 % of the sample, while those presided over by non-white judges account for 19%. While the number of trials presided by women is relatively low, it reflects the low number of women on the bench. In the sample, there are approximately 14 women among the 101 judges on the bench. Democrats judges oversee 65 % of trials, whereas Republicans judges are in charge of only 24 %. On average, judges possess 29 years of experience, and their workload is around 14% and 36% are former prosecutors.

Given the possible relevance of these characteristics on the main outcomes, I investigate if these variables are, on average, statistically different between judges' genders. In Table 2, I provide the results of the t-test conducted to compare various characteristics between female and male judges.[19]

As expected, the preliminary insights from Table 2 reveal some statistically significant distinctions between female and male judges. Overall, women judges appear to possess comparatively less experience, exhibit a lower proportion of white ethnicity and independent political affiliation, and are more likely to face an upcoming election in the following year. Considering the meaningful nature of these observed disparities between female and male judges and their potential repercussions on the primary outcomes, I introduce these variables as control factors within the main analysis (refer to Eq. (2)). This inclusion helps to mitigate the potential influence of these factors on the relationships under scrutiny and enhances the accuracy of the conclusions drawn from the study.

4. Research design

To identify the impact of judges' gender on the outcomes of jury trials, I employ a research design based on judges' rotation and fixed effects. Judges' rotation[20] is mandated by North Carolina Constitution (Article 4, Section 11).[21] In more detail, North Carolina is divided into 100 counties and 50 districts[22] and eight divisions, as shown in Fig. 1.

Superior Court Judges are assigned to one of the 50 districts, and every six months, they rotate across different districts within the same division (N.C. Gen. Stat. sec. 7). For instance, consider Judge A, initially assigned to Robson County, corresponding to district 16B in the fourth division. After six months, Judge A leaves District 16B and moves to District 16 A, which is also part of the fourth division and encompasses two counties: Scotland and Hoke. Judges' schedules are usually determined well in advance and are available online for consultation.[23] This mechanism was designed to eliminate possible conflicts of interest and corruption incentives, ensuring that judges do not have the discretion to select the trials they preside over.

Following Flanagan (2018), the identification strategy incorporates year and county fixed effects to account for factors such as crime trends, economic conditions and variations in demographic characteristics across time and counties. In line with prior literature (e.g., Anwar et al., 2012 and Foresta, 2022), I validate the identification strategy by conducting a regression of judges' attributes on observable trial characteristics, as illustrated in Eq. (1).

$$FemaleJudge_i = \alpha + \beta_1 Trial_i + \beta_2 Def_i + \beta_3 Jury_i + CountyFE_i + YearFE_i + \epsilon_i \quad (1)$$

Where $FemaleJudge_i$ is a binary variable denoting judges' gender for trial i, $Trial_i$ encompasses trial characteristics, Def_i represents defendant characteristics, and $Jury_i$ encapsulates jury pool characteristics.[24] $CountyFE_i$ and $YearFE_i$ denote county- and year-fixed effects. If judges' gender is indeed randomly assigned to each trial, there should be minimal or no correlation among the variables in Eq. (1). In Table 3, I present the results.

The findings in Table 3 are all non-statistically significant, in line with the research design, and the F-statistics are quite small. While the findings in Table 3 may not eliminate the possibility that judge selection is tied to unobservable characteristics, they strongly indicate that this concern is unlikely to be significant.

[18] I do not consider as former prosecutors a couple of judges doing some per-diem work as district attorneys. I removed those judges for which it was impossible to find information about the previous occupation (89 obs.).

[19] As a preliminary test, I also test if the dependent variables, *any guilty* and *prop. guilty* (as defined in Section 5), are statistically different based on gender. In both cases, I can reject the null hypothesis of equality at 10% (p-values 0.066 and 0.069, respectively).

[20] The rotation was suspended in 1990, 2002 and 2009 due to budgetary constraints. However, these years are not included in the sample. Source: Article.

[21] The Article states explicitly: "[t]he principle of rotating Superior Court judges among various districts of a division is a salutary one and shall be observed".

[22] A reform in 2016 reduced the number of districts from 50 to 48. All the trials present in the sample took place before 2016.

[23] Source: North Carolina Courts Website.

[24] Given that judges may influence the characteristics of the seated jury during the jury selection process, I use the jury pool's characteristics instead of those of the seated jury in the randomisation analysis. More on the relationship between jury selection and judges' gender can be found in Section 8.2.

Table 1
Descriptive Statistics.

	mean	sd	min	max
Defendant characteristics:				
Def. woman	0.1134	0.32	0.00	1.00
Def. non-white	0.6623	0.47	0.00	1.00
Def. over 30	0.6141	0.49	0.00	1.00
Def. age unknown	0.0261	0.16	0.00	1.00
Trials' characteristics:				
All charges	1.0782	0.27	1.00	2.00
Any Murder	0.0391	0.19	0.00	1.00
Any Robbery	0.0808	0.27	0.00	1.00
Any Sex crime	0.1108	0.31	0.00	1.00
Any Other violent crime	0.2034	0.40	0.00	1.00
Any Property crime	0.2855	0.45	0.00	1.00
Any Unknown crime	0.0143	0.12	0.00	1.00
Any Drug crime	0.1930	0.39	0.00	1.00
Any Other crime	0.0965	0.30	0.00	1.00
Dependent variable:				
Any Guilty	0.7445	0.44	0.00	1.00
Proportion of guilty convictions	0.7197	0.44	0.00	1.00
Seated juries characteristics:				
Women (seated jury)	0.5214	0.14	0.00	0.92
Men (seated jury)	0.4511	0.14	0.08	0.93
Unknown gender (seated jury)	0.0275	0.05	0.00	0.32
Black (seated jury)	0.1588	0.15	0.00	0.83
White (seated jury)	0.6219	0.19	0.00	1.00
Other (seated jury)	0.0205	0.04	0.00	0.23
Unknown *e*th. (seated jury)	0.1988	0.14	0.00	0.92
Republicans (seated jury)	0.2678	0.15	0.00	0.77
Other pol. aff (seated jury)	0.1773	0.12	0.00	0.58
Democrats (seated jury)	0.3052	0.16	0.00	0.85
Unknown political (seated jury)	0.2498	0.15	0.00	0.92
Jury pools' characteristics:				
Women (jury pool)	0.5093	0.11	0.14	0.79
Men (jury pool)	0.4589	0.11	0.18	0.81
Unknown gender (jury pool)	0.0318	0.04	0.00	0.25
Black (jury pool)	0.1563	0.13	0.00	0.67
White (jury pool)	0.6143	0.17	0.05	1.00
Other (jury pool)	0.0214	0.04	0.00	0.29
Unknown *e*th. (jury pool)	0.2080	0.12	0.00	0.90
Republicans (jury pool)	0.2632	0.13	0.00	0.72
Democrats (jury pool)	0.3069	0.14	0.00	0.83
Other pol. aff (jury pool)	0.1735	0.10	0.00	0.48
Unknown pol. aff. (jury pool)	0.2564	0.13	0.00	0.90
Judge characteristics:				
Female Judge	0.1004	0.30	0.00	1.00
Non-White Judge	0.1917	0.39	0.00	1.00
Workload	0.1364	0.11	0.01	1.00
Independent Judge	0.1160	0.32	0.00	1.00
Republican Judge	0.2386	0.43	0.00	1.00
Democratic Judge	0.6454	0.48	0.00	1.00
Election (Next year)	0.0887	0.28	0.00	1.00
Non-Elected judges	0.1291	0.34	0.00	1.00
Former prosecutor	0.3611	0.48	0.00	1.00
Experience (per 100)	0.2934	0.08	0.13	0.44
Observations	767			
N. county	61			
Year	2010–2012			

Notes: *Def.* is an abbreviation for defendant. *Experience (per 100)* is in years divided by 100. *Non-White Judge*, *Democratic Judge*, *Independent Judge*, *Republican Judge*, *Election (Next year)*, *Female Judge*, *Former prosecutor* and *Non-Elected Judge* are a series of dummies for ethnicity, political affiliation, elections, gender, previous work experience and non-elected judges. *Workload* is computed as indicated in Section 3.

5. Results

In this section, I will implement the main analysis and present the main findings. As mentioned in Sections 3 and 4, the regressions encompass county- and year-fixed effects and a set of control variables, as outlined in Eq. (2).

$$Guilty_i = \alpha + \beta_1 FemaleJudge_i + \beta_2 Controls_i + CountyFE_i + YearFE_i + \epsilon_i \quad (2)$$

Table 2
Differences between Judges by gender.

	Female (Mean)	Male (Mean)	Diff.	Std. Error
Non-White Judge	0.5065	0.1565	−0.3500***	0.0456
Workload	0.1290	0.1372	0.0082	0.0128
Independent Judge	0.0519	0.1232	0.0712*	0.0384
Democratic Judge	0.7143	0.6377	−0.0766	0.0575
Republican Judge	0.2338	0.2391	0.0054	0.0513
Election (Next year)	0.1818	0.0783	−0.1036***	0.0340
Non-Elected judges	0.1169	0.1304	0.0136	0.0403
Former prosecutor	0.4416	0.3522	−0.0894	0.0577
Experience (per 100)	0.2535	0.2978	0.0443***	0.0094

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. *Experience (per 100)* is in years divided by 100. *Non-White Judge*, *Independent Judge*, *Democratic Judge*, *Republican Judge*, *Election (Next year)*, *Female Judge*, *Former prosecutor* and *Non-Elected Judge* are a series of dummies for ethnicity, political affiliation, elections, gender, previous work experience and non-elected judges. *Workload* is computed as indicated in Section 3.

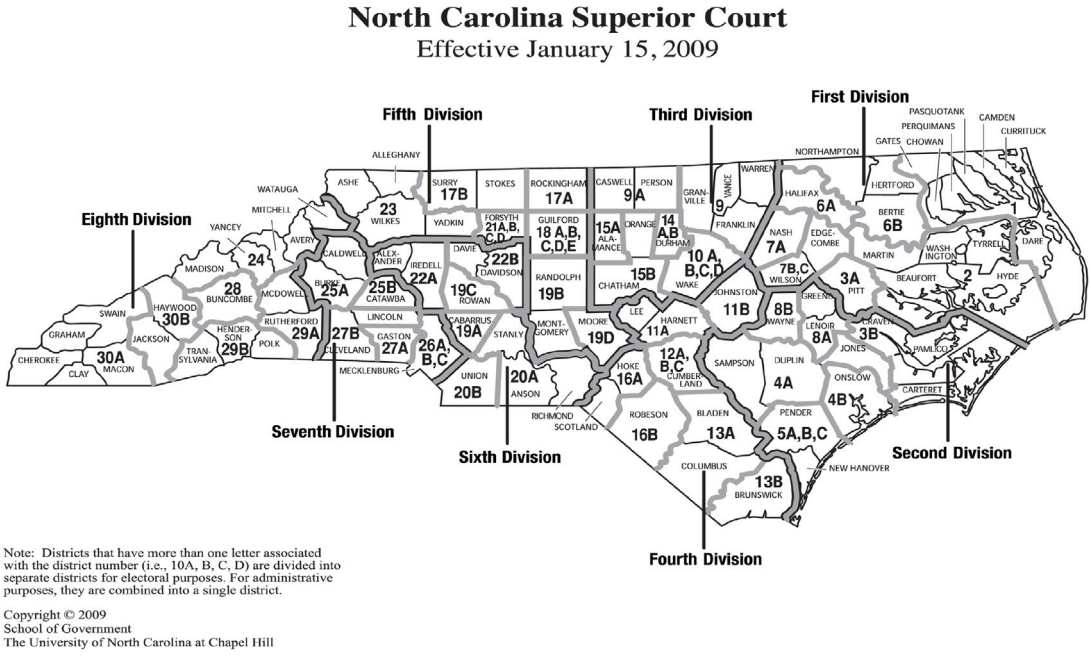

Fig. 1. North Carolina Superior Court Map.

where $Guilty_i$ represents the dependent variables for trial i: *proportion of guilty*, a ratio between the number of guilty verdicts and total charges, and *any guilty*, a dummy for at least a guilty verdict. $FemaleJudge_i$ represents a binary variable for judges' gender, and the set of control variables, $Controls_i$, is based on the characteristics of judge, jury pools, trials, and defendants. Furthermore, $CountyFE_i$ and $YearFE_i$ are county- and year-fixed effects. Given that trial distribution is not uniform across judges, to prevent potential over-rejections (as highlighted in Flanagan, 2018, Cameron et al., 2008, Carter et al., 2017, and Cameron and Miller, 2015), I adopt heteroscedasticity-robust standard errors instead of clustered standard errors at the judge level. The outcomes for Eq. (2) are outlined in Table 4.[25]

In Table 4, the coefficients for female judges are always positive and statistically significant, indicating that female judges systematically increase the probability of guilty outcomes. In terms of magnitude, having a female judge generates an increase between 9.64 and 13.50 percentage points in the probability of having at least a guilty verdict (Columns (1)-(3)) and between 9.5% and 13.45% the proportion of guilty verdicts (Columns (4)-(6)). The magnitudes of the findings are coherent with the previous literature. For example, Anwar et al. (2012) find a 16-percentage point conviction gap between all-white and juries where there is at least one black person in the jury pool. Moreover, there are other (slightly) significant coefficients. Non-white judges have a negative impact on guilty verdicts (Columns (3) and (6)), in line with the results for non-white jurors (e.g., Anwar et al., 2012 and Flanagan, 2018). In Column (6), there is a positive and slightly significant coefficient for next year's elections. The equivalent coefficient in Column (3) is similar but non-statistically significant.

[25] In Section A2 of the Online Appendix, I provide comprehensive results that include coefficients for all control variables (Table A3).

Table 3
Relationship between Judges demographics and Trials Characteristics.

dep. var.:	Female Judge	
	(1)	(2)
Women (jury pool)	−0.0872	−0.00929
	(0.11)	(0.11)
Unknown gender (jury pool)	−0.225	−0.0150
	(0.27)	(0.26)
Black (jury pool)	−0.0217	0.0397
	(0.36)	(0.38)
Unknown *e*th. (jury pool)	0.107	−0.0240
	(0.34)	(0.39)
White (jury pool)	−0.213	0.00318
	(0.32)	(0.35)
Republicans (jury pool)	−0.0656	−0.102
	(0.13)	(0.14)
Democrats (jury pool)	0.0955	0.187
	(0.12)	(0.13)
Unknown pol. aff. (jury pool)	−0.122	0.0665
	(0.15)	(0.20)
Def. woman	−0.0271	−0.0116
	(0.03)	(0.03)
Def. non-white	−0.00870	−0.0180
	(0.03)	(0.03)
Def. over 30	0.0220	−0.00208
	(0.02)	(0.02)
Def. age unknown	−0.0361	−0.0386
	(0.05)	(0.07)
All charges	0.0480	0.0753
	(0.05)	(0.05)
Any Murder	−0.0596	−0.0735
	(0.08)	(0.10)
Any Robbery	−0.123	−0.140
	(0.08)	(0.09)
Any Drug crime	−0.0713	−0.0848
	(0.08)	(0.09)
Any Sex crime	−0.0572	−0.0587
	(0.08)	(0.09)
Any Other violent crime	−0.0939	−0.0938
	(0.08)	(0.09)
Any Property crime	−0.0653	−0.0767
	(0.08)	(0.09)
Any Other crime	−0.0226	−0.0529
	(0.08)	(0.10)
Any Unknown crime	−0.0937	−0.0693
	(0.11)	(0.11)
Constant	0.301	0.0736
	(0.31)	(0.33)
County FE		✓
Year FE		✓
Observations	767	767
F-stat	1.132	0.940

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. *Female Judge* is a dummy for female judges. *Def.* is an abbreviation for defendant. All jury characteristics are measured in the *jury pool*.

To assess whether the estimated gender effect is statistically significant as a result of pure chance, I implement a permutation test, following the previous literature (e.g. Nagler et al., 2020, Bertrand et al., 2004 and Fujiwara and Wantchekon, 2013). This test randomly reassigns judges' gender (the treatment) in the sample and re-estimates β using this placebo assignment multiple times (1,000 in this case).[26] The randomisation inference test for the benchmark model (Columns (3) and (6) of Table 4) indicates that the estimated coefficients are statistically significant at 5% and larger in magnitude than almost all simulated effects, as shown in Figs. 2 and 3.

[26] To implement this analysis, I use the randomisation inference test (*ritest*) proposed by Heß (2017).

Table 4
Main results.

dep. var.:	Any Guilty			Prop. Guilty		
	(1)	(2)	(3)	(4)	(5)	(6)
Female Judge	0.09639**	0.1097**	0.1350**	0.09505**	0.09830*	0.1345**
	(0.05)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)
Non-White Judge			−0.1166**			−0.1119**
			(0.06)			(0.06)
Workload			0.04155			0.02151
			(0.18)			(0.18)
Independent Judge			0.09726			0.1089
			(0.07)			(0.07)
Democratic Judge			0.01643			0.01823
			(0.04)			(0.04)
Election (Next year)			0.1055			0.1069*
			(0.06)			(0.06)
Non-Elected judges			−0.07607			−0.08597
			(0.05)			(0.05)
Former prosecutor			0.06633			0.06135
			(0.04)			(0.04)
Experience (per 100)			−1.9180			−1.9743
			(1.77)			(1.78)
Experience sq. (per 100)			3.4020			3.4005
			(3.06)			(3.07)
County FE		✓	✓		✓	✓
Year FE		✓	✓		✓	✓
Controls			✓			✓
Observations	767	767	767	767	767	767
N. county	61	61	61	61	61	61
Mean Dep. Var	0.744	0.744	0.744	0.720	0.720	0.720
Mean Fem Judge	0.100	0.100	0.100	0.100	0.100	0.100

Notes: Robust standard errors in parenthesis. * p < 0.1, ** p < 0.05, *** p < 0.01. *Prop. guilty* is a ratio between the number of guilty verdicts and total charges, and *Any guilty* is a dummy for at least a guilty verdict. *Female Judge* is a dummy for female judges. *Controls* includes Women (seated jury), Unknown gender (seated jury), Black (seated jury), Unknown eth. (seated jury), White (seated jury), Republicans (seated jury), Democrats (seated jury), Unknown political affiliations (seated jury), Def. woman, Def. non-white, Def. over 30, Def. unknown age, All Charges, Any Murder, Any Robbery, Any Drug crime, Any Sex crime, Any Other violent crime, Any Property crime, Any Other crime, Any Unknown crime.

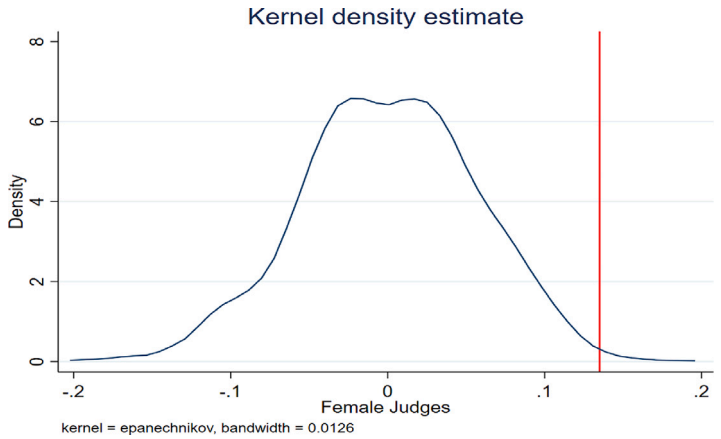

Fig. 2. Permutation test. Shown it is the Kernel density plot of a randomisation inference test for simulated judges' gender assigned using 1000 replications. The red vertical line shows the benchmark model estimated in Column (3) of Table 4.

6. Robustness checks

In Table 5, I implement a series of checks to evaluate the robustness of the findings in Table 4. In Columns (1)-(3) in Panel A and B of Table 5, I replicate the main analysis using different types of fixed effects. In Column (1), I use district-fixed effects instead of county-fixed effects to account for the fact that judges are allocated at the district level. In Column (2), I replicate the main analysis, including prosecutor fixed effects. As suggested by the previous literature (e.g., Bandyopadhyay and McCannon, 2014), prosecutors have significant powers, for example, in the scheduling of the trials. In Column (3), I replicate the main analysis using

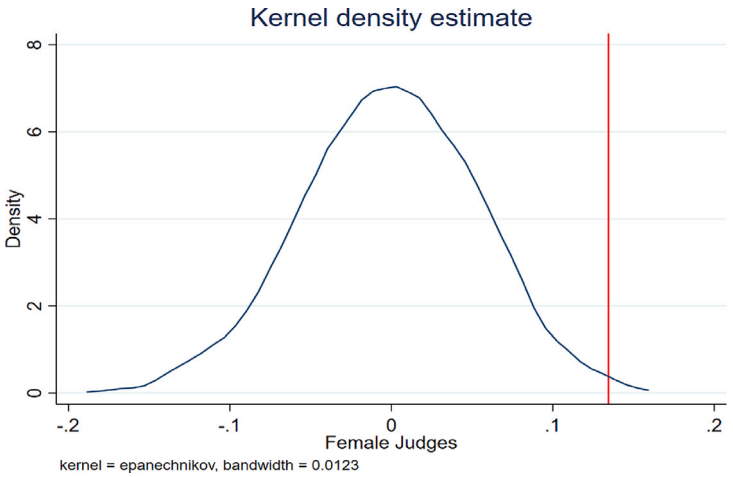

Fig. 3. Permutation test. Shown it is the Kernel density plot of a randomisation inference test for simulated judges' gender assigned using 1000 replications. The red vertical line shows the benchmark model estimated in Column (6) of Table 4.

term-fixed effects instead of year-fixed effects. Term-fixed effects are based on the division of the judges' calendar proposed by the North Carolina Administrative Office of the Courts.[27] This office divides the year into two sessions: a *Spring session* from January to June and a *Fall session* from July to December. Term-fixed effects are computed following this division.

In Column (4) of Panels A and B, I use the characteristics in the jury pools rather than in the seated jury.

In Columns (5)-(7) of Panels A and B, I implement the analysis in different sub-samples. The previous literature underlines that in-group/out-group bias is present in judiciary decisions. For example, Hoekstra and Street (2021) evaluate the relationship between jurors' and defendants' gender and find that own-gender jurors are significantly less likely to convict on drug charges. Another work is Ahrsjö et al. (2024). The authors investigate the role of shared identity in criminal cases in Stockholm District Court and find that if a juror belongs to the same identity-forming groups in the panel, the defendant is significantly less likely to get a prison sentence. To check for the possible in-group/out-group bias in terms of gender and given the relatively low number of female defendants (only 11% of my sample),[28] I replicate the analysis by simply removing the female defendants (Column (5)).[29]

Finally, the previous literature indicates that tenure on the bench might affect judges' attitudes (e.g., Johnson, 2014 and Steffensmeier and Hebert, 1999). I replicate the analysis by excluding the top 5% of judges' experience distribution (Column (7)) and the low 5% of judges' experience distribution (Column (6)). All the findings are presented in Table 5.[30]

Overall, the results for female judges in Table 5 are all positive and statistically significant, as in Table 4. The results in Columns (1)-(3) in both panels indicate that the main findings do not seem to be driven by the choice of fixed effects. As expected, the coefficients remain similar with a more "relaxed" choice of model (Column (1)) or a more "restrictive" one (Columns (2)-(3)). In Column (2) of Panel B, there is a loss of significance, and the significant drop in observations might partially drive this. Moreover, in the bottom panel of Table 5, I present the results for the equality test. I found that in all cases, I cannot reject the null hypothesis of equality between the coefficients in the benchmark models (Columns (3) and (6) in Table 4) and the coefficients in Table 5.

I also replicate the results by replacing the characteristics measured in the seated juries with those of the jury pools (Column (4) in Table 5). The coefficients are positive and statistically significant both in Panel A and in Panel B of Table 5, and, as shown in the bottom panel, it is not possible to reject the null hypothesis of equality with the benchmark models of Table 4.

The results in Columns (5) in Panels A and B of Table 5 are positive, statistically significant, and slightly bigger in magnitude. However, it is not possible to reject the equality test also in this case (bottom panels of Table 5). While the difference is not statistically significant, these outcomes point to the possibility of some in-group/out-group biases. The role of this potential bias is also hinted at by the negative, albeit non-significant, coefficients presented in Table A6 in the Online Appendix, where the analysis only for the female defendants in the subsample is presented. While the coefficients are non-significant and the reduction

[27] Source: Website.

[28] In Table A6 in the Online Appendix, I present the results in the female defendants subsample. While the results are non-statistically significant and the number of observations limited (only 7 trials have a female judge and a female defendant), the coefficients for female judges are always negative, which suggests that in-group/out-group bias might play a role in the main results of Table 4.

[29] Another possible category where in-group/out-group bias might play a role is connected with the victim's characteristics. In the sample, I have information on the gender of the victims only for 208 trials (and only for 22 trials for ethnicity), and the number of trials with a female victim and a female judge is quite low (only 8 trials). In Table A7 of the Online Appendix, I run a series of regressions with the interaction terms between judges' and victims' genders. While the limited number of observations does not allow me to perform a robust analysis, I still think the results might provide some interesting hints. The interaction terms between female victims and female judges, while mostly non-statistically significant, are all negative. Further research on the relationship between gender of judges, victims and defendants is need.

[30] Complete Tables with all control variables coefficients are presented in the Online Appendix (Tables A4 and A5).

Table 5
Robustness checks.

Panel A	Any Guilty						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Judge	0.1515***	0.1816**	0.1347**	0.1252**	0.1633***	0.1538**	0.1354**
	(0.06)	(0.09)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
District FE	✓						
Prosecutor FE		✓					
Term FE			✓				
Jury pool				✓			
No female def					✓		
No junior judges						✓	
No senior judges							✓
County FE		✓	✓	✓	✓	✓	✓
Year FE	✓	✓		✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	767	588	767	767	677	711	711
N. county	61	46	61	61	58	59	56
Mean Dep. Var	0.744	0.733	0.744	0.744	0.749	0.743	0.743
Mean Fem Judge	0.100	0.0816	0.100	0.100	0.103	0.103	0.105
Equality Test Benchmark (Col (3) Table 4):							
p-value	0.106	0.382	0.939	0.364	0.316	0.964	0.410
Panel B	Prop. Guilty						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Judge	0.1499**	0.1813*	0.1350**	0.1232**	0.1614***	0.1530**	0.1346**
	(0.06)	(0.09)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
District FE	✓						
Prosecutor FE		✓					
Term FE			✓				
Jury pool				✓			
No female def					✓		
No junior judges						✓	
No senior judges							✓
County FE		✓	✓	✓	✓	✓	✓
Year FE	✓	✓		✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	767	588	767	767	677	711	711
N. county	61	46	61	61	58	59	56
Mean Dep. Var	0.720	0.710	0.720	0.720	0.725	0.719	0.717
Mean Fem Judge	0.100	0.0816	0.100	0.100	0.103	0.103	0.105
Equality Test Benchmark (Col (6) Table 4):							
p-value	0.127	0.388	0.899	0.307	0.340	0.987	0.427

Notes: Robust standard errors in parenthesis. * p < 0.1, ** p < 0.05, *** p < 0.01. *Female Judge* is a dummy for female judges. All jury characteristics are measured in the *seated jury* except Column (4), where they are measured in the *jury pool*. *Controls* includes Non-white Judge, Workload, Independent Judge, Democratic Judge, Election (next-year), Non-Elected Judge, Former Prosecutor, Experience (per 100), Experience sq. (per 100), Women (Prop. jury), Unknown gender (Prop. jury), Black (Prop. jury), Unknown eth. (Prop. jury), White (Prop. jury), Republicans (Prop. jury), Democrats (Prop. jury), Unknown political affiliations (Prop. jury), Def. woman, Def. non-white, Def. over 30, Def. unknown age.

in the number of trials is important, the coefficients for female judges are always negative, in line with the in-group/out-group bias literature.

Finally, the coefficients in Columns (6)-(7) in Panels A and B of Table 5 indicate that more or less experienced judges do not drive the main findings. While some small variations in the magnitude of the coefficients are present, it is not possible to reject the null hypothesis of equality (bottom panels of Table 5).

Moreover, in the Online Appendix (Section A4) I also implement a series of sensitivity checks by removing each judge, each county and each type of crimes. The coefficients for female judges remain always positive and (mostly) statistically significant. While some coefficients lose significance (six out of 606 regressions), it is most likely to be driven by the small size of the sample and the strong imbalance between male and female judges.

7. Heterogeneity checks

In this section, I implement a series of heterogeneity checks. In Section 7.1, I replicate the main analysis by different types of crimes and in Section 7.2, I included the time-county trends.

Table 6
Heterogeneity checks: interactions with different types of crimes.

dep. var.:	Any Guilty (1)	Prop. Guilty (2)
Female Judges X Any Violent charge	0.2892***	0.2744***
	(0.10)	(0.10)
Female Judges X Any Property crime charge	0.1821*	0.1715*
	(0.09)	(0.10)
Female Judges X Any drug charge	0.1026	0.09165
	(0.13)	(0.13)
Female Judges X Any sex charge	−0.01472	0.02977
	(0.11)	(0.12)
Female Judges X Any other crimes charge	−0.05616	−0.03375
	(0.13)	(0.11)
Any Violent crime charge	−0.2147*	−0.2217**
	(0.11)	(0.11)
Any Property crime	−0.1158	−0.1231
	(0.11)	(0.11)
Any Drug crime	−0.1462	−0.1452
	(0.12)	(0.11)
Any Sex crime	−0.06104	−0.06259
	(0.12)	(0.11)
Any Other crime	0.04158	0.02981
	(0.10)	(0.10)
County FE	✓	✓
Year FE	✓	✓
Controls	✓	✓
Observations	756	756
N. county	61	61
Mean Dep. Var	0.742	0.717
Mean Fem Judge	0.101	0.101

Notes: Robust standard errors in parenthesis. * p < 0.1, ** p < 0.05, *** p < 0.01. *Female Judge* is a dummy for female judges. *Controls* includes Non-white Judge, Workload, Independent Judge, Democratic Judge, Election (next-year), Non-Elected Judge, Former Prosecutor, Experience (per 100), Experience sq. (per 100), Women (seated jury), Unknown gender (seated jury), Black (seated jury), Unknown eth. (seated jury), White (seated jury), Republicans (seated jury), Democrats (seated jury), Unknown political affiliations (seated jury), All Charges, Def. woman, Def. non-white, Def. over 30, Def. unknown age.

7.1. Different types of crimes

In this section, I investigate the role of different crimes in the main results. Given the limited number of observations per type of crime, I regroup *murder*, *robbery*, *other violent crimes* in a new category called *Violent Crimes*. I also remove the *unknown crime charges* from the sample (11 obs.). In Table 6, I present the results for the fully interacted model.[31]

Table 6, overall, indicates that the interactions between female judges and different types of crimes are generally positive, with two exceptions (*Any Other crime* and *Any Sex crime*). However, they are not always statistically significant; only the interactions between female judges with violent and property crimes are significant. The difference in the numerosity of these groups could partially drive this.

7.2. Time-county trends

In this section, I replicate the main analysis by allowing some heterogeneity in terms of time trends across different counties. In this way, I allow for potentially different time trends across different North Carolina counties. The findings are presented in Section 7.2.[32]

Overall, the findings in Columns (2) and (4) of Table 7 remain positive and statistically significant for female judges. However, there is a drop in the significance level when time-county trends are added. Also, in this case, the bottom panel indicate that I cannot reject the equality hypothesis with respect to the benchmark models (Columns (3) and (6) of Table 4).

8. Why does judges' gender play a role?

The previous sections document the importance of judges' gender on jury trial outcomes. In this section, I discuss the possible mechanisms. First, I investigate the possibility of different toughness of judges based on their gender, which, according to the

[31] In Table A8 of the Online Appendix, I present the results with all the control variables coefficients.
[32] In Table A9 of the Online Appendix, I present the results with all the coefficients.

Table 7
Heterogeneity checks: time-county trends.

dep. var.:	Any Guilty		Prop. Guilty	
	(1)	(2)	(3)	(4)
Female Judge	0.1350**	0.1135*	0.1345**	0.1113*
	(0.06)	(0.07)	(0.06)	(0.07)
Non-White Judge	−0.1166**	−0.1123*	−0.1119**	−0.1101*
	(0.06)	(0.06)	(0.06)	(0.06)
Workload	0.04155	−0.01112	0.02151	−0.04433
	(0.18)	(0.22)	(0.18)	(0.22)
Independent Judge	0.09726	0.1195	0.1089	0.1357*
	(0.07)	(0.08)	(0.07)	(0.08)
Democratic Judge	0.01643	−0.002395	0.01823	−0.0005021
	(0.04)	(0.05)	(0.04)	(0.05)
Election (Next year)	0.1055	0.1356*	0.1069*	0.1401**
	(0.06)	(0.07)	(0.06)	(0.07)
Non-Elected judges	−0.07607	−0.1002*	−0.08597	−0.1101*
	(0.05)	(0.06)	(0.05)	(0.06)
Former prosecutor	0.06633	0.06127	0.06135	0.05556
	(0.04)	(0.05)	(0.04)	(0.05)
Experience (per 100)	−1.9180	−1.4130	−1.9743	−1.5856
	(1.77)	(2.05)	(1.78)	(2.05)
Experience sq. (per 100)	3.4020	2.5545	3.4005	2.7401
	(3.06)	(3.52)	(3.07)	(3.53)
County FE	✓	✓	✓	✓
Year FE	✓		✓	
County X Year trend		✓		✓
Controls	✓	✓	✓	✓
Observations	767	767	767	767
N. county	61	61	61	61
Mean Dep. Var	0.744	0.744	0.720	0.720
Mean Fem Judge	0.100	0.100	0.100	0.100
Equality test Benchmark				
p-value		0.478		0.449

Notes: Robust standard errors in parenthesis. * p < 0.1, ** p < 0.05, *** p < 0.01. *Female Judge* is a dummy for female judges. *Controls* includes Women (seated jury), Unknown gender (seated jury), Black (seated jury), Unknown eth. (seated jury), White (seated jury), Republicans (seated jury), Democrats (seated jury), Unknown political affiliations (seated jury), Def. woman, Def. non-white, Def. over 30, Def. unknown age, All Charges, Any Murder, Any Robbery, Any Drug crime, Any Sex crime, Any Other violent crime, Any Property crime, Any Other crime, Any Unknown crime.

previous literature, can "leak" and influence the jury (Section 8.1). Second, I discuss the potential role of judges' gender in the jury selection process (Section 8.2).

8.1. Female vs Male judges

The previous literature underlines the influence of judges on juries. According to existing works, even the most restrained judges have opinions and "expectations" about the trial and its outcomes. Through their attitudes and non-verbal behaviours, these judges might "leak" their ideas and sway the jury in one direction or the other (e.g., Lenehan and O'Neill, 1981, Ekman and Friesen, 1969 and Burnett and Badzinski, 2005). According to Hart (1992, 1995), juries generally know what the judges think and tend to side with judges' positions. If female judges, on average, have a stricter attitude or are more likely to expect a guilty verdict, these attitudes and expectations might "leak" to the juries and help explain the findings in Table 4.

Unfortunately, the dataset does not provide information about judges' attitudes during the trials. However, it does provide information about sentencing in case of a guilty verdict. Following the approach proposed by Leslie and Pope (2017), I calculate a leave-out-mean measuring the degree to which each judge deviates from the average crime sentencing in terms of incarceration months. To control for systematic differences across time and place, I use the residuals from regressing the max and min sentence duration (in the log) on year and county fixed effects. Similar to Leslie and Pope (2017), I calculate this measure accounting for different crime types:

$$leniency_j = \frac{Res_{jc}}{N_{jc}} - \frac{Res_c}{N_c} \tag{3}$$

where $leniency_j$ is the measure for leniency for judge j, Res_{jc} the residuals of the regressions for judge j and type of crime c, N_{jc} the number of trials of judge j for type of crime c, Res_c the residuals of the regressions for type of crime c, N_c the number of trials for type of crime c. I plot the measure for female and male judges in Fig. 4 for minimum sentences and in Fig. 5 for maximum sentences.

The distributions in Figs. 4 and 5 indicate that female judges in the sample are, on average, less lenient than their male counterparts. While judges are not your "average characters", this evidence is in line with the general attitude of the population

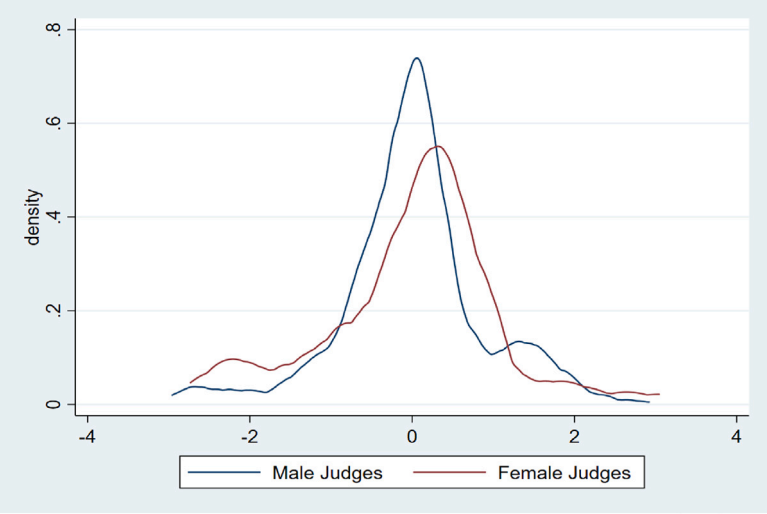

Fig. 4. Leniency measure using minimum sentences based on Leslie and Pope (2017).

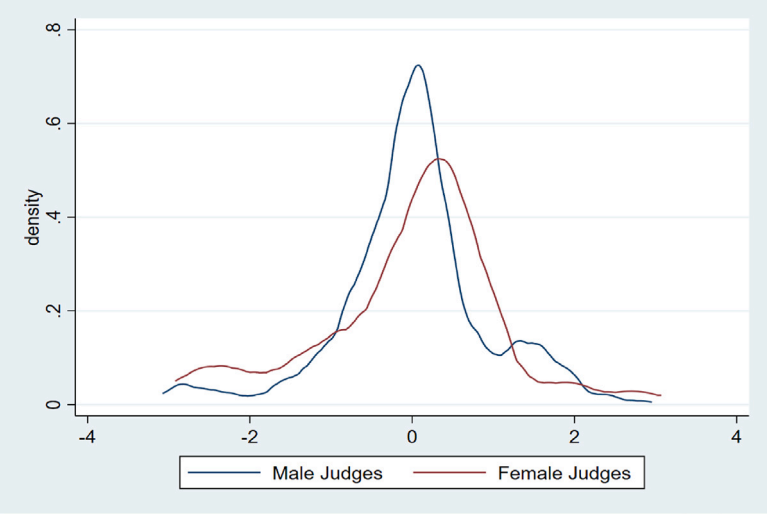

Fig. 5. Leniency measure using maximum sentences based on Leslie and Pope (2017).

towards the justice system, tested in GSS data (for more details, see Online Appendix Section A6). Moreover, in Section A7 in the Online Appendix, I replicate the main analysis using as a dependent variable the average months of sentencing. The coefficients are positive but non-statistically significant. Overall, while it is not defined evidence, it does point towards a difference in toughness between women and men on the bench and proposes a new avenue for further research.

8.2. Effect on the jury: Jury selection

In the previous section, I discuss the variation in sentencing between female and male judges. In this section, I check the potential role of judges' gender in shaping the seated jury. During the jury selection process, judges, defence attorneys, and prosecutors remove the jurors they deem biased during the jury selection process. Potential jurors can be excluded either *for cause* or thanks to *peremptory challenges*. *Removal for cause* is when the judge strikes a juror for apparent bias or hardship.[33] These removals are unlimited but are left to the judge's discretion. Prosecutors and defence attorneys can exclude jurors through the *peremptory challenges*. These removals do not require any explanations, but they are limited in numbers and cannot be based on ethnicity or gender (*Batson v. Kentucky*, [476

[33] Some examples are medical or financial hardship.

Table 8
Jury Selection Process and Female Judges.

dep. var.:	Remove		
	(1)	(2)	(3)
Democrat Juror	0.0172	0.0202*	0.0231*
	(0.01)	(0.01)	(0.01)
Independent Juror	0.00437	0.00419	0.00434
	(0.01)	(0.01)	(0.01)
Pol.Aff. Unknown Juror	0.0283**	0.0198*	0.00473
	(0.01)	(0.01)	(0.02)
Female Juror		−0.0264***	−0.0268***
		(0.01)	(0.01)
Gender Unknown Juror		0.0594**	0.0537**
		(0.02)	(0.03)
Black Juror			−0.00889
			(0.01)
Race other Juror			0.0109
			(0.03)
Race Unknown Juror			0.0215
			(0.02)
Female Judge x Democrat Juror	−0.0227	−0.0240	−0.0254
	(0.04)	(0.04)	(0.04)
Female Judge x Independent Juror	−0.0679	−0.0679	−0.0726*
	(0.04)	(0.04)	(0.04)
Female Judge x Pol.Aff. Unknown Juror	−0.0244	−0.0260	−0.0968*
	(0.04)	(0.04)	(0.05)
Female Juror x Female Judge		0.00597	0.00524
		(0.03)	(0.03)
Female Judge x Gender Unknown Juror		0.00882	−0.00741
		(0.08)	(0.08)
Female Judge x Black Juror			−0.00895
			(0.04)
Female Judge x Race other Juror			0.108
			(0.08)
Female Judge x Race Unknown Juror			0.0919*
			(0.05)
Constant	0.414***	0.426***	0.426***
	(0.01)	(0.01)	(0.01)
Trial FE X Female Judge	✓	✓	✓
Observations	16,719	16,719	16,719
N. Trials	745	745	745
N. Counties	60	60	60

Notes: robust standard errors in parenthesis. * p < 0.1, ** p < 0.05, *** p < 0.01. Observations are at the jurors' level. *Remove* is a dummy for being removed from the jury pool. *Excluded category*: whether the prospective jurors were seated on the jury.

U.S. 7US986]; *J.E.B. v. Alabama*, [511 U.S. 1US1994]). In North Carolina, prosecutors and defence attorneys have six peremptory challenges plus one for every alternative juror.

If female and male judges behave differently in their removal choices during the jury selection process, this could influence the seated juries' composition and, consequently, the verdicts. I implement a series of OLS regressions to evaluate the relationship between removals and the characteristics of jurors and judges. As dependent variables, I use a dummy for being removed from the jury pool. The regression is the following:

$$Remove_k = \alpha + \beta X_k FemaleJudge_k + \gamma X_k + TrialFE_k FemaleJudge_k + \epsilon_k \quad (4)$$

Where k represents the different jurors,[34] $Remove_k$ represents a dummy for being removed from the jury pool, X_k represents the characteristics of potential jurors, $TrialFE_k$ represents trials fixed effects, and $FemaleJudge_k$ represents a dummy for female judges. The findings for Eq. (4) are presented in Table 8.

Table 8 shows some slightly significant results, although they do not seem to be excessively robust. In Column (3), the interaction between female judges and independent jurors is negative and slightly statistically significant, while it is negative and non-statistically significant in Columns (1)-(2). Foresta (2022) indicates that the presence of one additional independent juror decreases the percentage of guilty verdicts by 2.93%, which seems to be at odds with the overall outcomes from the paper. However, the coefficients in Table 8 are only slightly statistically significant and only in Column (3). To summarise, while some statistically significant results in jury selection are present, they are not consistently robust across all specifications.

[34] Given the constraints for being called to perform jury duty (see note 2) and the short span of the panel, I do not have jurors that are called to perform jury duty multiple times in the sample.

Another relevant criterion in the juror's selection could be connected with *unobservable characteristics*. These characteristics may not be readily evident in the data collection process. Still, they could become apparent during jury selection through the tone of voice or potential jurors' attitudes. While an all-republican (or all democrats) jury might raise some eyebrows (and increase the possibility for the judge to be reversed later), female judges might use the *unobservable characteristics* to select harsher juries without the risks of impropriety or appeal. Unfortunately, I cannot test this hypothesis with this dataset, but it could be a new avenue of research.

9. Conclusion

This article investigates the impact of female judges on jury trial outcomes. The analysis was implemented in North Carolina between 2010 and 2012, and the research design relies on fixed effects and judges' random allocation to trials. The findings indicate that female judges are more likely to preside over jury trials that end with a guilty verdict. I test the robustness of these findings with respect to different fixed effects, the use of jury pool characteristics instead of seated jury characteristics, and removing female defendants, senior judges and junior judges. The results do not change. Furthermore, I check for potential heterogeneity due to differences in time trends across counties and across crimes.

Finally, I push the analysis a little bit further, and I test for some potential mechanisms behind the main findings. I first evaluate the differences in sentencing leniency between female and male judges. The results indicate that women on the bench tend to be less lenient than their male counterparts. Second, I evaluate the possible impact of judges' gender in the jury selection process, with somewhat mixed findings.

Declaration of competing interest

The author, Alessandra Foresta, declares that she has no relevant or material financial interests related to the research described in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ejpoleco.2025.102678.

Data availability

Data will be made available on request.

References

Abrams, David, Fackler, Ryan, 2018. To Plea or Not to Plea: Evidence from North Carolina. Tech. Rep. Mimeo, University of Pennsylvania.
Ahrsjö, Ulrika, Niknami, Susan, Palme, Mårten, 2024. Identity in court decision-making. Am. Econ. J.: Econ. Policy 16 (4), 142–164.
Anwar, Shamena, Bayer, Patrick, Hjalmarsson, Randi, 2012. The impact of jury race in criminal trials. Q. J. Econ. 127 (2), 1017–1055.
Anwar, Shamena, Bayer, Patrick, Hjalmarsson, Randi, 2019. A jury of her peers: The impact of the first female jurors on criminal convictions. Econ. J. 129 (618), 603–650.
Arantegui, Laura, Romero, Pablo, Tamarit, Josep M., 2024. Does the sex of judges matter? Analysis of sentencing in sex crimes in Spain. Criminol. Crim. Justice 17488958241239785.
Bandyopadhyay, Siddhartha, McCannon, Bryan C., 2014. The effect of the election of prosecutors on criminal trials. Public Choice 161 (1–2), 141–156.
Bertrand, Marianne, Duflo, Esther, Mullainathan, Sendhil, 2004. How much should we trust differences-in-differences estimates? Q. J. Econ. 119 (1), 249–275.
Blanck, Peter David, Rosenthal, Robert, Cordell, Doris Hazzard, 1985. The appearance of justice: judges' verbal and nonverbal behaviour in criminal jury trials. Stanf. Law Rev..
Boyd, Christina L., Epstein, Lee, Martin, Andrew D., 2010. Untangling the causal effects of sex on judging. Am. J. Political Sci. 54 (2), 389–411.
Burnett, Ann, Badzinski, Diane M., 2005. Judge nonverbal communication on trial: Do mock trial jurors notice? J. Commun. 55 (2), 209–224.
Cameron, A. Colin, Gelbach, Jonah B., Miller, Douglas L., 2008. Bootstrap-based improvements for inference with clustered errors. Rev. Econ. Stat. 90 (3), 414–427.
Cameron, A. Colin, Miller, Douglas L., 2015. A practitioner's guide to cluster-robust inference. J. Hum. Resour. 50 (2), 317–372.
Carter, Andrew V., Schnepel, Kevin T., Steigerwald, Douglas G., 2017. Asymptotic behavior of at-test robust to cluster heterogeneity. Rev. Econ. Stat. 99 (4), 698–709.
Cohen, Alma, Yang, Crystal S., 2019. Judicial politics and sentencing decisions. Am. Econ. J.: Econ. Policy 11 (1), 160–191.
Ekman, Paul, Friesen, Wallace V., 1969. Nonverbal leakage and clues to deception. Psychiatry 32 (1), 88–106.
Fischman, Joshua B., Schanzenbach, Max M., 2011. Do standards of review matter? The case of federal criminal sentencing. J. Leg. Stud. 40 (2), 405–437.
Flanagan, Francis X., 2018. Race, gender, and juries: Evidence from North Carolina. J. Law Econ. 61 (2), 189–214.
Foresta, Alessandra, 2022. Beyond Reasonable Doubt: The Impact of Politically Independent Jurors on Jury Trials in North Carolina. Tech. Rep. Discussion paper, University of York.
Fox, Richard, Van Sickel, Robert, 2000. Gender dynamics and judicial behavior in criminal trial courts: An exploratory study. Justice Syst. J. 21 (3), 261–280.
Fujiwara, Thomas, Wantchekon, Leonard, 2013. Can informed public deliberation overcome clientelism? Experimental evidence from Benin. Am. Econ. J.: Appl. Econ. 5 (4), 241–255.
Gilligan, Carol, 1982. In a Different Voice: Psychological Theory and Women's Development. Harvard University Press.
Halverson, Andrea M., Hallahan, Mark, Hart, Allen J., Rosenthal, Robert, 1997. Reducing the biasing effects of judges' nonverbal behavior with simplified jury instruction. J. Appl. Psychol. 82 (4), 590.
Hart, Allen Jay, 1992. On the sobriety of judges: Nonverbal influence in the courtroom.
Hart, Allen J., 1995. Naturally occurring expectation effects. J. Pers. Soc. Psychol. 68 (1), 109.
Heß, Simon, 2017. Randomization inference with Stata: A guide and software. Stata J. 17 (3), 630–651.

Hoekstra, Mark, Street, Brittany, 2021. The effect of own-gender jurors on conviction rates. J. Law Econ. 64 (3), 513–537.
Johnson, Brian D., 2014. Judges on trial: A reexamination of judicial race and gender effects across modes of conviction. Crim. Justice Policy Rev. 25 (2), 159–184.
Lenehan, Gregory E., O'Neill, Patrick, 1981. Reactance and conflict as determinants of judgment in a mock jury experiment. J. Appl. Soc. Psychol. 11 (3), 231–239.
Leslie, Emily, Pope, Nolan G., 2017. The unintended impact of pretrial detention on case outcomes: Evidence from New York City arraignments. J. Law Econ. 60 (3), 529–557.
Lim, Claire S.H., Silveira, Bernardo S., Snyder, James M., 2016. Do judges' characteristics matter? ethnicity, gender, and partisanship in texas state trial courts. Am. Law Econ. Rev. 18 (2), 302–357.
Nagler, Markus, Piopiunik, Marc, West, Martin R., 2020. Weak markets, strong teachers: Recession at career start and teacher effectiveness. J. Labor Econ. 38 (2), 453–500.
Philippe, Arnaud, 2020. Gender disparities in sentencing. Economica 87 (348), 1037–1077.
Schanzenbach, Max M., Tiller, Emerson H., 2008. Reviewing the sentencing guidelines: Judicial politics, empirical evidence, and reform. Univ. Chic. Law Rev. 75 (2), 715–760.
Songer, Donald R., Davis, Sue, Haire, Susan, 1994. A reappraisal of diversification in the federal courts: Gender effects in the courts of appeals. J. Politics 56 (2), 425–439.
Spitzer, Matthew, Talley, Eric, 2013. Left, right, and center: Strategic information acquisition and diversity in judicial panels. J. Law, Econ., & Organ. 29 (3), 638–680.
Spohn, Cassia, 1991. Decision making in sexual assault cases: Do black and female judges make a difference? Women & Crim. Justice 2 (1), 83–105.
Steffensmeier, Darrell, Hebert, Chris, 1999. Women and men policymakers: Does the judge's gender affect the sentencing of criminal defendants? Soc. Forces 77 (3), 1163–1196.

Further reading

Abrams, David, Galbiati, Roberto, Henry, Emeric, Philippe, Arnaud, 2022. When in Rome... on local norms and sentencing decisions. J. Eur. Econ. Assoc 20 (2), 700–738.
Anwar, Shamena, Bayer, Patrick, Hjalmarsson, Randi, 2014. The role of age in jury selection and trial outcomes. J. Law Econ. 57 (4), 1001–1030.
McGee, Robert W., Petrides, Yanira, Gelman, Wendy, 2022. Democrats v. Republicans: Is political party affiliation a significant demographic variable in jury selection? Charlest. L. Rev. 17, 153.
Silveira, Bernardo S., 2017. Bargaining with asymmetric information: An empirical study of plea negotiations. Econometrica 85 (2), 419–452.