

Discussion Paper Series

IZA DP No. 18639

May 2026

Gender Representation and Collective Decision-Making in Expert Committees

Rigissa Megalokonomou

Monash University, University of Queensland,
CESifo and IZA@LISER

The IZA Discussion Paper Series (ISSN: 2365-9793) ("Series") is the primary platform for disseminating research produced within the framework of the IZA@LISER Network, an unincorporated international network of labour economists coordinated by the Luxembourg Institute of Socio-Economic Research (LISER). The Series is operated by LISER, a Luxembourg public establishment (établissement public) registered with the Luxembourg Business Registers under number J57, with its registered office at 11, Porte des Sciences, 4366 Esch-sur-Alzette, Grand Duchy of Luxembourg.

Any opinions expressed in this Series are solely those of the author(s). LISER accepts no responsibility or liability for the content of the contributions published herein. LISER adheres to the European Code of Conduct for Research Integrity. Contributions published in this Series present preliminary work intended to foster academic debate. They may be revised, are not definitive, and should be cited accordingly. Copyright remains with the author(s) unless otherwise indicated.

Gender Representation and Collective Decision-Making in Expert Committees*

Abstract

This paper studies how gender representation affects collective decision-making in expert committees. I exploit quasi-random assignment of judges to panels in the Greek Supreme Court using newly digitized data on 3,700 criminal appeals. I find that panels with more female judges are more likely to reject appeals and less likely to delegate cases. Effects are nonlinear and emerge primarily once at least three of five judges are female; below this level, representation has no detectable effect. The mechanism appears to operate at the panel rather than the individual level—panels with a higher share of female judges take significantly longer to decide, especially in complex cases and in familiar panel compositions, consistent with more thorough deliberation rather than coordination costs. These findings suggest that diversity policies targeting modest increases in female representation will have limited impact unless they shift the deliberative composition of the group itself.

JEL classification

J16, D03, D71, J78

Keywords

panel decisions, gender composition, quasi-random assignment, Supreme Court

Corresponding author

Rigissa Megalokonomou

rigissa.megalokonomou@monash.edu

* Rigissa Megalokonomou acknowledges research support from the University of Queensland BEL Early Career Grant (No: UQECR1833757), the Monash University, Monash Establishment Grant (No: 1755774), and Monash University, Early Career Academic Research Grant (No: 1750840). I am grateful to Victor Lavy, Chetti Nicolleli, Giulio Zanella, Alessandra Foresta, and Evangelos Dioikitopoulos for insightful discussions and useful comments. I also thank participants at the University of Queensland, the University of Western Australia, the University of Cyprus, the University of Adelaide, University of Sydney, the 2026 Australian Gender Economics Workshop, the 2025 Australasian Meeting of the Econometric Society, 2024 European Society of Population Economics Conference at the University of Rotterdam, 2022 Labour Econometrics Workshop at the University of Wollongong, the University of Sydney, Deakin University, the 14th Annual Australasian Public Choice Conference, the 21st Conference on Research on Economic Theory and Econometrics at the Athens University of Economics and Business, the 2022 Gender, Norms, and Economics Workshop at the University of Melbourne, and the Australian Gender Economics Workshop at the Queensland University of Technology. I am grateful to Yi Zhang, Vu Minh Thu Hoang, Nicholas Umashev for superb research assistance and to Elias Arvanitakis, Konstantinos Zervos, Natalia Paraskeua, Giorgios Kambakis, Zoi Kwstaki, Stylianos Rousalis, Sofia Georgogala, and Panagiotis Sakelaridis, for helping compile the data and the data team manager, Zacharias Sousouris, for supervising the data collection process. There is no material or financial conflict of interest on the part of the author with respect to any of the research in this paper.

1 Introduction

Governments around the world are expanding female representation in courts, boardrooms, and expert committees on the premise that the composition of decision-makers influences decision outcomes ([Escobar-Lemmon, Hoekstra, Kang, and Kittilson, 2021](#)). Yet the causal evidence underlying this premise remains limited. A central question remains: whether representation affects collective behavior, and if so, through what mechanisms differences in composition translate into differences in outcomes. Most existing work focuses on individual decision-making or examines environments where the institutional structure of collective deliberation is difficult to observe directly ([Anwar, Bayer, and Hjalmarsson, 2012](#); [Knepper, 2018](#); [Dahl, Kostøl, and Mogstad, 2014](#)). This paper addresses this question using newly digitized data on more than 3,700 criminal appeals in the Greek Supreme Court between 2005 and 2017, where institutional rules generate quasi-random variation in panel gender composition and five-member professional panels provide a natural setting to study how expert group composition shapes both decisions and the deliberation process itself.

This issue is particularly relevant in light of the persistent underrepresentation of women in senior decision-making roles. Despite substantial progress at entry levels—women now constitute the majority of law students in many OECD countries—they remain underrepresented at the highest ranks ([Escobar-Lemmon, Hoekstra, Kang, and Kittilson, 2021](#)). Many countries have responded with explicit appointment targets and diversity mandates ([National Women’s Law Center, 2016](#); [OECD, 2018](#)). Yet identifying the causal effects of representation remains challenging: in most settings, team composition is endogenous, case selection is strategic, and it is difficult to separate who is in the room from what each individual would have done in isolation.

The Greek Supreme Court provides an unusually clean environment to study this question. First, appeals are decided exclusively by panels of professional judges, which eliminates interactions between experts and lay decision-makers ([Hoekstra and Street, 2021](#); [Anwar, Bayer, and Hjalmarsson, 2022](#)). Second, and most importantly for identification, cases are assigned to panels through the lexicographic ordering of defendant surnames—a rule determined at filing and unrelated to case characteristics or judicial preferences. This institutional feature generates quasi-random variation in panel gender composition that is orthogonal to the cases being decided. Third, panels typically consist of five judges, providing discrete and interpretable variation in composition and allowing the shape of composition effects—not just their average magnitude—to be identified. These features allow me to credibly identify the causal effect of group composition on collective decisions in a high-stakes, real-world setting.

I find that panels with a higher share of female judges are significantly more likely to reject appeals—upholding lower court decisions—and less likely to delegate cases. A 10 percentage-point increase in female representation raises the probability of rejection by approximately 1–1.5 percentage points and reduces delegation by about 0.4–1 percentage points. These effects increase with the share of female judges and are most pronounced among panels with high levels of female representation, particularly when women constitute a majority. By contrast, effects at lower levels of representation are smaller and imprecisely estimated, consistent with theoretical predictions that behavioral differences emerge only once women cross a threshold share (Kanter, 1977; Dahlerup, 1988; Collins, Manning, and Carp, 2010).

The results point toward mechanisms operating through collective interaction rather than the simple aggregation of individual preferences. I find no evidence that female judges differ systematically in their average panel-level outcomes, but instead show that panels with more female judges deliberate more intensively. A key contribution of the paper is to shed light on the mechanisms through which representation affects collective outcomes. I exploit variation in deliberation time—the number of days between the court hearing and the issuance of the final decision—as a proxy for the intensity of collective deliberation (Epstein, Landes, and Posner, 2013; Dimitrova-Grajzl, Grajzl, Sustersic, and Zajc, 2012). Panels with a higher share of female judges take systematically longer to decide cases, particularly in more complex appeals. This effect is amplified in panels with greater prior familiarity among judges, which is difficult to reconcile with a pure coordination-frictions interpretation. Additional evidence is consistent with this interpretation. Effects are concentrated in cases involving individual rather than institutional litigants, and are stronger in panels where women occupy key procedural roles, such as president or case introducer, which suggests that representation affects how deliberation unfolds within the panel. They are weaker in areas with higher economic development and stronger indicators of women’s empowerment, which suggests that local norms shape how representation translates into behavior (Bursztyn, González, and Yanagizawa-Drott, 2020; Bertrand, Kamenica, and Pan, 2015). These findings point toward group-mediated mechanisms—which include deliberative intensity, persuasion, and the structuring of collective discussion—rather than the simple aggregation of fixed individual preferences.

This distinction matters for policy. If representation affects outcomes solely through the direct aggregation of individual preferences, effects should scale approximately linearly with the share of women on the panel. By contrast, if representation operates through group-mediated processes—such as the depth of review, persuasion dynamics, and collective standards applied—effects may be nonlinear and depend on whether women attain sufficient presence to shape panel dynamics.

The evidence above is more consistent with the latter interpretation.

A large literature documents gender differences in preferences and evaluative behavior ([Adams and Funk, 2012](#); [Eckel and Grossman, 2008](#); [Cabral and Dillender, 2024](#); [Bordalo, Coffman, Genaioli, and Shleifer, 2019](#)), and studies across academia, peer review, politics, and judicial decision-making suggest that evaluator characteristics can affect outcomes, though findings remain context-dependent and mixed ([Bagues, Sylos-Labini, and Zinovyeva, 2017](#); [Abrevaya and Hamermesh, 2012](#); [Gagliarducci and Paserman, 2011](#); [Vega and Firestone, 1995](#); [Swers, 1998](#); [Chattopadhyay and Duflo, 2004](#); [Rehavi, 2007](#); [Clots-Figueras, 2012](#); [Card, DellaVigna, Funk, and Iriberri, 2020](#)). Much of this evidence, however, relies on observational variation or settings where assignment is not clearly exogenous ([Boyd, Epstein, and Martin, 2010](#); [Feenan, 2008](#); [Walker and Barrow, 1985](#); [Songer and Crews-Meyer, 2000](#)), which makes it difficult to separate identity effects from selection.

This paper contributes to two strands of the literature. First, this paper provides causal evidence on the effect of gender representation in high-stakes expert settings by exploiting quasi-random variation in panel composition to isolate the effect of representation from selection. Most existing work relies on observational variation or non-random assignment ([Coontz, 2000](#); [Boyd, 2016](#); [Feenan, 2008](#); [Walker and Barrow, 1985](#); [Moyer, Szmer, Haire, and Christensen, 2021](#)), while a related literature uses quasi-random assignment to identify causal effects at the level of individual decision-makers ([Hoekstra and Street, 2021](#); [Flanagan, 2018](#); [Ash, 2022](#); [Hanemaaijer, Ketel, and Marie, 2025](#); [Foresta, 2025](#); [Ash, Chen, and Ornaghi, 2024](#)). Although the latter provides credible evidence on individual evaluator effects, it typically treats panels as collections of independent actors and offers limited insight into how group composition shapes collective outcomes ([Steffensmeier and Hebert, 1999](#); [Johnson, 2014](#); [Songer and Crews-Meyer, 2000](#)). By focusing on panel-level behavior, this paper shifts attention from individual to team-level decision-making and provides evidence on how the composition of decision-making bodies shapes collective outcomes through group interaction and deliberative processes.

Second, the paper contributes to the literature on organizations and group decision-making. Economic models predict that individual characteristics shape group outcomes through differences in risk preferences, punishment attitudes, and evaluative standards ([Bagues, Sylos-Labini, and Zinovyeva, 2017](#); [Austen-Smith and Feddersen, 2006](#)). The evidence here suggests that this view is incomplete. The deliberative patterns documented here—where familiar panels with more female judges deliberate more thoroughly rather than more contentiously—highlights how representation shapes collective decisions in expert teams beyond the judiciary, including hiring committees, promotion boards, and policy panels ([Bhuller, Dahl, Løken, and Mogstad, 2020](#); [Arnold, Dobbie, and Hull, 2022](#); [Bourreau-Dubois, Doriat-Duban, Jeandidier, and Ray, 2020](#)).

The findings have direct implications for the design of diversity reform. Current policy debates focus almost entirely on increasing the share of women in judiciaries and expert committees—moving from underrepresentation toward a threshold typically framed as 30 to 40 percent. The evidence here suggests this framing is misspecified in two ways. First, the relevant margin is not the share of women in the institution overall but the composition of the decision-making unit itself—the panel or committee that actually deliberates and decides. Second, incremental increases in that composition, while valuable on equity grounds, are unlikely to change how groups decide unless they cross the level at which deliberative dynamics shift. For five-member panels, that level is a majority. Diversity policies that do not shift the deliberative composition of expert groups are likely to expand representation without changing collective outcomes.

2 Institutional Setting and Data

2.1 Local Courts

Local courts adjudicate different types of criminal cases and are located across Greece. Figure 1 shows their geographic distribution. In my sample, cases originate from 364 distinct local courts, which I group into 11 broad categories.¹ Table A5 shows the different types of local courts, the number of courts of each type in the sample, and the number of cases originating from each. Appeals can be filed within the local court system following a decision. Once no further appeal is available under the applicable law, the case reaches its final local decision. At that stage, a litigant (typically the defendant) or the district attorney may seek further review by filing an appeal to the Supreme Court.²

Appeals to the Supreme Court are typically filed when a litigant disputes the local court's decision or considers the penalty excessive. A defendant filing such an appeal must have been found guilty at the local level. Defendants are often in custody during the period between filing the appeal and the Supreme Court hearing. Decisions of the Supreme Civil and Criminal Court are final within the Greek judicial system.³ No juries are involved at this stage; cases are decided by panels of professional judges.

¹Local courts differ by case type and panel composition, which includes magistrate, misdemeanor, appeal, jury, and juvenile courts (single-, three-, and five-member panels).

²Full name: Civil and Criminal Court of Greece.

³As a member of the Council of Europe, Greece allows cases to be brought before the European Court of Human Rights after the Supreme Court decision.

2.2 The Supreme Court

The Supreme Court (Areios Pagos) is the highest court in Greece for civil and criminal matters. It is responsible for ensuring the uniform interpretation and application of the law across the country and for overseeing the legal consistency of lower court decisions.⁴ Unlike many common-law courts, the Supreme Court of Greece does not follow the principle of precedent and evaluates each case on its own legal merits. Courts and judges are constitutionally independent of parliament and the government.

The basis for appeal is typically that legal or procedural errors may have occurred, or that there are doubts about the consistency with which the law was applied. Grounds for appeal include: perceived errors in the interpretation or application of law by the lower court; procedural errors during the trial that may have affected the fairness of the proceedings or the outcome; a decision that contradicts existing case law or established legal principles; a belief that fundamental rights or constitutional provisions were violated; or a need to ensure uniformity in the application of the law. The Supreme Court does not typically re-examine the facts of the case but focuses on legal and procedural questions to determine whether the lower court's decision was lawfully reached.

A case reaches the Supreme Court when it has received a final decision at the local level and no further local appeal is available. When the Supreme Court accepts an appeal—partially or in full—it sets aside the local court's decision and may: delegate the case to a different local court type or to the same local court for a new trial before a different judge; reject only selected parts of the judgment; or fully accept the appeal and terminate all criminal proceedings against the defendant, which converts the matter to a civil one. Full acceptance implies the initial conviction was erroneous or that a different sentence is warranted. When the Supreme Court finds no grounds to accept the appeal, it rejects it, which renders the local court's decision final and valid. Decisions are never issued on the day of the trial; the committee deliberates and issues its ruling in the period following the hearing.

The judging committee that decides whether to reject or not an appeal case typically consists of five judges: the head, the introducer, and three additional Supreme Court judges.⁵ The introducer is responsible for the preparatory work prior to the trial and presents the case to the full committee on the trial day. The district attorney of the Supreme Court receives every trial application. Decisions are reached by majority vote. The head—who may be the President of the Supreme

⁴The Supreme Court does not directly review the constitutionality of laws; this role is primarily held by other high courts in Greece.

⁵While the vast majority of cases (approximately 90 percent) are decided by five-member panels—with the remaining cases decided by three-member panels—committee size varies in a small number of cases; I control for the number of judges in all specifications.

Court, a Deputy President, or a senior committee member—presides over the trial.

2.3 Career Trajectories of Judges

The Greek judiciary operates with a high degree of independence, with judges serving as career civil servants under constitutional protections that limit political involvement and government interference. Promotions and disciplinary decisions are handled primarily by judicial councils composed of senior judges, and publicly available biographical information on judges is limited.

Entry into the judiciary requires passing competitive examinations at the National School of Judges, followed by training and a probationary period. Judges advance through a hierarchical system based on seniority and merit, with Supreme Court judges selected from among the most senior members of the judiciary.⁶ The Supreme Court consists of roughly 50–60 judges at any point, who sit in panels—typically of five judges—to decide cases. There are 142 unique judges in the dataset for the sample period.

2.4 Allocation of Judges to Cases

Cases are assigned to one of three Supreme Court divisions based on the lexicographic order of the defendant's surname.⁷ This ordering is determined prior to the initial filing of the case and cannot be altered at later stages.⁸ This implies that assignment to divisions is predetermined and unrelated to case characteristics at the Supreme Court stage. Judges are split into three groups based on scheduling preferences, with each group including at least one senior judge. The Court operates on specific trial days, which generates variation in panel composition across cases.⁹ Judges are assigned to groups before the realization of incoming cases and have no influence over case allocation.

⁶Supreme Court judges are selected from among senior career judges and formally appointed by Presidential Decree following a Cabinet decision. They serve until mandatory retirement and can be removed only in cases of serious misconduct or incapacity.

⁷Presidential decree 35 and N.1756/1988 states strict rules about how committees in the Supreme Court should be formed. Document Title: “New Code of Courts Organization and Status of Judges” https://www.elsyn.gr/el/neos_kwdikas.

⁸If the surname of the first-listed defendant starts with any letter in the Greek alphabet from A to KA, then the case is assigned to the first division. If it starts from KB to ΠA, the case is assigned to the second division. Remaining cases are assigned to the third division.

⁹The Supreme Court usually holds trials only on Tuesdays, Wednesdays, and Friday. In my sample, 36% of cases in the sample were decided on a Tuesday, 29% on a Wednesday, and 34% on a Friday. No decisions are made between July 1 and September 15, when the Supreme court is closed. Since 2020, an amendment by the Justice Ministry has shortened the Supreme Court's summer recess by four weeks. As a result, the court does not operate from July 16 to August 31. During this period, regular court sessions are suspended. However, the Supreme Court remains operational for urgent matters, which ensures that essential judicial functions continue even during the recess. Rejections of appeals cannot be filed during this period.

Judges assigned to each division are drawn from these scheduling-based groups. Judges do not select into cases, and case assignment does not depend on panel composition. These institutional features imply that the characteristics of the case—which includes the type of crime and characteristics of the litigants—are plausibly exogenous to the characteristics of the judging committee and cases are as-good-as randomly assigned to panels. Thus, the gender composition of the assigned panel is orthogonal to case characteristics, which supports a causal interpretation of the effect of female representation on judicial outcomes. I provide direct empirical evidence which support this assumption in Section 5.

3 Data Sources and Digitization

I assembled a large-scale data digitization effort to construct a case-level dataset from the online archives of the Supreme Court of Greece which covers the period 2005–2017. I hired and supervised a team of research assistants who manually extracted information from more than 3,700 criminal cases reviewed by the Supreme Court. Data collection began in 2018 and lasted several years. Figures A1 and A2 show examples of case records.¹⁰ To the best of my knowledge, this is the first time these data have been systematically compiled and used for research purposes. The source archive is publicly available and maintained on the Supreme Court’s website.¹¹

The dataset includes detailed case-level information: case characteristics such as the type of case (e.g., homicide, rape, calumny, etc.), the date the appeal was filed, and the trial date; the procedural outcome of each case, including whether the appeal was accepted or rejected and any monetary compensation awarded (in Euros); and information on case participants, including the total number of defendants, plaintiffs, and lawyers, along with the names of all assigned judges, lawyers, the court secretary, and the district attorney. The dataset also indicates whether the plaintiff is an individual or an institutional entity, such as the government, a bank, or

¹⁰A version of this dataset was also more recently used in [Foresta, Megalokonomou, and Vlassopoulos \(2026\)](#), which investigates whether crisis narratives affect judicial enforcement of tax evasion.

¹¹The website of the Supreme Court of Greece is available at <http://areiospagos.gr/en/INDEX.htm>. I used the Greek-language version of the site and extracted all accessible cases listed under “Criminal Cases” that were posted up to and including 2017. Cases that were still pending or protected due to privacy restrictions were not clickable and were therefore excluded. The version of the system in place in 2018, when digitization occurred, displayed cases vertically in chronological order, with the most recent cases appearing at the top and the oldest at the bottom. Each page listed 50 cases, with multiple pages available. Most cases were clickable and provided access to detailed case information. I archived the links for all cases included in the dataset to ensure reproducibility. The full sample consists of 3,753 criminal appeal cases. The baseline sample excludes 754 cases in which the plaintiff is a non-individual entity such as a bank, the government, or another organization, which yields 2,998 cases used in the multiple-litigant specifications. The single-litigant sample further restricts to cases with exactly one defendant or one plaintiff, which yields 2,562 observations. All three samples are used throughout the analysis, with the relevant sample size reported in each table.

another organization. Gender is identified based on Greek naming conventions, which provide unambiguous gender markers. I observe cases that come from 364 different local courts, which provides broad geographic coverage across Greece.

Because I observe the full names of judges, I construct complete participation histories for each judge over the sample period. Although litigants' names are partially anonymized, gender can be identified based on grammatical markers in Greek. Thus, I am able to distinguish whether the defendant or plaintiff is of male or female gender without ambiguity. The most common names appear in word clouds by gender in Figure A3. The total number of cases included in the sample is 3,753. Table A1 summarizes the institutional setting and key features of the dataset. The annual distribution of cases filed and decided is reported in Table A2, which shows broad coverage across the full sample period with no obvious clustering in particular years.

Table 1 reports summary statistics for trials, litigants, committees. On average, panels consist of five judges, and have on average 2 female judges. Panel B of Table 1 reports the distribution of cases across 19 categories, with tax evasion, forgery, theft, and homicide among the most common types.¹² Table A3 reports additional summary statistics. Cases are evenly distributed across trial days and months, and female judges are broadly represented across roles within committees. On average, 67% of appeals are rejected. Figure 2 shows the distribution of the share of female judges in Supreme Court panels.¹³ Table A4 provides detailed definitions of case types and the keywords used in their classification.

Figure A5 shows the share of female judges by case type. Differences across case types are relatively small. Figures A6 and A7 show the distribution of plaintiff and defendant gender, as well as rejection rates, across case types, with limited variation across categories.

4 Estimation Strategy

I exploit the quasi-random assignment of judges to panels to identify the causal effect of female judge representation on appeal outcomes. The key concern is that the gender composition of the assigned committee may be correlated with unobserved case characteristics that independently affect outcomes. Such confounding may arise for several reasons: Litigants may influence assignment, judges may self-select into committees, or committees may hear systematically different cases along unobserved dimensions. The empirical strategy addresses these concerns by comparing

¹²All cases have a primary case type; 189 also have a secondary case type, which follows the case-type classifications.

¹³Figure A4 provides additional background on gender representation in legal education in Greece using data obtained from the Hellenic Ministry of Education. It shows that around 3% of female students and 1% of male students enter law school in the sample period.

cases that are similar in observable features—such as the number of defendants, lawyers, and plaintiffs—and originate from the same type of local court, but differ in the gender composition of the assigned committee due to plausibly random variation.

I use the following specification to study whether the share of female judges on a randomly assigned judging committee affects appeal outcomes:

$$y_{icgt} = \alpha_0 + \delta \text{ShareFemale}_{icgt} + \text{LocalCourtTypeFE}_c + \text{YearFE}_t + \kappa X_i + \text{MonthFE}_t + \text{DayFE}_t + \epsilon_{icgt} \quad (1)$$

where y_{icgt} denotes the appeal outcomes for case i , originally decided in local court c , heard by judging committee g , at time t . The variable $\text{ShareFemale}_{icgt}$ captures the share of female judges on the assigned panel, measured as a share ranging from 0 to 1. g indexes the judging committee assigned to case i . Since each case is decided by exactly one committee, g is uniquely defined at the case level. I consider three outcome variables that are available for each case: a binary indicator equal to 1 if the appeal request is rejected, a binary indicator equal to 1 if the case is delegated to another court or judge, and a placebo outcome. The latter is the monetary trial expenses (in Euros)—an outcome determined solely by the number of courtroom participants rather than by the judging committee. $\text{LocalCourtTypeFE}_c$ captures differences in case severity and procedural complexity associated with the court of origin. By including it, I am able to compare cases that entered the appeals process from similar institutional settings.¹⁴ The vector of case-specific controls (X_i) includes the number of defendants, judges, and lawyers; binary indicators for whether the president is female, the district attorney is male, and the introducer is male; indicators for violent crimes, gender-sensitive cases, and juvenile defendants; the experience of the introducer measured in number of trials; and missing value indicators for all covariates. YearFE_t , MonthFE_t and DayFE_t control for time trends in judicial behavior and case filing. δ is the coefficient of interest, which represents the change in each appeal outcome associated with a one-unit increase in the share of female judges—that is, moving from an all-male to an all-female judge panel. To interpret magnitudes in terms of a 10 percentage point increase in female representation, the relevant quantity is $\delta \times 0.1$.¹⁵

However, even if δ is statistically different from zero, this does not necessarily indicate that committees with different gender compositions judge cases systematically differently. Appeal

¹⁴The classification of local court types is reported in Table A5.

¹⁵A recent study by Sigstad (2026) documents that monotonicity—a key assumption in leniency IV designs—is violated in up to 50% of non-unanimous panel cases. The design in this paper differs fundamentally from leniency IV designs: rather than using individual judge leniency as an instrument for a treatment variable, I exploit quasi-random variation in the share of female judges assigned to a panel as the treatment variable itself. Monotonicity concerns specific to leniency IV designs therefore do not apply to this setting. The identification here requires only that the gender composition of judges in assigned panels is orthogonal to unobserved determinants of the trial outcome, which I validate directly through balancing exercises, placebo tests, and the lead-lag design reported in Sections 5 and 6.

requests assigned to panels with a higher share of female judges may differ along unobserved dimensions—for instance, such committees may disproportionately handle cases of lower severity or complexity. To address this concern, I estimate the following augmented specification:

$$y_{icgt} = \alpha_1 + \delta \text{ShareFemale}_{icgt} + \text{CaseTypeFE}_i + \text{LocalCourtTypeFE}_c + \text{YearFE}_t + \text{MonthFE}_t + \text{DayFE}_t + \epsilon_{icgt}, \quad (2)$$

The augmented specification adds fixed effects for both the primary and secondary case types (CaseTypeFE_i), which absorb unobserved heterogeneity in the nature and severity of offenses, with cases classified into 19 broad categories (e.g., homicide, rape, forgery). These fixed effects ensure that comparisons are made across cases of the same legal type. Heteroskedasticity-robust standard errors are used.¹⁶ To further validate the identifying assumption, Section 5 also implements a lead-lag design that augments the baseline specification with the female share of committees judging on adjacent trial days; if the identifying assumption holds, only the contemporaneous female share should predict appeal outcomes, while past and future shares should have no predictive power.

5 Validity of The Identification Strategy

Based on the institutional setting described above, the identification strategy assumes that the assignment of judges to cases is orthogonal to case characteristics. I test this assumption directly by examining whether the gender composition of the judging committee assigned to a case is related to case type, number of judges, the gender of the district attorney, the number of defendants, plaintiffs, and lawyers involved, and the gender of the presiding judge. To do so, I regress the share of female judges on each case characteristic separately, while I include local court type, day-of-the-week, month, and year fixed effects, in order to assess whether judge assignment is conditionally random. The coefficient of interest in these regressions is β . The unit of observation is the case. The balancing specification is as follows:

$$\text{PercentageFemaleJudges}_{icgt} = \alpha_0 + \beta \text{CaseCharacteristics}_{ict} + \text{LocalCourtTypeFE}_c + \text{YearFE}_t + \text{MonthFE}_t + \text{DayFE}_t + \epsilon_{icgt} \quad (3)$$

Estimates are reported in Table 2, which presents the results of 33 separate regressions testing whether each covariate is associated with the share of female judges on the assigned committee. In the presence of multiple hypothesis testing, some coefficients are expected to appear statistically significant by chance alone; under quasi-random assignment, approximately 10% of coefficients

¹⁶I later report robustness checks with standard errors clustered at the local court and panel levels.

should be significant at the 10% level, 5% at the 5% level, and 1% at the 1% level purely due to random variation. Panel A reports the estimated coefficients β for the case-level characteristics discussed above: 2 out of 14 coefficients are statistically significant, at the 10% and 1% levels (female president and non-individual plaintiff). Panel B examines whether the share of female judges is related to the primary case type; only 2 out of 19 coefficients are statistically significant (fraudulent checks and forgery). Case type is nevertheless included as a control in all specifications. Across both panels, the share of statistically significant coefficients is below 14%. This pattern is broadly consistent with what would be expected under random assignment. Overall, these balancing tests support the identifying assumption that judge assignment is plausibly exogenous conditional on the reported fixed effects. Figure 3 presents these results visually. Each dot reports the OLS coefficient from a separate regression of the share of female judges on the indicated covariate, with thick and thin bars denoting 90% and 95% confidence intervals, respectively. Across all three panels, the estimates cluster tightly around zero, with no systematic pattern of significant deviations, which is consistent with quasi-random assignment of cases to judging committees.

As an additional diagnostic of the identifying assumption, I conduct a randomization check to assess whether observable case characteristics systematically predict panel gender composition. I first estimate a specification in which I use as independent variables only observable case characteristics and while I exclude all judge-related variables. The observable case characteristics include case type indicators, number of litigants, defendants, and lawyers, case length, indicators for immigrant involvement, underage individuals, gender-sensitive or violent cases, whether the district attorney appealed, and missing plaintiff information. I then residualize both the predicted rejection probability and the share of female judges of the panel with respect to time and court-type fixed effects and plot the residualized means by quintile of the female share. Figure 4 shows that the predicted rejection rate is essentially flat across all five quintiles of female representation. Point estimates cluster tightly around zero with narrow confidence intervals, and no quintile exhibits a statistically or economically meaningful deviation. The values on the vertical axis are an order of magnitude smaller than the main effects reported in the baseline results, which indicates that there is no evidence of systematic sorting of cases to panels by gender composition. These results support the view that panel composition is plausibly random conditional on the fixed effects used in the baseline specification, which supports the causal interpretation of the main findings.

6 Results

6.1 Baseline Results

Table 3 reports baseline estimates of the effect of representation of female judges on appeal outcomes. Columns (1)–(2) present results for the full sample, while columns (3)–(4) restrict the sample to cases with a single litigant (either one defendant or one plaintiff). The main outcomes are whether the appeal is rejected and whether the case is delegated to another court or judge. The lower panel reports results for the placebo outcome of monetary trial expenses.¹⁷ The estimated coefficients for the rejection outcome are positive and statistically significant across specifications. A 10 percentage point increase in the share of female judges raises the probability of rejecting an appeal by approximately 1–1.5 percentage points, depending on the specification.¹⁸ These estimates suggest that committees with a higher share of female judges are more likely to uphold first-instance rulings, which lead to outcomes that are systematically less favorable to defendants and more favorable to plaintiffs. Higher rejection rates should not be interpreted as greater punitiveness per se; rather, they likely reflect stricter enforcement of procedural standards and lower tolerance for weak appeals.

The second outcome considers whether the Supreme Court committee delegates the case to another local court or judge. Committees with a higher share of female judges are less likely to delegate cases. A 10 percentage point increase in the share of female judges reduces the probability of delegation by approximately 0.5–1.0 percentage points, consistent with female-majority panels being more likely to reach definitive rulings. The estimates are stable across specifications with and without additional controls, which indicates that the results are not driven by omitted covariates.

Three placebo exercises support the identification. First, the lower panel of Table 3 shows no relationship between the share of female judges and monetary trial expenses, a placebo outcome determined by the number of courtroom participants rather than by the judging committee. All estimates are small, have inconsistent signs, and are insignificant. Second, the last four columns of Table 3 replace the actual female share with the mean coefficient obtained from 1,000 random within-month–year shuffles of the female share. For each shuffle, the female share is randomly

¹⁷The baseline sample excludes cases in which the plaintiff is a non-individual entity such as a bank, the government, or another organization, which reduces the sample to 2,998 cases and 2,562 cases when restricting to single-litigant cases.

¹⁸This finding is consistent with [Foresta \(2022\)](#), who finds that trials presided over by female judges are more likely to result in guilty verdicts at the first-instance level. While that setting differs from the present one—which examines appeal panel composition rather than individual trial judges—both findings suggest that greater female representation is associated with outcomes less favorable to defendants. Table A6 replicates the baseline estimates for cases with a single defendant or a single plaintiff; results are stable across all specifications.

reassigned across cases decided in the same month–year, which breaks the link between committee composition and case outcomes. All placebo coefficients are small, inconsistent in sign, and statistically indistinguishable from zero, in sharp contrast to the baseline rejection estimates of 0.120 (s.e.=0.050) and 0.145 (s.e.=0.054). Third, Figure 5 plots the full distribution of these shuffle coefficients separately for each of the outcomes. For the outcomes of rejection and delegation, the actual estimate lies in the extreme tail of the placebo distribution—which indicates that the observed effect is unlikely to arise by chance—while the monetary expenses estimate sits near the center, consistent with a null effect. This provides strong visual confirmation that the baseline effects reflect genuine differences in judicial behavior rather than spurious time variation or mechanical scheduling patterns.

I next examine whether the estimated effects differ when the plaintiff is not an individual. The hypothesis is that judges may respond differently to individual litigants than to institutional parties, as individual litigants may trigger stronger behavioral responses ([Eisenberg and Farber, 1997](#)). Table 4 reports the baseline estimates for the 754 cases in which the plaintiff is a bank, the government, or another organization. For non-individual plaintiffs, estimated coefficients are smaller, statistically imprecise, and negative for the rejection outcome. The placebo estimates remain statistically insignificant. The evidence is consistent with the interpretation that the baseline effects primarily reflect behavioral responses arising when an individual litigant is involved.

6.2 Heterogeneous Effects

Table 5 examines heterogeneous effects by defendant age.¹⁹ Columns 1–2 and 5–6 correspond to cases with multiple litigants, and columns 3–4 and 7–8 to single-litigant cases. For juvenile defendants (columns 1–4), committees with a higher share of female judges are significantly less likely to reject appeal requests, though the estimates are less precise. This is consistent with a more protective stance toward younger offenders. For adult defendants (columns 5–8), the pattern reverses: committees with a higher share of female judges are significantly more likely to reject appeals, consistent with the baseline results. This heterogeneity suggests that judicial gender composition affects outcomes differently depending on defendant age, with committees with higher female representation adopting a more protective stance toward juvenile defendants while applying stricter procedural standards toward adults. The delegation result mirrors the baseline for adult defendants, while it is attenuated and statistically insignificant for juvenile cases.

¹⁹Defendant age status is directly observed in the data and can be verified by the type of local court that initially decided the case: cases decided by juvenile courts involve juvenile defendants, while cases decided by regular criminal courts involve adult defendants.

Table A7 further disaggregates the baseline results by the gender of the individual litigant. The rejection effect is concentrated in cases with male defendants and male plaintiffs, while estimates for female defendants are negative and imprecisely estimated, due to the small sample of female defendants.

Next, I examine heterogeneity in the effect of representation of female judges by the gender of key judicial roles: the president of the Supreme Court and the introducer judge. I focus on the president and introducer rather than the head judge because these roles carry distinct institutional responsibilities that directly shape panel dynamics. The president sets the broader institutional agenda of the Supreme Court, while the introducer prepares and presents the case to the full committee. Table 6 reports the results. The baseline rejection effect is concentrated in cases where the president of the Supreme Court is female: the estimated coefficient is 0.188 (s.e.=0.070) when the president is female, compared with 0.067 (s.e.=0.093) when the president is male. A similar pattern holds for the introducer judge, with the point estimates are larger when the introducer is female (0.132, s.e.=0.097) compared to a male introducer (0.076, s.e.=0.071), though neither estimate reaches conventional levels of statistical significance. These results indicate that the effect of female representation interacts with the gender of leadership roles within the panel rather than reflecting a uniform compositional effect.

Table A8 examines heterogeneity by the experience of the judging committee, the head judge, and the introducer judge. The rejection effect is essentially zero or negative for below-median experience committees across all three dimensions, while it is positive and statistically significant for above-median experience committees, with the contrast being particularly stark for the head judge. This is consistent with evidence that gender differences in decision-making emerge more strongly at senior career stages ([Adams and Funk, 2012](#); [Eagly and Karau, 2002](#)). The delegation result remains negative across experience subgroups.

Table A9 examines heterogeneity by characteristics of the originating court.²⁰ The rejection effect is larger for cases originating from single-judge courts and smaller and less precisely estimated for multi-judge origin courts. By contrast, the coefficients for first-time and previously appealed cases are similar in magnitude and significance, which suggests little heterogeneity along the appeal-history dimension. The delegation effect remains negative across all subsamples, and no systematic pattern emerges for the placebo outcome.²¹

²⁰Information on the gender composition of judges in the originating court is not available in the case archives.

²¹I also examine heterogeneity across broad crime categories, which includes cases grouped into violent (e.g., homicide, injuries, rape), financial (e.g., tax evasion, fraud, forgery), drugs, property (theft), regulatory (e.g., environmental and immigration offenses), and other categories. As shown in Figure A8, the estimated effects for rejection (panel a) and delegation (panel b) are relatively imprecise and do not exhibit a systematic pattern across groups. This suggests the results are not driven by a particular type of cases.

6.3 Allowing for Non-Linearity in the Female Judges Effect

To explore the potential for non-linear effects of the share of female judges on appeal outcomes, I computed quintiles of the treatment variable and replaced the continuous treatment variable in the regression with a set of quintile indicators. Table 7 presents the estimated effects of being in the second, third, fourth, or fifth quintile of the share of female judges (relative to the first quintile) on each of the appeal outcomes. The bottom row of the table presents some summary statistics on the quintiles. The first quintile includes cases with an average share of female judges equal to 0. The second (Q2), third (Q3), fourth (Q4), and fifth (Q5) quintiles are defined for the following average values, respectively: 0.2, 0.39, 0.6, and 0.86.

I exploit the substantial variation in the share of female judges across cases of the same type, which originate from similar local court types and are adjudicated on the same weekday at the Supreme Court. All estimates reported vertically are derived from the same regression, with outcomes indicated in the column headings. Most estimated effects intensify with higher quintiles of female representation. Statistically significant effects are concentrated in the fourth and fifth quintiles, which correspond to committees with, on average, four or five female judges. For instance, in cases with a share of female judges higher than 60% (fourth quintile) and 67% (fifth quintile), the judging committee is 10 and 13 percentage points more likely, respectively, to reject the appeal request, and 6 to 7.5 percentage points less likely to delegate the case to another court, relative to cases in the first quintile (with no female judges). These patterns imply fewer reversals of lower-court decisions and are consistent with stricter adherence to procedural and legal standards. The placebo outcome—monetary trial expenses—is statistically insignificant for all quintiles and does not follow the pattern identified in other outcomes.

Overall, my findings support the hypothesis that a higher share of female judges in a committee has a progressively larger impact on appeal outcomes. The effect is mainly evident when the share of female judges in the committee exceeds 60 percent. In particular, as the share of female judges increases, panels become more likely to reject appeals, which is consistent with stricter adherence to procedural and legal standards.

One may hypothesize that male judges change their behavior in the presence of a higher share of female judges in the judging committee. These findings are consistent with female judges directly driving the results, rather than male judges changing behavior in response to female presence. However, the nonlinear analysis alone cannot fully distinguish between these channels.

6.4 Falsification Exercise

I conduct a falsification exercise to assess whether the baseline estimates capture causal effects of the judicial gender composition rather than spurious correlations driven by serial correlation, court-level shocks, or persistent scheduling patterns. To test this, I replace the contemporaneous share of female judges with the share of female judges assigned on the previous trial day and on the following trial day. Because these placebo measures are mechanically unrelated to the composition of the committee deciding the focal case, they should not predict judicial outcomes if the identifying assumptions hold. Figure A9 shows the distribution of the share of female judges assigned to the case and the placebo share of females assigned to cases in the previous and following day.

Table A10 reports the results of the falsification exercise. The outcome variable is whether the appeal request is rejected in columns (1)–(4), and whether the trial is delegated to another local court or judge in columns (5)–(8). Across all specifications, the placebo coefficients are small in magnitude and statistically indistinguishable from zero. For example, the estimated coefficients on the share of female judges on the previous day range from -0.010 to 0.020 , with standard errors between 0.017 and 0.036 , while the corresponding coefficients for the following day range from -0.035 to -0.013 , with standard errors between 0.016 and 0.035 . None of these estimates approaches conventional levels of statistical significance. The results are stable across outcomes and for both multiple-litigant and single-litigant cases.²²

Taken together, the absence of placebo effects provides reassurance that the baseline estimates are not driven by unobserved court-level trends, temporal dependence in scheduling, or mechanical correlations in judge assignment. Instead, the evidence supports the interpretation that contemporaneous variation in the gender composition of panels is the relevant source of identifying variation.

6.5 Lead-Lag Design

As a further test of the identifying assumption that the gender composition of the assigned committee—rather than unobserved time trends in case characteristics or scheduling patterns—drives the baseline results, I implement a lead-lag design. Specifically, I augment the baseline

²²Figure A9 complements these regressions by plotting the distributions of the actual share of female judges on the trial day together with the distributions constructed from the share of female judges assigned to the previous and following trial days. The placebo distributions closely overlap with the support and overall shape of the actual distribution, which indicates that the placebo measures exhibit comparable variation but lack predictive power for case outcomes.

specification with the average female share of committees judging on the three trial days before and the three trial days after the focal case, using the following regression:

$$y_{it} = \sum_{k=-3}^3 \beta_k \text{FemaleShare}_{t+k} + \kappa X_i + \text{LocalCourtTypeFE}_c + \text{YearFE}_t + \text{MonthFE}_t + \text{DayFE}_t + \varepsilon_{it} \quad (4)$$

where FemaleShare_{t+k} denotes the average share of female judges assigned to cases decided k trial days relative to the focal case, with $k = 0$ corresponding to the contemporaneous committee. The identifying assumption implies that only β_0 should be statistically different from zero, while the lead and lag coefficients $\beta_{-3}, \beta_{-2}, \beta_{-1}, \beta_{+1}, \beta_{+2}, \beta_{+3}$ should be indistinguishable from zero. Intuitively, if the result were driven by court-level time trends, unobserved case-scheduling patterns, or persistent variation in judicial behavior across days, the female share on adjacent trial days would also predict outcomes. By contrast, if the effect is genuinely causal and operates only through the composition of the assigned committee, past and future female shares should have no predictive power.

Figure 6 plots the estimated coefficients β_k for $k \in \{-3, -2, -1, 0, +1, +2, +3\}$ together with 95% confidence intervals, separately for the two main outcomes: the probability that the appeal request is rejected (Panel A) and the probability that the case is delegated to another court or judge (Panel B). The results are consistent with the identifying assumption across both outcomes. The contemporaneous coefficient β_0 is positive, economically meaningful, and statistically significant in Panel A, closely matching the baseline estimate in Table 3, and negative and significant in Panel B, consistent with the baseline delegation result. All six lead and lag coefficients are small in magnitude, inconsistent in sign, and statistically indistinguishable from zero across both panels. The pre-period coefficients β_{-3}, β_{-2} , and β_{-1} are of particular importance, as they constitute a falsification test: the female share of committees judging in the days before a case is decided cannot causally affect that case's outcome, so any significant pre-trend would indicate the presence of confounding. The absence of pre-trends in both panels, combined with precisely estimated contemporaneous effects that closely match the baseline, provides strong visual confirmation that the results reflect the causal effect of committee gender composition rather than spurious time variation or mechanical scheduling correlations.

7 Robustness Checks

In this section, I report a series of robustness checks that support the causal interpretation of the main findings. Table A12 re-estimates the baseline specifications with standard errors clustered at the local court level instead of heteroskedasticity-robust standard errors used in the main analysis. The magnitude and statistical significance of the coefficients in the robustness exercise remain very similar to the baseline results.

Table A13 re-estimates the baseline specifications for the rejection, delegation, and monetary expenses (placebo) outcomes with standard errors clustered at the panel level to account for potential correlation in outcomes across cases decided by the same panel. Among the cases in the sample, I identify 837 unique panel compositions, with the same combination of judges sitting together on average 4.5 times over the sample period. The magnitude and statistical significance of the coefficients remain very similar to the baseline results.

Table A14 augments the baseline specifications with prosecutor fixed effects to absorb unobserved heterogeneity in prosecutorial behavior (e.g., charging practices). Columns (1) and (3) reproduce the baseline estimates for the multiple and single litigant samples respectively. Columns (2) and (4) add prosecutor fixed effects. The sample includes 62 prosecutors. Across all specifications, the estimated effects remain stable in magnitude and standard errors change little, which indicates that the main results in Table 3 are not driven by systematic differences across prosecutors.

Table A15 stacks the data at the judge–case level and includes judge fixed effects to control for time-invariant unobserved heterogeneity across judges. Each case contributes as many observations as there are judges assigned to it, so the stacked sample is larger than the case-level sample. Because individual votes are unavailable, outcomes remain measured at the case level and are repeated across all judges assigned to the same case. This approach allows judge fixed effects to absorb permanent differences in behavior across judges while leaving the identifying variation in panel gender composition intact. This approach increases the sample size and allows me to account for judge-specific, time-invariant unobserved heterogeneity. Results are reported with robust standard errors and with clustering at both the local court and judge levels. The significance patterns remain consistent across specifications for both outcomes. These patterns closely align with the baseline results, which further supports the robustness of the findings.

As an additional robustness check, I examine whether the baseline results are sensitive to extreme values of the treatment variable. In Table A16, I trim the distribution of the share of female judges by removing observations in the bottom 5th percentile and the top 5th percentile (and drop

cases that had no female or only female judges in their committees). This procedure mitigates the potential influence of outliers that may have a disproportionate impact on the estimates. Re-estimating the baseline specifications on the trimmed sample yields coefficients that are very similar in magnitude and statistical significance to the main results, which indicates that the findings are not driven by extreme realizations of female representation on judging committees.

I also assess whether the baseline results are driven by a small number of influential judges whose gender or behavior might systematically affect both committee composition and outcomes. To do this, I implement a leave-one-judge-out robustness test. For each of the judges in the sample, I drop all cases in which that judge participated in any role—as head, introducer, or committee member—and re-estimate the baseline specification from column (2) of Table 3. Figure 7 plots the resulting leave-one-out coefficients, ranked by magnitude, together with the full sample estimate and its 95% confidence interval. The estimates are strikingly stable: nearly all leave-one-out coefficients cluster tightly around the full sample estimate of 0.120, remain positive, and are statistically significant at conventional levels. The two estimates with wide confidence intervals at the left of the figure correspond to judges who appeared in very few cases, so that dropping them substantially reduces the sample size and lowers precision; their point estimates nonetheless remain positive and close to the baseline. Taken together, these results confirm that the main finding is not driven by any individual judge and is robust to the exclusion of any single participant from the sample.

A further concern is that the baseline result reflects the aggregation of individual judge strictness rather than a genuine panel-level effect—that is, panels with a higher share of female judges may simply tend to include judges who are, on average, individually more likely to reject appeals. To address this, I construct a leave-one-out measure of judge harshness, defined as the mean rejection rate of judge j across all cases excluding case i , to avoid a mechanical correlation between the control and the outcome. I then compute the average of this harshness measure across all judges on the committee and include it as an additional control in the baseline specification. Table A11 shows that the estimated coefficient on the female share remains positive, statistically significant, and changes only slightly in magnitude after controlling for average committee harshness. This confirms that the baseline result is not driven by systematic differences in the individual strictness of judges assigned to female-majority committees.

8 Mechanisms Behind Gender Composition Effects

Having established that panels with a higher share of female judges are significantly more likely to reject appeal requests and less likely to delegate cases, I now turn to mechanisms. The central claim of this section is that the panel-level effects are more consistent with mechanisms operating through collective deliberation and group interaction than with the simple aggregation of individually stricter judges. I develop this argument in three steps. I first show that female and male judges sit on panels with statistically indistinguishable rejection rates when measured at the judge level, which is difficult to reconcile with systematic sorting of individually harsher female judges as the primary driver. I then develop and test the deliberation mechanism directly, which uses variation in case complexity and panel familiarity to distinguish genuine deliberative thoroughness from coordination costs. Finally, I discuss the limits of what the administrative data can establish and situate the findings in the context of complementary evidence from the World Values Survey, which I treat as explicitly illustrative background rather than a direct mechanism test.

8.1 Evidence Against Simple Aggregation of Individual Preferences

The most natural alternative to a panel-level mechanism is an individual-level one: panels with more female judges may produce higher rejection rates simply because female judges are individually stricter, so that adding more of them to a panel mechanically raises the group's rejection rate through aggregation. I cannot directly test this claim because individual votes are unobservable in the administrative data—decisions are recorded at the case level, not the judge level. What I can test is whether female judges systematically sit on panels that produce different outcomes, which would be consistent with individual-level differences in voting behavior.

To do this, I construct a judge-level measure of panel outcomes by computing, for each judge, their mean rejection rate across all cases in which they participated in the sample period. This measure captures the average case outcomes of the panels a judge sat on, not the judge's individual vote. I compute an analogous measure for delegation. This yields a sample of 142 judges—75 male and 67 female. Figure A10 plots the kernel density of these two judge-level measures separately for male and female judges. For rejection rates (panel a), the distribution for female judges is more concentrated, which peaks sharply in the range 0.65—0.75, whereas the male distribution is flatter and more dispersed. For delegation rates (panel b), both distributions are heavily concentrated near zero, with male judges displaying a longer right tail.

Despite these distributional differences, two-sample t-tests confirm that neither the difference in mean rejection rates (0.688 vs 0.712, $t=-0.58$, $p=0.56$) nor the difference in mean delegating rates

(0.067 vs 0.070, $t=-0.15$, $p=0.88$) is statistically distinguishable from zero. Female and male judges therefore sit on panels with similar average outcomes. This evidence is inconsistent with the specific concern that female judges systematically sort onto panels that are more rejection-prone for reasons unrelated to gender composition, and is inconsistent with a story in which female judges are individually so much stricter that their presence mechanically drives up panel rejection rates through simple aggregation.

These results are difficult to reconcile with an account in which outcomes simply reflect the additive aggregation of fixed individual preferences. They redirect attention toward a mechanism that operates at the group level, which I develop in the following section.

8.2 Panel-Level Deliberation: The Affirmative Account

If the effects do not arise solely from the sorting or additive aggregation of individually stricter judges, what panel-level process produces them? The affirmative claim is that panels with more female judges engage in more thorough collective deliberation, and that this investment in deliberation is consistent with stricter enforcement of procedural standards. I build this case in three stages, each designed to sharpen the interpretation and address the most natural alternative explanation at each step.

8.3 Deliberation Time and Female Representation

Most of the literature focuses on judicial outcomes such as conviction rates, sentencing, or case disposition (Hoekstra and Street, 2021; Bindler and Hjalmarsson, 2020; Anwar, Bayer, and Hjalmarsson, 2019; Collins, Manning, and Carp, 2010). I exploit an underexplored but directly informative margin: deliberation time, measured as the number of days between the court trial and the issuance of the final decision. While imperfect, this measure captures variation in the amount of time panels devote to post-hearing decision-making. A longer time to decision may reflect greater caution, diligence, or more intensive internal discussion among panel members (Epstein, Landes, and Posner, 2013; Dimitrova-Grajzl, Grajzl, Sustersic, and Zajc, 2012)—all signatures of more thorough deliberation rather than individual strictness. If panels with a higher share of female judges are more deliberative, they should take longer to decide, and the gap should be larger precisely where deliberation matters most: in complex cases.

Using the newly compiled data from the Greek Supreme Court, I measure deliberation time as the number of days between the trial and the final decision. I estimate specification 2 using deliberation time as the outcome. Table 8 reports the results. A 10 percentage point increase in

female representation raises deliberation time by 6.8–7.8 days across specifications. Stratifying by case complexity—proxied by case length relative to the sample median of four pages—the delay is present for both simple and complex cases but is substantially larger for complex ones: moving from an all-male to an all-female panel raises deliberation time by approximately 81–88 days in complex cases, compared to roughly 60 days in simple cases. Panels with a higher share of female judges invest greater scrutiny precisely where cases demand it.²³ This pattern is consistent with prior evidence which suggests that women tend to adopt more cautious or deliberative decision-making styles, particularly in environments characterized by uncertainty and constitutes the first piece of affirmative evidence for the deliberation account.

An alternative explanation is that the estimated deliberation effects reflect administrative congestion rather than differences in collective deliberation. To address this concern, I construct several measures of court workload and congestion, which include the number of cases heard per day, per day originating from the same local court, and at monthly aggregation levels. Table A21 shows the results. In Panel A, I find no evidence of a systematic relationship between female representation and these congestion measures. In Panels B and C, I find that controlling for court congestion leaves the estimated effects on appeal and delegation outcomes unchanged. These findings suggest that the results are unlikely to be driven by scheduling patterns or administrative workload.

8.4 Deliberative Thoroughness or Coordination Costs?

The deliberation time finding admits two interpretations that must be distinguished, because they carry opposite implications for the welfare interpretation of the main results. Under a deliberative thoroughness interpretation, panels with a higher share of female judges invest more effort in reviewing cases and reach better-considered, procedurally stricter decisions—the additional time is a signal of quality. Under a coordination cost interpretation, panels with a higher share of female judges simply take longer because they face greater internal disagreement or higher consensus costs—the additional time is a symptom of friction, not diligence. I use variation in panel familiarity to test between these accounts.

I define panel familiarity as the number of times the same combination of judges has sat together prior to the focal case, with a sample median of three prior joint sittings. The coordination cost

²³In Table A20 I examine whether the outcomes of rejecting the appeal request and delegating the case to another court exhibit non-linearities with respect to the share of female judges on the panel. Both outcomes reproduce the pattern observed in the main specifications, with effects concentrated in panels that belong to the fourth and fifth quintiles of the female-judge share. This suggests that behavioral differences emerge only once female representation reaches relatively high levels, albeit imprecisely estimated.

and deliberative thoroughness interpretations generate opposite predictions about how familiarity should interact with the deliberation delay. Under coordination costs, the delay should be largest in less-familiar panels, where judges have had fewer opportunities to develop shared norms and efficient procedures, and should attenuate as panels accumulate joint experience. Under deliberative thoroughness, the delay should be larger in more-familiar panels, where established working relationships allow female judges to more fully express and reinforce deliberative norms—familiarity amplifies rather than erodes the effect.

Table A18 reports the results. The deliberation delay is approximately twice as large in above-median familiarity panels—around 104–108 days—compared to below-median familiarity panels—around 58–59 days—and is statistically significant in both subgroups. This pattern directly contradicts the coordination cost account. If unfamiliarity were the driver, familiarity should reduce the gap. Instead, the effect grows with familiarity, consistent with a social dynamics account in which female judges in established panels are better positioned to shape deliberative norms. This pattern is more consistent with enhanced deliberative engagement than with a pure coordination-cost interpretation. The familiarity test provides particularly strong evidence in favor of the deliberative-engagement interpretation.

8.5 Deliberation and Nonlinear Effects of Panel Composition

An additional test of the deliberation account comes from comparing how deliberation time and appeal outcomes each respond to changes in panel composition across the full female representation distribution. If deliberation is the operative channel, the two should diverge in a specific way: deliberation intensity should increase continuously with female representation, while appeal outcomes should remain negligible at modest levels of representation and emerge only once women constitute a majority and can shift the collectively supported outcome. Table A19 examines whether this pattern holds. Unlike the rejection and delegation outcomes—where statistically significant effects are concentrated in the fourth and fifth quintiles, which corresponds to a female share above 60 percent—the deliberation delay emerges already at the second quintile and increases monotonically through the fifth, with statistically significant effects from the second quintile onward. The delay is substantially larger for complex cases—panels in the highest quintile take approximately 63 additional days relative to all-male panels, compared to approximately 38 days for simple cases with multiple litigants.

This divergence is itself informative and directly consistent with the deliberation account. Deliberative thoroughness operates more continuously across the female representation distribution

— even one or two additional female judges changes how the panel processes cases. But translating deliberative differences into different verdicts requires overcoming the inertia of collective decision-making, which requires shifting the majority-supported outcome of the panel. Part of this nonlinearity is naturally consistent with majority voting rules, under which changes in panel outcomes require shifts in the effective median voter. However, a pure median-voter interpretation alone does not explain why deliberation time increases continuously across the full distribution of female representation, which includes at modest levels where verdict effects remain absent. The evidence instead suggests that changes in panel composition affect how cases are processed throughout the distribution, while changes in collective outcomes emerge only once deliberative differences become sufficiently influential to alter majority decisions. The nonlinear pattern in the main outcomes therefore appears to reflect both the mechanics of collective voting and the increasing influence of deliberative dynamics as female representation rises. One interpretation is that modest increases in female representation affect discussion and review intensity continuously, but larger representation shares are required before these changes alter majority-supported outcomes within the panel. This provides a more precise interpretation of the observed nonlinearity than a generic critical-mass account.

Overall, panels with a higher share of female judges take longer to decide, especially in complex cases. This delay grows with panel familiarity rather than shrinking, which is inconsistent with coordination costs and supports deliberative thoroughness as the operative channel. Deliberative differences emerge at modest levels of female representation but translate into stricter appeal outcomes only once female judges become the majority. These three pieces of evidence provide a coherent and internally consistent account of the panel-level mechanism: more thorough deliberation produces stricter enforcement of procedural standards, not greater individual punitiveness toward defendants.

Supplementary evidence from the World Values Survey shows that among Greek respondents, women report significantly lower risk tolerance and lower patience than men, while men report substantially higher negative reciprocity (Table A22). Although these respondents are not judges and the estimates are not causally identified, the pattern is broadly consistent with the behavioral dimensions most relevant to deliberative decision-making under uncertainty [Epstein et al. \(2013\)](#); [Eckel and Grossman \(2008\)](#); [Adams and Funk \(2012\)](#). The deliberation-time evidence in Section 8.2, together with the familiarity analysis and the nonlinear outcome patterns, remains the primary empirical foundation for the mechanism interpretation developed in the paper.

8.6 Local Social and Economic Context

I examine whether the effect of committee gender composition varies with local social and economic conditions. The motivation draws on a signaling argument: in less egalitarian environments, female judges may face stronger pressures to demonstrate professional authority and impartiality, which leads them to adopt stricter decision-making standards as a credibility signal (Eagly and Karau, 2002). In more empowered contexts, these pressures diminish and behavior converges toward the baseline. To test this, I interact the share of female judges with measures of women's educational attainment, regional economic activity (log GDP), and social norms proxied by civil marriages. These variables are drawn from the 2001 Population Census and matched to the municipality of the local court. I focus on the probability that appeal requests are rejected.

Table A17 shows that all interaction terms are negative and statistically significant. The effect of female judges declines with the prevalence of civil marriages: a one-standard deviation increase reduces the baseline effect by about 40 percent (0.05 relative to 0.125). The interaction with log GDP is also negative; a one-unit increase reduces the effect by roughly 10 percent (0.053 relative to 0.545). Similarly, the interactions with women's educational attainment—upper secondary, tertiary, and postgraduate—are negative, which indicates that the marginal effect declines as education rises. The main effect of female judges remains positive and statistically significant across specifications.²⁴

These results imply that the effect of female representation is stronger in less developed and less socially liberal environments, and weaker in more supportive contexts. This suggests that judicial behavior is shaped not only by who judges are (their gender composition) but also by the social environment in which they operate. One plausible interpretation is that in less empowered or more traditional environments, female judges face stronger incentives to signal authority, impartiality, or professional credibility, which leads to relatively stricter decision-making (Alesina and Giuliano, 2015). This interpretation is consistent with evidence that social context shapes how judicial identity characteristics translate into decisions Shayo and Zussman (2011). In more empowered contexts, these pressures diminish and behavior converges toward the baseline (Goldin, 2014; Bursztny, González, and Yanagizawa-Drott, 2020). Overall, these findings highlight the importance of accounting for local norms and socioeconomic conditions when evaluating the behavioral consequences of gender composition in decision-making committees.

²⁴Figure A11 shows that the effect is strongest in low-education regions, attenuates as women's educational attainment rises, and turns negative in the highest-education regions (where approximately 90% of women hold upper secondary education).

9 Conclusion and Discussion

In this paper, I examine how gender representation shapes collective decision-making in expert committees. Using newly digitized data on more than 3,700 appeals in the Greek Supreme Court and exploiting quasi-random panel assignment, I show that panels with greater female representation are more likely to reject appeals and less likely to delegate cases. These effects are not linear. They arise primarily when female representation exceeds roughly 60 percent, which indicates that modest changes in composition may not alter outcomes unless they shift the dynamics of deliberation. The effects are stronger when women hold leadership roles and in cases involving individual litigants, and are attenuated for juvenile defendants, which suggests that representation interacts with both institutional structure and case characteristics. These results suggest that policies aimed at increasing representation may have limited effects unless they reach levels that meaningfully alter group dynamics. In a five-member panel, this requires appointing at least three female judges—a substantially higher target than current average representation levels in the Greek Supreme Court and in many comparable judiciaries.

The evidence points to a mechanism operating at the group level rather than through individual preferences. Female judges are not individually stricter than their male counterparts, as their rejection rates are statistically indistinguishable when measured at the judge level. Instead, panels with more women take longer to reach decisions—especially in complex cases—consistent with more intensive deliberation at the group level. The higher rejection rates associated with greater female representation therefore likely reflect stricter enforcement of procedural standards rather than greater punitiveness toward defendants.

The mechanism I document—how representation shapes deliberation in expert teams—is likely to extend beyond panels to other collective decision-making bodies, including hiring committees, promotion boards, and policy panels. In these settings, as in courts, decisions are made collectively under uncertainty, and group composition can influence both deliberative processes and final outcomes. The findings are therefore most directly informative for institutions that operate through formal deliberation and majority rule in high-stakes professional environments. The results suggest that representation matters not only through who participates in decision-making, but also through how groups deliberate once representation reaches levels sufficient to reshape internal dynamics. This highlights the importance of considering both composition and deliberation when evaluating the consequences of diversity in professional organizations. At the same time, the evidence comes from appellate decision-making in a specific institutional setting, and caution is warranted in extrapolating the magnitudes of these effects to other contexts.

References

- Abrevaya, J. and D. S. Hamermesh (2012). Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)? *Review of Economics and Statistics* 94(1), 202–207.
- Adams, R. B. and P. Funk (2012). Beyond the Glass Ceiling: Does Gender Matter? *Management Science* 58(2), 219–235.
- Alesina, A. and P. Giuliano (2015). Culture and Institutions. *Journal of Economic Literature* 53(4), 898–944.
- Anwar, S., P. Bayer, and R. Hjalmarsson (2012). The Impact of Jury Race in Criminal Trials. *Quarterly Journal of Economics* 127(2), 1017–1055.
- Anwar, S., P. Bayer, and R. Hjalmarsson (2019). A Jury of Her Peers: The Impact of the First Female Jurors on Criminal Convictions. *The Economic Journal* 129(619), 603–650.
- Anwar, S., P. Bayer, and R. Hjalmarsson (2022, June). Unequal Jury Representation and Its Consequences. *American Economic Review: Insights* 4(2), 159–74.
- Arnold, D., W. Dobbie, and P. Hull (2022, September). Measuring Racial Discrimination in Bail Decisions. *American Economic Review* 112(9), 2992–3038.
- Ash, E. (2022). Are Judges Political? An Empirical Analysis of the Federal Judiciary. *American Law and Economics Review* 24(2), 359–403.
- Ash, E., D. L. Chen, and A. Ornaghi (2024). Gender Attitudes in the Judiciary: Evidence from US Circuit Courts. *American Economic Journal: Applied Economics* 16(1), 314–350.
- Austen-Smith, D. and T. J. Feddersen (2006). Deliberation, Preference Uncertainty, and Voting Rules. *American Political Science Review* 100(2), 209–217.
- Bagues, M., M. Sylos-Labini, and N. Zinovyeva (2017). Does the Gender Composition of Scientific Committees Matter? *American Economic Review* 107(4), 1207–38.
- Bertrand, M., E. Kamenica, and J. Pan (2015). Gender Identity and Relative Income within Households. *The Quarterly Journal of Economics* 130(2), 571–614.
- Bhuller, M., G. B. Dahl, K. V. Løken, and M. Mogstad (2020). Incarceration, Recidivism, and Employment. *Journal of Political Economy* 128(4), 1269–1324.

- Bindler, A. and R. Hjalmarsson (2020). The Persistence of the Criminal Justice Gender Gap: Evidence from 200 Years of Judicial Decisions. *Journal of Law and Economics* 63(2), 297–339.
- Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019, March). Beliefs about Gender. *American Economic Review* 109(3), 739–73.
- Bourreau-Dubois, C., M. Doriat-Duban, B. Jeandidier, and J.-C. Ray (2020). Does Gender Diversity in Panels of Judges Matter? Evidence from French Child Support Cases. *International Review of Law and Economics* 63, 105929.
- Boyd, C. L. (2016). Representation on the Courts? The Effects of Trial Judges' Sex and Race. *Political Research Quarterly* 69(4), 788–799.
- Boyd, C. L., L. Epstein, and A. D. Martin (2010). Untangling the Causal Effects of Sex on Judging. *American Journal of Political Science* 54(2), 389–411.
- Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2020). Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia. *American Economic Review* 110(10), 2997–3029.
- Cabral, M. and M. Dillender (2024, February). Gender Differences in Medical Evaluations: Evidence from Randomly Assigned Doctors. *American Economic Review* 114(2), 462–99.
- Card, D., S. DellaVigna, P. Funk, and N. Iriberry (2020). Are Referees and Editors in Economics Gender Neutral? *Quarterly Journal of Economics* 135(1), 269–327.
- Chattopadhyay, R. and E. Duflo (2004). Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica* 72(5), 1409–1443.
- Clots-Figueras, I. (2012). Are Female Leaders Good for Education? Evidence from India. *American Economic Journal: Applied Economics* 4(1), 212–44.
- Collins, P. M., K. L. Manning, and R. A. Carp (2010). Gender, Critical Mass, and Judicial Decision Making. *Law and Policy* 32(2), 260–281.
- Coontz, P. (2000). Gender and Judicial Decisions: Do Female Judges Decide Cases Differently than Male Judges? *Gender Issues* 18(4), 59–73.

- Dahl, G. B., A. R. Kostøl, and M. Mogstad (2014). Family Welfare Cultures. *The Quarterly Journal of Economics* 129(4), 1711–1752.
- Dahlerup, D. (1988). From a Small to a Large Minority: Women in Scandinavian Politics. *Scandinavian Political Studies* 11(4), 275–298.
- Dimitrova-Grajzl, V., P. Grajzl, J. Sustersic, and K. Zajc (2012). Court Output, Judicial Staffing, and the Demand for Court Services: Evidence from Slovenian Courts. *International Review of Law and Economics* 32(1), 19–29.
- Eagly, A. H. and S. J. Karau (2002). Role Congruity Theory of Prejudice Toward Female Leaders. *Psychological Review* 109(3), 573–598.
- Eckel, C. C. and P. J. Grossman (2008). Men, Women and Risk Aversion: Experimental Evidence. *Handbook of Experimental Economics* 1, 1061–1073.
- Eisenberg, T. and H. S. Farber (1997). The Litigious Plaintiff Hypothesis: Case Selection and Resolution. *RAND Journal of Economics* 28, S92–S112.
- Epstein, L., W. M. Landes, and R. A. Posner (2013). *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*. Harvard University Press.
- Escobar-Lemmon, M. C., V. J. Hoekstra, A. J. Kang, and M. C. Kittilson (2021). Breaking the Judicial Glass Ceiling: The Appointment of Women to High Courts Worldwide. *The Journal of Politics* 83(2), 662–674.
- Feenan, D. (2008). Women Judges: Gendering Judging, Justifying Diversity. *Journal of Law and Society* 35(4), 490–519.
- Flanagan, F. X. (2018). Race, Gender, and Juries: Evidence from North Carolina. *The Journal of Law and Economics* 61(2), 189–214.
- Foresta, A. (2022). Lady Justice: The Impact of Female Judges on Trials’ Verdicts in North Carolina. *Discussion Papers* 22/04, Department of Economics, University of York.
- Foresta, A. (2025). Beyond a Reasonable Doubt: The Impact of Jurors’ Political Affiliations on Trials: Evidence from North Carolina. *Journal of Law and Economics* 68(2).
- Foresta, A., R. Megalokonomou, and M. Vlassopoulos (2026). Crisis Narratives and Judicial Enforcement: Evidence from the Greek Fiscal Crisis. (26088).

- Gagliarducci, S. and M. D. Paserman (2011). Gender Interactions within Hierarchies: Evidence from the Political Arena. *The Review of Economic Studies* 79(3), 1021–1052.
- Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review* 104(4), 1091–1119.
- Hanemaaijer, A., N. Ketel, and O. Marie (2025). Minority Salience and Criminal Justice Decisions. *American Economic Review: Insights*. Forthcoming.
- Hoekstra, M. and B. Street (2021). The Effect of Own-Gender Jurors on Conviction Rates. *The Journal of Law and Economics* 64(3), 513–537.
- Johnson, B. D. (2014). Judges on Trial: A Reexamination of Judicial Race and Gender Effects across Modes of Conviction. *Criminal Justice Policy Review* 25(2), 159–184.
- Kanter, R. M. (1977). Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women. *American Journal of Sociology* 82(5), 965–990.
- Knepper, M. M. (2018). When the Shadow Is the Substance: Judge Gender and the Outcomes of Workplace Sex Discrimination Cases. *Journal of Labor Economics* 36(3), 623–664.
- Moyer, L. P., J. Szmer, S. Haire, and R. K. Christensen (2021). “All Eyes Are on You”: Gender, Race, and Opinion Writing on the U.S. Courts of Appeals. *Law and Society Review* 55(3), 452–472.
- National Women’s Law Center (2016). Judges and the Courts: Women in the Federal Judiciary: Still a Long Way to Go. *Technical Report*.
- OECD (2018). *Better Policies for Better Lives: Gender Equality*. OECD.
- Rehavi, M. M. (2007). Sex and Politics: Do Female Legislators Affect State Spending? *Unpublished manuscript, University of Michigan*.
- Shayo, M. and A. Zussman (2011). Judicial Ingroup Bias in the Shadow of Terrorism. *Quarterly Journal of Economics* 126(3), 1447–1484.
- Sigstad, H. (2026). Monotonicity Among Judges: Evidence from Judicial Panels and Consequences for Judge IV Designs. *American Economic Review* 116(1), 189–208.

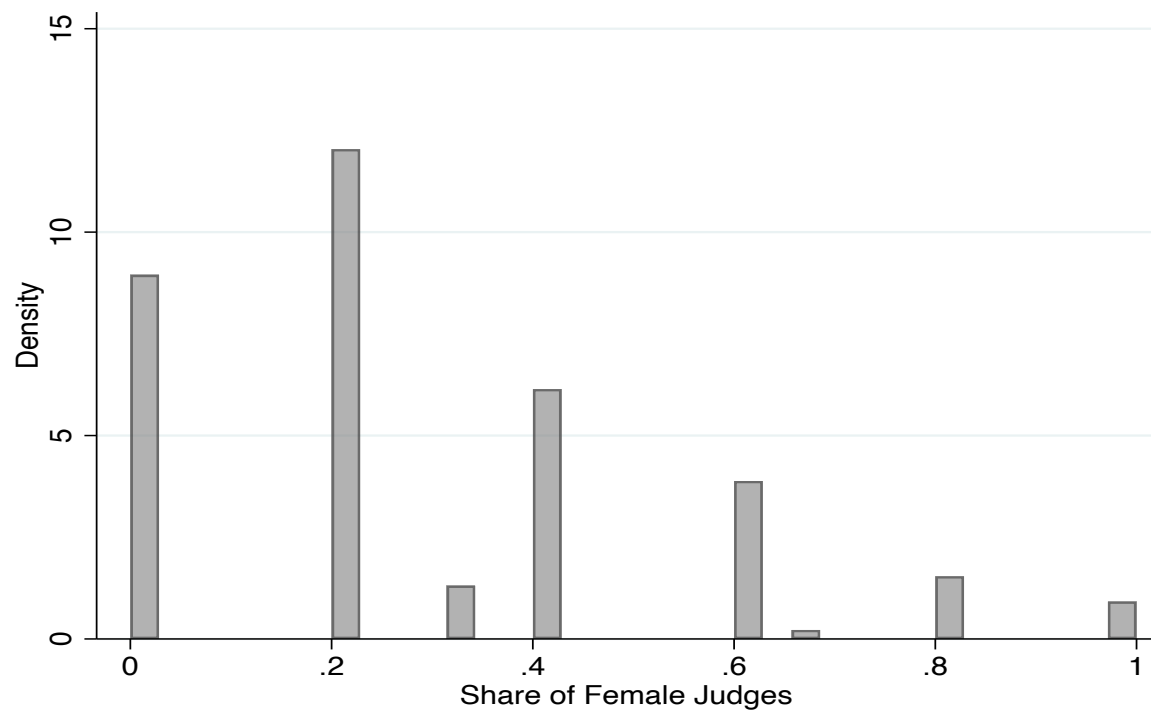
- Songer, D. R. and K. A. Crews-Meyer (2000). Does Judge Gender Matter? Decision Making in State Supreme Courts. *Social Science Quarterly*, 750–762.
- Steffensmeier, D. and C. Hebert (1999). Women and Men Policymakers: Does the Judge's Gender Affect the Sentencing of Criminal Defendants? *Social Forces* 77(3), 1163–1196.
- Swers, M. L. (1998). Are Women More Likely to Vote for Women's Issue Bills than Their Male Colleagues? *Legislative Studies Quarterly*, 435–448.
- Vega, A. and J. M. Firestone (1995). The Effects of Gender on Congressional Behavior and the Substantive Representation of Women. *Legislative Studies Quarterly*, 213–222.
- Walker, T. G. and D. J. Barrow (1985). The Diversification of the Federal Bench: Policy and Process Ramifications. *The Journal of Politics* 47(2), 596–617.

Figure 1: MAP OF LOCAL COURTS



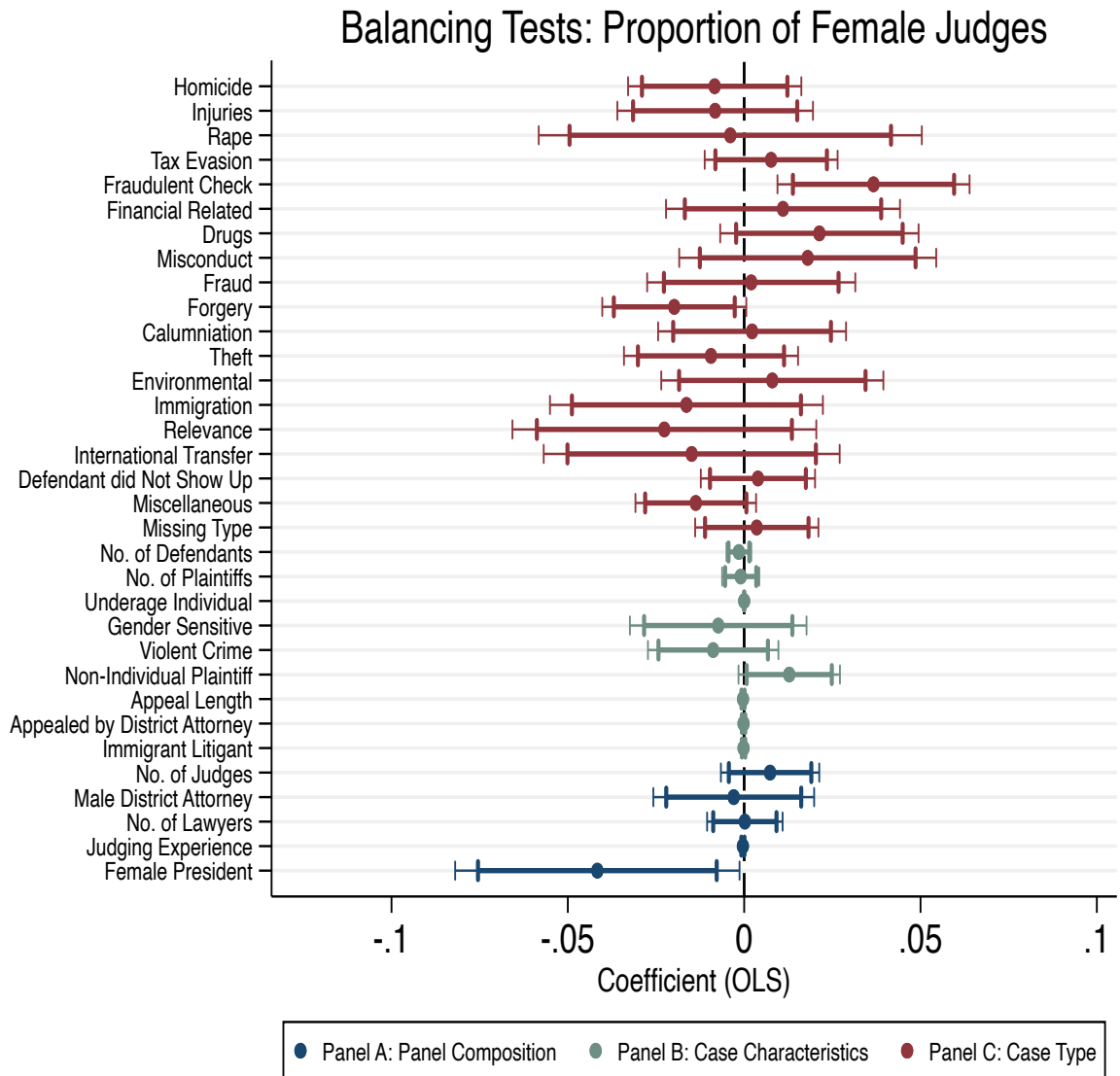
Notes: The map illustrates the locations of local courts in Greece. Some local courts may be situated in the same building, so the size of each dot indicates the concentration of courts in that area.

Figure 2: SHARE OF FEMALE JUDGES IN JUDGING COMMITTEES



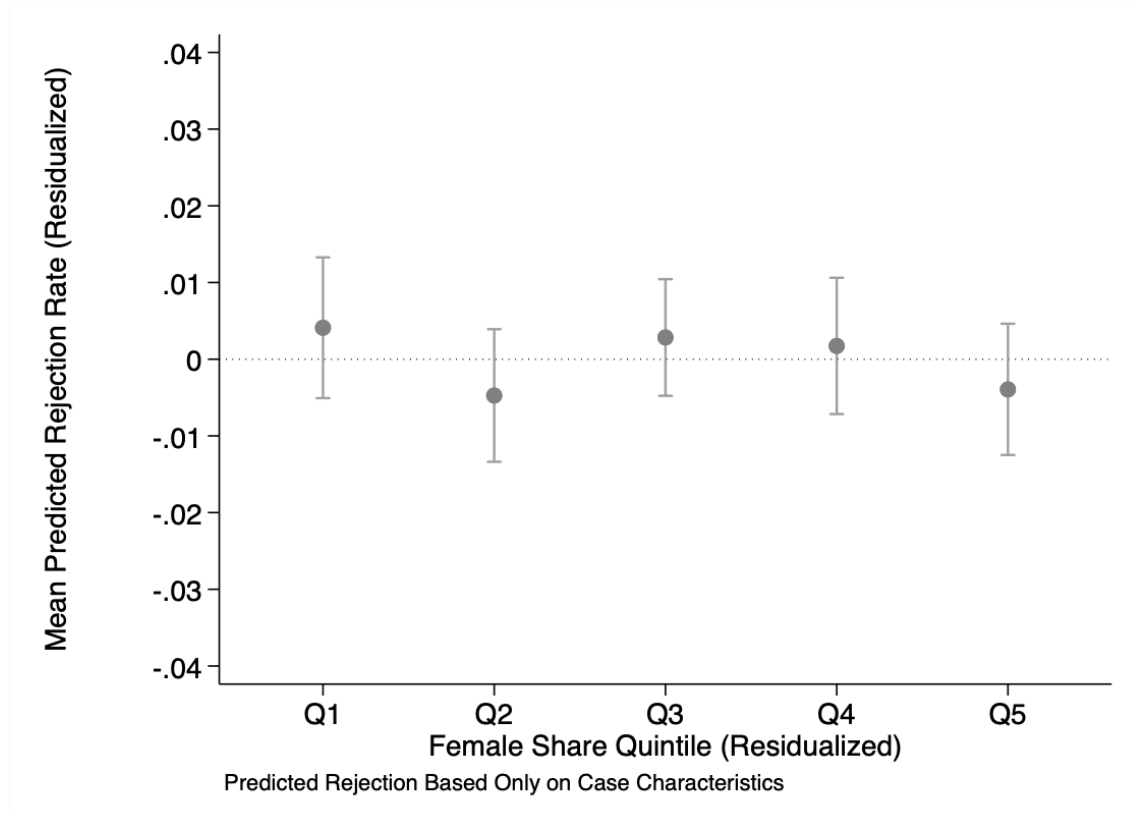
Notes: The figure shows the share of female judges in the judging committees in the Supreme Court. The average share of female judges is 0.28 (SD=0.25). Around 26% of judging committees have no female judges, and around 3% of judging committees have only female judges. Around 35% of cases have exactly 20% female judges (i.e., one female judge out of five).

Figure 3: Balancing Tests for Share of Female Judges



Notes: Each dot represents the OLS coefficient from a separate regression of the share of female judges on the indicated covariate, conditional on trial month, trial day, case filing year, and court type fixed effects. Robust standard errors. Thick bars indicate 90% confidence intervals; thin bars indicate 95% confidence intervals. Panel A includes variables capturing the composition of the panel. Panel B includes case-level characteristics. Panel C includes indicators for case type. Sample restricted to panels of five or fewer judges.

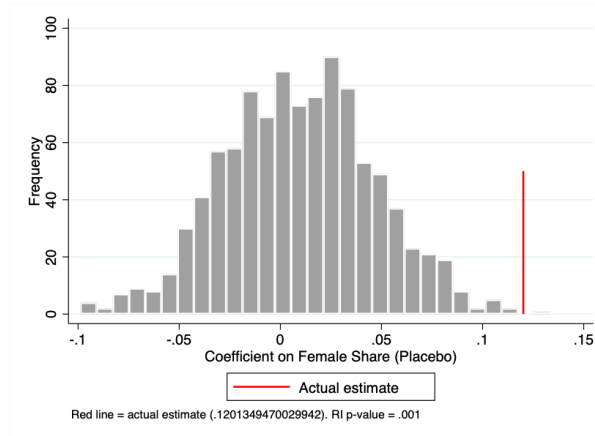
Figure 4: RANDOMIZATION EXERCISE: PREDICTED REJECTION RATES BY QUINTILE OF FEMALE PANEL SHARE USING ONLY CASE CHARACTERISTICS



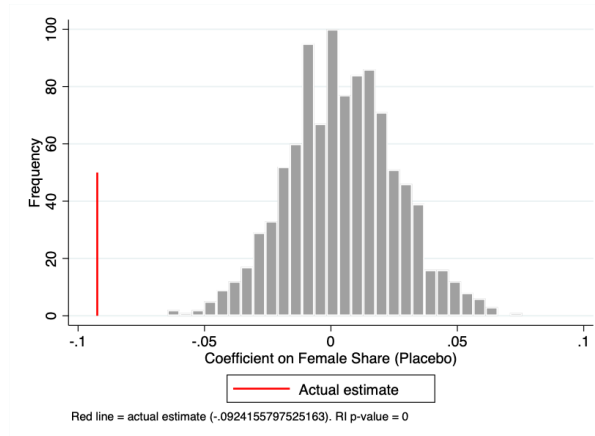
Notes: The figure plots the residualized means of predicted rejection rates across quintiles of the residualized female share of the panel. Predicted rejection is constructed using only observable case characteristics—which include case type indicators, number of litigants, number of defendants, number of lawyers, case length, whether an immigrant is involved, whether the case involves underage individuals, whether the case is gender-sensitive or violent, whether the district attorney appealed, and whether plaintiff information is missing—while excluding all judge-related variables. Both the predicted rejection rate and the female share of the panel are residualized on time and court-type fixed effects. Bars denote 95% confidence intervals.

Figure 5: PERMUTATION EXERCISE: PLACEBO DISTRIBUTION OF FEMALE SHARE COEFFICIENT

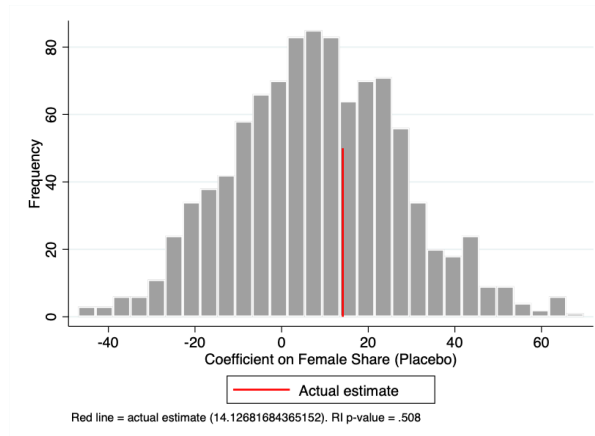
(a) Reject Appeal Request



(b) Delegates Trial to Another Court



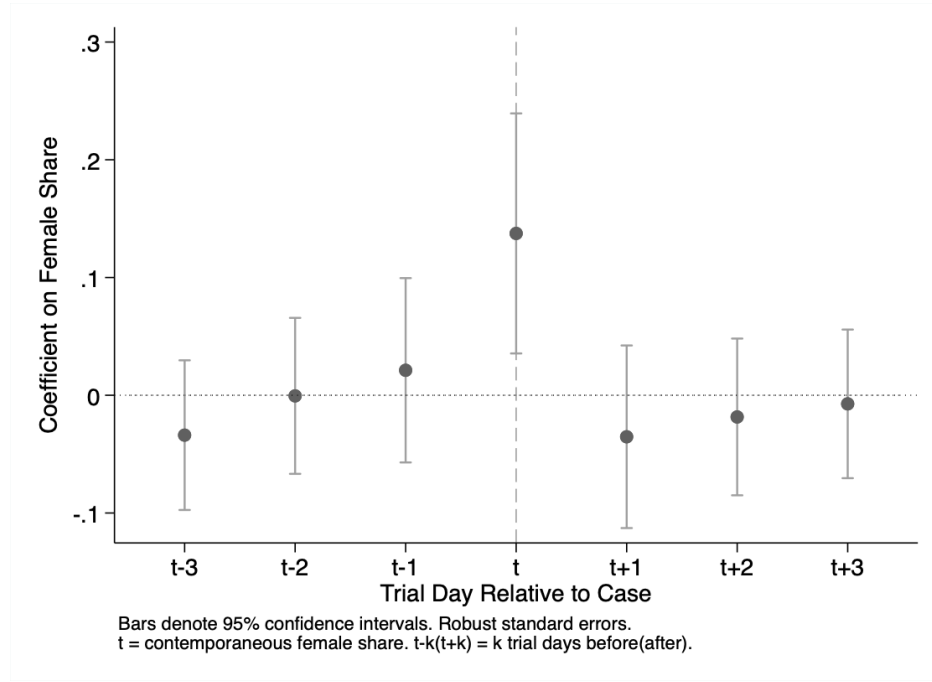
(c) Monetary Trial Expenses (Placebo Outcome)



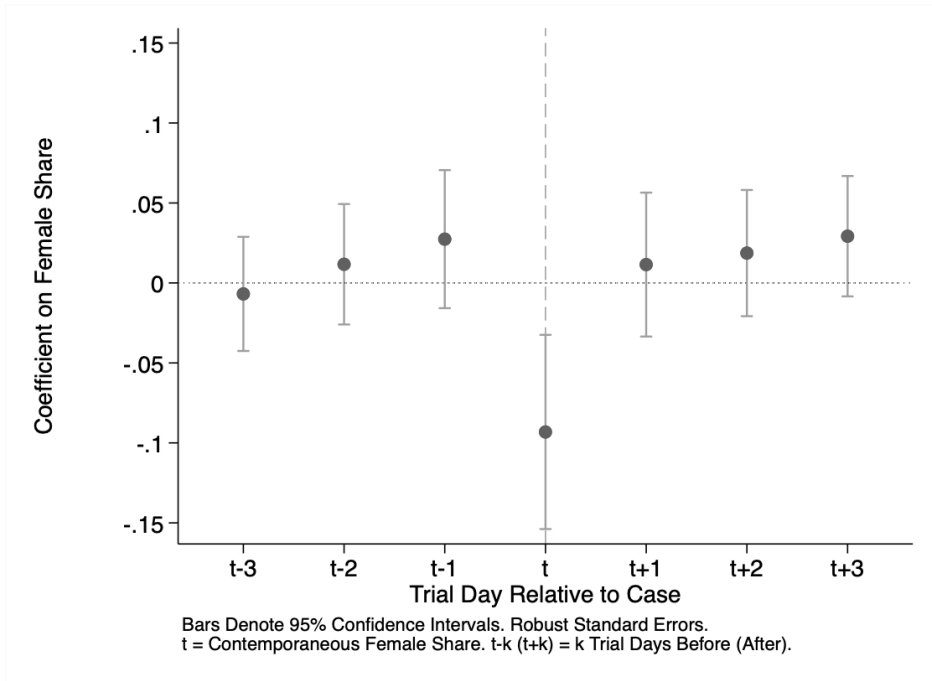
Notes: Each panel plots the distribution of placebo coefficients on the female share of the judging committee obtained from 1,000 random shuffles of the female share across committees within the same month–year cell. The red vertical line marks the actual estimate from the baseline specification. The randomization inference p-value reports the share of placebo draws that produce a coefficient at least as large in absolute value as the actual estimate. If the actual estimate lies in the extreme tail of the placebo distribution, this provides evidence that the baseline effect reflects genuine variation in committee gender composition rather than spurious correlation.

Figure 6: LEAD-LAG DESIGN: EFFECT OF FEMALE SHARE ON APPEAL OUTCOMES

(a) Outcome: Reject Appeal Request

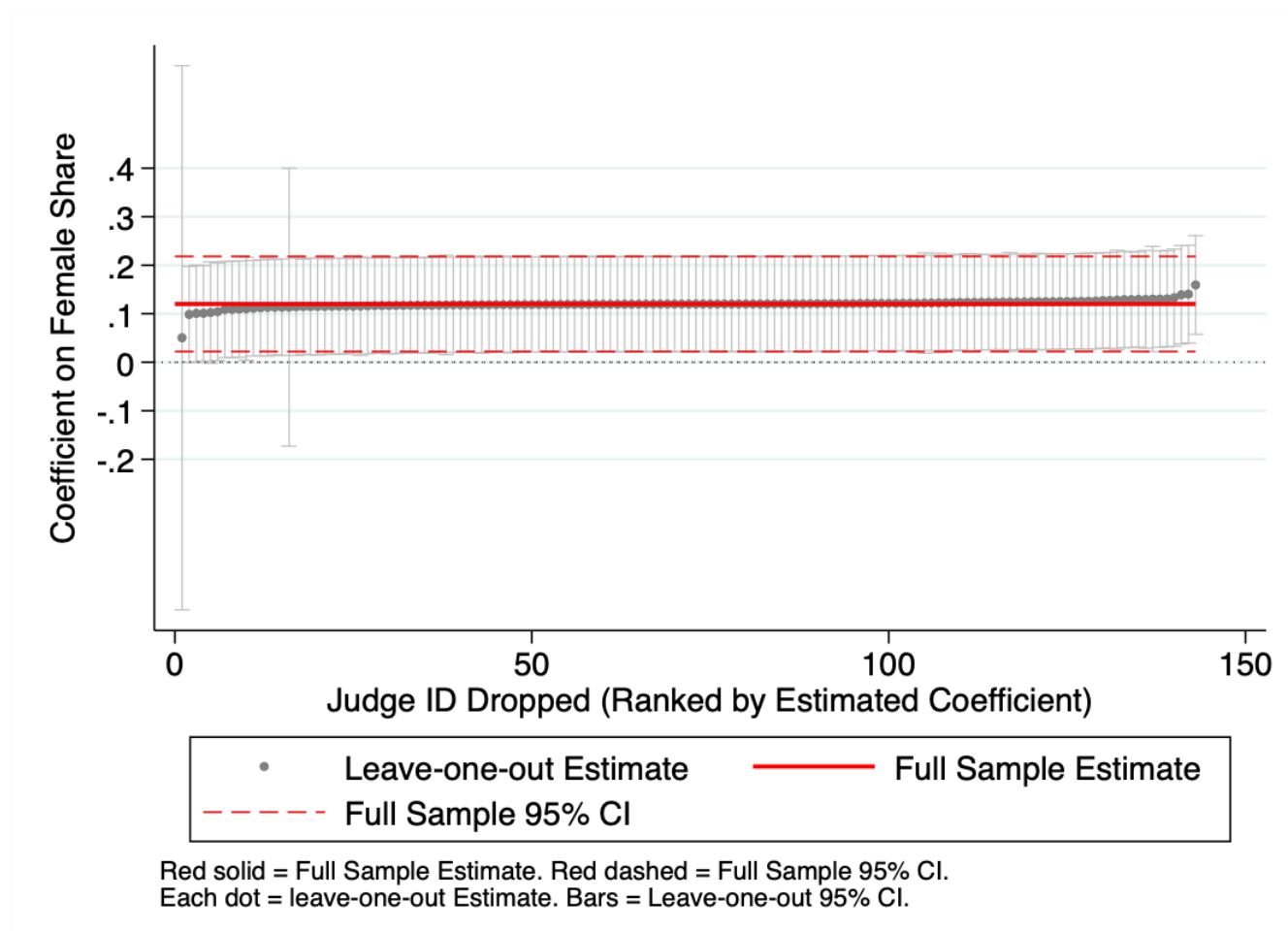


(b) Outcome: Delegate Trial to Another Court



Notes: Each panel plots estimated coefficients from equation (4), which augments the baseline specification (Table 3, column 2) with the average female share of committees judging on the three trial days before ($t-1$, $t-2$, $t-3$) and the three trial days after ($t+1$, $t+2$, $t+3$) the focal case. The top panel reports results for the probability that the appeal request is rejected; the bottom panel reports results for the probability that the case is delegated to another court or judge. The horizontal axis denotes the trial day relative to the focal case, with $t=0$ corresponding to the contemporaneous committee. Bars denote 95% confidence intervals based on heteroskedasticity-robust standard errors. The dotted horizontal line marks zero and the dashed vertical line separates the pre-period from the post-period.

Figure 7: LEAVE-ONE-JUDGE-OUT ROBUSTNESS TEST



Notes: The figure plots coefficient estimates from the baseline specification obtained by dropping all cases involving one judge at a time. For each of the 142 judges in the sample, all cases in which that judge participated in any role—as head, introducer, or committee member—are removed and the main regression is re-estimated on the remaining sample. Each grey dot represents the point estimate on the share of female judges when the corresponding judge is excluded, with judges ranked on the horizontal axis by the magnitude of the leave-one-out estimate. Grey bars denote the 95% confidence interval for each leave-one-out estimate. The red solid line marks the full sample estimate and the red dashed lines mark its 95% confidence interval. The dotted black line at zero is shown for reference. The stability of the estimates across all leave-one-out samples indicates that the baseline result is not driven by any single influential judge.

Table 1: DESCRIPTIVE STATISTICS OF MAIN STUDY SAMPLE

	Mean	Std.Dev	Min	Max
	(1)	(2)	(3)	(4)
Panel A: Trial, Litigant, and Committee Characteristics				
<u>Trial and Committee Characteristics</u>				
Prop. of Female Judges in Committee	0.284	0.250	0	1
No. of Judges in the Committee	4.816	0.578	3	5
Male District Attorney (1=yes)	0.908	0.289	0	1
Total Number of Lawyers	0.572	0.570	0	5
Judging Experience of Committee (No. of Trials)	349.543	124.080	5	704.5
Female President of Supreme Court (1=yes)	0.372	0.483	0	1
Total Number of Defendants	1.344	1.436	1	22
Total Number of Plaintiffs	1.263	1.031	1	22
Case Involves Underage Individual (1=yes)	0.084	0.277	0	1
Gender Sensitive Case (1=yes)	0.050	0.217	0	1
Violent Crime (1=yes)	0.101	0.301	0	1
Non-Individual Plaintiff (1=yes)	0.227	0.419	0	1
Length of the Appeal (in pages)	6.424	11.017	2	167
Immigrant Litigant Involved (1=yes)	2.217	14.321	0	100
Appealed by District Attorney (1=yes)	1.176	10.065	0	100
Panel B: Case-Type Characteristics				
Homicide	0.051	0.220	0	1
Injuries	0.037	0.190	0	1
Rape	0.012	0.111	0	1
Tax Evasion	0.106	0.308	0	1
Fraudulent Check	0.056	0.230	0	1
Financial Related	0.031	0.173	0	1
Drugs	0.033	0.180	0	1
Misconduct	0.028	0.166	0	1
Fraud	0.035	0.184	0	1
Forgery	0.064	0.245	0	1
Calumny	0.057	0.232	0	1
Theft	0.055	0.228	0	1
Environmental	0.030	0.171	0	1
Immigration	0.016	0.125	0	1
Relevance	0.029	0.169	0	1
International Transfer	0.021	0.142	0	1
Defendant Did Not Show Up	0.069	0.253	0	1
Miscellaneous	0.119	0.324	0	1
Missing Type	0.149	0.356	0	1

Notes: The table shows summary statistics for trial, litigant, committee, and case-type characteristics. The sample includes appeal cases that were brought up for re-evaluation to the Supreme Court in Greece in the period 2005-2017. Committee experience is defined as the average number of cases previously handled by its judges over the sample period. "Relevance" refers to cases in which the court of first instance lacked jurisdiction. "Miscellaneous" includes rare or uncategorized offenses, while "Missing" denotes cases without a reported type.

Table 2: BALANCING TESTS

Panel A: Case, Litigants, and Committees Characteristics	
No. of Judges in the Committee	0.007 (0.007)
Male District Attorney (1=yes)	-0.003 (0.012)
Total Number of Lawyers	0.000 (0.005)
Judging Experience of Committee (No. of Trials)	-0.000 (0.000)
Female President of Supreme Court (1=yes)	-0.042 (0.021)**
Total Number of Defendants	-0.002 (0.002)
Total Number of Plaintiffs	-0.001 (0.003)
Case Involves Underage Individual (1=yes)	0.005 (0.077)
Gender Sensitive Case (1=yes)	-0.007 (0.013)
Violent Crime (1=yes)	-0.009 (0.009)
Non-Individual Plaintiff (1=yes)	0.013 (0.007)*
Length of the Appeal (in pages)	-0.000 (0.000)
Appealed by District Attorney (1=yes)	-0.000 (0.000)
Immigrant Litigant Involved (1=yes)	-0.000 (0.000)
Panel B: Case Type Characteristics	
Homicide	-0.008 (0.013)
Injuries	-0.008 (0.014)
Rape	-0.004 (0.028)
Tax Evasion	0.008 (0.010)
Fraudulent Check	0.037 (0.014)***
Financial Related	0.011 (0.017)
Drugs	0.021 (0.014)
Misconduct	0.018 (0.019)
Fraud	0.002 (0.015)
Forgery	-0.020 (0.010)*
Calumniation	0.002 (0.014)
Theft	-0.009 (0.013)
Environmental	0.008 (0.016)
Immigration	-0.016 (0.020)
Relevance	-0.023 (0.022)
International Transfer	-0.015 (0.021)
Litigant Did Not Show Up	0.004 (0.008)
Miscellaneous	-0.014 (0.009)
Missing Type	0.004 (0.009)
Share Significant at 1 %	0.033
Share Significant at 5 %	0.066
Share Significant at 10 %	0.133
Basic Controls	✓

Notes: The table reports estimated coefficients from separate OLS regressions where the dependent variable is the share of female judges on the committee. Each variable is entered separately as the dependent variable of interest in its own regression, alongside a set of basic controls. *Basic Controls* include trial-month fixed effects, case-filing-year fixed effects, day-of-week fixed effects, and court-type fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 3: BASELINE ESTIMATES OF THE EFFECT OF SHARE OF FEMALE JUDGES ON APPEAL OUTCOMES

	Baseline Regressions				Placebo Regressions			
	Multiple Litigants		One Litigant		Multiple Litigants		One Litigant	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<u>Outcomes Decided by Judging Committee</u>								
Reject Appeal Request	0.101	0.120	0.139	0.145	0.000	0.011	0.006	0.015
	(0.046)**	(0.050)**	(0.056)**	(0.054)***	(0.042)	(0.040)	(0.044)	(0.042)
N	2,998	2,998	2,562	2,562	2,998	2,998	2,562	2,562
Delegates Trial to Another Court	-0.052	-0.092	-0.084	-0.097	0.008	0.005	0.004	0.000
	(0.027)**	(0.029)***	(0.032)***	(0.031)***	(0.024)	(0.023)	(0.025)	(0.025)
N	2,998	2,998	2,562	2,562	2,998	2,998	2,562	2,562
<u>Placebo Outcome Decided by Court</u>								
Monetary Trial Expenses	-1.730	14.127	20.206	27.436	0.906	8.257	6.463	8.987
	(23.238)	(23.626)	(23.240)	(22.556)	(23.213)	(20.506)	(19.694)	(18.795)
N	2,998	2,998	2,562	2,562	2,998	2,998	2,562	2,562
Basic Controls	✓	✓	✓	✓	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓	✓	✓	✓	✓
Additional Controls		✓		✓		✓		✓

Notes: Columns (1)–(4) present estimates for the effects of the share of female judges in a judging committee on appeal outcomes. Columns (5)–(8) report placebo estimates in which the actual share of female judges is replaced by a simulated measure. This measure corresponds to the mean coefficient obtained from 1,000 random shuffles of the female share across committees within the same month–year group. The unit of observation is the case. *Basic Controls* include trial-month fixed effects, case-filing-year fixed effects, day-of-week fixed effects, and court-type fixed effects. Additional controls include the number of defendants, the number of judges, and the number of lawyers; binary indicators for whether the president is female, the district attorney is male, the introducer is male; indicators for violent crimes, gender-sensitive cases, and juvenile defendants; the average judging experience of the committee (measured in number of trials); experience of the introducer; indicators for missing values in these covariates; and primary and secondary case-type fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 4: ESTIMATES OF THE EFFECT OF THE SHARE OF FEMALE JUDGES ON APPEAL OUTCOMES FOR INDIVIDUAL AND NON-INDIVIDUAL LITIGANTS

	Multiple		One			
	Individual Litigant/s				Non-Individual	
	(1)	(2)	(3)	(4)	(5)	(6)
Outcomes Decided by Judging Committee						
Reject Appeal Request	0.114 (0.054)**	0.120 (0.050)**	0.155 (0.059)***	0.145 (0.054)***	-0.144 (0.121)	-0.211 (0.120)*
<i>N</i>	2,998	2,998	2,562	2,562	754	754
Delegates Trial to Another Court	-0.107 (0.033)***	-0.092 (0.029)***	-0.103 (0.035)***	-0.097 (0.031)***	-0.030 (0.029)	-0.032 (0.029)
<i>N</i>	2,998	2,998	2,562	2,562	754	754
Outcome Decided by Court						
Monetary Trial Expenses	-2.396 (24.353)	12.944 (23.313)	12.330 (23.389)	23.869 (22.251)	-23.276 (59.533)	-52.037 (55.596)
<i>N</i>	2,562	2,562	2,562	2,562	754	754
Basic Controls	✓	✓	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓	✓	✓
Additional Controls		✓		✓		✓

Notes: The table reports baseline estimates (columns 1–4) and placebo estimates (columns 5–6) of the effect of the share of female judges on appeal outcomes. The non-individual plaintiff sample includes 754 cases where the plaintiff is a bank, an organization, or the government. These cases predominantly involve crime types where institutional entities are the natural plaintiff, such as fraudulent checks, financial fraud, and tax evasion. *Basic Controls* include trial-month fixed effects, case-filing-year fixed effects, day-of-week fixed effects, and court-type fixed effects. The unit of observation is the case. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 5: HETEROGENEOUS EFFECTS: JUVENILE VS. ADULT DEFENDANTS

	Juvenile Defendant				Adult Defendant			
	Multiple		Single		Multiple		Single	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<u>Outcomes Decided by Judging Committee</u>								
Reject Appeal Request	-0.269 (0.161)*	-0.251 (0.155)	-0.313 (0.179)*	-0.329 (0.176)*	0.172 (0.058)***	0.166 (0.053)***	0.215 (0.062)***	0.188 (0.057)***
<i>N</i>	510	510	400	400	2,487	2,487	2,161	2,161
Delegates Trial to Another Court	-0.076 (0.063)	-0.067 (0.053)	-0.050 (0.077)	-0.039 (0.066)	-0.108 (0.036)***	-0.100 (0.032)***	-0.104 (0.038)***	-0.106 (0.034)***
<i>N</i>	510	510	400	400	2,487	2,487	2,161	2,161
<u>Outcome Decided by Court</u>								
Monetary Trial Expenses	-54.882 (75.184)	19.559 (75.931)	-80.234 (74.522)	-27.752 (80.930)	9.931 (25.984)	14.452 (25.316)	29.498 (24.887)	32.138 (23.587)
<i>N</i>	510	510	400	400	2,487	2,487	2,161	2,161
Basic Controls	✓	✓	✓	✓	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓	✓	✓	✓	✓
Additional Controls		✓		✓		✓		✓

Notes: The table presents estimates for the effects of the share of female judges on appeal outcomes for juvenile defendants and adult defendants. Cases that were initially decided by local juvenile court judges involve juvenile defendants. The unit of observation is the case. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 6: HETEROGENEOUS EFFECTS BY PRESIDENT AND INTRODUCER GENDER

	President		Introducer Judge	
	Male (1)	Female (2)	Male (3)	Female (4)
<u>Outcomes Decided by Judging Committee</u>				
Reject Appeal Request	0.067 (0.093)	0.188 (0.070)***	0.076 (0.071)	0.132 (0.097)
<i>N</i>	955	1,607	1,792	770
Delegates Trial to Another Court	-0.090 (0.059)	-0.095 (0.038)**	-0.081 (0.039)**	-0.075 (0.060)
<i>N</i>	955	1,607	1,792	770
<u>Outcome Decided by Court</u>				
Monetary Trial Expenses	-15.801 (40.896)	49.073 (28.082)*	4.483 (29.559)	46.178 (42.694)
<i>N</i>	955	1,607	1,792	770
Basic Controls	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓

Notes: The table presents estimates for the effects of the share of female judges on appeal outcomes for different subsamples: male (column 1) and female (column 2) president of the Supreme Court; male (column 3) and female (column 4) introducer. The unit of observation is the case. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 7: NON-LINEAR EFFECTS OF THE SHARE OF FEMALE JUDGES ON APPEAL OUTCOMES

	Outcomes Decided by Judging Committee		Placebo Outcome Decided by Court
	Reject Appeal Request	Delegates Trial to Another Court	Monetary Trial Expenses
	(1)	(2)	(3)
Q2: Mean = 0.20 Range = 0.20–0.20			
Q2	0.042 (0.023)*	-0.022 (0.013)*	8.664 (9.786)
Q3: Mean = 0.39 Range = 0.33–0.40			
Q3	0.045 (0.030)	-0.017 (0.016)	15.042 (12.463)
Q4: Mean = 0.60 Range = 0.60–0.60			
Q4	0.100 (0.042)**	-0.058 (0.024)**	-0.395 (16.853)
Q5: Mean = 0.86 Range = 0.67–1.00			
Q5	0.133 (0.046)***	-0.080 (0.026)***	27.431 (20.974)
Obs.	2,562	2,562	2,562
Basic Controls	✓	✓	✓
Case Type FE	✓	✓	✓
Additional Controls	✓	✓	✓

Notes: The table reports nonlinear effects of the share of female judges in the Supreme Court committee on appeal outcomes. Estimates presented in each column are derived from the same regression. *Basic Controls* include trial-month fixed effects, case-filing-year fixed effects, day-of-week fixed effects, and court-type fixed effects. The model replaces the single treatment variable with a set of quintile indicators for the share of female judges. The omitted category is quintile 1 (Q1). The mean proportion female judges in quintile 1 is 0. The unit of observation is the case. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table 8: MECHANISM: EFFECT OF FEMALE JUDGE REPRESENTATION ON DECISION DELIBERATION TIME

	Multiple Litigants		One Litigant	
	(1)	(2)	(3)	(4)
Days Between Trial and Final Decision	75.448 (8.681)***	78.304 (8.181)***	68.035 (8.959)***	73.697 (8.564)***
N	2,998	2,998	2,562	2,562
Basic Controls	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓
Additional Controls		✓		✓
	Simple Cases		Complicated Cases	
	<i>Multiple Litigants</i>	<i>One Litigant</i>	<i>Multiple Litigants</i>	<i>One Litigant</i>
Days Between Trial and Final Decision	59.872 (10.807)***	58.499 (11.527)***	81.937 (14.570)***	87.975 (15.301)***
N	1,180	1,054	1,204	978
Basic Controls	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓

Notes: The table presents estimates for the effects of the share of female judges in a judging committee on the days between the trial date and the final decision. The unit of observation is the case. Basic controls include year, day-of-week, and court-type fixed effects. Additional Controls include number of defendants, number of judges, number of lawyers involved, a binary indicator for whether the president is female, a binary indicator for whether the district attorney is male, a binary indicator for whether the introducer is male, the committee overall experience, whether it is a violent crime, experience of the introducer, whether it is a gender-sensitive case, whether it is a juvenile defendant, indicators for missing values in these covariates, and primary and secondary case-type fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Online Appendix:

Additional Tables and Figures

10 Figure Appendices

Figure A1: EXAMPLES OF SUPREME COURT CASES

Απόφαση 190 / 2008 (Ε, ΠΟΙΝΙΚΕΣ) Θέμα <i>Αναρτήσεις απαράδεκτο, Εκκρεμοδικία.</i> Περίληψη: Απαράδεκτος λόγος αναρτήσεις που πλήττει απόφαση πρωτοβαθμίου δικαστηρίου. Πότε υφίσταται εκκρεμοδικία. ΑΡΙΘΜΟΣ 190/2008 ΤΟ ΔΙΚΑΣΤΗΡΙΟ ΤΟΥ ΑΡΕΙΟΥ ΠΑΓΟΥ Ε' ΠΟΙΝΙΚΟ ΤΜΗΜΑ Συγκροτήθηκε από τους Δικαστές: Ηρακλή Κωνσταντινίδη, Αντιπρόεδρο Αρείου Πάγου, Κωνσταντίνο Κούκλη, Ελευθέριο Νικολάου, Αναστάσιο Λιανό - Εισηγητή και Βιολέττα Κυτία, Αρεοπαγίτες. Συνήλθε σε δημόσια συνεδρίαση στο Κατάστημά του στις 19 Οκτωβρίου 2007, με την παρουσία του Αντεισαγγελέα του Αρείου Πάγου Βασιλείου Μαρτί (γιατί κωλύεται ο Εισαγγελέας) και της Γραμματέως Ευδοκίας Φραγκίδη, για να δικάσει την αίτηση του αναρσεισίντος - κατηγορουμένου και ήδη κρατούμενου στο Γενικό Κατάστημα Κράτησης Α' Τύπου Μαλανδρινού, που εκπροσωπήθηκε από τον πληρεξούσιο δικηγόρο του Φραγκίσκο Ραγκούση, περί αναρσεισιν της 75/2006 αποφάσεως του Πενταμελούς Εφετείου Καλαμάτας. Το Πενταμελές Εφετείο Καλαμάτας, με την ως άνω απόφασή του διέταξε όσα λεπτομερώς αναφέρονται σ' αυτή, και ο αναρσεισιν - κατηγορούμενος ζητεί την ανάρτηση αυτής, για τους λόγους που αναφέρονται στην από 14 Δεκεμβρίου 2006 αίτηση του αναρσεισιν, η οποία καταχωρίστηκε στο οικείο πινάκιο με τον αριθμό 28/2007. Αφού άκουσε Τον πληρεξούσιο δικηγόρο του αναρσεισίντος, που ζήτησε όσα αναφέρονται στα σχετικά πρακτικά και τον Αντεισαγγελέα, που πρότεινε να απορριφθεί η προκείμενη αίτηση αναρσεισιν. ΣΚΕΦΘΗΚΕ ΣΥΜΦΩΝΑ ΜΕ ΤΟ ΝΟΜΟ 1. Κατά τη διάταξη του άρθρου 504 παρ. 1 του ΚΠΔ όπου ο νόμος δεν ορίζει ειδικά κάτι άλλο, το ένδικο μέσο της απήσεως	Απόφαση 1321 / 2017 (ΣΤ, ΠΟΙΝΙΚΕΣ) Αριθμός 1321/2017 ΤΟ ΔΙΚΑΣΤΗΡΙΟ ΤΟΥ ΑΡΕΙΟΥ ΠΑΓΟΥ ΣΤ' ΠΟΙΝΙΚΟ ΤΜΗΜΑ Συγκροτήθηκε από τους Δικαστές: Ευφημία Λαμπροπούλου, Αντιπρόεδρο του Αρείου Πάγου, (σύμφωνα με την υπ' αριθμ. 42/2017 πράξη της Προέδρου του Αρείου Πάγου), Αριστείδη Πελεκάνο - Εισηγητή, Αρετμισία Παναγιώτου, Χρήστο Βρυώνη και Ιωάννη Μαγγίνα, Αρεοπαγίτες. Συνήλθε σε δημόσια συνεδρίαση στο Κατάστημά του στις 14 Φεβρουαρίου 2017, με την παρουσία του Αντεισαγγελέα του Αρείου Πάγου Αθανάσιου Κασιρωδή (γιατί κωλύεται η Εισαγγελέας) και του Γραμματέα Χαράλαμπος Αθανασίου, για να δικάσει την αίτηση του αναρσεισίντος-κατηγορουμένου D. M. του D., κατοίκου ..., ο οποίος εκπροσωπήθηκε από τον πληρεξούσιο δικηγόρο του Σωτήριο Σζούκο, για αναρσεισιν της υπ' αριθ. 118/2015 απόφασης του Τριμελούς Εφετείου Κακουρηγμάτων Ιωαννίνων. Το Τριμελές Εφετείο Κακουρηγμάτων Ιωαννίνων, με την ως άνω απόφασή του διέταξε όσα λεπτομερώς αναφέρονται σ' αυτή, και ο αναρσεισιν-κατηγορούμενος ζητεί την ανάρτηση αυτής, για τους λόγους που αναφέρονται στην από 23 Νοεμβρίου 2015 αίτηση του αναρσεισιν, η οποία καταχωρίστηκε στο οικείο πινάκιο με τον αριθμό .../2015. Αφού άκουσε Τον πληρεξούσιο δικηγόρο του αναρσεισίντος, που ζήτησε όσα αναφέρονται στα σχετικά πρακτικά και τον Αντεισαγγελέα, που πρότεινε να απορριφθεί η προκείμενη αίτηση αναρσεισιν. ΣΚΕΦΘΗΚΕ ΣΥΜΦΩΝΑ ΜΕ ΤΟ ΝΟΜΟ Κατά το άρθρο 20 παρ. 1 του ν. 3459/2006 (Κ.Ν.Ν.), "Με κάθετη τουλάχιστον δέκα (10) ετών και με χρηματική ποινή δύο χιλιάδων εννακασίων (2.900) μέχρι διακοσίων ενενήντα χιλιάδων (290.000) ευρώ τιμωρείται όποιος α)... β) πωλεί, αγοράζει... ναρκωτικά ... γ) κατέχει ή μεταφέρει ναρκωτικά με οποιονδήποτε τρόπο ή μέσο είτε στα έδαφος της επικράτειας είτε ...". Εξάλλου, στο αντίστοιχο άρθρο του νεότερου ν. 4139/2013 ορίζονται τα εξής: Άρθρο 20 (διακίνηση ναρκωτικών) παρ. 1. "Όποιος, εκτός από τις περιπτώσεις που προβλέπονται στα άρθρα 21, 22 και 23, διακινεί παράνομα ναρκωτικά, τιμωρείται με κάθετη τουλάχιστον οκτώ (8) ετών και με χρηματική ποινή μέχρι τριακόσιες χιλιάδες (300.000) ευρώ" και παρ. 2 "Με την επιφύλαξη των διατάξεων του άρθρου 29, ως έγκλημα διακίνησης ναρκωτικών νοείται κάθε πράξη με την οποία συντελείται η κυκλοφορία ναρκωτικών ουσιών και ιδίως ... η πώληση, η αγορά ... η κατοχή, η μεταφορά ...". Από τις προαναφερόμενες διατάξεις προκύπτει ότι εκάστη του άρθρου 20 παρ. 1 του ν. 4139/2013 (αφού προβλέπει μικρότερο κατώτατο όριο κάθετησης) είναι ευμενέστερη από την αντίστοιχη του άρθρου 20 παρ. 1 του ν. 3459/2006 και, σύμφωνα με τη θεμελιώδη αρχή της αναδρομικής ισχύος του ηπιότερου ποινικού νόμου (άρθρο 2 παρ. 1 Π Κ), καταλαμβάνει και εγκληματικές πράξεις που τελέστηκαν υπό το νομικό καθεστώς του ν. 3459/2006 και δεν εκδόθηκε γι' αυτές αμετάκλητη απόφαση μέχρι της θέσπισης του ν. 4139/2013, όπως συμβαίνει και στην προκείμενη υπόθεση. Εξάλλου, η καταδικαστική απόφαση έχει την απαιτούμενη από τα άρθρα 93 παρ. 3 του Συντάγματος και 139 του ΚΠΔ, ειδική και
Απόφαση 1164 / 2010 (ΣΤ, ΠΟΙΝΙΚΕΣ) Θέμα <i>Απολογία επίταξη, Ακυρότητα απόλυτη, Φοροδιαφυγή.</i> Περίληψη: Αποδοχή ειδικών φορολογικών στοιχείων. Ο χρόνος παραγραφής αρχίζει από τη θεώρηση του πορίσματος οικονομικού ελέγχου. Επαρκής προσδιορισμός της συνθέσεως του εφετίου, ως προς τους κωλυόμενους δικαστές. Για την υποβολή ερωτημάτων σε συγκατηγορούμενα απαιτείται αίτημα. Επαρκής προσδιορισμός αναγνωσθέντων εγγράφων. Επαρκής απολογία. Απορρίπτεται αίτηση. ΑΡΙΘΜΟΣ 1164/2010 ΤΟ ΔΙΚΑΣΤΗΡΙΟ ΤΟΥ ΑΡΕΙΟΥ ΠΑΓΟΥ Στ' ΠΟΙΝΙΚΟ ΤΜΗΜΑ Συγκροτήθηκε από τους Δικαστές Δημήτριο Πατινίδη, Αντιπρόεδρο του Αρείου Πάγου, Νικόλαο Κωνσταντόπουλο, Παναγιώτη Ρουμπή, Χριστόφορο Κοσμίδη - Εισηγητή και Κυριακούλα Γεροστάθη, Αρεοπαγίτες. Συνήλθε σε δημόσια συνεδρίαση στο Κατάστημά του, την 11η Μαΐου 2010, με την παρουσία του Αντεισαγγελέως του Αρείου Πάγου Γεωργίου Παντελή (γιατί κωλύεται ο Εισαγγελέας) και της Γραμματέως Πελαγίας Λόδου, για να δικάσει την αίτηση αναρσεισιν της 1273/2009 αποφάσεως του Τριμελούς Πλημμελειοδικείου Τρικάλων, του αναρσεισίντος - κατηγορουμένου Χ, ο οποίος παραιτήθηκε δια του πληρεξούσιου δικηγόρου Γεωργίου Σαμαρά (ΑΜ ΔΣΑ 20304). Το Τριμελές Πλημμελειοδικείο Τρικάλων, που δίκασε κατ' έφεση, με την ως άνω απόφαση, διέταξε όσα λεπτομερώς αναφέρονται σ' αυτή. Ο αναρσεισιν - κατηγορούμενος ζητεί την ανάρτηση της εν λόγω αποφάσεως, για τους λόγους που αναφέρονται στην από 12-2-2010 αίτηση αναρσεισιν και το από 23-4-2010 πρόσθετο δικόγραφο, που καταχωρήθηκαν στο οικείο πινάκιο με τον αριθμό 260/2010. Αφού άκουσε τον πληρεξούσιο δικηγόρο του αναρσεισίντος, που ζήτησε όσα αναφέρονται στα σχετικά πρακτικά και τον Αντεισαγγελέα, που πρότεινε να απορριφθεί η προκείμενη αίτηση, καθώς και οι πρόσθετοι λόγοι αναρσεισιν.	Απόφαση 1503 / 2008 (ΣΤ, ΠΟΙΝΙΚΕΣ) Θέμα <i>Απολογία ανεπίταξη, Διατροφής υποχρέωση.</i> Περίληψη: Διατροφή. Έλλειψη ειδικής και εμπειραστατωμένης απολογίας υπάρχει αν δεν αναφέρονται στην απόφαση περιστατικά, από τα οποία προκύπτει η κακοβουλία του κατηγορουμένου, που παραλείπει να καταβάλει τη διατροφή στους δικαιούχους. Ανεπαρκής η απόφαση. ΑΡΙΘΜΟΣ 1503/2008 ΤΟ ΔΙΚΑΣΤΗΡΙΟ ΤΟΥ ΑΡΕΙΟΥ ΠΑΓΟΥ ΣΤ' ΠΟΙΝΙΚΟ ΤΜΗΜΑ Συγκροτήθηκε από τους Δικαστές: Κωνσταντίνο Κούκλη, Προεδρεύοντα Αρεοπαγίτη (κωλυόμενου του Αντιπροέδρου Γεωργίου Σαραντινού), ο οποίος ορίστηκε με την 571/4.08 Πράξη του Προέδρου του Αρείου Πάγου, Βασίλειο Λικουδή, Ανδρέα Τσάλα - Εισηγητή, Ιωάννη Παπουτσή και Νικόλαο Ζαΐρη, Αρεοπαγίτες. Συνήλθε σε δημόσια συνεδρίαση στο Κατάστημά του στις 6 Μαΐου 2008, με την παρουσία του Αντεισαγγελέα του Αρείου Πάγου Ιωάννη Χρυσού (γιατί κωλύεται ο Εισαγγελέας) και της Γραμματέως Πελαγίας Λόδου, για να δικάσει την αίτηση του αναρσεισίντος - κατηγορουμένου χ'1, που παραιτήθηκε με την πληρεξούσια δικηγόρο του Αλίκη Αντωνία, περί αναρσεισιν της 1620/2007 αποφάσεως του Τριμελούς Πλημμελειοδικείου Ηρακλείου. Με πολιτικώς ενάγουσα την ψ1, που δεν παραιτήθηκε. Το Τριμελές Πλημμελειοδικείο Ηρακλείου, με την ως άνω απόφασή του διέταξε όσα λεπτομερώς αναφέρονται σ' αυτή, και ο αναρσεισιν - κατηγορούμενος ζητεί την ανάρτηση αυτής, για τους λόγους που αναφέρονται στην από 21 Ιανουαρίου 2008 αίτηση του αναρσεισιν, η οποία καταχωρίστηκε στο οικείο πινάκιο με τον αριθμό 235/2008. Αφού άκουσε Τον πληρεξούσιο δικηγόρο του αναρσεισίντος, που ζήτησε όσα αναφέρονται στα σχετικά πρακτικά και τον Αντεισαγγελέα, που πρότεινε να απορριφθεί η προκείμενη αίτηση αναρσεισιν.

Notes: This figure presents the first rows from four cases judged in the Supreme Court in Greece. I have highlighted the relevant places where the names of the judges are reported. Judges’ assigned roles are also reported (e.g., president, deputy president, introducer). The first names of the judges were used to determine their gender.

Figure A2: EXAMPLE OF A SUPREME COURT CASE, ZOOMED IN

Απόφαση 1763 / 2017 (Ζ, ΠΟΙΝΙΚΕΣ)

ΑΡΙΘΜΟΣ 1763/2017

ΤΟ ΔΙΚΑΣΤΗΡΙΟ ΤΟΥ ΑΡΕΙΟΥ ΠΑΓΟΥ

Ζ' ΠΟΙΝΙΚΟ ΤΜΗΜΑ - ΣΕ ΣΥΜΒΟΥΛΙΟ

Συγκροτήθηκε από τους Δικαστές: Αγγελική Αλιφεροπούλου, Αντιπρόεδρο Αρείου Πάγου, Δημήτριο Γεώργα, Δημήτριο Τζιούβα, Μαρία Γκανιάτσου και Μαρία Παπασωτηρίου - Εισηγήτρια, Αρεοπαγίτες.

Με την παρουσία και της Αντεισαγγελέως του Αρείου Πάγου Ελένης Μετσοβίτου - Φλουρή (γιατί κωλύεται η Εισαγγελέας) και της Γραμματέως Αικατερίνης Σιταρά.

Συνήλθε σε Συμβούλιο στο Κατάστημά του στις 4 Οκτωβρίου 2017, προκειμένου να αποφανθεί για την αίτηση του αναιρεσιπόντος- κατηγορουμένου Ι. Π. του Γ. κατοίκου ..., που δεν παρέστη στο συμβούλιο, για αναίρεση της υπ' αριθμ. 1552/2016 αποφάσεως του Τριμελούς Πλημμελειοδικείου Χαλκιδικής. Με πολιτικώς ενάγοντα τον Χ. Μ. του Γ., κάτοικο

Το Τριμελές Πλημμελειοδικείο Χαλκιδικής, με την ως άνω απόφασή του διέταξε όσα λεπτομερώς αναφέρονται σ' αυτή, και ο αναιρεσιώνων-κατηγορούμενος ζητεί τώρα την αναίρεση της απόφασης αυτής, για τους λόγους που αναφέρονται στην από 17 Φεβρουαρίου 2017 αίτησή του, η οποία καταχωρίστηκε στο οικείο πινάκιο με τον αριθμό ...2017.

Έπειτα η Αντεισαγγελέας του Αρείου Πάγου Ελένη Μετσοβίτου - Φλουρή, εισήγαγε για κρίση στο Συμβούλιο τη σχετική δικογραφία με την πρόταση του Αντεισαγγελέα του Αρείου Πάγου Αθανάσιου Κατσιρώδη με αριθμό ...-4-2017, στην οποία αναφέρονται τα ακόλουθα: "Ι. Εισάγω στο Δικαστήριο Σας σε Συμβούλιο, σύμφωνα με τα άρθρα 473 παρ. 2 και 476 παρ. 1 του Κ.Π.Δ. όπως το τελευταίο αντ. από το άρθρο 2 παρ. 18 του Ν. 3160/2003, την ...2-2017 αίτηση αναίρεσεως του Ι. Γ. Π. κατοίκου ... κατά της 1552/2016 αποφάσεως του Τριμελούς Πλημμελειοδικείου Χαλκιδικής, με την οποία καταδικάστηκε σε φυλάκιση έξι μηνών για πλαστογραφία με τριετή αναστολή, και εκθέτω τα ακόλουθα: ΙΙ. Σύμφωνα με το άρθρο 473 παρ. 2 και 3 του Κ.Π.Δ., η αναίρεση κατά της τελεσίδικης καταδικαστικής απόφασης μπορεί να ασκηθεί από εκείνον που κρίθηκε ένοχος και με δήλωση του στον Γραμματέα του δικαστηρίου που εξέδωσε την προσβαλλόμενη απόφαση εντός προθεσμίας δέκα ημερών από την ημέρα που η απόφαση αυτή καταχωρήθηκε καθαρογραμμένη στο ειδικό βιβλίο που προβλέπεται από την παρ. 3 του άρθρου αυτού. Η ημέρα καταχώρησης δεν υπολογίζεται στην προθεσμία αυτή δηλαδή η προθεσμία αυτή αρχίζει από την επόμενη της καταχώρησης (ΑΠ 1095/1985, ΑΠ 426/1984). Σε περίπτωση που η αίτηση αναίρεσεως ασκηθεί εκπρόθεσμα αυτή, χωρίς να αναφέρονται σε αυτή οι λόγοι ανωτέρας βίας που δικαιολογούν την εκπρόθεσμη άσκηση της αναίρεσεως, πρέπει αυτή, σύμφωνα με το άρθρο 476 παρ. 1 του Κ.Π.Δ. ίδιου Κώδικα, να απορριφθεί ως απαράδεκτη από το Δικαστήριο του Αρείου Πάγου σε Συμβούλιο, ύστερα από πρόταση του Εισαγγελέα και αφού ακούσει τους διαδίκους που εμφανιστούν, ενώ ταυτόχρονα αυτό: α) διατάσσει την εκτέλεση της απόφασης που έχει προσβληθεί και β) καταδικάζει τον αναιρεσιέοντα στα σχετικά δικαστικά έξοδα σύμφωνα με το άρθρο 583 παρ. 1 του Κ.Π.Δ. (ΑΠ 738/2016, ΑΠ 614/2015, ΑΠ 29/2013, ΑΠ 76/2011, ΑΠ 1639/2010). ΙΙΙ. Η παραπάνω αίτηση αναίρεσεως ασκήθηκε στις 17-2-2017 με δήλωση του αναιρεσιέοντα στον Γραμματέα Πλημμελειοδικών Χαλκιδικής και η καταδικαστική απόφαση κατά της οποίας στρέφεται καταχωρήθηκε καθαρογραμμένη στο ειδικό βιβλίο στις 2-2-2017, ενώ έπρεπε η αναίρεση αυτή, σύμφωνα με όσα παραπάνω εκθέσαμε, να ασκηθεί μέχρι και τις 13-2-2017 ημέρα Δευτέρα. Επειδή η αίτηση αυτή αναίρεσεως ασκήθηκε εκπρόθεσμα, χωρίς επίκληση λόγων ανωτέρας βίας για την εκπρόθεσμη αυτή άσκηση, πρέπει το Δικαστήριο Σας σε Συμβούλιο, σύμφωνα με τα άρθρα 476 παρ.1 και 583 παρ. 1 του Κ.Π.Δ. όπως αντ. από το άρθρο 55 παρ. 1 του Ν. 3160/2003 και το άρθρο 3 του Ν. 663/77 όπως αντ. από το άρθρο 18 του Ν. 969/79 και τις 134423/1992, 58553/2006 και 123827/23-12-2010 Α.Υ. Οικονομικών και Δικαιοσύνης: α) να απορρίψει την αίτηση αυτή αναίρεσεως ως απαράδεκτη β) να διατάξει την εκτέλεση της προσβαλλόμενης απόφασης και γ) να επιβάλλει σ' αυτόν τα σχετικά δικαστικά έξοδα.

ΓΙΑ ΤΟΥΣ ΛΟΓΟΥΣ ΑΥΤΟΥΣ Προτείνω: Α) Να απορριφθεί ως απαράδεκτη η ...2-2017 αίτηση αναίρεσεως του Ι. Γ. Π. κατοίκου ... κατά της 1552/2016 αποφάσεως του Τριμελούς Πλημμελειοδικείου Χαλκιδικής, με την οποία καταδικάστηκε σε φυλάκιση έξι μηνών για πλαστογραφία Β) Να διαταχθεί η εκτέλεση της προσβαλλόμενης αποφάσεως και Γ) Να καταδικασθεί ο αναιρεσιώνων στα δικαστικά έξοδα από 250 Ευρώ. Ο Αντεισαγγελέας του Αρείου Πάγου Αθανάσιος Κ. Κατσιρώδης".

Αφού άκουσε

Notes: This figure shows the first rows from one case judged in the Supreme Court in Greece in 2017. The top box indicates the names of the judges who judged the case. Some key aspects of the case are highlighted: the gender of the defendant, the gender of the plaintiff, the first-level court name, imprisonment duration, and case type. The bottom box highlights the court decision by the judges. In the remaining pages of the case which are not shown here more details about the timeline and the prior decisions are reported.

Figure A3: Word Clouds with Judges' First Names by Gender



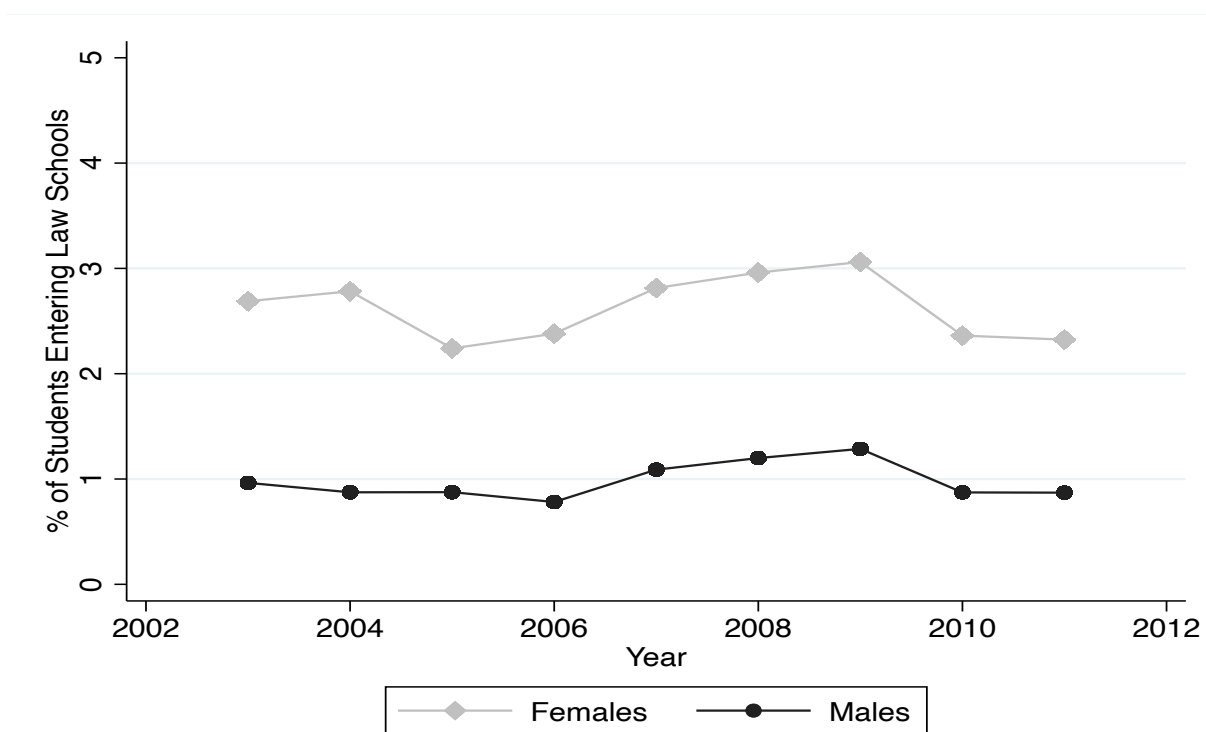
(a) Male Judges



(b) Female Judges

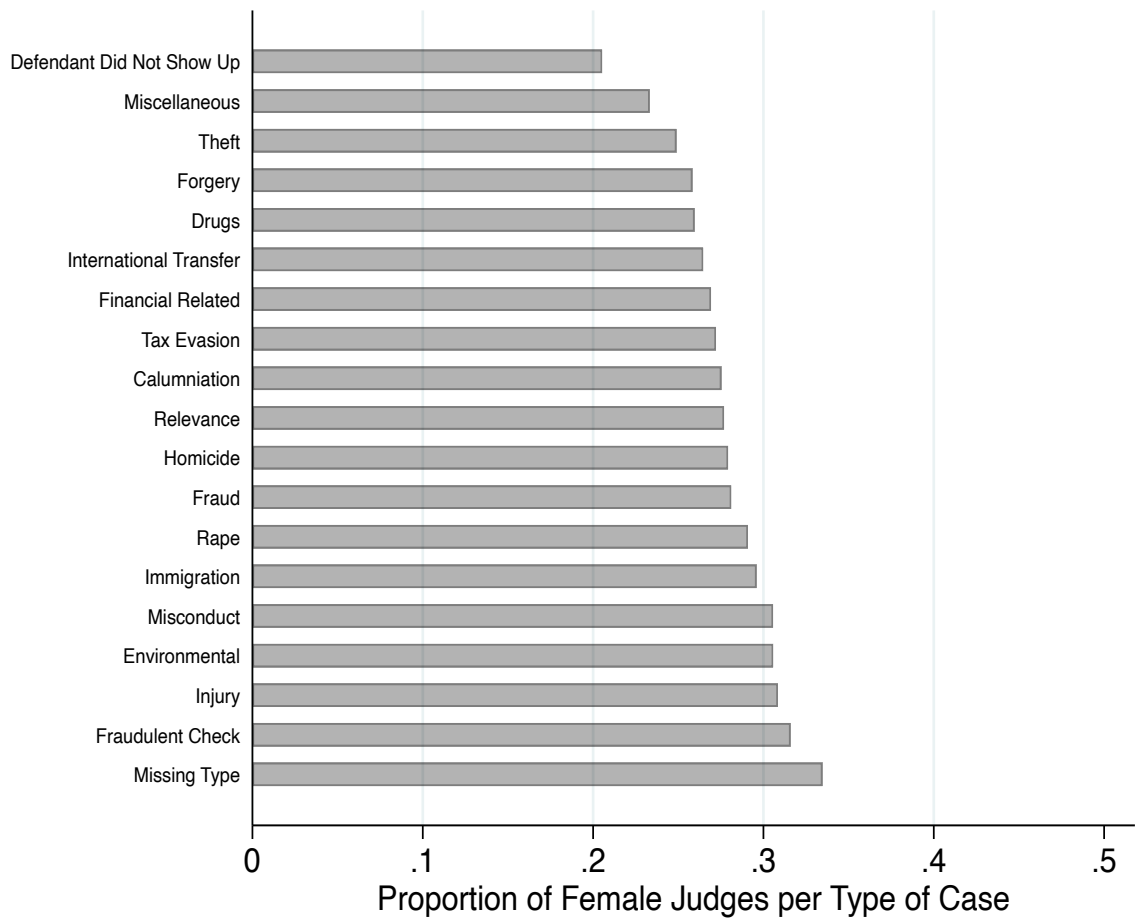
Notes: The figure displays word clouds of first names for judges appearing in the sample, separately for male and female judges. The size of each name is proportional to its frequency in the data. Gender is identified based on Greek naming conventions, which provide unambiguous gender markers. The figure is intended to illustrate the distribution and commonality of names by gender in the dataset.

Figure A4: PROPORTION OF FEMALE AND MALE STUDENTS ENTERING LAW SCHOOLS PER YEAR



Notes: This figure shows the proportion of male and female students entering law schools every year.

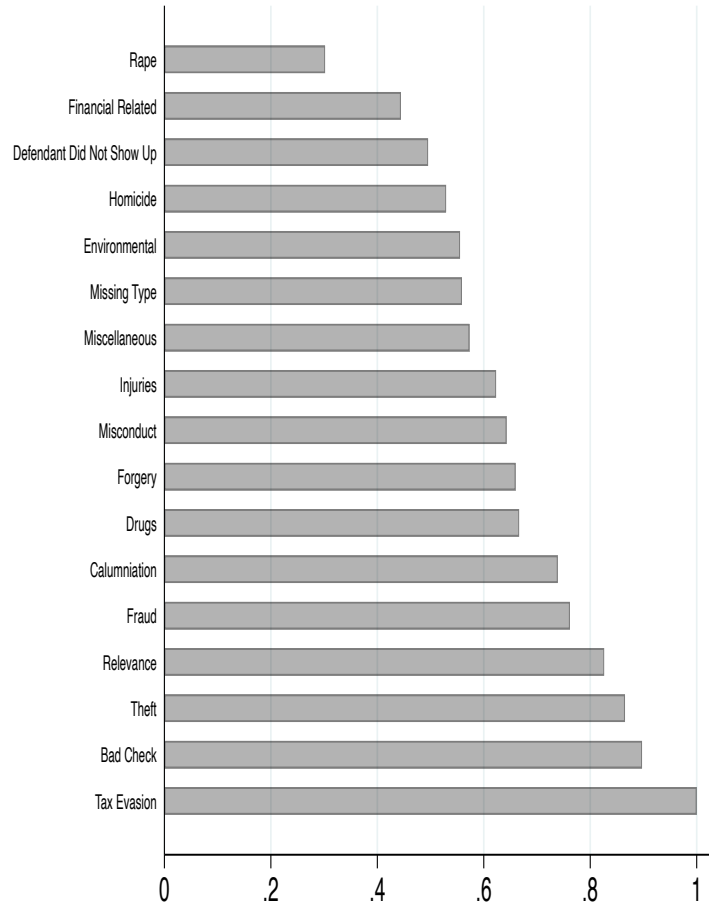
Figure A5: SHARE OF FEMALE JUDGES BY CASE TYPE



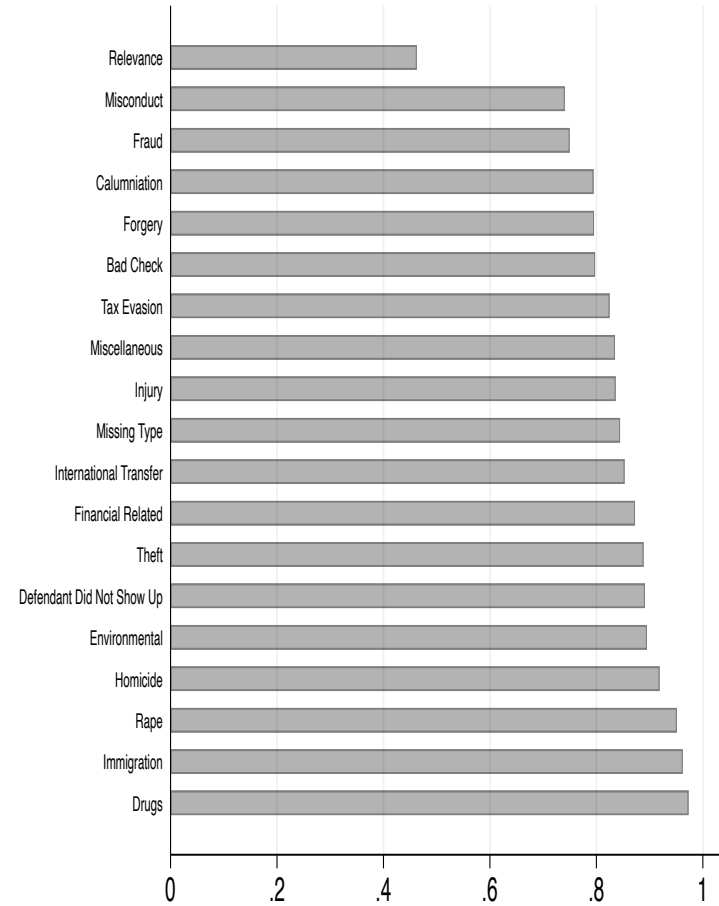
Notes: The figure shows the share of female judges assigned to judging committees at the Supreme Court for different types of cases. Case types on the vertical axis are ranked from the lowest to the highest share of female judges.

Figure A6: SHARE OF MALE PLAINTIFFS AND DEFENDANTS BY CASE TYPE

(a) Proportion of Male Plaintiffs

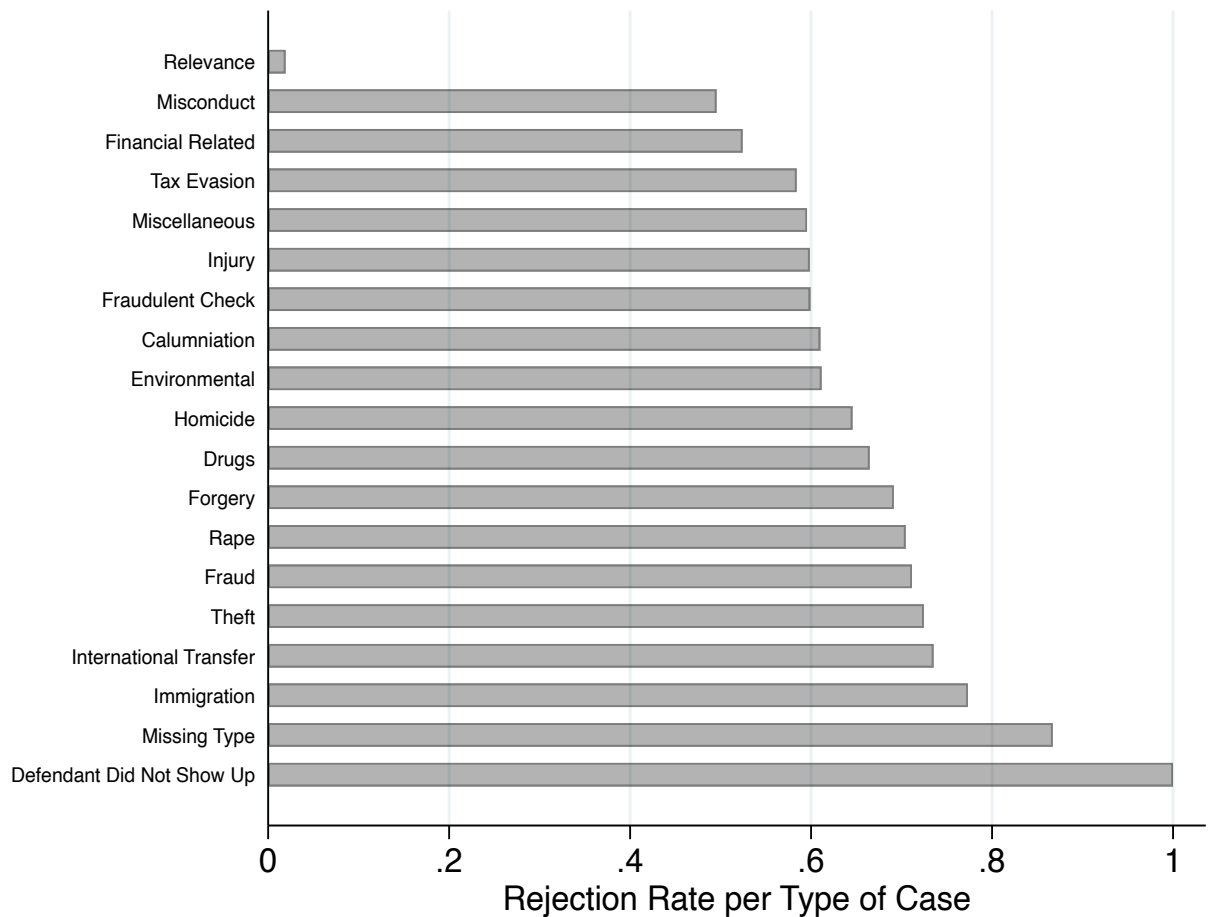


(b) Proportion of Male Defendants



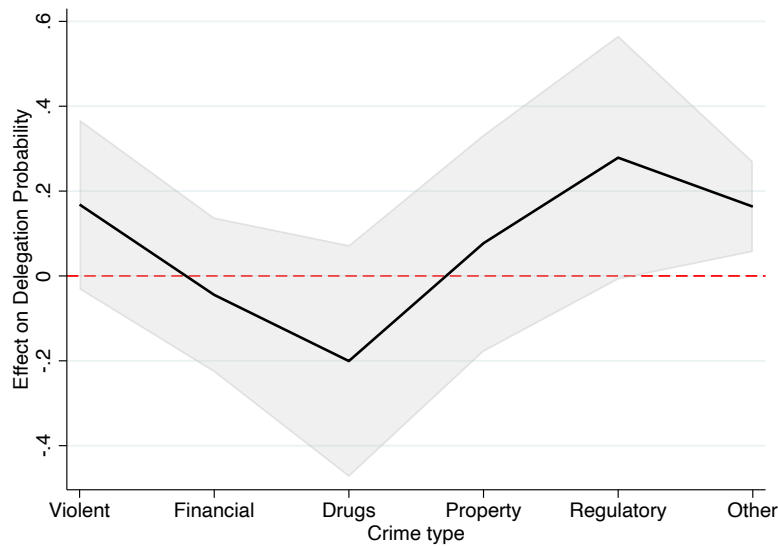
Notes: The figure shows the share of male plaintiffs (Panel A) and male defendants (Panel B) by case type. Cases in which the plaintiff is a bank, organization, or government entity are excluded. Case types on the vertical axis are ranked by the share of male plaintiffs or defendants, from lowest to highest, respectively.

Figure A7: REJECTION RATES OF APPEAL REQUESTS BY CASE TYPE



Notes: The figure shows the rejection rates for all classified case types in the Supreme Court. Case types on the vertical axis are ordered from the lowest to the highest rejection rate.

Figure A8: Heterogeneity by Broad Crime Categories



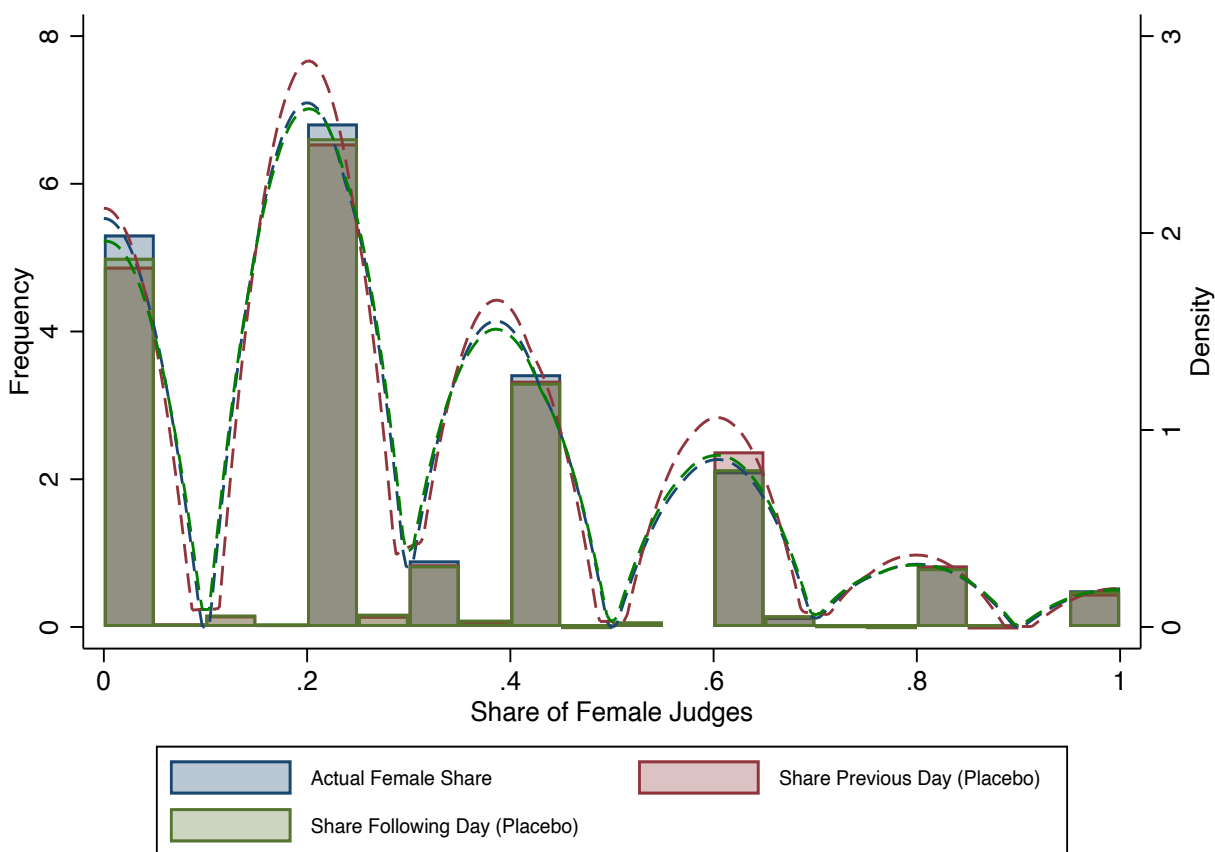
(A) Effects on Reject Appeal Request by Broad Crime Categories



(B) Effects on Delegating Trial by Broad Crime Categories

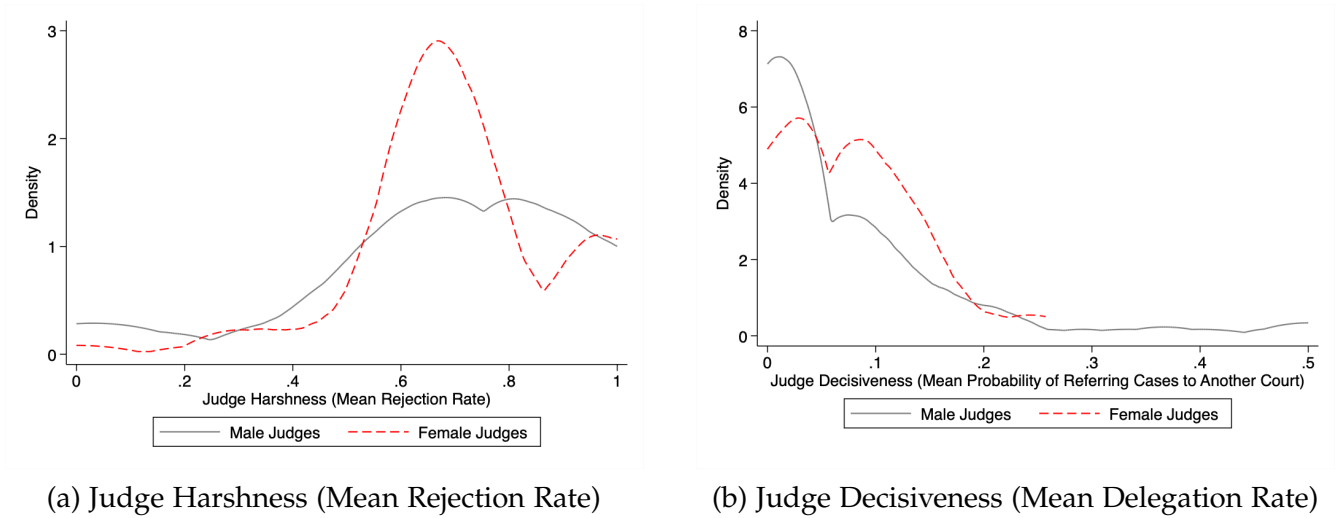
Notes: The figure reports marginal effects of the share of female judges across broad crime categories. Crime types are grouped as follows: violent (homicide, injuries, rape), financial (tax evasion, fraudulent checks, fraud, forgery, and related offenses), drugs, property (theft), regulatory (environmental and immigration offenses), and other (including misconduct, calumny, relevance, international transfer, non-appearance of the defendant, and miscellaneous cases). The estimates reported in Panels A and B are obtained from specifications that interact the share of female judges with crime-type indicators and include the full set of controls and fixed effects used in the main regressions. Shaded areas denote 95% confidence intervals.

Figure A9: DISTRIBUTIONS OF ACTUAL AND PLACEBO SHARE OF FEMALE JUDGES



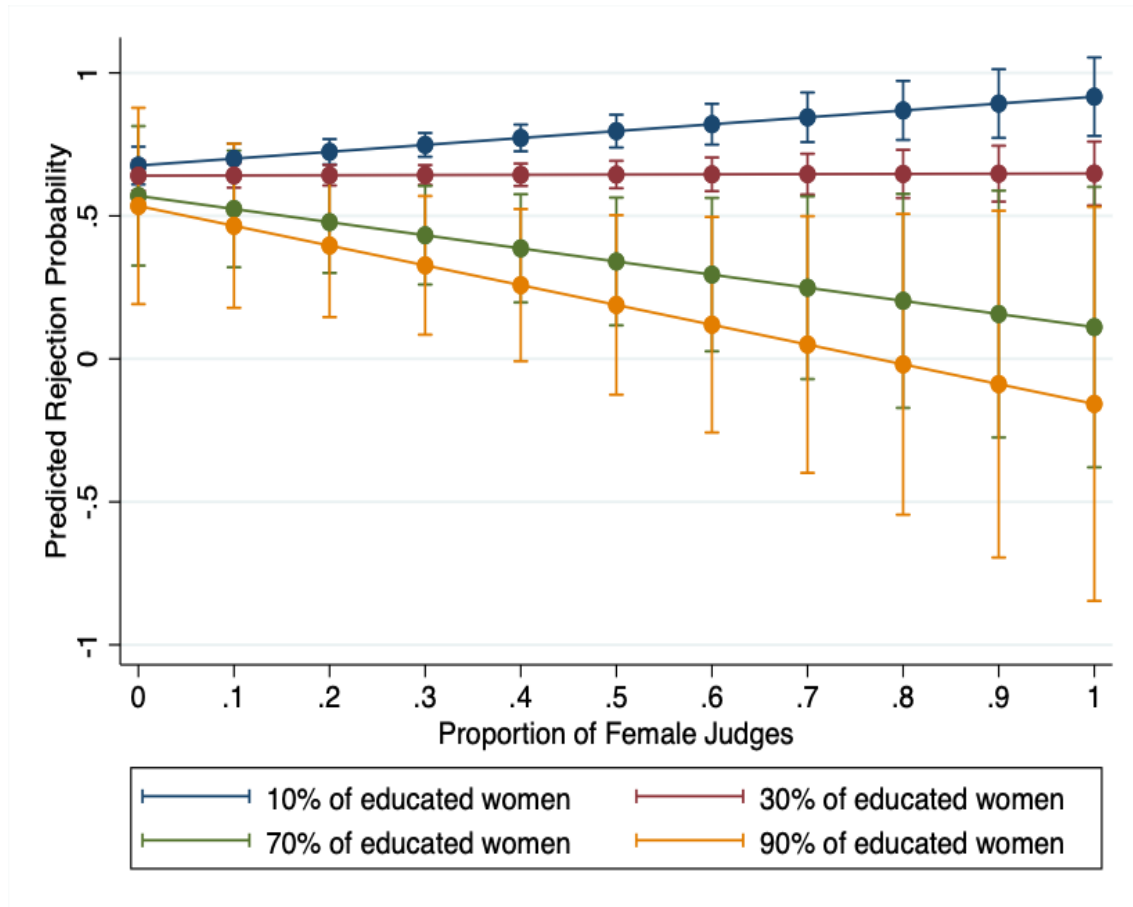
Notes: The figure plots the distribution of the share of female judges assigned to Supreme Court cases on the trial day (Actual Female Share) and the corresponding placebo distributions constructed using the share of female judges assigned on the previous (Share Previous Day, Placebo) and following (Share Following Day, Placebo) trial days. Histograms are shown on the left axis, while kernel density estimates are shown on the right axis.

Figure A10: JUDGE-LEVEL HARSHNESS AND DECISIVENESS BY GENDER



Notes: The figure plots kernel density estimates of judge-level harshness and decisiveness separately for male and female judges. The left panel (a) shows the distribution of each judge's mean rejection rate across all cases in which they participated. The right panel (b) shows the distribution of each judge's mean probability of referring a case to another court. Both measures are computed at the judge level, where each judge contributes one observation corresponding to their average behavior across all cases. A two-sample t-test of the difference in mean rejection rates between male and female judges yields a difference of 2.4 percentage points (0.712 vs. 0.688), which is statistically indistinguishable from zero ($t = -0.58$, $p = 0.56$), indicating that male and female judges do not differ meaningfully in individual harshness. The gray solid line corresponds to male judges and the red dashed line corresponds to female judges.

Figure A11: HETEROGENEITY IN THE EFFECT OF FEMALE JUDGES BY WOMEN'S EDUCATIONAL ATTAINMENT



Notes: The figure plots predicted appeal rejection probabilities as a function of the share of female judges on the Supreme Court committee, separately for regions where 10%, 30%, 70%, and 90% of women hold upper secondary education. These values span the observed range of women's upper secondary educational attainment across local court municipalities, drawn from the 2001 Population Census. Vertical bars denote 95 percent confidence intervals.

Table A1: The Current Setting

Court	Types of Cases	Time	Judges	Panels	Cases	Judges	Local	
		Coverage	per Panel				Courts	Divisions
Greek Supreme Court	Criminal Appeals	2005–2017	Mostly Five	837	3,753	142	364	3

Note: The vast majority of cases ($\approx 90\%$) are decided by five-member panels; the remainder by three-member panels. Panels refer to unique panel compositions identified in the dataset. Trial days are Tuesday, Wednesday, and Friday. Data digitized from the online archives of the Supreme Court of Greece.

Table A2: NUMBER OF CASES PER YEAR FILED AND DECIDED IN THE SAMPLE

Panel A			
Year Case Filed	No. of cases	Perc.	Cum.
2005	55	1.53	1.53
2006	412	11.43	12.95
2007	439	12.17	25.12
2008	429	11.9	37.02
2009	366	10.15	47.17
2010	330	9.15	56.32
2011	258	7.15	63.48
2012	169	4.69	68.16
2013	290	8.04	76.21
2014	271	7.52	83.72
2015	285	7.9	91.63
2016	200	5.55	97.17
2017	102	2.83	100

Panel B			
Decision Year			
2005	6	0.17	0.17
2006	416	11.54	11.7
2007	382	10.59	22.3
2008	406	11.26	33.56
2009	443	12.29	45.84
2010	332	9.21	55.05
2011	275	7.63	62.67
2012	171	4.74	67.42
2013	277	7.68	75.1
2014	274	7.6	82.7
2015	268	7.43	90.13
2016	228	6.32	96.45
2017	127	3.52	99.97
2018	1	0.03	100

Notes: The table presents the annual distribution of cases in the dataset, categorized by the year of case filing (Panel A) and the year of decision (Panel B).

Table A3: ADDITIONAL DESCRIPTIVE STATISTICS FOR THE MAIN STUDY SAMPLE

	Mean	Std.Dev	Min	Max
	(1)	(2)	(3)	(4)
Panel A: Trial Month and Day				
January	0.136	0.343	0	1
February	0.126	0.332	0	1
March	0.107	0.309	0	1
April	0.101	0.301	0	1
May	0.125	0.331	0	1
June, July, or September	0.080	0.271	0	1
October	0.111	0.314	0	1
November	0.119	0.324	0	1
December	0.096	0.294	0	1
Tuesday	0.371	0.483	0	1
Wednesday	0.285	0.451	0	1
Friday	0.341	0.474	0	1
Panel B: Gender Share for Each Judge				
Female Head Judge (1=yes)	0.216	0.411	0	1
Female Introducer Judge (1=yes)	0.301	0.459	0	1
Female Other Judge 1 (1=yes)	0.304	0.460	0	1
Female Other Judge 2 (1=yes)	0.296	0.456	0	1
Female Other Judge 3 (1=yes)	0.291	0.454	0	1
Panel C: Appeal Outcomes				
Rejection of Appeal Case	0.670	0.470	0	1
Days Between Trial and Final Trial Outcome	61.742	74.955	0	1132
Delegates to Another Court and Judge	0.090	0.286	0	1
Monetary Trial Expenses (in Euro)	191.671	200.647	0	2750

Notes: The table presents summary statistics on the trial month and day (Panel A), the role composition of judges in the Supreme Court committee by gender (Panel B) and appeal outcomes (Panel C). The head judge is the most senior judge in the committee. The introducer is the judge assigned to prepare the case for trial in the Supreme Court. Judges 1, 2, and 3 serve as committee members without specific additional responsibilities. In three-member panels (approximately 10% of cases), only one additional committee member beyond the head and introducer is present. All specifications control for the number of judges on the panel to account for this variation in committee size. *Rejection of Appeal Case* is a binary indicator that takes the value of 1 if the Supreme Court judging committee rejects the appeal request. *Delegates to Another Court and Judge* is a binary indicator that takes the value of 1 if the Supreme Court judging committee delegates the case to another court and judge.

Table A4: DATA DICTIONARY FOR TYPES OF SUPREME COURT CASES

Offense	Keywords
Homicide	Homicide, murder, manslaughter, attempted homicide, negligent homicide, fatal assault
Injuries	Bodily Harm, Violence, Traffic Offenses, Human Trafficking
Rape	Rape, seduction, indecency, sexual abuse
Tax Evasion	Tax evasion, delay or non-payment of taxes, employer contributions, fictitious invoices
Fraudulent Check	Dishonoured checks, Bad Checks
Financial Related	Smuggling, usury, counterfeiting, financial fraud, money laundering, misuse of funds, non-payment of wages
Drugs	Drugs, narcotics, illegal substances, drug possession, possession of narcotics, drug use, drug trafficking, drug dealing, distribution of drugs, drug production, cultivation of drugs, import or export of drugs
Misconduct	Violation of duty, abuse or misuse of power, bribery
Fraud	Fraud, deception, misrepresentation, fraudulent practices
Forgery	Forgery, perjury, false certification, false documents, document exportation
Calumny	Slander, false accusation, moral harm, distortion of facts
Theft	Theft, robbery, embezzlement
Environmental	Pollution, illegal logging, unauthorized construction, arson, illegal dumping, environmental degradation, urban planning violations
Immigration	Illegal immigration, illegal entry, facilitation of foreigner transfer
Relevance	Jurisdiction regulation
International Transfer of Defendant	Extradition, extradition of foreigners
Defendant Did Not Show Up	Desertion of the appellant
Miscellaneous	Cases not fitting into the above categories

Notes: The table describes case types and the corresponding keywords used in the data classification process. These classifications served as a protocol during data collection. Cases with no reported type are classified as “Missing Type” and are captured by a separate missing value indicator in all regression specifications.

Table A5: SUMMARY STATISTICS FOR LOCAL COURT TYPES

Type of Local Court	Total Local Courts	Total Cases
Magistrates Court	8	11
Single-member Court for Misdemeanors	90	334
Three-member Court for Misdemeanors	61	1327
Single-member Appeal Courts	36	458
Three-member Appeal Courts	47	859
Mixed Jury Appeal Court	18	24
Five-member Appeal Courts	11	299
Single-member Juvenile Court	1	1
Three-member Appeal Juvenile Courts	11	41
Three-member Juvenile Courts	71	358
Mixed Jury Courts of Appeal	6	41
Total	364	3,753

Notes: The table shows the different local court types in which cases are initially decided at the local level (column 1), the number of total local courts that belong in each local court type (column 2) and the number of total cases decided by each local court type (column 3).

Table A6: BASELINE ESTIMATES OF THE EFFECT OF SHARE OF FEMALE JUDGES ON APPEAL OUTCOMES BY SUBGROUP

	One Defendant		One Plaintiff	
	(1)	(2)	(3)	(4)
<u>Outcomes Decided by Judging Committee</u>				
Reject Appeal Request	0.129 (0.062)**	0.141 (0.058)**	0.222 (0.108)**	0.167 (0.103)
<i>N</i>	2,330	2,330	742	742
Delegates Trial to Another Court	-0.088 (0.033)***	-0.098 (0.033)***	-0.112 (0.064)*	-0.118 (0.056)**
<i>N</i>	2,330	2,330	742	742
<u>Placebo Outcome Affected by Court</u>				
Monetary Trial Expenses	2.900 (22.240)	17.406 (21.016)	57.841 (63.888)	33.191 (62.184)
<i>N</i>	2,330	2,330	742	742
Basic Controls	✓	✓	✓	✓
Additional Controls		✓		✓

Notes: The table presents estimates for the effects of the share of female judges in a judging committee on appeal outcomes for one defendant (columns 1–2) and one plaintiff (columns 3–4). The unit of observation is the case. Basic controls include year, day-of-week, and court-type fixed effects. Additional Controls include number of defendants, number of judges, number of lawyers involved, a binary indicator for whether the president is female, a binary indicator for whether the district attorney is female, a binary indicator for whether the introducer is female, the committee overall experience, whether it is a violent crime, experience of the introducer, whether it is a gender-sensitive case, whether it is a juvenile defendant, indicators for missing values in these covariates, and primary and secondary case-type fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A7: ESTIMATES OF THE EFFECT OF THE SHARE OF FEMALE JUDGES ON APPEAL OUTCOMES BY LITIGANT GENDER

	One Defendant			One Plaintiff		
	All (1)	Male (2)	Female (3)	All (4)	Male (5)	Female (6)
<u>Outcomes Decided by Judging Committee</u>						
Reject Appeal Request	0.117 (0.052)**	0.139 (0.056)**	-0.126 (0.148)	0.211 (0.100)**	0.198 (0.119)*	0.123 (0.198)
N	2,959	2,557	401	769	528	241
Delegates Trial to Another Court	-0.087 (0.026)***	-0.075 (0.028)***	-0.123 (0.075)	-0.107 (0.054)**	-0.116 (0.067)*	-0.098 (0.104)
N	2,959	2,557	401	769	528	241
<u>Outcome Affected by Court</u>						
Monetary Trial Expenses	19.285 (20.373)	18.030 (21.768)	-8.649 (65.217)	28.206 (62.030)	48.130 (77.841)	-58.161 (112.934)
N	2,959	2,557	401	769	528	241
Basic Controls	✓	✓	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓	✓	✓

Notes: The table presents estimates for the effects of the share of female judges on appeal outcomes by the gender of the litigant: all cases that have one defendant (column 1), cases with one male defendant (column 2), cases with one female defendant (column 3), all cases with one plaintiff (column 4), cases with one male plaintiff (column 5), and cases with one female plaintiff (column 6). The unit of observation is the case. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A8: HETEROGENEOUS EFFECTS BY COMMITTEE AND JUDGES EXPERIENCE

	By Judging Experience of							
	Judging Committee				Head Judge		Introducer Judge	
	(1) Below Median Multiple Litigants	(2) Above Median	(3) Below Median One Litigant	(4) Above Median	(5) Below Median	(6) Above Median	(7) Below Median	(8) Above Median
<u>Outcomes Decided by Judging Committee</u>								
Reject Appeal Request	0.036 (0.087)	0.108 (0.071)	0.072 (0.092)	0.135 (0.078)*	-0.023 (0.085)	0.190 (0.078)**	0.071 (0.073)	0.137 (0.075)*
<i>N</i>	2,998	2,998	2,562	2,562	1,276	1,286	1,316	1,246
Delegates Trial to Another Court	-0.097 (0.046)**	-0.085 (0.045)*	-0.093 (0.048)*	-0.103 (0.047)**	0.016 (0.048)	-0.137 (0.047)***	-0.029 (0.042)	-0.123 (0.046)***
<i>N</i>	2,998	2,998	2,562	2,562	1,276	1,286	1,316	1,246
<u>Placebo Outcome Affected by Court</u>								
Monetary Trial Expenses	-21.800 (43.547)	20.201 (30.650)	1.621 (39.504)	46.768 (28.121)*	33.360 (47.679)	10.366 (38.282)	4.110 (30.607)	22.020 (39.018)
<i>N</i>	2,998	2,998	2,562	2,562	1,276	1,286	1,316	1,246
Basic Controls	✓	✓	✓	✓	✓	✓	✓	✓
Additional Controls		✓		✓		✓		✓

Notes: The table presents estimates for the effects of the share of female judges in a judging committee on appeal outcomes. The unit of observation is the case. Basic controls include year, day-of-week, and court-type fixed effects. Additional Controls include number of defendants, number of judges, number of lawyers involved, a binary indicator for whether the president is female, a binary indicator for whether the district attorney is male, a binary indicator for whether the introducer is male, the committee overall experience, whether it is a violent crime, experience of the introducer, whether it is a gender-sensitive case, whether it is a juvenile defendant, indicators for missing values in these covariates, and primary and secondary case-type fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A9: HETEROGENEOUS EFFECTS BY ORIGIN COURT TYPE

	Single-Judge First Court	Multiple-Judges First Court	Appealed Before	First Time Appealed
	(1)	(2)	(3)	(4)
<u>Outcomes Decided by Judging Committee</u>				
Reject Appeal Request	0.325 (0.126) ^{***}	0.086 (0.060)	0.146 (0.072) ^{**}	0.165 (0.080) ^{**}
<i>N</i>	567	1,745	1,152	1,160
Delegates Trial to Another Court	-0.155 (0.085) [*]	-0.079 (0.032) ^{**}	-0.108 (0.043) ^{**}	-0.075 (0.044) [*]
<i>N</i>	567	1,745	1,152	1,160
<u>Outcome Affected by Court</u>				
Monetary Trial Expenses	43.395 (33.227)	12.657 (27.221)	11.961 (31.007)	27.673 (35.704)
<i>N</i>	567	1,745	1,152	1,160
Basic Controls	✓	✓	✓	
Additional Controls	✓	✓	✓	✓

Notes: The table presents heterogeneous estimates of the effects of the share of female judges in a judging committee on appeal outcomes by original local court type. I examine the following subgroups of local courts: single-judge local courts (column 1), multiple-judges local courts (column 2), appeal first-level courts (column 3) and non-appeal first level courts (column 4). The unit of observation is the case. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A10: FALSIFICATION TEST, SHARE OF FEMALE JUDGES IN PREVIOUS AND FOLLOWING DAY

	Reject Appeal Request				Delegates Trial to Another Court			
	Multiple Litigants (1)	(2)	One Litigant (3)	(4)	Multiple Litigants (5)	(6)	One Litigant (7)	(8)
Share Female Judges, Previous Day (Placebo)	-0.001 (0.034)	0.001 (0.032)	-0.010 (0.036)	0.002 (0.035)	0.014 (0.017)	0.014 (0.017)	0.020 (0.019)	0.020 (0.019)
<i>N</i>	2,976	2,976	2,545	2,545	2,976	2,976	2,545	2,545
Share Female Judges, Following Day (Placebo)	-0.024 (0.032)	-0.025 (0.031)	-0.033 (0.035)	-0.035 (0.034)	-0.025 (0.017)	-0.018 (0.016)	-0.021 (0.018)	-0.013 (0.018)
<i>N</i>	2,978	2,978	2,543	2,543	2,978	2,978	2,543	2,543
Basic Controls	✓	✓	✓	✓	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓	✓	✓	✓	✓
Additional Controls		✓		✓		✓		✓

Notes: The table reports estimates for falsification tests using the share of female judges in the previous and following trial day. Columns (1)-(4) show the estimates when the outcome is the likelihood to reject the appeal request, and columns (5)-(8) show the estimates when the outcome is the probability of delegating the trial to another local court or judge. The unit of observation is the case. *Basic Controls* include trial-month fixed effects, case-filing-year fixed effects, day-of-week fixed effects, and court-type fixed effects. *Additional controls* include the number of defendants, the number of judges, and the number of lawyers; binary indicators for whether the president is female, the district attorney is male, the introducer is male; indicators for violent crimes, gender-sensitive cases, and juvenile defendants; the average judging experience of the committee (measured in number of trials); experience of the introducer; indicators for missing values in these covariates; and primary and secondary case-type fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A11: ROBUSTNESS EXERCISE: BASELINE RESULTS WHILE CONTROLLING FOR JUDGE HARSHNESS

	Reject Appeal Request				Delegates Trial			
	Multiple Litigants		One Litigant		Multiple Litigants		One Litigant	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proportion of Female Judges	0.120 (0.050)**	0.085 (0.051)*	0.145 (0.054)***	0.111 (0.055)**	-0.092 (0.029)***	-0.070 (0.029)**	-0.097 (0.031)***	-0.080 (0.031)**
Avg. Committee Harshness		1.006 (0.309)***		0.957 (0.326)***		-0.636 (0.182)***		-0.493 (0.186)***
Harshness Control		✓		✓		✓		✓
Basic Controls	✓	✓	✓	✓	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓	✓	✓	✓	✓
N	2,998	2,998	2,565	2,565	2,998	2,998	2,565	2,565

Notes: The table reports estimates of the effect of the proportion of female judges on appeal outcomes, augmented with a leave-one-out measure of average committee harshness. For each judge j and case i , judge harshness is defined as the mean rejection rate of judge j across all cases in which judge j participated *excluding* case i , to avoid a mechanical correlation between the control variable and the outcome. The average committee harshness is computed as the mean of the leave-one-out harshness scores across all five judges assigned to each case. Columns (1), (3), (5), and (7) reproduce the baseline estimates from Table 3 without the harshness control; Columns (2), (4), (6), and (8) add the harshness control. The unit of observation is the case. *Basic Controls* include trial-month fixed effects, case-filing-year fixed effects, day-of-week fixed effects, and court-type fixed effects. Additional controls include the number of the number of defendants and lawyers, the size of the judging committee, indicators for president gender, district attorney gender, introducer gender, violent crime, gender-sensitive case, and juvenile defendant; experience of the introducer; missing-value indicators; and primary and secondary case-type fixed effects. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A12: ROBUSTNESS ESTIMATES WITH STANDARD ERRORS CLUSTERED AT THE LOCAL COURT LEVEL

	Multiple Litigants		One Litigant	
	(1)	(2)	(3)	(4)
<u>Outcomes Decided by Judging Committee</u>				
Reject Appeal Request	0.114 (0.055)**	0.120 (0.047)**	0.155 (0.059)***	0.145 (0.054)***
<i>N</i>	2,998	2,998	2,562	2,562
Delegates Trial to Another Court	-0.107 (0.031)***	-0.092 (0.029)***	-0.103 (0.034)***	-0.097 (0.031)***
<i>N</i>	2,998	2,998	2,562	2,562
<u>Placebo Outcome Affected by Court</u>				
Monetary Trial Expenses	18.158 (23.232)	27.436 (22.286)	18.158 (23.232)	27.436 (22.286)
<i>N</i>	2,998	2,998	2,562	2,562
Basic Controls	✓	✓	✓	✓
Additional Controls		✓		✓

Notes: The table presents estimates for the effects of the share of female judges in a judging committee on appeal outcomes. The unit of observation is the case. Basic controls include year, day-of-week, and court-type fixed effects. Additional Controls include number of defendants, number of judges, number of lawyers involved, a binary indicator for whether the president is female, a binary indicator for whether the district attorney is female, binary indicator for whether the introducer is female, the committee overall experience, whether it is a violent crime, experience of the introducer, whether it is a gender-sensitive case, whether it is a juvenile defendant, indicators for missing values in these covariates, and primary and secondary case-type fixed effects. Standard errors are reported in parentheses and are clustered at the local court level. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A13: Robustness Estimates with Standard Errors Clustered at the Panel Level

	Multiple Litigants		One Litigant	
	(1)	(2)	(3)	(4)
<i>Outcomes Decided by Judging Committee</i>				
Reject Appeal Request	0.101 (0.049)**	0.120 (0.052)**	0.139 (0.059)**	0.145 (0.059)**
N	2,998	2,998	2,562	2,562
Delegates Trial to Another Court	-0.052 (0.027)*	-0.092 (0.032)***	-0.084 (0.033)**	-0.097 (0.033)***
N	2,998	2,998	2,562	2,562
<i>Placebo Outcome Decided by Court</i>				
Monetary Trial Expenses	-1.730 (25.386)	14.127 (26.778)	20.206 (26.469)	27.436 (25.958)
N	2,998	2,998	2,562	2,562
Basic Controls	✓	✓	✓	✓
Additional Controls		✓		✓

Notes: The table presents estimates for the effects of the share of female judges in a judging committee on appeal outcomes with standard errors clustered at the panel level to account for potential correlation in outcomes across cases decided by the same panel. Among the 3,753 cases in the sample, I identify 837 unique panel compositions, with the same combination of judges sitting together on average 4.5 times over the sample period. The unit of observation is the case. Basic controls include year, day-of-week, and court-type fixed effects. Additional Controls include number of defendants, number of judges, number of lawyers involved, a binary indicator for whether the president is female, a binary indicator for whether the district attorney is male, a binary indicator for whether the introducer is male, the committee overall experience, whether it is a violent crime, experience of the introducer, whether it is a gender-sensitive case, whether it is a juvenile defendant, indicators for missing values in these covariates, and primary and secondary case-type fixed effects. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A14: ROBUSTNESS ESTIMATES WITH PROSECUTOR FIXED EFFECTS

	Multiple Litigant		One Litigant	
	(1)	(2)	(3)	(4)
Outcomes Decided by Judging Committee				
Reject Appeal Request	0.120 (0.050)**	0.111 (0.053)**	0.145 (0.054)***	0.138 (0.056)**
<i>N</i>	2,998	2,998	2,562	2,562
Delegates Trial to Another Court	-0.092 (0.029)***	-0.083 (0.031)***	-0.097 (0.031)***	-0.087 (0.033)***
<i>N</i>	2,998	2,998	2,562	2,562
Placebo Outcome Affected by Court				
Monetary Trial Expenses	27.436 (22.556)	32.551 (22.398)	27.436 (22.556)	32.551 (22.398)
<i>N</i>	2,998	2,998	2,562	2,562
Basic Controls	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓
Prosecutor FE		✓		✓

Notes: The table presents estimates for the effects of the share of female judges in a judging committee on appeal outcomes while I include prosecutor fixed effects. The unit of observation is the case. I use the same *Basic Controls* and *Additional Controls* as in the main analysis. Columns (1) and (3) show the baseline effects for multiple and single litigants. In columns (2) and (4), I add prosecutor fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A15: STACKED DATA, ESTIMATES WITH JUDGE FIXED EFFECTS

	Multiple Litigant (1)	One Litigant (2)
Reject Appeal Request		
Robust Standard Errors	0.117 (0.030)***	0.132 (0.032)***
Cluster at the Local Court Level	(0.051)**	(0.056)**
Cluster at the Judge Level	(0.032)*** 14,282 (1)	(0.039)*** 12,297 (2)
Delegates Trial to Another Court		
Robust Standard Errors	-0.092 (0.017)***	-0.096 (0.018)***
Cluster at the Local Court Level	(0.032)***	(0.033)***
Cluster at the Judge Level	(0.015)*** 14,282	(0.015)*** 12,297
Basic Controls	✓	✓
Additional Controls	✓	✓
Judge FE	✓	✓

Notes: The table reports a single point estimate per outcome, followed by three rows of standard errors in parentheses corresponding to: (i) robust standard errors, (ii) standard errors clustered at the local court level, and (iii) standard errors clustered at the judge level, in that order. The table presents estimates for the effects of the share of female judges in a judging committee on appeal outcomes using a stacked dataset. All specification include a judge fixed effect. Each observation corresponds to a specific judge and case combination. I use the same *Basic Controls* and *Additional Controls* as in the main analysis. I present estimates using a) robust standard errors, b) standard errors clustered at the local court level, and c) standard errors clustered at the judge level. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A16: ROBUSTNESS EXERCISE: BASELINE ESTIMATES WHEN TRIMMING TREATMENT VARIABLE BY 5% ON EACH SIDE

	Baseline Regressions				Placebo Regressions			
	Multiple Litigants		One Litigant		Multiple Litigants		One Litigant	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<u>Outcomes Decided by Judging Committee</u>								
Reject Appeal Request	0.082 (0.050)*	0.103 (0.054)*	0.133 (0.064)**	0.134 (0.059)**	0.017 (0.046)	0.021 (0.044)	0.007 (0.049)	0.010 (0.047)
<i>N</i>	2,927	2,927	2,497	2,497	2,927	2,927	2,497	2,497
Delegates Trial to Another Court	-0.041 (0.028)	-0.090 (0.033)***	-0.099 (0.037)***	-0.090 (0.033)***	0.002 (0.025)	0.005 (0.024)	0.003 (0.025)	0.008 (0.025)
<i>N</i>	2,998	2,562	2,562	2,562	2,998	2,562	2,998	2,562
<u>Placebo Outcome Decided by Court</u>								
Monetary Trial Expenses	6.937 (25.636)	20.059 (26.321)	21.899 (26.807)	28.929 (25.835)	2.151 (23.796)	1.383 (21.391)	26.435 (22.623)	21.414 (21.722)
<i>N</i>	2,998	2,562	2,562	2,562	2,998	2,562	2,998	2,562
Basic Controls	✓	✓	✓	✓	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓	✓	✓	✓	✓
Additional Controls		✓		✓		✓		✓

Notes: The table presents estimates for the effects of the share of female judges in a judging committee on appeal outcomes. The unit of observation is the case. *Basic Controls* include trial-month fixed effects, case-filing-year fixed effects, day-of-week fixed effects, and court-type fixed effects. Additional controls include the number of defendants, the number of judges, and the number of lawyers; binary indicators for whether the president is female, the district attorney is male, the introducer is male; indicators for violent crimes, gender-sensitive cases, and juvenile defendants; the average judging experience of the committee (measured in number of trials); experience of the introducer; indicators for missing values in these covariates; and primary and secondary case-type fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A17: HETEROGENEITY IN THE EFFECT OF FEMALE JUDGES BY LOCAL SOCIOECONOMIC CONTEXT

	Outcome: Reject Appeal Request				
Proportion of Female Judges	0.125 (0.050)**	0.550 (0.203)***	0.391 (0.151)***	0.298 (0.108)***	0.183 (0.060)***
Proportion of Female Judges × Civil Marriages (std.)	-0.050 (0.027)*				
Proportion of Female Judges × Log GDP		-0.054 (0.025)**			
Proportion of Female Judges × Women Upper Secondary			-1.253 (0.655)*		
Proportion of Female Judges × Women Tertiary				-1.534 (0.819)*	
Proportion of Female Judges × Women Masters or PhD					-18.593 (9.618)*
Civil Marriages (std.)	0.015 (0.011)				
Log GDP		0.012 (0.010)			
Women Upper Secondary			0.044 (0.263)		
Women Tertiary				0.265 (0.327)	
Women Masters or PhD					2.911 (3.974)
N	2,997	2,999	2,999	2,999	2,999
Basic Controls	✓	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓	✓

Notes: The table reports estimates from specifications in which the share of female judges is interacted with local socioeconomic characteristics. The interacted variables include a standardized measure of civil marriages, log GDP, the share of women with upper secondary education, the share of women with tertiary education, and the local employment rate. These characteristics are obtained from the 2001 Population Census conducted by the National Statistical Authority and matched to the municipality in which each local court is located. The unit of observation is the case. Basic controls include year, day-of-week, and court-type fixed effects. Additional controls include the number of defendants and lawyers, the size of the judging committee, indicators for president gender, district attorney gender, introducer gender, committee experience, violent crime, gender-sensitive case, and juvenile defendant; experience of the introducer; missing-value indicators; and primary and secondary case-type fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table A18: DELIBERATION TIME BY PANEL FAMILIARITY

	Multiple Litigants		One Litigant	
	Below Median Familiarity (1)	Above Median Familiarity (2)	Below Median Familiarity (3)	Above Median Familiarity (4)
<u>Outcome: Days Between Trial and Final Decision</u>				
Days Between Trial and Final Decision	58.388 (11.300)***	107.358 (12.119)***	57.135 (12.120)***	102.300 (12.462)***
N	1522	1476	1296	1266
Basic Controls	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓

Notes: The table presents estimates for the effects of the share of female judges in a judging committee on deliberation time (days between the trial date and the final decision), split by panel familiarity. Panel familiarity is defined as the number of times the exact same combination of judges has sat together prior to each given case. The median familiarity is 3. Below-median familiarity panels have sat together fewer times than 3; above-median familiarity panels have sat together more frequently. The unit of observation is the case. *Basic Controls* include trial-month fixed effects, case-filing-year fixed effects, day-of-week fixed effects, and court-type fixed effects. Additional controls include the number of defendants, the number of judges, and the number of lawyers; binary indicators for whether the president is female, the district attorney is male, the introducer is male; indicators for violent crimes, gender-sensitive cases, and juvenile defendants; the average judging experience of the committee (measured in number of trials); experience of the introducer; indicators for missing values in these covariates; and primary and secondary case-type fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A19: NON-LINEAR EFFECTS OF THE PROPORTION OF FEMALE JUDGES ON DELIBERATION TIME

	All Cases		Simple Cases		Complex Cases	
	Multiple Litigants	One Litigant	Multiple Litigants	One Litigant	Multiple Litigants	One Litigant
	(1)	(2)	(3)	(4)	(5)	(6)
Q2	12.275 (3.200)***	11.864 (3.430)***	8.846 (4.367)**	7.908 (4.427)*	15.811 (4.783)***	15.196 (5.224)***
Q3	30.541 (4.878)***	29.878 (5.174)***	28.548 (8.877)***	30.291 (9.617)***	34.668 (6.256)***	33.214 (6.270)***
Q4	51.228 (6.217)***	49.888 (6.498)***	41.442 (8.060)***	43.284 (9.073)***	57.359 (8.739)***	54.880 (9.150)***
Q5	53.159 (6.986)***	49.705 (7.301)***	38.282 (7.520)***	37.700 (8.318)***	62.899 (10.550)***	58.846 (11.218)***
N	2998	2562	1145	1026	1853	1536
Basic Controls	✓	✓	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓	✓	✓

Notes: The table presents nonlinear estimates of the effects of the share of female judges in a judging committee on deliberation time (days between the trial date and the final decision). Columns (1)–(2) use the full sample. Columns (3)–(4) restrict to simple cases (opinion length below the median of 4 pages) and columns (5)–(6) to complex cases (opinion length at or above the median), mirroring the sample split in Table A18. The model replaces the continuous female share variable with quintile indicators for the share of female judges. The omitted category is quintile 1 (Q1), which corresponds to panels with no female judges. The unit of observation is the case. *Basic Controls* include trial-month fixed effects, case-filing-year fixed effects, day-of-week fixed effects, and court-type fixed effects. Additional controls include the number of defendants, the number of judges, and the number of lawyers; binary indicators for whether the president is female, the district attorney is male, the introducer is male; indicators for violent crimes, gender-sensitive cases, and juvenile defendants; the average judging experience of the committee (measured in number of trials); experience of the introducer; indicators for missing values in these covariates; and primary and secondary case-type fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A20: NON-LINEAR EFFECTS FOR SIMPLE AND COMPLEX CASES

	Outcome: Reject Appeal Request			
	Simple Cases		Complicated Cases	
	Multiple Litigants		One Litigant	
Q2	-0.008 (0.029)	-0.008 (0.029)	-0.035 (0.030)	-0.035 (0.030)
Q3	-0.033 (0.036)	-0.033 (0.036)	-0.043 (0.038)	-0.043 (0.038)
Q4	0.030 (0.047)	0.030 (0.047)	0.008 (0.050)	0.008 (0.050)
Q5	0.049 (0.053)	0.049 (0.053)	0.033 (0.054)	0.033 (0.054)
Basic Controls	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓
Additional Controls		✓		✓
	Outcome: Delegates Trial to Another Court			
	Simple Cases		Complicated Cases	
	Multiple Litigants		One Litigant	
Q2	-0.001 (0.020)	-0.001 (0.020)	0.015 (0.020)	0.015 (0.020)
Q3	-0.006 (0.022)	-0.006 (0.022)	0.009 (0.023)	0.009 (0.023)
Q4	-0.010 (0.036)	-0.010 (0.036)	0.001 (0.036)	0.001 (0.036)
Q5	-0.063 (0.039)	-0.063 (0.039)	-0.056 (0.039)	-0.056 (0.039)
Basic Controls	✓	✓	✓	✓
Case Type FE	✓	✓	✓	✓
Additional Controls		✓		✓

Notes: The table presents nonlinear estimates of the effects of the share of female judges in a judging committee on appeal outcomes for simple and complex cases, separately. Simple and complex cases are defined based on opinion length, with complex cases having above-median length (4 pages). The model replaces the single treatment variable with a set of quintile indicators for the share of female judges. The omitted category is quintile 1 (Q1). The unit of observation is the case. Basic controls include year, day-of-week, and court-type fixed effects. Additional Controls include number of defendants, number of judges, number of lawyers involved, a binary indicator for whether the president is female, a binary indicator for whether the district attorney is male, a binary indicator for whether the introducer is male, the committee overall experience, whether it is a violent crime, experience of the introducer, whether it is a gender-sensitive case, whether it is a juvenile defendant, indicators for missing values in these covariates, and primary and secondary case-type fixed effects. Robust standard errors are reported. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A21: FEMALE REPRESENTATION AND COURT CONGESTION

Panel A: Relationship between Share of Female Judges and Court Congestion				
Cases per day	0.002 (0.001)			
Cases per court-day		0.001 (0.002)		
Cases per month-year			0.000 (0.000)	
Cases per court-month-year				-0.000 (0.001)
N	2,998	2,998	2,998	2,998
Panel B: Main Results on Rejection while Controlling for Court Congestion				
Share of Female Judges	0.120 (0.050)**	0.121 (0.050)**	0.120 (0.050)**	0.120 (0.050)**
Cases per day	0.000 (0.003)			
Cases per court-day		-0.005 (0.005)		
Cases per month-year			0.000 (0.001)	
Cases per court-month-year				-0.001 (0.002)
N	2,998	2,998	2,998	2,998
Panel C: Main Results on Delegation while Controlling for Court Congestion				
Share of Female Judges	-0.091 (0.029)***	-0.092 (0.029)***	-0.092 (0.029)***	-0.092 (0.029)***
Cases per day	-0.003 (0.002)*			
Cases per court-day		-0.001 (0.002)		
Cases per month-year			-0.001 (0.000)*	
Cases per court-month-year				-0.000 (0.001)
N	2,998	2,998	2,998	2,998
Basic Controls	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓

Notes: Panel A reports regressions of the share of female judges on alternative measures of court congestion, including the number of cases heard per day, per court-day, and at monthly aggregation levels. Panels B and C report the main specifications for appeal rejection and case delegation, respectively, controlling for these congestion measures. The unit of observation is the case. All specifications include court fixed effects, calendar fixed effects (trial month, trial day, and filing year), and case-type fixed effects. Additional controls include case characteristics, committee composition, experience measures, and indicators for missing values. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table A22: GENDER DIFFERENCES IN PREFERENCES AND ATTITUDES, WVS

	Men	Women	Diff. (F–M)	SE	p-value	N
Willingness to Take Risks	0.016	-0.292	-0.308***	(0.056)	0.000	1000
Patience (Willingness to Wait)	-0.276	-0.421	-0.146***	(0.049)	0.003	999
Negative Reciprocity (Revenge/Punish)	0.529	0.282	-0.247***	(0.058)	0.000	998
Positive Reciprocity (Return a Favor)	-0.050	-0.090	-0.040	(0.056)	0.470	1000
Altruism (Donation/Good Causes)	-0.136	-0.180	-0.044	(0.055)	0.432	1000
Trust (People Have Best Intentions)	-0.340	-0.381	-0.041	(0.053)	0.436	996

Notes: The table reports mean outcomes for men and women, the difference (Female–Male), the standard error of the difference, the p-value, and the number of observations. The difference is obtained from a regression of each variable on a female indicator; standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The measures capture risk-taking, patience, and reciprocity. The data come from the 7th wave of the World Value Survey and the analysis is restricted to Greece.